Two-way communication with adaptive data acquisition

Behzad Ahmadi* and Osvaldo Simeone
CWCSPR, ECE Dept., New Jersey Institute of Technology, Newark, NJ 07102, USA

ABSTRACT
Motivated by computer networks and machine-to-machine communication applications, a bidirectional link is studied in which two nodes, Node 1 and Node 2, communicate to fulfill generally conflicting informational requirements. Node 2 is able to acquire information from the environment, for example, via access to a remote database or via sensing. Information acquisition is expensive in terms of system resources, for example, time, bandwidth and energy, and thus should be carried out efficiently by adapting the acquisition process to the needs of the application. As a result of the forward communication from Node 1 to Node 2, the latter wishes to compute some function, such as a suitable average, of the data available at Node 1 and of the data obtained from the environment. The forward link is also used by Node 1 to query Node 2 with the aim of retrieving suitable information from the environment on the backward link. The problem is formulated in the context of multi-terminal rate distortion theory and the optimal trade-off between communication rates, distortions of the information produced at the two nodes and costs for information acquisition at Node 2 is derived. The issue of robustness to possible malfunctioning of the data acquisition process at Node 2 is also investigated. The results are illustrated via an example that demonstrates the different roles played by the forward communication, namely data exchange, query and control. Copyright © 2013 John Wiley & Sons, Ltd.

*Correspondence
Behzad Ahmadi, CWCSPR, ECE Dept., New Jersey Institute of Technology, Newark, NJ 07102, USA.
E-mail: behzad.ahmadi@njit.edu
Received 24 September 2012; Revised 31 March 2013; Accepted 22 April 2013

1. 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 illustrative example, consider the set-up in Figure 1 in which the terminals labeled Node 1 and Node 2 communicate on bidirectional links. Node 2 has access to a database or, more generally, is able to acquire information from the environment, for example, through sensors. As a result of the communication on the forward link (Figure 1(a)), Node 2 wishes to compute some functions $\hat{X}_2^n$, for example, a suitable average, of the data $X^n$ available at Node 1 and of the information it can retrieve from the environment*. The latter is measured as $Y^n$ by Node 2 (Figure 1(b)). The forward link (Figure 1(a)) is also used by Node 1 to query Node 2 with the aim of retrieving some information $\hat{X}_1^n$ about $Y^n$ from the environment through the backward link (Figure 1(c)).

Information acquisition from the environment at Node 2 is generally expensive in terms of system resources, for example, time, bandwidth or energy. For instance, accessing a remote database as in Figure 1(b) requires interfacing with a server by following the appropriate protocol, and activating sensors entail 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, that is, to the requirements on the calculation of $\hat{X}_1^n$ and $\hat{X}_2^n$, respectively.

To summarise the previous discussion, in the system of Figure 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 about data $X^n$ for the latter to compute the desired quantity; (ii) Query: Node 1 informs Node 2 about its own informational requirements with regard to $\hat{X}_1^n$; to be met via the backward link; (iii) Control: Node 1 instructs Node 2 on the most effective way to perform data acquisition.
Figure 1. Two-way communication with adaptive data acquisition.

(i.e. acquisition of $Y^n$) from the environment to satisfy Node 1’s query and to allow Node 2 to perform the desired computation (of $X^n_2$).

This work sets out to analyse the setting in Figure 1 from a fundamental theoretical standpoint via information theory. Specifically, the problem is formulated within the context of network rate distortion theory, and the optimal communication strategy, involving the elements of data exchange, query and control, is identified. Examples are worked out to illustrate the relevance of the developed theory. Finally, the issue of robustness is tackled by assuming that, unbeknownst to Node 1, Node 2 may be unable to acquire information from the environment, due to, for example, energy shortages or malfunctioning. The optimal robust strategy is derived, and the examples are extended to account for this generalised model.

1.1. Related work

The work in this paper builds on the long line of research within network information theory that deals with source coding with side information (see, e.g. [1]). A very short introduction is provided in Section 2. In this section, we instead briefly review the most related works. We first recall that the conventional formulation of the source coding problem with side information assumes that the relationship between the source $X$ available at the encoder and the side information $Y$ available at the decoder is determined by a given conditional distribution $p_{Y|X}(y|x)$. In contrast, in this work, we adopt the model of a side information ‘vending machine’ that has been introduced in [2]. This model accounts for source coding scenarios in which acquiring the information $Y$ at the receiver entails some cost and thus should be carried out efficiently. Specifically, in this model, the quality of the side information $Y$ can be controlled at the decoder by selecting an action $A$ that affects the effective channel between the source $X$ and the side information $Y$ through a conditional distribution $p_{Y|X,A}(y|x,a)$. The distribution $p_{Y|X,A}(y|x,a)$ defines the side information ‘vending machine’ as per the nomenclature of [2]. Each action $A$ is associated with a cost, and the problem studied in [2] is that of characterising the available trade-offs among rate, distortion and action cost.
Various works have extended the results in [2]. Extensions to multi-terminal models can be found in [3]. Specifically, references [3–10] considered a set-up analogous to the Heegard–Berger problem [11,12], in which the side information may or may not be available at the decoder (Section 2). In [5], a distributed source coding setting that generalises [13] to the case of a decoder with a side information ‘vending machine’ is investigated. Multi-hop models were studied in [5,6]. In [8], a related problem is considered in which the sequence to be compressed is dependent on the actions taken by a separate encoder. Other extensions include [9,10] where the model of [2] is revisited under the additional constraints of common reconstruction [14] or of secrecy with respect to an ‘eavesdropping’ node.

