Two-Way Communication with Adaptive Data Acquisition

Behzad Ahmadi and Osvaldo Simeone
CWCSPR, NJIT
Newark, NJ, 07102, USA
Email: {behzad.ahmadi,osvaldo.simeone}@njit.edu

***Abstract*—A bidirectional link between two nodes, Node 1 and Node 2, is studied in which Node 2 is able to acquire information from the environment, e.g., via access to a remote data base or via sensing. Information acquisition is expensive in terms of system resources, e.g., time, bandwidth and energy, and thus should be done efficiently. As a result of the forward communication from Node 1 to Node 2, the latter wishes to obtain information about the the data available at Node 1 and of the data obtained from the environment. The forward link is, however, also used by Node 1 to query Node 2 with the aim of retrieving information from the environment on the backward link. The problem is formulated using the concept of side information "vending machine", and the optimal trade-off between communication rates, distortions of the estimates produced at the two nodes and costs for information acquisition at Node 2 is derived.**

***Index Terms*—Source coding, side information, interactive communication.**

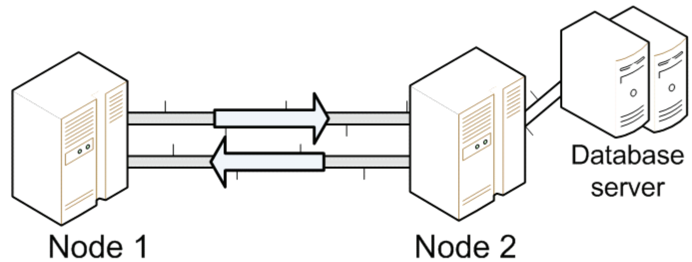

Figure 1. Two-way communication with adaptive data acquisition.

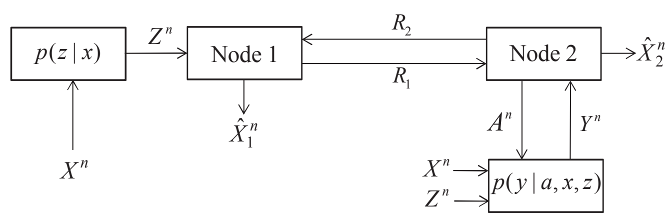

Figure 2. Indirect two-way source coding with a side information vending machine at Node 2.

I. Introduction

In computer networks and machine-to-machine links, communication is often interactive and serves a number of integrated functions, such as data exchange, query and control. As an exemplifying example, consider the set-up in Fig. 1 in which the terminals labeled Node 1 and Node 2 communicate on bidirectional links. Node 2 has access to a data base or, more generally, is able to acquire information from the environment, e.g., through sensors. As a result of the communication on the forward link, Node 2 wishes to compute some function, e.g., a suitable average, of the data available at Node 1 and of the information retrievable from the environment. At the same time, Node 1 queries Node 2 on the forward link with the aim of retreiving some information from the environment through the backward link.

Information acquisition from the environment is generally expensive in terms of system resources, e.g., time, bandwidth or energy. For instance, accessing a remote data base requires interfacing with a server by following the appropriate protocol, and activating sensors entails some energy expenditure. Therefore, data acquisition by Node 2 should be performed efficiently by adapting to the informational requirements of Node 1 and Node 2.

To summarize the discussion above, in the system of Fig. 1 the forward communication from Node 1 to Node 2 serves three integrated purposes: *i*) *Data exchange*: Node 1 provides Node 2 with the information necessary for the latter to compute the desired quantities; *ii*) *Query*: Node 1 informs Node 2 about its own informational requirements, to be met via the backward link; *iii*) *Control*: Node 1 instructs Node 2 on the most effective way to perform data acquisition from the environment in order to satisfy Node 1's query and to allow Node 2 to perform the desired computation.

In this work, to account for the cost of side information acquisition at Node 2, we adopt the model of a side information "vending machine" introduced in [1] for a point-to-point unidirectional link. Extensions of [1] to multi-terminal models, with unidirectional communication, can be found in [2]-[5] and references therein. Instead, the problem of characterizing the rate-distortion region for a two-way source coding models, with conventional action-independent side information sequences at Node 2 has been addressed in [6], [7], [8] and references therein.

