Distributed and Cascade Lossy Source Coding with a Side Information "Vending Machine"

Behzad Ahmadi
CWCSPR, ECE Dept.
New Jersey Institute of Technology
Newark, NJ, 07102, USA
Email: behzad.ahmadi@njit.edu

Osvaldo Simeone
CWCSPR, ECE Dept.
New Jersey Institute of Technology
Newark, NJ, 07102, USA
Email: osvaldo.simeone@njit.edu

***Abstract*—Source coding with a side information "vending machine" is a recently proposed framework in which the statistical relationship between the side information available at the decoder and the source sequence can be controlled by the decoder based on the message received from the encoder. In this paper, the characterization of the optimal rate-distortion performance as a function of the cost associated with the control actions is extended from the previously studied point-to-point set-up to two multiterminal models. First, a distributed source coding model is studied, in which two encoders communicate over rate-limited links to a decoder, whose side information can be controlled based on the control actions selected by one of the encoders. The rate-distortion-cost region is characterized under the assumption of lossless reconstruction of the source encoded by the node that does not control the side information. Then, a three-node cascade scenario is investigated, in which the last node has controllable side information. The rate-distortion-cost region is derived for general distortion requirements and under the assumption of "causal" availability of side information at the last node.**

Keywords: Distributed source coding, cascade source coding, observation costs, side information, rate-distortion theory.

I. INTRODUCTION

Reference [1] introduced the notion of a side information "vending machine". In this framework, unlike the conventional Wyner-Ziv set-up, the joint distribution of the side information Y available at the decoder and of the source X observed at the encoder can be controlled through the selection of an "action" A. Action A is selected by the decoder based on the message M, of R bits per source symbol, received from the encoder, and is subject to a cost constraint. The performance of the system is thus expressed in terms of the interplay among three metrics, namely the rate R, the cost budget Γ on the action A, and the distortion D of the reconstruction $\hat{X}$ at the decoder. The *rate-distortion-cost* function $R(D,\Gamma)$ is derived in [1] for the case in which the side information Y is available "non-causally" to the decoder, as in the standard Wyner-Ziv model, and in the case in which it is available "causally", as introduced in [2].

Recent works [3] and [4] generalized the characterization of the rate-distortion-cost function in [1] to a multi-terminal set-up analogous to the so called Heegard-Berger problem [5], in which the side information vending machine may or may not be available at the decoder. This entails the presence of *two decoders*, one with access to the vending machine and one without any side information. Reference [4] also solved the more general case in which both decoders have access to the same vending machine, and either the side informations produced by the vending machine at the two decoders satisfy a degradedness condition, or lossless source reconstructions are required at the decoders. Instead, the work [3] considered a set-up that generalizes the Heegard-Berger problem mentioned above by allowing for functional reconstructions of the source X and of an additional sequence measured only through the vending machine at the decoder. A further related work is [6], where the model in [1] is extended to include secrecy constraints.

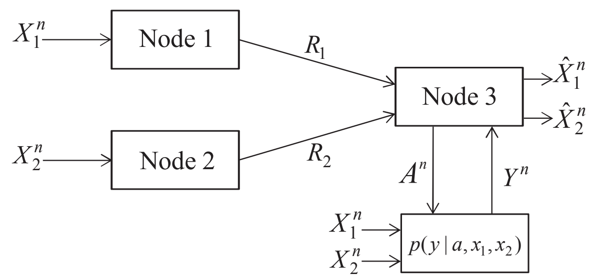

Figure 1. Distributed source coding with a side information vending machine at the decoder. Side information is assumed to be available "non-causally" to the decoder.

Contributions: In this paper, we study two multi-terminal extensions of the set-up of [1], namely the *distributed source coding* setting of Fig. 1 and the cascade model of Fig. 2. In the *distributed source coding* setting of Fig. 1, two encoders (Node 1 and Node 2), which measure correlated sources X_1 and X_2, respectively, communicate over rate-limited links, of rates R_1 and R_2, to a single decoder (Node 3). The action sequence controlling the side information at Node 3 is selected by Node 3 based on the message M_1 (of rate R_1) received from Node 1. In Sec. II, we characterize the set $\mathcal{R}(D_1,\Gamma)$ of all achievable rates (R_1,R_2) for a given distortion constraint D_1 on the reconstruction[1] $\hat{X}_1$ and an action cost constraint of Γ, under the requirement that source X_2 must be decoded (near)

[1]Reconstruction $\hat{X}_1$ may be (a lossy version of) an arbitrary function of sources X_1 and X_2.

