Semantic Coding: Partial Transmission

Semantic Coding: Partial Transmission

Frans M.J. Willems
Eindhoven University of Technology
Electrical Engineering Department
Eindhoven, The Netherlands

Ton Kalker
Hewlett-Packard Research
Mobile and Media Systems Lab
Palo Alto, CA, USA

***Abstract*—Shannon wrote in 1948: "The semantic aspects of communication are irrelevant to the engineering problem." He demonstrated indeed that the information generated by a source depends only on its statistics and not on the meaning of the source output. The authors derived the fundamental limits for semantic compaction, transmission and compression systems recently. These systems have the property that the codewords are semantic however, i.e. close to the source sequences. In the present article we determine the minimum distortion for semantic partial transmission systems. In these systems only a quantized version of each source source symbol is transmitted to the receiver. It should be noted that our achievability proof is based on weak instead of strong typicality. This is unusual for Gelfand-Pinsker [1980] related setups as e.g. semantic coding and embedding.**

I. Introduction

In [1] Shannon wrote: "The semantic aspects of communication are irrelevant to the engineering problem." Indeed Shannon demonstrated that the information that is generated by a source depends only on the statistics of the source, and not on the meaning of the source output. In contrast with this we have investigated in [2] whether in a compaction system the codewords can be (almost) as meaningful as the source output sequences. We required the codewords to be close to the source sequences for some given distortion measure. Moreover we considered semantic transmission. Now the encoded source output sequence is transmitted over a memoryless noisy channel to a decoder. Semantic transmission requires the codewords, i.e. the channel input sequences, to be close to the source sequences again. Finally we investigated semantic compression. A semantic compression system is a vector quantizer for which the codeword, i.e. the index to the reproduction vector, resembles the source sequence. For semantic compaction, transmission, and compression we determined the fundamental limits for the i.i.d. case in [2].

Here we consider semantic *partial* transmission, which is the transmission over a memoryless noisy channel of a "quantized" version of the source sequence, in such a way that the channel input sequence is *semantic, i.e. close to the source sequence*. The quantized symbols could e.g. represent the most significant bits of the source symbols. Since the receiver is only interested in the quantized source outputs we speak about partial transmission. For this model we determine the fundamental limit, i.e. set of achievable distortions. The obtained result is an extension of the semantic transmission problem, but can also be considered as a solution for the problem that combines semantic transmission and semantic compression, however only for a special distortion measure. In this combined problem the decoder at the output of the noisy channel has to produce a sequence whose distortion to the source sequence is small. In addition the channel input sequence has to semantic. For the case where the distortion measure is such that for each source output only its quantized value is allowed as receiver output, we found the solution. It should be noted that for arbitrary distortion measures the problem is still unsolved.

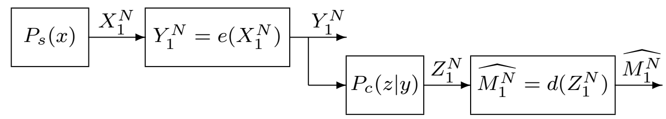

Fig. 1. A semantic partial transmission system.

The achievability parts in [2] are based on to the Gelfand-Pinsker proof for the side-information channel [3]. For semantic partial transmission we need a more general achievability proof than the one in [2]. Important is also that the proof that is presented here, is not based on strong typicality (developed by Wolfowitz [4], Berger [5], etc.) as in [3], but rather on weak typicality (developed by Forney [6] and Cover [7], etc.).

Semantic coding is closely related to reversible embedding. In fact semantic compaction is identical to zero-rate reversible embedding [8], semantic transmission is the same as robust reversible embedding [9], and semantic compression is zero-rate partially reversible embedding [10]. For an overview see [11]. Reversible embedding work has the same flavor as the work of Sutivong et al. [12], [13], however there semantic distortion is absent. We conclude this article by demonstrating how it relates to a result of Yang and Sun on embedding correlated watermarks [14].

II. Definitions, statement of result

A. Definitions

In Figure 1 a model of a semantic partial transmission system is shown. The source is assumed to be independent and identically distributed (i.i.d.). It produces the sequence $x_1^N = (x_1, x_2, \cdots, x_N)$ with probability

$$\Pr\{X_1^N = x_1^N\} = \Pi_{n=1}^N P_s(x_n) \quad (1)$$

for all $x_1^N \in \mathcal{X}^N$. Here $\mathcal{X}$ is the finite source alphabet and $\{P_s(x), x \in \mathcal{X}\}$ is probability distribution of the source. N is the block length. We are now interested in transmitting a

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

quantized version $m_1^N = (m_1, m_2, \cdots, m_N)$ of the source sequence x_1^N. The mapping $\mu(\cdot)$ from $\mathcal{X}$ on finite alphabet $\mathcal{M}$ determines the *partial source sequence* $m_1^N = (m_1, m_2, \cdots, m_N)$ of x_1^N component by component, as follows:

$$m_n = \mu(x_n), \text{ for } n = 1, N. \quad (2)$$

The mapping $\mu(\cdot)$ defines the joint probability of a source symbol $x \in \mathcal{X}$ and its quantized version $m \in \mathcal{M}$ as follows

$$P_s'(x, m) = P_s(x)\delta_{m,\mu(x)}, \quad (3)$$

where $\delta_{i,j} = 1$ if $i = j$ and zero otherwise (Kronecker delta).