The main contribution of this paper is the derivation of the optimal trade-off between communication rates, distortions of the information produced at the two nodes and costs for information acquisition at Node 2 assuming two communication rounds (Sec. III). Moreover, an example is worked out in order to demonstrate an application of the theory (Sec. IV).

II. System Model

The two-way source coding problem of interest, sketched in Fig. 2, is formally defined by the probability mass functions (pmfs) $p_{XZ}(x,z)$ and $p_{Y|AXZ}(y|a,x,z)$, and by the discrete alphabets $\mathcal{X},\mathcal{Y},\mathcal{Z},\mathcal{A},\hat{\mathcal{X}}_1,\hat{\mathcal{X}}_2$, along with distortion and cost metrics to be discussed below. The source sequences $X^n = (X_1,...,X_n) \in \mathcal{X}^n$ and $Z^n = (Z_1,...,Z_n) \in \mathcal{Z}^n$ consist of n independent and identically distributed (i.i.d.) entries X_i, Z_i for $i \in [1,n]$ with pmf $p_{XZ}(x,z)$. Node 1 measures sequence Z^n and encodes it in a message M_1 of nR_1 bits, which is delivered to Node 2.

Node 2 wishes to estimate a sequence $\hat{X}_2^n \in \hat{\mathcal{X}}_2^n$ within given distortion requirements. To this end, Node 2 receives message M_1 and based on this, it 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 the sequences X^n and Z^n obtained at the Node 2. Specifically, given A^n, X^n and Z^n, the sequence Y^n is distributed as $p(y^n|a^n,x^n,z^n) = \prod_{i=1}^n p_{Y|A,X,Z}(y_i|a_i,x_i,z_i)$. The cost of the action sequence is defined by a cost function Λ: $\mathcal{A} \rightarrow [0,\Lambda_{\max}]$ with $0 \leq \Lambda_{\max} < \infty$, as $\Lambda(a^n) = \sum_{i=1}^n \Lambda(a_i)$. The estimated sequence $\hat{X}_2^n$ with $\hat{X}_2^n \in \hat{\mathcal{X}}_2^n$ is then obtained as a function of M_1 and Y^n.

Upon reception on the forward link, Node 2 maps the message M_1 received from Node 1 and the locally available sequence Y^n in a message M_2 of nR_2 bits, which is delivered back to Node 1. Node 1 estimates a sequence $\hat{X}_1^n \in \hat{\mathcal{X}}_1^n$ as a function of M_2 and Z^n within given distortion requirements.

The quality of the estimated sequence $\hat{X}_j^n$ is assessed in terms of the distortion metrics $d_j(x,y,z,\hat{x}_j)$:$\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \hat{\mathcal{X}}_j \rightarrow \mathbb{R}_+ \cup \{\infty\}$, for $j = 1,2$. Note that this implies that $\hat{X}_j^n$ is allowed to be a lossy version of any function of the source and side information sequences. A more general model is studied in Sec. III. It is assumed that $D_j = \min_{\hat{x}_j \in \hat{\mathcal{X}}_j} E[d(X,Y,Z,\hat{X}_j)] < \infty$ for $j = 1,2$. A formal description of the operations at encoder and decoder follows.

Definition 1. An $(n,R_1,R_2,D_1,D_2,\Gamma,\epsilon)$ code for the set-up of Fig. 2 consists of a source encoder for Node 1

$$\mathrm{g}_1\colon \mathcal{Z}^n \rightarrow [1,2^{nR_1}], \tag{1}$$

which maps the sequence Z^n into a message M_1; an "action" function

$$\ell\colon [1,2^{nR_1}] \times \mathcal{Y}^{i-1} \rightarrow \mathcal{A}, \tag{2}$$

which maps the message M_1 and the previously observed sequence into an action sequence A^n; a source encoder for Node 2

$$\mathrm{g}_2\colon \mathcal{Y}^n \times [1,2^{nR_1}] \rightarrow [1,2^{nR_2}], \tag{3}$$

which maps the sequence Y^n and message M_1 into a message M_2; two decoders, namely

$$\mathrm{h}_1\colon [1,2^{nR_2}] \times \mathcal{Z}^n \rightarrow \hat{\mathcal{X}}_1^n, \tag{4}$$