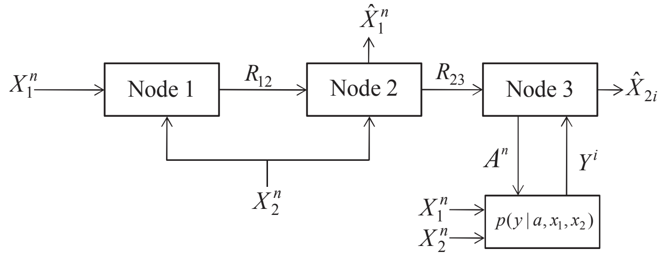

Figure 2. Cascade source coding with a side information vending machine. Side information is assumed to be available "causally" to the decoder.

losslessly at the decoder. We also provide a numerical example to obtain further insight into the role of control information in achieving optimal performance. This result generalizes the classical result of [7] for distributed source coding with "one distortion criterion".

In the *cascade* model of Fig. 2, Node 1 is connected via a rate-limited link, of rate R_{12}, to Node 2, which is in turn communicates with Node 3 with rate R_{23}. Source X_1 is measured by Node 1 and the correlated source X_2 by both Node 1 and Node 2. Node 3 has side information Y, which can be controlled via an action A selected by Node 3 based on the message received from Node 2. In Sec. III, we derive the set $\mathcal{R}(D_1, D_2, \Gamma)$ of all achievable rates (R_{12}, R_{23}) for given distortion constraints (D_1, D_2) on the reconstructions $\hat{X}_1$ and $\hat{X}_2$ at Node 2 and Node 3, respectively,[2] and for cost constraint Γ. This characterization is obtained under the assumption that the side information Y be available causally at Node 3. This result extends the characterization of the rate-distortion achievable performance for the cascade model studied in [9] to the set-up at hand with a side information vending machine at Node 3.

II. Distributed Source Coding with a Side Information Vending Machine

In this section, we first detail the system model for the setting of Fig. 1 of distributed source coding with a vending machine. We then present the characterization of the corresponding rate-distortion-cost performance in Sec. II-B. An example follows in Sec. II-C.

A. System Model

The problem of distributed lossy source coding with a vending machine and non-causal side information is defined by the probability mass functions (pmfs) $p_{X_1X_2}(x_1, x_2)$ and $p_{Y|AX_1X_2}(y|a, x_1, x_2)$ and discrete alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, \mathcal{A}$ and $\hat{\mathcal{X}}_1$ as follows. The source sequences X_1^n and X_2^n with $X_1^n \in \mathcal{X}_1^n$ and $X_2^n \in \mathcal{X}_2^n$, respectively, are such that the pairs (X_{1i}, X_{2i}) for $i \in [1, n]$ are independent and identically distributed (i.i.d.) with joint pmf $p_{X_1X_2}(x_1, x_2)$. Node 1 measures sequences X_1^n and encodes it into message M_1 of nR_1 bits, while Node 2 measures sequences X_2^n and encodes it into message M_2 of nR_2 bits. Node 3 wishes to reconstruct the two sources within given distortion requirements, to be discussed below, as $\hat{X}_1^n \in \hat{\mathcal{X}}_1^n$ and $\hat{X}_2^n \in \hat{\mathcal{X}}_2^n$.