An encoder $e(\cdot)$ transforms the source sequence x_1^N into a channel input sequence $y_1^N = (y_1, y_2, \cdots, y_N) \in \mathcal{Y}^N$. The modified sequence y_1^N is close to the original sequence x_1^N in the sense that the so-called *average semantic distortion* $\overline{D_{xy}}$ between X_1^N and Y_1^N is not too large. Here

$$\overline{D_{xy}} \triangleq \sum_{x_1^N} \Pr\{X_1^N = x_1^N\} D(x_1^N, e(x_1^N)) \text{ with}$$

$$D(x_1^N, y_1^N) \triangleq \frac{1}{N}\sum_{n=1}^{N} D_{xy}(x_n, y_n), \quad (4)$$

where $D_{xy}(\cdot,\cdot)$ is a matrix consisting of $|\mathcal{X}||\mathcal{Y}|$ non-negative values. The semantic sequence y_1^N is now transmitted over a discrete memoryless channel with input alphabet $\mathcal{Y}$, output alphabet $\mathcal{Z}$, and transition probability matrix $\{P_c(z|y), y \in \mathcal{Y}, z \in \mathcal{Z}\}$. The probability that output sequence $z_1^N = (z_1, z_2, \cdots, z_N)$ occurs when y_1^N is the channel input sequence is

$$\Pr\{Z_1^N = z_1^N | Y_1^N = y_1^N\} = \Pi_{n=1}^N P_c(z_n|y_n). \quad (5)$$

From the channel output sequence z_1^N a decoder $d(\cdot)$ constructs an estimate $\widehat{M_1^N}$ of the partial source sequence m_1^N. The *error probability* $P_\mathcal{E}$ is defined as

$$P_\mathcal{E} \triangleq \Pr\{\widehat{M_1^N} \neq M_1^N\}. \quad (6)$$

B. Statement of result

An $(N, \overline{D_{xy}}, P_\mathcal{E})$-code consists of an encoding function $e(\cdot)$ and a decoding function $d(\cdot)$, both operating on sequences of length N, resulting in an average semantic distortion $\overline{D_{xy}}$ and error probability $P_\mathcal{E}$. Distortion level Δ_{xy} is now said to be *achievable* if for all $\epsilon > 0$ there exists for all large enough N, codes $(N, \overline{D_{xy}}, P_\mathcal{E})$ such that

$$\overline{D_{xy}} \leq \Delta_{xy} + \epsilon, \text{ and } P_\mathcal{E} \leq \epsilon. \quad (7)$$

In sections III and IV we will prove the next theorem.

Theorem 1: For semantic partial transmission the set of achievable distortions is equal to $\mathcal{D}$ which is defined as

$$\begin{aligned}\mathcal{D} \triangleq \{\Delta_{xy} \;:\; & \Delta_{xy} \geq \sum_{x,y} P(x,y) D_{xy}(x,y), \text{ for} \\ & P(x,m,u,y,z) = P_s'(x,m)P_t(u,y|x)P_c(z|y) \\ & \text{for some auxiliary } U \text{ with } |\mathcal{U}| \leq |\mathcal{X}||\mathcal{Y}| \\ & \text{and test channel } P_t(u,y|x) \\ & \text{such that } I(M,U;Z) \geq I(M,U;X)\}. \quad (8)\end{aligned}$$

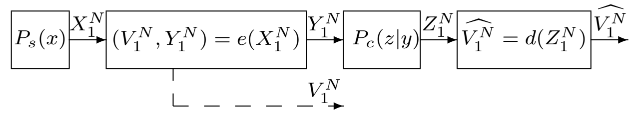

Fig. 2. An underlying model for semantic partial transmission.

The smallest possible distortion $\Delta_{\text{min}} = \min_{\Delta_{xy} \in \mathcal{D}} \Delta_{xy}$. Semantic transmission turns out to be impossible if $H(M)$ is larger than the capacity of the channel.