which maps the message M_2 and the sequence Z^n into the estimated sequence $\hat{X}_1^n$;

$$\mathrm{h}_2\colon [1,2^{nR_1}] \times \mathcal{Y}^n \rightarrow \hat{\mathcal{X}}_2^n, \tag{5}$$

which maps the message M_1 and the sequence Y^n into the estimated sequence $\hat{X}_2^n$; such that the action cost constraint Γ and distortion constraints D_j for $j = 1,2$ are satisfied, i.e.,

$$\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[\Lambda(A_i)\right] \leq \Gamma \tag{6}$$

$$\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[d_j(X_i,Y_i,Z_i,\hat{X}_{ji})\right] \leq D_j \text{ for } j = 1,2. \tag{7}$$

Given a distortion-cost tuple (D_1,D_2,Γ), a rate tuple (R_1,R_2) is said to be achievable if, for any $\epsilon > 0$, and sufficiently large n, there exists a $(n,R_1,R_2,D_1+\epsilon,D_2+\epsilon,\Gamma+\epsilon)$ code. The *rate-distortion-cost region* $\mathcal{R}(D_1,D_2,\Gamma)$ is defined as the closure of all rate tuples (R_1,R_2) that are achievable given the distortion-cost tuple (D_1,D_2,Γ).

In the following sections, for simplicity of notation, we drop the subscripts from the definition of the pmfs, thus identifying a pmf by its argument.

III. Rate-Distortion-Cost Region

The next proposition derives a single-letter characterization of the rate-distortion-cost region.

Proposition 1. *The rate-distortion-cost region $\mathcal{R}(D_1,D_2,\Gamma)$ is given by the union of all rate pairs (R_1,R_2) that satisfy the conditions*

$$R_1 \geq I(Z;A) + I(Z;U|A,Y) \tag{8a}$$

$$\text{and } R_2 \geq I(Y;V|A,Z,U), \tag{8b}$$

where the mutual information terms are evaluated with respect to the joint pmf

$$p(x,y,z,a,u,v) = p(x,z)p(a,u|z)p(y|a,x,z)p(v|a,u,y), \tag{9}$$

for some pmfs $p(a,u|x)$ and $p(v|a,u,y)$ such that the inequalities

$$\mathrm{E}[d_1(X,Y,Z,\mathrm{f}_1(V,Z))] \leq D_1, \tag{10a}$$

$$\mathrm{E}[d_2(X,Y,Z,\mathrm{f}_2(U,Y))] \leq D_2, \tag{10b}$$

$$\text{and } \mathrm{E}[\Lambda(A)] \leq \Gamma, \tag{10c}$$

are satisfied for some function f_1: $\mathcal{V} \times \mathcal{Z} \rightarrow \hat{\mathcal{X}}_1$ and f_2: $\mathcal{U} \times \mathcal{Y} \rightarrow \hat{\mathcal{X}}_2$. Finally, U and V are auxiliary random variables whose alphabet cardinality can be constrained as $|\mathcal{U}| \leq |\mathcal{Z}||\mathcal{A}| + 3$ and $|\mathcal{V}| \leq |\mathcal{U}||\mathcal{Y}||\mathcal{A}| + 1$ without loss of optimality.

Remark 1. For the special case in which the side information Y is independent of the action A given X, i.e., for $p(y|a,x) = p(y|x)$, the rate-distortion region $\mathcal{R}(D_1,D_2,\Gamma)$ in Proposition 1 reduces to that derived in [6], [7]. Instead, if $D_1 = D_{1,\max}$, the result reduces to that in [1].

The proof of the converse is provided in [9]. The achievability follows as a combination of the techniques proposed in [1] and [6], and requires the forward link to be used, in an integrated manner, for data exchange, query and control. Specifically, for the forward link, similar to [1], Node 1 uses a successive refinement codebook. Accordingly, the base layer is used by Node 1 to instruct Node 2 on which actions are best tailored to fulfill the informational requirements of both Node 1 and Node 2. This base layer thus represents control information that also serves the purpose of querying Node 2 in view of the backward communication. We observe that Node 1 selects this base layer as a function of the source Z^n, thus allowing Node 2 to adapt its actions for information acquisition to the current realization of the source Z^n. The refinement layer of the code used by Node 1 is leveraged, instead, to provide additional information to Node 2 in order to meet Node 2's distortion requirement. Node 2 then employs standard Wyner-Ziv coding (i.e., binning) [10] for the backward link to satisfy Node 1's distortion requirement.