To this end, Node 3 selects an action sequence A^n, where $A^n \in \mathcal{A}^n$, based on the message M_1 received from Node 1. The side information sequence Y^n is then realized as the output of a memoryless channel with inputs (A^n, X_1^n, X_2^n). Specifically, given A^n, X_1^n and X_2^n, the sequence Y^n is distributed as

$$p(y^n|a^n, x_1^n, x_2^n) = \prod_{i=1}^{n} p_{Y|AX_1X_2}(y_i|a_i, x_{1i}, x_{2i}). \quad (1)$$

The overall cost of an action sequence a^n is defined by a per-symbol cost function Λ: $\mathcal{A} \rightarrow [0, \Lambda_{\max}]$ with $0 \leq \Lambda_{\max} < \infty$. The estimated sequences $\hat{X}_1^n$ and $\hat{X}_2^n$ are obtained as a function of both messages M_1 and M_2 and of the side information Y^n. The estimates $\hat{X}_1^n$ and $\hat{X}_2^n$ are constrained to satisfy distortion constraints defined by two per-symbol distortion metrics, namely $d_1(x_1, x_2, y, \hat{x}_1)$: $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \hat{\mathcal{X}}_1 \rightarrow [0, D_{\max}]$ with $0 \leq D_{\max} < \infty$ and $d_2(x_2, \hat{x}_2) = d_H(x_2, \hat{x}_2)$: $\mathcal{X}_2 \times \mathcal{X}_2 \rightarrow \{0, 1\}$. Note that, while metric $d_1(x_1, x_2, y, \hat{x}_1)$ is arbitrary, metric $d_2(x_2, \hat{x}_2)$ is assumed to be the *Hamming* distortion $d_H(x_2, \hat{x}_2)$, where $d_H(x_2, \hat{x}_2) = 0$ if $x_2 = \hat{x}_2$ and $d_H(x_2, \hat{x}_2) = 1$ otherwise.

More formally, an $(n, R_1, R_2, D_1, D_2, \Gamma)$ code for the set-up of Fig. 1 consists of two source encoders

$$\begin{aligned} &\mathrm{g}_1\colon \mathcal{X}_1^n \rightarrow [1, 2^{nR_1}], \\ \text{and } &\mathrm{g}_2\colon \mathcal{X}_2^n \rightarrow [1, 2^{nR_2}], \end{aligned} \quad (2)$$

which map the sequences X_1^n and X_2^n into messages M_1 and M_2, respectively; an "action" function

$$\ell\colon [1, 2^{nR_1}] \rightarrow \mathcal{A}^n, \quad (3)$$

which maps the message M_1 into an action sequence A^n; and a decoding function

$$\mathrm{h}\colon [1, 2^{nR_1}] \times [1, 2^{nR_2}] \times Y^n \rightarrow \hat{\mathcal{X}}_1^n \times \mathcal{X}_2^n, \quad (4)$$

which maps the messages M_1 and M_2, and the side information sequence Y^n into the estimated sequences $\hat{X}_1^n$ and $\hat{X}_2^n$; such that the action cost constraint Γ is satisfied as

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\Lambda(A_i)\right] \leq \Gamma, \quad (5)$$

and the distortion constraints D_1 and D_2 hold, namely

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[d_1(X_{1i}, X_{2i}, Y_i, \hat{X}_{1i})\right] \leq D_1 \quad (6a)$$

$$\text{and } \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[d_H(X_{2i}, \hat{X}_{2i})\right] = \frac{1}{n}\sum_{i=1}^{n} \Pr[\hat{X}_{2i} \neq X_{2i}] \leq D_2. \quad (6b)$$

Given a distortion-cost tuple (D_1, D_2, Γ), a rate pair (R_1, R_2) is said to be achievable if, for any $\epsilon > 0$ and sufficiently large n, there exists a $(n, R_1, R_2, D_1+\epsilon, D_2+\epsilon, \Gamma+\epsilon)$

[2]Reconstructions $\hat{X}_1$ and $\hat{X}_2$ may be arbitrary functions of both X_1 and X_2.

code. The *rate-distortion-cost region* $\mathcal{R}(D_1, D_2, \Gamma)$ is defined as closure of all rate pairs (R_1, R_2) that are achievable given the distortion-cost tuple (D_1, D_2, Γ). We focus on characterizing the rate-distortion-cost function $\mathcal{R}(D_1, \Gamma)$, which is defined as $\mathcal{R}(D_1, \Gamma) = \mathcal{R}(D_1, 0, \Gamma)$, that is, we impose the constraint $\frac{1}{n}\sum_{i=1}^{n} \Pr[\hat{X}_{2i} \neq X_{2i}] \to 0$ for $n \to \infty$.

B. Rate-Distortion-Cost Region

In this section, a single-letter characterization of the rate-distortion-cost region $\mathcal{R}(D_1, \Gamma)$ is derived.