III. ACHIEVABILITY PROOF

A. Introduction

We will first consider a slightly different problem and prove a lemma concerning a joint distribution $P(x, v, y, z) = P_s(x)P_t(v, y|x)P_c(z|y)$ where $P_s(\cdot)$ is the source distribution, $\mathcal{V}$ a finite auxiliary alphabet, $P_t(\cdot,\cdot|\cdot)$ some fixed test-channel between $\mathcal{X}$ and $\mathcal{V} \times \mathcal{Y}$, and $P_c(\cdot|\cdot)$ the channel. We claim the existence of a set of sequences $v_1^N \in \mathcal{V}^N$ having certain properties within the following scenario. Consider Figure 2. The memoryless source produces a sequence x_1^N. An encoder $e(\cdot)$ observing x_1^N transmits over the channel $\{P_c(z|y), y \in \mathcal{Y}, z \in \mathcal{Z}\}$ a sequence $v_1^N \in \mathcal{V}$ that results in a channel input sequence y_1^N and whose (semantic) distortion to x_1^N should be acceptable. Moreover the sequence v_1^N should be the unique sequence jointly-typical with z_1^N such that a decoder $d(\cdot)$ can find v_1^N using typicality. An error occurs if the semantic distortion is too large or if v_1^N is not decoded.

Lemma 1: In the scenario described above, for each $\epsilon > 0$, for all N large enough, there exists a set of M sequences $v_1^N \in \mathcal{V}^N$ such that the error probability is not larger than 4ϵ if

$$I(V;X) + 4\epsilon \leq \frac{1}{N}\log_2 M \leq I(V;Z) - 4\epsilon. \quad (9)$$

B. Definition and properties of typical sets $\mathcal{A}_\epsilon^N$ and $\mathcal{B}_\epsilon^N$

First let K be a positive integer and fix an $0 < \epsilon < 1$.

Definition 1: The set $\mathcal{A}_\epsilon^N(V_1V_2\cdots V_K)$ of ϵ-typical N-sequences $(\underline{v}_1, \underline{v}_2, \cdots, \underline{v}_K)$ with respect to joint distribution $P(v_1, v_2, \cdots, v_K)$ is defined by

$$\begin{aligned}&\mathcal{A}_\epsilon^N(V_1V_2\cdots V_K) \\ &\quad \triangleq \left\{(\underline{v}_1, \underline{v}_2, \cdots, \underline{v}_K) : \left|\frac{1}{N}\log_2\frac{1}{P(\underline{w})} - H(W)\right| \leq \epsilon, \right. \\ &\qquad\qquad \left. \forall W \subseteq \{V_1, V_2, \cdots, V_K\}\right\}, \quad (10)\end{aligned}$$

where $P(\underline{w}) = \prod_{n=1}^N P(w_n)$.

For the properties of $\mathcal{A}_\epsilon^N$ we refer to Cover and Thomas [15].

Definition 2: For given $\underline{v}_1, \cdots \underline{v}_{K-1}$ we define

$$\begin{aligned}&\mathcal{A}_\epsilon^N(V_K|\underline{v}_1, \cdots, \underline{v}_{K-1}) \quad (11) \\ &\quad \triangleq \{\underline{v}_K : (\underline{v}_1, \cdots, \underline{v}_{K-1}, \underline{v}_K) \in \mathcal{A}_\epsilon^N(V_1V_2\cdots V_K)\},\end{aligned}$$

which is the set of sequences $\underline{v}_K$ conditionally ϵ-typical on $\underline{v}_1, \cdots \underline{v}_{K-1}$ with respect to $P(v_1, v_2, \cdots, v_K)$.

Note that the test-channel $P_t(v,y|x)$ determines the joint probability distribution $P(x,v,y,z) = P_s(x)P_t(v,y|x)P_c(z|y)$. The following definition is crucial. It allows us to avoid using strong typicality.

Definition 3: Consider the sets $\mathcal{B}_\epsilon^N(XV)$ defined as

$$\begin{aligned}&\mathcal{B}_\epsilon^N(XV) \\ &\triangleq \{(\underline{x},\underline{v}) : \Pr\{\underline{Z} \in \mathcal{A}_\epsilon^N(Z|\underline{x},\underline{v}) \\ &\quad \wedge d(\underline{Y},\underline{x}) \le D_{\text{exp}} + \epsilon \mid (\underline{X},\underline{V}) = (\underline{x},\underline{v})\} \ge 1-\epsilon\}\end{aligned} \tag{12}$$

where $\underline{Z}$ is the output of a "channel" $P(z|x,v) = P(x,v,z)/\sum_z P(x,v,z)$, with fixed inputs $\underline{x}$ and $\underline{v}$. Moreover $D_{\text{exp}} = \sum_{x,y} P(x,y)D_{xy}(x,y)$.