We now briefly outline the main technical aspects of the achievability proof, since the details follow from standard arguments and do not require further elaboration here. To be more precise, Node 1 first maps sequence Z^n into the action sequence A^n using the standard joint typicality criterion. This mapping requires a codebook of rate $I(Z;A)$ (see, e.g., [10, pp. 62-63]). Given the sequence A^n, the description of sequence Z^n is further refined through mapping to a sequence U^n. This requires a codebook of size $I(Z;U|A,Y)$ for each action sequence A^n using Wyner-Ziv binning with respect to side information Y^n [10, pp. 62-63]. In the reverse link, Node 2 employs Wyner-Ziv coding for the sequence Y^n by leveraging the side information Z^n available at Node 1 and conditioned on the sequences U^n and A^n, which are known to both Node 1 and Node 2 as a result of the communication on the forward link. This requires a rate equal to the right-hand side of (8b). Finally, Node 1 and Node 2 produce the estimates $\hat{X}_1^n$ and $\hat{X}_2^n$ as the symbol-by-symbol functions $\hat{X}_{1i} = \mathrm{f}_1(V_i, Z_i)$ and $\hat{X}_{2i} = \mathrm{f}_2(U_i, Y_i)$ for $i \in [1,n]$, respectively.

Remark 2. The achievability scheme discussed above uses actions that do not adapt to the past values of the side information Y. The fact that this scheme attains the optimal performance characterized in Proposition 1 shows that, as demonstrated in [11] for the one-way model with $R_2 = 0$, adaptive actions do not improve the rate-distortion performance.

IV. Example

Consider a sensor network consisting of two sensors deployed to monitor a given phenomenon of interest (i.e., the concentration of a given chemical). Assume that the state of the observed phenomenon is described by a random source $X \sim$ Bern(0.5) (e.g., $X = 0$ indicates a low concentration of the chemical and $X = 1$ a high concentration). Due to malfunctioning or environmental causes, at each time i, Node 1 measures X_i with probability $1-\epsilon$, and reports instead an unusual event e (referred to as "erasure") with probability ϵ. This implies that we have $Z_i = \mathrm{e}$ with probability ϵ, and $Z_i = X_i$ with probability $1-\epsilon$, for $i \in [1,n]$.

Node 2 has the double purpose of monitoring the operation of Node 1 and of assisting Node 1 in case of measurement failures (erasures). To this end, if necessary, Node 2 can measure the phenomenon X_i by investing a unit of energy. This is modeled by assuming that the vending machine at Node 2 operates as follows:

$$Y = \begin{cases} X & \text{for } A = 1 \\ \phi & \text{for } A = 0 \end{cases}, \tag{11}$$

with cost constraint $\Lambda(a) = a$, for $a \in \{0,1\}$, where ϕ is a dummy symbol representing the case in which no useful information is acquired by Node 2. This model implies that a cost budget of Γ limits the average number of samples of the sequence Y that can be measured by Node 2 to around $n\Gamma$ given the constraint (6).

Node 1 wishes to reconstruct the source X^n, while Node 2 is interested in recovering Z^n in order to monitor the operation of Node 1. The distortion functions are the Hamming metrics $d_1(x,\hat{x}_1) = 1_{\{x \neq \hat{x}_1\}}$ and $d_2(z,\hat{x}_2) = 1_{\{z \neq \hat{x}_2\}}$. Therefore, the maximum distortions are easily seen to be given by $D_{1,\max} = 0.5$ and $D_{2,\max} = 1 - \max\{\epsilon, (1-\epsilon)/2\}$.

To obtain analytical insight into the rate-distortion-cost region, in the following we focus on a number of special cases.