Proposition 1. *The rate-distortion-cost region $\mathcal{R}(D_1, \Gamma)$ for the model in Fig. 1 is given by union of the set of all of rate tuples (R_1, R_2) that satisfy the inequalities*

$$R_1 \geq I(X_1; A|Q) + I(X_1; V|A, X_2, Y, Q) \quad \text{(7a)}$$

$$R_2 \geq H(X_2|A, Y, V, Q) \quad \text{(7b)}$$

$$\text{and } R_1 + R_2 \geq I(X_1; A|Q) + H(X_2|A, Y, Q) + I(X_1; V|A, X_2, Y, Q), \quad \text{(7c)}$$

for some joint pmfs that factorizes as

$$p(q, x_1, x_2, y, v, a, \hat{x}_1) = p(q)p(x_1, x_2)p(a, v|x_1, q) \cdot p(y|a, x_1, x_2)\delta(\hat{x}_1 - \hat{\mathrm{x}}_1(v, y, q)), \quad \text{(8)}$$

with pmfs $p(q)$ and $p(a, v|x_1, q)$ and deterministic function $\hat{\mathrm{x}}_1(v, y, q)$, such that the action cost and the distortion constraints

$$\mathrm{E}\left[\Lambda(A)\right] \leq \Gamma \quad \text{(9)}$$

$$\text{and } \mathrm{E}\left[d_1(X_1, X_2, Y, \hat{X}_1)\right] \leq D_1 \quad \text{(10)}$$

hold. Finally, Q and V are auxiliary random variables whose alphabet cardinality can be constrained as $|\mathcal{Q}| \leq 6$ and $|\mathcal{V}| \leq 6\,|\mathcal{X}_1|\,|\mathcal{A}| + 3$ without loss of optimality.

Remark 2. If we set $p(y|a, x_1, x_2) = p(y|x_1, x_2)$, Proposition 1 reduces to [7, Theorem 1]. If, instead, X_1 is independent of X_2, the minimum rate R_1, given by the right-hand side of (7a), recovers [1, Theorem 1].

For the proof of converse, we refer to [8]. As for achievability, the scheme at hand combines the distributed Wyner-Ziv approach of [10, Theorem II] with the scheme proposed in [1, Sec. II-B]. Specifically, Node 1 first maps the input sequence X_1^n into an action sequence A^n, so that the two sequences are jointly typical. Conveying sequence A^n to the receiver requires $I(X_1; A)$ bits per source sample, as follows easily from standard rate-distortion theory results. The sequences (A^n, Y^n) are now regarded as side information available at the decoder. Based on this, the distributed Wyner-Ziv scheme proposed in [10, Theorem 2] is used to convey an auxiliary codeword V^n from Node 1 and sequence X_2^n from Node 2[3]. Note that the fact that sequences (A^n, Y^n) are not i.i.d. does not affect achievability of the rate region derived in [10]. Finally, the decoder estimates $\hat{X}_1^n$ sample by sample by using function $\hat{\mathrm{x}}_1(v, y, q)$ as $\hat{X}_{1i} = \hat{\mathrm{x}}_1(V_i, Y_i, Q_i)$.

We remark that an extension of Proposition 1 to any number of encoders can be found in [8].

C. A Binary Example

We now focus on a specific example in order to illustrate the result derived in Proposition 1. Specifically, we assume that all alphabets are binary and that (X_1, X_2) is a doubly symmetric binary source (DSBS) characterized by probability p, with $0 \leq p \leq 1/2$, so that $p(x_1) = p(x_2) = 1/2$ for $x_1, x_2 \in \{0, 1\}$ and $\Pr[X_1 \neq X_2] = p$. Moreover, the decoder wishes to reconstruct both X_1 and X_2 losslessly in the sense discussed above. This implies that we set $d_1(x_1, x_2, y, \hat{x}_1) = d_H(x_1, \hat{x}_1)$ and $D_1 = 0$. The side information Y_i is such that

$$Y_i = \begin{cases} \mathrm{f}(X_{1i}, X_{2i}) & \text{if } A_i = 1 \\ 1 & \text{if } A_i = 0 \end{cases}, \quad \text{(11)}$$