Property 1: For $(\underline{x},\underline{v}) \in \mathcal{B}_\epsilon^N(XV)$ there is at least one $\underline{z}$ such that $(\underline{x},\underline{v},\underline{z}) \in \mathcal{A}_\epsilon^N(XVZ)$, and therefore $(\underline{x},\underline{v}) \in \mathcal{A}_\epsilon^N(XV)$.

Property 2: Let $\underline{X},\underline{V},\underline{Y},\underline{Z}$ be i.i.d. with respect to $P(x,v,y,z)$. Then observe that

$$\begin{aligned}Q &\triangleq \Pr\{(\underline{X},\underline{V},\underline{Z}) \in \mathcal{A}_\epsilon^N(XVZ) \wedge d(\underline{X},\underline{Y}) \le D_{\text{exp}}+\epsilon\} \\ &\le \sum_{(\underline{x},\underline{v})\in\mathcal{B}_\epsilon^N(XV)} p(\underline{x},\underline{v}) + \sum_{(\underline{x},\underline{v})\notin\mathcal{B}_\epsilon^N(XV)} p(\underline{x},\underline{v})(1-\epsilon) \\ &= 1-\epsilon\Pr\{(\underline{X},\underline{V}) \notin \mathcal{B}_\epsilon^N(XV)\},\end{aligned} \tag{13}$$

or

$$\Pr\{(\underline{X},\underline{V}) \notin \mathcal{B}_\epsilon^N(XV)\} \le \frac{1-Q}{\epsilon}. \tag{14}$$

The weak law of large numbers implies that $Q \ge 1-\epsilon^2$ for large enough N. Using (14) this leads to the statement that for large enough N

$$\sum_{(\underline{x},\underline{v})\in\mathcal{B}_\epsilon^N(XV)} p(\underline{x},\underline{v}) \ge 1-\epsilon. \tag{15}$$

C. Random code construction

- *Random coding:* Generate M sequences $\underline{v}(w)$ for $w \in \{1,2,\cdots,M\}$ at random according to $p(v) = \sum_{x,y} P_s(x)P_t(v,y|x)$.
- *Encoding:* The encoder chooses an index w such that $(\underline{x},\underline{v}(w)) \in \mathcal{B}_\epsilon^N(XV)$. If such an index cannot be found an error is declared. The channel input sequence $\underline{y}$ now results from applying the "channel" $p(y|x,v) = p(v,y|x)/\sum_y p(v,y|x)$ to $(\underline{x},\underline{v}(w))$.
- *Decoding:* The decoder upon receiving $\underline{z}$, looks for the index $\widehat{w}$ such that $(\underline{v}(\widehat{w}),\underline{z}) \in \mathcal{A}_\epsilon^N(VZ)$. If a unique index does not exists an error is declared.

D. Error probability

Let $\underline{X}$ be the source sequence, W the index to $\underline{V}$, and $\underline{Z}$ the result of $\underline{X}$ and $\underline{V}(W)$. Then for $w \in \{1,2,\cdots,M\}$ we define the events:

$$\begin{aligned}B_w &\triangleq \{(\underline{X},\underline{V}(w)) \in \mathcal{B}_\epsilon^N(XV)\}, \\ A_w &\triangleq \{(\underline{V}(w),\underline{Z}) \in \mathcal{A}_\epsilon^N(VZ)\}, \\ C_w &\triangleq \{(\underline{X},\underline{V}(w),\underline{Z}) \in \mathcal{A}_\epsilon^N(XVZ)\}.\end{aligned} \tag{16}$$

The error probability (averaged over the ensemble of codes) is now:

$$\begin{aligned}\overline{P_\mathcal{E}} &= \Pr\{(\cap_w B_w^c) \cup A_W^c \cup (\cup_{w\ne W} A_w)\} \\ &\le \Pr\{\cap_w B_w^c\} + \Pr\{(\cup_w B_w)\cap C_W^c\} + \sum_{w\ne W}\Pr\{A_w\},\end{aligned} \tag{17}$$

where we used the fact that $C_w \Rightarrow A_w$. We will investigate these three terms now. First for all $\underline{x}$ let $\mathcal{B}_\epsilon^N(V|\underline{x}) \triangleq \{\underline{v} : (\underline{x},\underline{v}) \in \mathcal{B}_\epsilon^N(XV)\}$. Note that $M \ge 2^{N(I(X;V)+4\epsilon)}$. Then, see Gallager [16], p. 454,