A. $D_1 = D_{1,\max}$ *and* $D_2 = 0$

Consider the distortion requirements $D_1 = D_{1,\max}$ and $D_2 = 0$. As a result, Node 1 requires no backward communication from Node 2, while Node 2 wishes to recover Z^n losslessly. In the context of the example, here the only functionality of the network is the monitoring of the operation of Node 1 by Node 2. For the given distortions, the rate-cost region in Proposition 1 can be evaluated as

$$R_1 \geq H_2(\epsilon) + (1-\epsilon-\Gamma)^+ \tag{12a}$$

$$\text{and } R_2 \geq 0, \tag{12b}$$

for any cost budget $\Gamma \geq 0$, where $H_2(\alpha) = -\alpha\log_2\alpha - (1-\alpha)\log_2(1-\alpha)$ is the binary entropy function.

A formal proof of this result can be found in Appendix. The rate region (12) shows that, as the cost budget Γ for information acquisition increases, the required rate R_1 decreases down to the rate $H_2(\epsilon)$ that is required to describe only the erasures process E^n with $E_i = 1_{\{Z_i = \mathrm{e}\}}$, $i = 1, ..., n$, losslessly to Node 2. This can be explained by noting that the time-sharing strategy discussed next achieves region (12) and is thus optimal.

In the mentioned time-sharing strategy, Node 1 describes the process E^n losslessly to Node 2 with $H_2(\epsilon)$ bits per symbol. In addition to E^n, in order to obtain a lossless reconstruction of Z^n, Node 2 needs to be informed about $Z_i = X_i$ for all i in which $E_i = 0$. Note that we have around $n(1-\epsilon)$ such samples of Z_i. Node 1 describes $Z_i = X_i$ for $n(1-\epsilon-\Gamma)^+$ of these samples, while the remaining $n\min(\Gamma, 1-\epsilon)$ are measured by Node 2 through the vending machine. An alternative strategy based directly on Proposition 1 can be found in Appendix.

Fig. 3 illustrates the rate R_1 in (12a) versus the cost budget Γ for $\epsilon = 0.2$ (line with circles). The first observation is that for $\Gamma = 0$, since there is no side information available at Node 2, we have that $R_1 = H_2(\epsilon) + 1 - \epsilon = 1.52$, which is the rate required to transmit the sequence Z^n losslessly to Node 2. Moreover, as mentioned, as the cost budget Γ increases, the rate R decreases down to the value $H_2(\epsilon) = 0.72$ needed to describe only the sequence E^n. Finally, we observe that if

$\Gamma \geq 1-\epsilon = 0.8$ no further improvement of the rate is possible since Node 2 only needs to measure a fraction $(1-\epsilon)$ of values of X^n in order to recover Z^n losslessly.

B. $D_1 = 0$ and $D_2 = D_{2,\max}$

Here we consider the dual case in which Node 1 wishes to reconstruct sequence X^n losslessly ($D_1 = 0$), while Node 2 does not have any distortion requirements ($D_2 = D_{2,\max}$). In the context of the example, the network thus operates with the aim of allowing Node 1 to measure the phenomenon of interest X^n reliably. As shown in Appendix, if $\Gamma \geq \epsilon$, the rate-cost region is given by the union of all rate pairs (R_1, R_2) such that

$$R_1 \geq H_2(\epsilon) - \Gamma H\left(\frac{\epsilon}{\Gamma}\right) \quad (13a)$$
$$\text{and } R_2 \geq \epsilon. \quad (13b)$$

Moreover, for $\Gamma < \epsilon$, the region is empty as the lossless reconstruction of X at Node 1 is not feasible.

A proof of this result based on Proposition 1 can be found in Appendix. In the following, we argue that a natural time-sharing strategy, akin to that used for the case $D_1 = D_{1,\max}, D_2 = 0$ above, would be suboptimal, implying that the optimal strategy requires a more sophisticated approach based on the successive refinement code presented in Sec. III.