where $\mathrm{f}(x_1, x_2)$ is a deterministic function to be specified. Therefore, when a unitary action, $A_i = 1$, is selected, then $Y_i = \mathrm{f}(X_{1i}, X_{2i})$ is measured at the receiver, while with $A_i = 0$ no useful information is collected by the decoder. The action sequence A^n must satisfy the cost constraint (5), where the cost function is defined as $\Lambda(A_i) = 1$ if $A_i = 1$ and $\Lambda(A_i) = 0$ if $A_i = 0$. It follows that, given (11), a cost Γ implies that the decoder can observe $\mathrm{f}(X_{1i}, X_{2i})$ only for at most $n\Gamma$ symbols. As for the function $\mathrm{f}(x_1, x_2)$, we consider two cases, namely $\mathrm{f}(x_1, x_2) = x_1 \oplus x_2$, where $\oplus$ is the binary sum and $\mathrm{f}(x_1, x_2) = x_1 \odot x_2$, where $\odot$ is the binary product.

Under the requirement of lossless reconstruction for both X_1 and X_2 (i.e., $D_1 = 0$ along with $D_2 = 0$), it can be easily shown from Proposition 1 that the minimum *sum-rate* $R_1 + R_2$ for a given cost constraint Γ, which we denote as $R_{sum}(\Gamma)$ is given by the right-hand side of (7c) with $V = X_1$, namely[4]

$$R_{sum}(\Gamma) = \min I(X_1; A) + H(X_1, X_2|A, Y), \quad \text{(12)}$$

where the mutual informations are calculated with respect to the distribution $p(x_1, x_2, y, a) = p(x_1, x_2)p(a|x_1)p(y|a, x_1, x_2)$, and the minimum is taken over all distributions $p(a|x_1)$ such that $\mathrm{E}\left[\Lambda(A)\right] = \mathrm{E}\left[A\right] \leq \Gamma$. Note that, by its definition, function $R_{sum}(\Gamma)$ is non-increasing for all $\Gamma \geq 0$ (and constant for $\Gamma \geq 1$) so that in particular $R_{sum}(1) = \min_{\Gamma \geq 0} R_{sum}(\Gamma)$. Given the function $\mathrm{f}(x_1, x_2)$, evaluation of (12) requires solving a simple convex optimization problem. We do not provide a more explicit expression here, as it can be easily derived. Instead, we discuss some numerical results for the two functions $\mathrm{f}(x_1, x_2)$ mentioned above, namely (binary) sum and product.

To start with, we evaluate the sum-rate (12) $R_{sum}(1)$, which provides the minimum value of $R_{sum}(\Gamma)$ over Γ, as discussed above. With $\Gamma = 1$, it is clearly optimal to set $A = 1$, irrespective of the value of X_1. It is not difficult to see that

[3]More precisely, since A^n is known to Node 1 as well, the codebook used to map X_1^n into V^n is generated conditioned on A^n.

[4]The entire rate region $\mathcal{R}(0, \Gamma)$ also follows immediately by setting $V = X_1$.

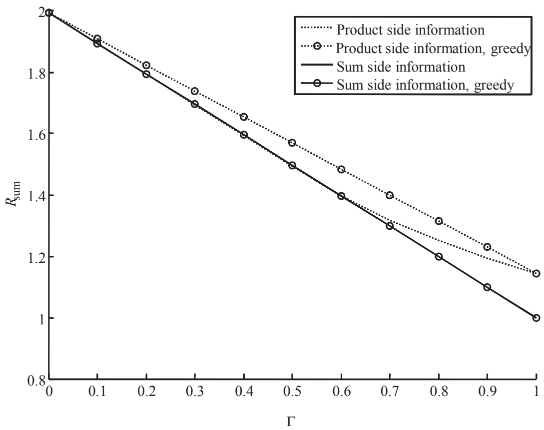

Figure 3. Sum-rates (12) and (14) versus the action cost Γ for sum and product side informations ($p = 0.4$).

we have $R^{\oplus}_{sum}(1) = 1$ for the sum side information, while for the product side information we obtain

$$R^{\odot}_{sum}(1) = H\left(\frac{1-p}{1+p}, \frac{p}{1+p}, \frac{p}{1+p}\right)\left(\frac{1+p}{2}\right), \quad (13)$$

where we have used the definition $H(p_1, p_2, ..., p_k) = -\sum_{i=1}^{k} p_i \log_2 p_i$. By comparing these two expressions, it can be seen that, if p is sufficiently small, namely if $p \lesssim 0.33$, we have $R^{\odot}_{sum}(1) < R^{\oplus}_{sum}(1)$ and thus product side information is more informative than the sum, while for $p \gtrsim 0.33$ the opposite is true (and for $p = 1$, they are equally informative).