$$\begin{aligned}\Pr\{\cap_{w=1}^M B_w^c\} &= \sum_{\underline{x}\in\mathcal{X}^N} p(\underline{x}) \prod_{w=1}^M \left(1 - \sum_{\underline{v}\in\mathcal{B}_\epsilon^N(V|\underline{x})} p(\underline{v})\right) \\ &\overset{(a)}{\le} \sum_{\underline{x}\in\mathcal{X}^N} p(\underline{x}) \left(1 - 2^{-N(I(X;V)+3\epsilon)} \sum_{\underline{v}\in\mathcal{B}_\epsilon^N(V|\underline{x})} p(\underline{v}|\underline{x})\right)^M \\ &\overset{(b)}{\le} \sum_{\underline{x}\in\mathcal{X}^N} p(\underline{x}) \Bigg(1 - \sum_{\underline{v}\in\mathcal{B}_\epsilon^N(V|\underline{x})} p(\underline{v}|\underline{x}) \\ &\qquad + \exp(-M2^{-N(I(X;V)+3\epsilon)})\Bigg) \\ &\le \sum_{(\underline{x},\underline{v})\notin\mathcal{B}_\epsilon^N(XV)} p(\underline{x},\underline{v}) + \sum_{\underline{x}\in\mathcal{X}^N} p(\underline{x})\exp(-2^{N\epsilon}) \overset{(c)}{\le} 2\epsilon,\end{aligned} \tag{18}$$

for N large enough. Here (a) follows from the fact that for $(\underline{x},\underline{v}) \in \mathcal{B}_\epsilon^N(XV)$, using Property 1,

$$p(\underline{v}) = p(\underline{v}|\underline{x})\frac{p(\underline{x})p(\underline{v})}{p(\underline{x},\underline{v})} \ge p(\underline{v}|\underline{x})2^{-N(I(X;V)+3\epsilon)}, \tag{19}$$

(b) from the inequality $(1-\alpha\beta)^M \le 1-\alpha+\exp(-M\beta)$, which holds for $0 \le \alpha,\beta \le 1$ and $M > 0$, and (c) from Property 2. Secondly we consider

$$\begin{aligned}&\Pr\left\{\left(\cup_{w=1}^M B_w\right)\cap C_W^c\right\} \\ &\le \max_{(\underline{x},\underline{v})\in\mathcal{B}_\epsilon^N(XV)} \Pr\{\underline{Z} \notin \mathcal{A}_\epsilon^N(Z|\underline{x},\underline{v}) | (\underline{V},\underline{X}) = (\underline{x},\underline{v})\} \\ &\overset{(d)}{\le} \epsilon.\end{aligned} \tag{20}$$

Here (d) follows directly from definition 3 of the set $\mathcal{B}_\epsilon^N(XV)$. Thirdly, for a fixed $\underline{z}$

$$\begin{aligned}\Pr\{\underline{V} \in \mathcal{A}_\epsilon^N(V|\underline{z})\} &= \sum_{\underline{v}\in\mathcal{A}_\epsilon^N(V|\underline{z})} p(\underline{v}|\underline{z})\frac{p(\underline{v})p(\underline{z})}{p(\underline{v},\underline{z})} \\ &\le 2^{-N(I(V;Z)-3\epsilon)} \sum_{\underline{v}\in\mathcal{A}_\epsilon^N(V|\underline{z})} p(\underline{v}|\underline{z}) \\ &\le 2^{-N(I(V;Z)-3\epsilon)}.\end{aligned} \tag{21}$$

From $M \le 2^{N(I(V;Z)-4\epsilon)}$ we now get for N large enough

$$\begin{aligned}\sum_{w\ne W}\Pr\{A_w\} &\le \sum_{w\ne W}\max_{\underline{y}}\Pr\{\underline{V} \in \mathcal{A}_\epsilon^N(V|\underline{z})\} \\ &\le \sum_{w\ne W} 2^{-N(I(V;Z)-3\epsilon)} \\ &\le M2^{-N(I(V;Z)-3\epsilon)} \le 2^{-N\epsilon} \le \epsilon.\end{aligned} \tag{22}$$

E. Last part achievability proof

In the ensemble of codes, for all N large enough, there now *exists* a a set of M sequences such that $P_{\mathcal{E}} \leq 2\epsilon + \epsilon + \epsilon = 4\epsilon$, as long as $I(V;X) + 4\epsilon \leq \frac{1}{N}\log_2 M \leq I(V;Z) - 4\epsilon$, for our fixed $0 < \epsilon < 1$. This follows from combining (24), (20), and (22). This finishes the proof of the lemma.