A natural time-sharing strategy would be the following. Node 1 describes $n\eta$ samples of the erasure process E^n, for some $0 \leq \eta \leq 1$, losslessly to Node 2, using rate $R_1 = \eta H_2(\epsilon)$. This information is used by Node 1 to query Node 2 about the desired information. Specifically, Node 2 sets $A_i = 1$ if $E_i = 1$, thus observing around $n\eta\epsilon$ samples $Y_i = X_i$ from the vending machine. These samples are needed to fulfill the distortion requirements of Node 1. For all the remaining $n(1-\eta)$ samples, for which Node 2 does not have control information from Node 1, Node 2 sets $A_i = 1$, thus acquiring all the side information samples. Again, this is necessary given Node 1's requirements. Node 2 conveys losslessly the $n\eta\epsilon$ samples $Y_i = X_i$ obtained when $E_i = 1$, which requires $\eta\epsilon$ bits per sample, along with the $n(1-\eta)$ samples Y_i in the second set, which amount instead to $(1-\eta)H(X|Z)$ bits per sample. Note that we have the rate $H(X|Z)$ by the Slepian-Wolf theorem [10, Chapter 10], since Node 1 has side information Z_i for the second set of samples. Overall, we have $R_2 = \eta\epsilon + (1-\eta)\epsilon = \epsilon$ bits/source symbol. This entails a cost budget of $\Gamma = \eta\epsilon + 1 - \eta$, and thus $\eta = {(1-\Gamma)}/{(1-\epsilon)}$.

Fig. 3 compares the rate R_1 as in (13a) (line with squared markers) with the corresponding rate obtained via time-sharing (dashed line), for $\epsilon = 0.2$. As seen, in this second case the time-sharing strategy is strictly suboptimal (except for the two extreme case of $\Gamma = 0$ and $\Gamma = 1$). Moreover, as discussed above, achieving $D_1 = 0$ is impossible for $\Gamma \leq \epsilon$, since Node 2 must obtain the fraction ϵ of samples of X^n that Node 1 fails to measure in order to guarantee lossless reconstruction at Node 1.

C. $D_1 = D_2 = 0$

We now consider the case in which both nodes wish to achieve lossless reconstruction, i.e., $D_1 = D_2 = 0$. In this case, in the context of the example, both measurement of Node 1 and monitoring at Node 2 are required to operate correctly. As seen in the previous case, achieving $D_1 = 0$ is not possible if $\Gamma < \epsilon$ and thus this is a fortiori true for $D_1 = D_2 = 0$. For $\Gamma \geq \epsilon$, the rate-cost region is given by

$$R_1 \geq H_2(\epsilon) + (1-\Gamma) \quad (14a)$$
$$\text{and } R_2 \geq \epsilon, \quad (14b)$$

as shown in Appendix.

A time-sharing strategy that achieves (14) is as follows. Node 1 describes the process E^n losslessly to Node 2 with $H_2(\epsilon)$ bits per symbol. This information serves the functions of query and control for Node 2. In order to satisfy its distortion requirement, Node 2 now needs to be informed about $Z_i = X_i$ for all i in which $E_i = 0$. Note that we have $n(1-\epsilon)$ such samples of Z_i. Node 1 describes $Z_i = X_i$ for $n(1-\Gamma) \leq n(1-\epsilon)$ of these samples, while the remaining $n(\Gamma-\epsilon)$ are measured by Node 2 through the vending machine. Node 2 compresses losslessly the sequence of around $n\epsilon$ samples of X_i with i such that $E_i = 1$ which requires $R_2 = \epsilon$ bits per sample.

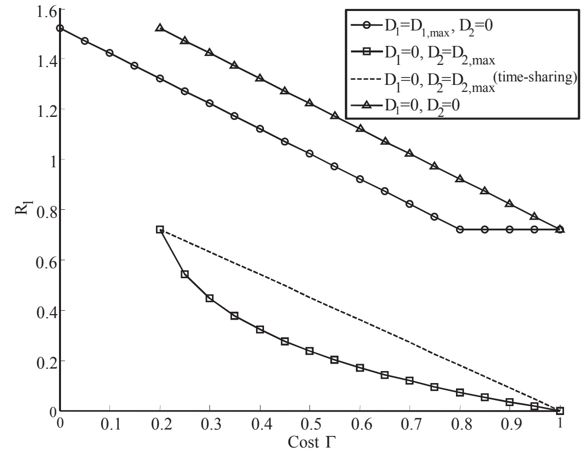

Figure 3. Rate R_1 versus cost Γ for the examples in Sec. IV with $\epsilon = 0.2$.