We now evaluate the sum-rate (12) for a general cost budget $0 \leq \Gamma \leq 1$ for both sum and product side information. We are interested in emphasizing the role of data and control information in achieving the optimal sum-rate (12). To this end, we compare (12) with the performance attainable by imposing that the action A be selected by Node 3 a priori, that is, without any control from Node 1. The sum-rate attainable by such a scheme, which is referred to as "greedy", following [1], can be easily seen to be given by[5]

$$R_{sum,\ greedy}(\Gamma) = \Gamma H(X_1, X_2|Y) + (1-\Gamma)(1 + H(p)). \quad (14)$$

We use, as above, the notations $R^{\oplus}_{sum,\ greedy}(\Gamma)$ and $R^{\odot}_{sum,\ greedysum}(\Gamma)$ for (14) as evaluated with sum and product side informations.

A first observation is that, with sum side information, we have that (see [8] for details)

$$R^{\oplus}_{sum,\ greedy}(\Gamma) = R^{\oplus}_{sum}(\Gamma). \quad (15)$$

This shows that a "greedy" approach, in which only data information is conveyed by Node 1, is optimal. Instead, this is not the case with product side information and we generally have $R^{\odot}_{sum,\ greedy}(\Gamma) \geq R^{\odot}_{sum}(\Gamma)$, where inequality can be strict (unless, of course, $\Gamma = 1$). The reason is that, if $X_1 = 0$, the value of the side information is always $Y = X_1 \odot X_2 = 0$ irrespective of the value of X_2. Therefore, if $X_1 = 0$, the side information is less informative than if $X_1 = 1$ and thus it may be advantageous to save on the action cost by setting $A = 0$. Consequently, choosing actions based on the message received from Node 1 can result in a lower sum-rate.

To illustrate the discussion above, Fig. 3 depicts the sum-rates (12) and (14) versus the action cost Γ for $p = 0.4$. It can be seen that, for sufficiently large probability p (here, $p = 0.4$), while a product side information is less advantageous that sum side information for $\Gamma = 1$, as per the discussion above, this may not be the case for smaller costs. Moreover, in this case, the greedy approach suffers from a significant performance loss for product side information.

[5]It can be obtained from (12) by setting $p(a|x_1) = \Gamma$ for $a = 1$ irrespective of the value of X_1.

III. Cascade Source Coding with a Side Information Vending Machine

In this section, we first describe the system model for the setting of Fig. 2 of cascade source coding with a side information vending machine. We recall that side information Y is here assumed to be available causally at the decoder (Node 3). We then present the characterization of the corresponding rate-distortion-cost performance in Sec. III-B.

A. System Model

The problem of cascade lossy computing with causal observation costs at second user is defined by the pmfs $p_{X_1X_2}(x_1, x_2)$ and $p_{Y|AX_1X_2}(y|a, x_1, x_2)$ and discrete alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, \mathcal{A}, \hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2$. The source sequences X_1^n and X_2^n with $X_1^n \in \mathcal{X}_1^n$ and $X_2^n \in \mathcal{X}_2^n$, respectively, are such that the pairs (X_{1i}, X_{2i}) for $i \in [1, n]$ are i.i.d. with joint pmf $p_{X_1X_2}(x_1, x_2)$. Node 1 measures sequences X_1^n and X_2^n and encodes them in a message M_{12} of nR_{12} bits, which is delivered to Node 2. Node 2 estimates a sequence $\hat{X}_1^n \in \hat{\mathcal{X}}_1^n$ within given distortion requirements. Moreover, Node 2 encodes the message M_{12}, received from Node 1, and the locally available sequence X_2^n in a message M_{23} of nR_{23} bits, which is delivered to node 3.