Let $D_{\max} \stackrel{\Delta}{=} \max_{x,y} D_{x,y}(x,y)$ then we obtain for the average semantic distortion of this code

$$\begin{aligned}\overline{D_{xy}} &\leq (1-P_{\mathcal{E}})(D_{\exp}+\epsilon) + P_{\mathcal{E}} D_{\max} \\ &\leq D_{\exp} + \epsilon + 4\epsilon D_{\max}, \end{aligned} \quad (23)$$

Now returning to the achievability proof we assume that $V = (M, U)$. If $I(M,U;X) < I(M,U;Z)$ there should be an $\epsilon > 0$ such that $I(M,U;X) + 4\epsilon \leq I(M,U;Z) - 4\epsilon$. The lemma implies for large enough N the existence of a code $(N, \overline{D_{xy}}, P_{\mathcal{E}})$ with $P_{\mathcal{E}} \leq 4\epsilon$ and $\overline{D_{xy}} \leq D_{\exp} + \epsilon + 4\epsilon D_{\max}$. Letting $\epsilon \downarrow 0$ proves the achievability part of the theorem.

Observe that we did not consider test-channels for which $I(M,U;X) = I(M,U;Z)$. Also for such a test-channel achievability can be proved, using the idea that this test-channel can be adapted a little bit without increasing the distortion too much. We will not work out this idea here.

IV. Converse

A. Mutual information part

Consider an $(N, \overline{D_{xy}}, P_{\mathcal{E}})$-code. From $H(M_1^N|X_1^N) = 0$, and $H(M_1^N|Z_1^N) \leq 1 + P_{\mathcal{E}}\log_2(|\mathcal{M}|^N)$ we obtain

$$0 \leq I(M_1^N;Z_1^N) - I(M_1^N;X_1^N) + 1 + P_{\mathcal{E}}\log_2(|\mathcal{M}|^N). \quad (24)$$

For the difference $I(M_1^N;Z_1^N) - I(M_1^N;X_1^N)$ we find

$$\begin{aligned} &I(M_1^N;Z_1^N) - I(M_1^N;X_1^N) \\ &= \sum_{n=1,N} \big[H(Z_n|Z_1^{n-1}) - H(Z_n|M_1^N, Z_1^{n-1}, X_{n+1}^N) \\ &\qquad -I(Z_n;X_{n+1}^N|M_1^N,Z_1^{n-1}) - I(M_1^N;X_n|X_{n+1}^N)\big] \\ &\stackrel{(a)}{=} \sum_{n=1,N} \big[H(Z_n|Z_1^{n-1}) - H(Z_n|M_1^N, Z_1^{n-1}, X_{n+1}^N) \\ &\qquad -I(X_n;Z_1^{n-1}|M_1^N,X_{n+1}^N) - I(M_1^N;X_n|X_{n+1}^N)\big] \\ &= \sum_{n=1,N} \big[H(Z_n|Z_1^{n-1}) - H(Z_n|M_1^N, Z_1^{n-1}, X_{n+1}^N) \\ &\qquad -H(X_n|X_{n+1}^N) + H(X_n|M_1^N,Z_1^{n-1},X_{n+1}^N)\big] \\ &\stackrel{(b)}{\leq} \sum_{n=1,N} \big[H(Z_n) - H(Z_n|M_1^N, Z_1^{n-1}, X_{n+1}^N) \\ &\qquad -H(X_n) + H(X_n|M_1^N,Z_1^{n-1},X_{n+1}^N)\big] \\ &\stackrel{(c)}{=} \sum_{n=1,N} \left[I(Z_n;M_n,V_n) - I(X_n;M_n,V_n)\right], \end{aligned} \quad (25)$$

with probability distribution

$$P(x_n,m_n,v_n,y_n,z_n) = P_s'(x_n,m_n)P(v_n,y_n|x_n)P_c(z_n|y_n) \quad (26)$$

for some $P(v_n,y_n|x_n)$ for $n = 1, N$. Here (a) follows from the "summation by parts"-lemma in Csiszar and Körner [17], (b) from the fact that $H(X_n|X_{n+1}^N) = H(X_n)$ since the source symbols are i.i.d., and (c) from the substitution

$$V_n \stackrel{\Delta}{=} (M_1^{n-1}, Z_1^{n-1}, X_{n+1}^N) \text{ for } n = 1, N, \quad (27)$$

where we should note that M_{n+1}^N is contained in X_{n+1}^N. We continue with

$$\begin{aligned} &\sum_{n=1,N} \left[I(Z_n;M_n,V_n) - I(X_n;M_n,V_n)\right] \\ &\stackrel{(d)}{=} N\left[I(Z;M,V|T) - I(X;M,V|T)\right] \\ &\stackrel{(e)}{\leq} N\left[H(Z) - H(Z|M,V,T) \right.\\ &\qquad \left. -H(X) + H(X|M,V,T)\right] \\ &\stackrel{(f)}{=} N\left[I(Z;M,U) - I(X;M,U)\right], \end{aligned} \quad (28)$$

with $P(x,m,u,y,z) = P_s'(x,m)P(u,y|x)P_c(z|y)$ for some $P(u,y|x)$. Here (d) follows from defining a time sharing-variable T, independent of all the other variables, assuming value $n \in \{1,2,\cdots,N\}$ with probability $1/N$, and $X = X_T$, $M = M_T$, $U = U_T$, $Y = Y_T$ and $Z = Z_T$, (e) from the fact that $H(X|T) = H(X)$ since the symbols X_n are assumed to be i.i.d., and (f) from the substitution $U = (V,T)$.