Fig. 3 illustrates the rate R_1 in (14a) versus the cost budget Γ for $\epsilon = 0.2$ (line with triangular markers). As discussed above, achieving $D_1 = 0$ is impossible for $\Gamma \leq \epsilon$. Moreover, for $\Gamma = 1$ the performance of system is identical to that with $D_1 = D_{1,\max}$ and $D_2 = 0$, since in this case the informational requirements of both Node 1 and Node 2 are satisfied if Node 1 conveys the locations of the erasures to Node 2 (which requires rate $R_1 = H_2(\epsilon) = 0.72$).

V. Concluding Remarks

For applications such as complex communication networks for cloud computing or machine-to-machine communication, the bits exchanged by two parties serve a number of integrated functions, including data transmission, control and query. In this work, we have considered a baseline two-way communication scenario that captures some of these aspects. The problem is addressed from a fundamental theoretical standpoint using an information theoretic formulation. The analysis reveals the structure of optimal communication strategies and can be applied to elaborate on specific examples, as illustrated in the paper. This work opens a number of possible avenues for future research, including the analysis of scenarios in which more than one round of interactive communication is possible [8].

Appendix: Proofs for the Example in Sec. IV

1) $D_1 = D_{1,\max}$ and $D_2 = 0$: Here, we prove that the rate-cost region in Proposition 1 is given by (12) for $D_1 = D_{1,\max}$ and $D_2 = 0$. We begin with the converse part. Starting from (8a), we have

$$\begin{aligned} R_1 &\overset{(a)}{\geq} I(A;Z) + H(Z|A,Y) \\ &= H(Z) - I(Z;Y|A) \\ &\overset{(b)}{\geq} H(Z) - \Gamma I(Z;X|A=1) \qquad (15) \\ &\overset{(c)}{\geq} H(Z) - \Gamma H(X|A=1) \\ &\overset{(d)}{\geq} H(Z) - \Gamma \\ &\overset{(e)}{=} H_2(\epsilon) + 1 - \epsilon - \Gamma, \qquad (16) \end{aligned}$$

where (a) follows from (8a) and since Z has to be recovered losslessly at Node 2; (b) follows since $\Pr[A=1] = \mathrm{E}[\Lambda(A)] \leq \Gamma$; (c) follows because entropy is non-negative; (d) follows since $H(X|A=1) \leq 1$; and (e) follow because $H(Z) = H_2(\epsilon) + 1 - \epsilon$. Achievability follows by setting $U = Z$, $V = \emptyset$, $\Pr(A=1|Z=0) = \Pr(A=1|Z=1) = \Gamma/(1-\epsilon)$ and $\Pr(A=0|Z=\mathrm{e}) = 1$ in (8).

2) $D_1 = 0$ and $D_2 = D_{2,\max}$: Here, we turn to the case $D_1 = 0$ and $D_2 = D_{2,\max}$. We start with the converse. Since X is to be reconstructed losslessly at Node 1, we have the requirement $H(X|V,Z) = 0$ from (10a). It easy to see that this requires that the equalities $A = 1$ and $V = Y = X$ be met if $Z = \mathrm{e}$. In fact, otherwise, X could not be a function of (V,Z) as required by the equality $H(X|V,Z) = 0$. The condition that $A = 1$ if $Z = \mathrm{e}$ requires that the pmf $p(a|z)$ be such that $\Pr(A=1|Z=\mathrm{e}) = 1$, which entails $\Gamma = \Pr[A=1] \geq \Pr[Z=\mathrm{e}] = \epsilon$. Moreover, we can set $\Pr(A=1|Z=0) = \Pr(A=1|Z=1) = (\gamma-\epsilon)/(1-\epsilon)$, for some $0 \leq \gamma \leq \Gamma$, by leveraging the symmetry of the problem on the selection of the actions given $Z = 0$ and $Z = 1$. Starting from (8a), we can thus write

$$\begin{aligned} R_1 &\overset{(a)}{\geq} I(Z;A) \\ &= H(Z) - H(Z|A) \\ &= H_2(\epsilon) + 1 - \epsilon - \gamma H(Z|A=1) \\ &- (1-\gamma) H(Z|A=0) \\ &\overset{(b)}{=} H_2(\epsilon) + 1 - \epsilon - \gamma H\Big(\frac{\epsilon}{\gamma}, \frac{\gamma-\epsilon}{2\gamma}, \frac{\gamma-\epsilon}{2\gamma}\Big) \\ &\quad -(1-\gamma) \\ &= H_2(\epsilon) - \gamma H_2\big(\frac{\epsilon}{\gamma}\big) \\ &\overset{(a)}{\geq} H_2(\epsilon) - \Gamma H_2\big(\frac{\epsilon}{\Gamma}\big), \qquad (17) \end{aligned}$$