Node 3 wishes to estimate a sequence $\hat{X}_2^n \in \hat{\mathcal{X}}_2^n$ within given distortion requirements. To this end, Node 3 receives message M_{23} and based on this, selects an action sequence A^n, where $A^n \in \mathcal{A}^n$. The action sequence affects the quality of the measurement Y^n of sequence X_1^n and X_2^n obtained at the Node 3. Specifically, given A^n, X_1^n and X_2^n, the sequence Y^n is distributed as in (1). The estimated symbol $\hat{X}_{2i}$ with $\hat{X}_{2i} \in \hat{\mathcal{X}}_2$ is then obtained as a function of M_{23} and Y^i for $i \in [1, n]$. Estimated sequences $\hat{X}_j^n$ for $j = 1, 2$ must satisfy distortion constraints defined by functions $d_j(x_1, x_2, y, \hat{x}_j)$: $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y} \times \hat{\mathcal{X}}_j \to [0, D_{\max}]$ with $0 \leq D_{\max} < \infty$ for $j = 1, 2$, respectively.

The formal description of an $(n, R_{12}, R_{23}, D_1, D_2, \Gamma)$ code for the set-up of Fig. 2 can be constructed similar to Sec. 2 following the discussion above and is fully detailed in [8]. Here we remark that the "action" function at Node 3

$$\ell\colon [1, 2^{nR_{23}}] \to \mathcal{A}^n, \quad (16)$$

maps the message M_{23} into an action sequence A^n; and that we have the decoding function at Node 2

$$\mathrm{h}_1\colon [1, 2^{nR_{12}}] \times \mathcal{X}_2^n \rightarrow \hat{\mathcal{X}}_1^n, \tag{17}$$

which maps the message M_{12} and the measured sequence X_2^n into the estimated sequence $\hat{X}_1^n$; and a sequence of decoding functions at Node 3

$$\mathrm{h}_{2i}\colon [1, 2^{nR_{23}}] \times \mathcal{Y}^i \rightarrow \hat{\mathcal{X}}_2, \tag{18}$$

for $i \in [1, n]$ which maps the message M_{23} and the measured sequence Y^i into the ith estimated symbol $\hat{X}_{2i} = \mathrm{h}_{2i}(M_{23}, Y^i)$. We also note that the action cost constraint (5) and distortion constraints D_j

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[d_j(X_{1i}, X_{2i}, Y_i, \hat{X}_{ji})\right] \leq D_j \text{ for } j = 1, 2, \tag{19}$$

must be satisfied. Achievability and the *rate-distortion-cost region* $\mathcal{R}(D_1, D_2, \Gamma)$ are defined similar to Sec. 2.

B. Rate-Distortion-Cost Region

We have the following characterization of the rate-distortion-cost region.

Proposition 3. *The rate-distortion-cost region $\mathcal{R}(D_1, D_2, \Gamma)$ for the set-up of Fig. 2 is given by the union of all rate pairs (R_{12}, R_{23}) satisfying the inequalities*

$$R_{12} \geq I(X_1; U, A, \hat{X}_1|X_2) \tag{20a}$$

$$\text{and } R_{23} \geq I(X_1, X_2; U, A), \tag{20b}$$

for some joint pmf that factorizes as

$$p(x_1, x_2, y, a, u, \hat{x}_1, \hat{x}_2) = p(x_1, x_2)p(a, u, \hat{x}_1|x_1, x_2) \cdot p(y|a, x_1, x_2)\delta(\hat{x}_2 - \mathrm{x}_2(u, y)), \tag{21}$$

with pmf $p(a, u, \hat{x}_1|x_1, x_2)$ and deterministic function $\hat{\mathrm{x}}_2(u, y)$, such that the action and the distortion constraints

$$\mathrm{E}\left[\Lambda(A)\right] \leq \Gamma \tag{22}$$

$$\text{and } \mathrm{E}[d_j(X_1, X_2, Y, \hat{X}_j)] \leq D_j, \text{for } j = 1, 2, \tag{23}$$

respectively, hold. Finally, U is an auxiliary random variable whose alphabet cardinality can be constrained as $|\mathcal{U}| \leq |\mathcal{X}_1||\mathcal{X}_2| + 4$, without loss of optimality.