B. Distortion part

Next we study the average semantic distortion

$$\begin{aligned} \overline{D_{xy}} &= \sum_{x_1^N,y_1^N} P(x_1^N,y_1^N)\frac{1}{N}\sum_{n=1,N} D_{xy}(x_n,y_n) \\ &= \sum_{n=1}^N \frac{1}{N}\sum_{x_n,y_n} P(x_n,y_n)D_{xy}(x_n,y_n) \\ &= \sum_{x,y} P(x,y)D_{xy}(x,y), \end{aligned} \quad (29)$$

where $P(x,y) = \sum_{m,u,z} P_s'(x,m)P(u,y|x)P_c(z|y)$ for the same $P(u,y|x)$ that satisfies (28).

C. Last part converse

We now conclude that our $(N, \overline{D_{xy}}, P_{\mathcal{E}})$-code satisfies

$$\begin{aligned} 0 &\leq I(Z;M,U) - I(X;M,U) + \frac{1}{N} + P_{\mathcal{E}}\log_2(|\mathcal{M}|), \\ \overline{D_{xy}} &= \sum_{x,y} P(x,y)D_{xy}(x,y), \end{aligned} \quad (30)$$

for some

$$P(x,m,y,z) = P_s'(x,m)P(u,y|x)P_c(z|y). \quad (31)$$

This implies that for an achievable Δ_{xy} for any $\epsilon > 0$ and N large enough

$$\begin{aligned} 0 &\leq I(Z;M,U) - I(X;M,U) + \frac{1}{N} + \epsilon\log_2(|\mathcal{M}|), \\ \Delta_{xy} &\geq \overline{D_{xy}} - \epsilon = \sum_{x,y} P(x,y)D_{xy}(x,y) - \epsilon. \end{aligned}$$

for some $P(x,m,u,y,z) = P_s'(x,m)P(u,y|x)P_c(z|y)$. Letting $\epsilon \downarrow 0$ and $N \to \infty$, proves the converse part of the theorem.

D. Cardinality bounds

To find a bound on the cardinality of the auxiliary variable U let $\mathcal{D}$ be the set of probability distributions on $\mathcal{X} \times \mathcal{Y}$ and consider the $|\mathcal{X}||\mathcal{Y}|$ continuous functions of $P \in \mathcal{D}$ defined as

$$\begin{aligned} \phi_{xy}(P) &= P(x,y) \text{ for all but one pair } (x,y), \\ \phi_h(P) &= H_P(X|M) - H_P(Z|M). \end{aligned} \quad (32)$$

By the Fenchel-Eggleston strengthening of the Caratheodory lemma (see Wyner and Ziv [18]) there are $|\mathcal{X}||\mathcal{Y}|$ elements $P_u \in \mathcal{D}$ and α_u that sum to one, such that

$$\begin{aligned} P(x,y) &= \sum_{u=1,|\mathcal{X}||\mathcal{Y}|} \alpha_u \phi_{xy}(P_u) \text{ for} \\ & \quad \text{all but one pair } (x,y), \\ H(X|M,U) - H(Z|M,U) &= \sum_{u=1,|\mathcal{X}||\mathcal{Y}|} \alpha_u \phi_h(P_u). \end{aligned} \quad (33)$$

The entire probability distribution $\{P(x,y), x \in \mathcal{X}, y \in \mathcal{Y}\}$ and consequently the entropy $H(X)$ is now specified. Observe that now also the distribution $\{P(y), y \in \mathcal{Y}\}$ and therefore also $H(Z)$, and $\sum_{x,y} P(x,y) D_{xy}(x,y)$ are specified. Hence also $I(X;M,U) - I(Z;M,U)$. This implies that $|\mathcal{U}| = |\mathcal{X}||\mathcal{Y}|$ suffices.

V. Embedding watermarks correlated to the covertext

Suppose that we have an i.i.d. source S to which M is correlated (as in Yang and Sun [14]). The correlated sequence m_1^N now has to be embedded into s_1^N. The sequence y_1^N (semantic to s_1^N) is conveyed via a memoryless channel $\{P_c(z|y), \mathcal{Y}, \mathcal{Z}\}$. The decoder has to reconstruct m_1^N. The fundamental limit for this problem was determined by Yang and Sun [14]. We show here that this problem can also be cast into our "semantic partial transmission" setup. This leads to the characterization of the set of achievable distortions.