where (a) follows from (8a) and since there is no distortion requirement at Node 2; (b) follows by direct calculation; and (c) follows since $H_2(\epsilon) - \gamma H_2(\frac{\epsilon}{\gamma})$ is minimzed at $\gamma = \Gamma$ over all $0 \leq \gamma \leq \Gamma$.

The bound (13b) follows immediately by providing Node 2 with the sequence X^n and then using the bound $R_2 \geq H(X|Z) = \epsilon$.

Achievability follows by setting $U = \emptyset$ and the pmf $p(a|z)$ be such that $\Pr(A=1|Z=\mathrm{e}) = 1$ and $\Pr(A=1|Z=0) = \Pr(A=1|Z=1) = \frac{\Gamma-\epsilon}{1-\epsilon}$. Moreover, let $V = Y = X$ if $Z = \mathrm{e}$ and $V = Y = \phi$ otherwise. Evaluating (8) with these choices leads to (13).

3) $D_1 = D_2 = 0$: Here, we prove the rate-cost region (14) for the case $D_1 = D_2 = 0$. Starting from (8a), we have

$$\begin{aligned} R_1 &\overset{(a)}{\geq} H(Z) - \Gamma I(Z;X|A=1) \\ &\overset{(b)}{=} H(Z) - \Gamma H(X|A=1) \\ &\quad + \Gamma H(X|A=1,Z=\mathrm{e}) \Pr(Z=\mathrm{e}|A=1) \\ &\overset{(c)}{\geq} H(Z) - \Gamma + \Gamma . \frac{\epsilon}{\Gamma} \\ &= H_2(\epsilon) + 1 - \Gamma, \qquad (18) \end{aligned}$$

where (a) follows as in (15); (b) follows because $H(X|A=1,Z=0) = H(X|A=1,Z=1) = 0$; (c) follows since $H(X|A=1) \leq 1$, $H(X|A=1,Z=\mathrm{e}) = 1$ and because $p(Z=\mathrm{e}|A=1) = \frac{\epsilon}{\Gamma}$, where latter follows from the the requirement $H(X|V,Z) = 0$ as per discussion provided in the previous section.

For the achievability, let $U = Z$, $\Pr(A=1|Z=\mathrm{e}) = 1$ and $\Pr(A=1|Z=0) = \Pr(A=1|Z=1) = \frac{\Gamma-\epsilon}{1-\epsilon}$. Moreover, let $V = Y = X$ if $Z = \mathrm{e}$ and $V = Y = \emptyset$ otherwise. Evaluating (8) with these choices leads to (14).

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] 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.

[3] B. Ahmadi and O. Simeone, "Distributed and cascade lossy source coding with a side information "Vending Machine"," *IEEE Trans. Inform. Theory,* http://arxiv.org/abs/1109.6665.

[4] B. Ahmadi, C. Choudhuri, O. Simeone and U. Mitra, "Cascade Source Coding with a Side Information "Vending Machine"," http://arxiv.org/abs/1207.2793.

[5] L. Zhao, Y. K. Chia, and T. Weissman, "Compression with actions," in *Proc. Allerton conference on communications, control and computing*, Monticello, Illinois, Sep. 2011.

[6] A. H. Kaspi., "Two-way source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 735–740, Nov. 1985.

[7] H. Permuter, Y. Steinberg, and T. Weissman, "Two-way source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2905–2919, Jun. 2010.

[8] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6180–6195, Sep. 2011.

[9] B. Ahmadi and O. Simeone, "Two-way Communication with Adaptive Data Acquisition," http://arxiv.org/abs/1209.5978.

[10] A. El Gamal and Y. Kim, *Network Information Theory*, Cambridge University Press, Dec. 2011.

[11] C. Choudhuri and U. Mitra, "How useful is adaptive action?," in *Proc. IEEE GLOBECOM 2012*, Anaheim, CA, Dec. 2012.