Remark 4. If $p(y|a, x_1, x_2) = p(y|x_1, x_2)$, Proposition 3 reduces to [9, Theorem 1].

The proof of converse is provided in [8]. The coding strategy that proves achievability is a combination of the techniques proposed in [1] and [9, Theorem 1]. Specifically, Node 1 first maps sequences X_1^n and X_2^n into the action sequence A^n and an auxiliary codeword U^n using the standard joint typicality criterion. This mapping operation requires a codebook of rate $I(X_1, X_2; U, A)$. Then, given the so obtained sequences A^n and U^n, source sequences X_1^n and X_2^n are further mapped into the estimate $\hat{X}_1^n$ for Node 2 so that the sequences $(X_1^n, X_2^n, A^n, U^n, \hat{X}_1^n)$ are jointly typical. This requires rate $I(X_1, X_2; \hat{X}_1|U, A)$. Leveraging the side information X_2^n available at Node 2, conveying the codewords A^n, $\hat{X}_1^n$ and U^n to Node 2 requires rate $I(X_1, X_2; U, A) + I(X_1, X_2; \hat{X}_1|U, A) - I(U, A, \hat{X}_1; X_2)$, which equals the right-hand side of (20a). Node 2 conveys U^n and A^n to Node 3 by simply forwarding the index received from Node 1 (of rate $I(X_1, X_2; U, A)$). Finally, Node 3 estimates $\hat{X}_2^n$ through a symbol-by-symbol function as $\hat{X}_{2i} = \hat{\mathrm{x}}_2(U_i, Y_i)$ for $i \in [1, n]$.

IV. Concluding Remarks

As a concluding remark, one aspect worth emphasizing of the two problems solved in this paper concerns the way the side information is assumed to be available at the decoder. For distributed source coding, we have in fact assumed that side information is available in a non-causal fashion in the conventional sense of the Wyner-Ziv problem. Adaptation of the results given here to a model where side information is available only causally, in the sense of [2], proved challenging and is left open by this work. On the contrary, for the cascade/triangular model, we have assumed causal side information at the decoder. In this case, adaption of the given results to the set-up of non-causal side information proved difficult, and is again left as an open problem.

V. Acknowledgement

This work was supported in part by the U.S. National Science Foundation under Grant No. 0914899.

References

[1] H. Permuter and T. Weissman, "Source coding with a side information "vending machine"," *IEEE Trans. Inf. Theory*, vol. 57, pp. 4530–4544, Jul 2011.

[2] T. Weissman and A. El Gamal, "Source coding with limited-look-ahead side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5218–5239, Dec. 2006.

[3] B. Ahmadi and O. Simeone, "Robust coding for lossy computing with receiver-side observation costs," in *Proc. IEEE International Symposium on Information Theory* (ISIT 2011), July 31-Aug. 5, Saint Petersburg, Russia, 2011 (see also arXiv:1108.1535).

[4] Y. Chia, H. Asnani, and T. Weissman, "Multi-terminal source coding with action dependent side information," in *Proc. IEEE International Symposium on Information Theory* (ISIT 2011), July 31-Aug. 5, Saint Petersburg, Russia, 2011.

[5] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727-734, Nov. 1985.

[6] K. Kittichokechai, T. J. Oechtering and M. Skoglund, "Secure Source Coding with Action-dependent Side Information," in *Proc. IEEE International Symposium on Information Theory* (ISIT 2011), July 31-Aug. 5, Saint Petersburg, Russia, 2011.

[7] T. Berger and R. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. Inform. Theory*, vol. 35, pp. 228–236, Mar 1989.

[8] B. Ahmadi and O. Simeone, "Distributed and Cascade Lossy Source Coding with a Side Information "Vending Machine"," submitted [arXiv:1109.6665].

[9] Y.-K. Chia and T. Weissman, "Cascade and triangular source coding with causal side information," in *Proc. IEEE International Symposium on Information Theory* (ISIT 2011), July 31-Aug. 5, Saint Petersburg, Russia, 2011.

[10] M. Gastpar, "The Wyner–Ziv problem with multiple sources," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2762–2767, Nov. 2004.