Take $X = (S, M)$ and let $\mu(X) = M$. The set of achievable distortion is now

$$\begin{aligned} &\mathcal{D} \triangleq \\ &\{\Delta_{sy} \; : \; \Delta_{sy} \geq \sum_{s,y} P(s,y) D_{sy}(s,y), \text{ for} \\ &\qquad P(s,m,u,y,z) = P_s(s,m) P_t(u,y|s,m) P_c(z|y) \\ &\qquad \text{for test channel } P_t(u,y|s,m) \text{ such} \\ &\qquad \text{that } I(M,U;Z) \geq I(M,U;S,M)\}. \end{aligned} \quad (34)$$

It turns out that this result also can be obtained from [14] if take an auxiliary random variable U that includes M. Inspection of the converse in [14] shows that this is justified, and consequently our characterization is more specific. For the U-cardinality bound we obtain that $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{M}||\mathcal{Y}|$.

VI. Conclusion

In this article we have proposed the semantic partial transmission setup. We could determine the set of achievable distortions for this situation. By introducing a typical set with some additional properties we were able to formulate an achievability proof based on weak instead of strong typicality. Strong typicality proofs are standard in semantic-coding and embedding problems, which are based on the (strong typicality) Gelfand-Pinsker proof. We have investigated the connection between semantic partial transmission and embedding a watermark that is correlated to the cover-text. Yang and Sun [14] determined the fundamental limit for this embedding setup. We could refine their formulation using our results for semantic partial transmission.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.

[2] F.M.J. Willems and T. Kalker, "Semantic Compaction, Transmission, and Compression Codes," *Proc. IEEE ISIT,* Adelaide, South Australia, Australia, 4-9 September, 2005. CD-ROM, pp. 214-218.

[3] S. Gelfand and M. Pinsker, "Coding for a Channel with Random Parameters", *Prob. of Control and Inform. Th.,* vol. 9, pp. 19-31, 1980.

[4] J. Wolfowitz, *"Coding theorems of information theory"* , Berlin, Springer-Verlag, 1961. Third edition 1978.

[5] T. Berger, "Multiterminal Source Coding," *The Information Theory Approach to Communications,* CISM Courses and Lectures, G. Longo, Ed. Springer-Verlag, 1978, vol. 229, pp. 171-231.

[6] G.D. Forney, Jr., *Information Theory,* course notes (unpublished), Stanford University, 1972.

[7] T.M. Cover, "A Proof of the Data Compression Theorem of Slepian and Wolf for Ergodic Sources," *IEEE Trans. Inform. Th.,* vol. IT - 21, pp. 226 - 228, March 1975.

[8] T. Kalker and F.M.J. Willems, "Capacity Bounds and Constructions for Reversible Data-hiding," *Proc. Int. Conf. DSP,* Santorini, Greece, July 1-3, 2002.

[9] T. Kalker and F.M.J. Willems, "Capacity Bounds and Code Constructions for Reversible data-hiding," *IS & T / SPIE's 15th Ann. Symp. Electronic Imaging,* January 20-24, 2003. Santa Clara, California. CDROM.

[10] F.M.J. Willems and T. Kalker, "Methods for Reversible Embedding," *Proc. 40th Ann. Allerton Conf. Communication, Control, and Computing,* Allerton House, Monticello, Illinois, Oct. 2-4, 2002. CDROM.

[11] F.M.J. Willems and T. Kalker, "Coding Theorems for Reversible Embedding," *DIMACS Series in Discrete Math. and Theoretical Comp. Sc.,* vol. 66, American Mathematical Society, pp. 61 - 76, 2004.

[12] A. Sutivong, T.M. Cover, and M. Chiang, "Trade-off Between Message and State Information Rates," *Proc. ISIT 2001,* Washington-DC, June 24 - 29, 2001, p. 303.

[13] A. Sutivong, T.M. Cover, M. Chiang, and Young-Han Kim, "Rate vs. Distortion Trade-off for Channels with State Information," *Proc. ISIT 2002,* Lausanne, Switzerland, June 30 - July 5, 2002, p. 226.

[14] En-hui Yang and Wei Sun, "On Information Embedding When Watermark and Covertext are Correlated," *ISIT 2006,* Seattle, USA, July 9-14, 2006, pp. 346- 350.

[15] T.M. Cover and Y. Thomas, *"Elements of Information Theory,"* Wiley & Sons, New York, 1991. Second edition, 2006.

[16] R.G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[17] I. Csiszar and J. Korner, "Broadcast Channels with Confidential Messages," *IEEE Trans. Inform. Th.,* vol. IT-24, pp. 339-348, May 1978.

[18] A.D. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. Inform. The.,* vol. IT-22, no. 1, pp. 1-10, Jan. 1976.