In this paper, the model of a side information ‘vending machine’ is used to model the information acquisition process at Node 2 in Figure 1. Unlike [2] and the previous work discussed previously, communication between Node 1 and Node 2 is assumed to be bidirectional. The problem of characterising 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 [15–17] and references therein.

1.2. Contributions and organization of the paper

This work studies the model in Figure 1, which is detailed in terms of a block diagram in Figure 2. The system model is introduced in Section 3. The optimal trade-off between the rates of the bidirectional communication, the distortions of the reconstructions of the desired quantities at the two nodes and the budget for information acquisition at Node 2 is derived in Section 4. An example that illustrates the application of the developed theory is discussed in Section 5. Finally, in Section 6, the results are extended to the scenario in Figure 10 in which, unbeknownst to Node 1, Node 2 may be unable to perform information acquisition.

Notation: Throughout the paper, a random variable is denoted by an upper case letter (e.g. $X$, $Y$, $Z$), and its realisation is denoted by a lower case letter (e.g. $x$, $y$, $z$). All the random variables are discrete and take values in a finite discrete set represented by a script letter corresponding to the random variable, for example, $X \in \mathcal{X}$. Moreover, the shorthand notation $X^n$ is used to denote the tuple (or the column vector) of random variables $(X_1, \ldots, X_n)$, and $x^n$ is used to denote a realisation. We define $[a, b] = [a, a+1, \ldots, b]$ for $a \leq b$ and $[a, b] = \emptyset$ otherwise. The probability mass function (pmf) of a random variable $X$ is denoted as $p_X(x)$ or $p(x)$ where clear from the context. Similar notations are used for joint and conditional pmfs. We say that $X = Y = Z$ forms a Markov chain if $p(x, y, z) = p(x)p(y|x)p(z|y)$, that is, if $X$ and $Z$ are conditionally independent of each other given $Y$.

2. PRELIMINARIES

In this section, we review some preliminary information theoretic results on related problems in the field of source coding with side information.

From an information theoretic perspective, the baseline setting for the class of source coding problems with side information is one in which a memoryless source $X^n = (X_1, \ldots, X_n)$, where $X_i \in \mathcal{X}$ has pmf $p(x)$ for $i = 1, \ldots, n$, is to be communicated by an encoder via a message $M$ of rate $R$ bits per source symbol to a decoder. The decoder has available correlated sequence $Y^n$, with $Y_i \in \mathcal{Y}_i$, that is related to $X^n$ via a memoryless channel $p(y|x)$ (Figure 3). The optimal trade-off between rate $R$ and the average distortion $D$ between the source $X^n$ and reconstruction $\hat{X}_n$ was obtained by Wyner and Ziv in [7] for any given distortion metric $d(x, \hat{x}): \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}_+ \cup \{\infty\}$. It was shown that the rate distortion function is given by

$$R(D) = \min I(X; U | Y)$$

where the minimum is taken over all pmfs $p(u|x)$, with $u \in \mathcal{U}$, and deterministic function $\hat{x}(u, y)$ such that $E[d(X, \hat{x}(U, Y))] \leq D$.

The optimal performance can be achieved as follows. The sequence $X^n$ is quantised using a randomly generated codebook of codewords $U^n$ using the standard joint typicality criterion. For quantisation to be successful, $2^{nI(X; U)}$ codewords $U^n$ are sufficient. Thanks to the side information $Y^n$ available at Node 2, the resulting rate $I(X; U)$ (bits per source symbol) can be further decreased to $I(X; U | Y)$ using the technique of binning. The idea is that all the $2^{nI(X; U)}$ codewords $U^n$ are divided into $2^{nI(X; U | Y)}$ bins. An example of a bin is shown in gray in Figure 4. The codewords $U^n$ in the same bin are mapped to an identical message $M$. The bins contain $2^{nI(X; U)} / 2^{nI(X; U | Y)} = 2^{nI(U; Y)}$ codewords each, and thus, by the channel coding theorem, the decoder can

![Figure 2. Two-way source coding with a side information vending machine at Node 2.](image)

![Figure 3. Source coding with side information.](image)
distinguish among the codewords $U^n$ in the bin based on $Y^n$. This scheme is known as ‘Wyner-Ziv’ coding in information theory.

In sensor networks and cloud computing, reliability of all the computing devices (e.g. sensors or servers) cannot be guaranteed all the time. Therefore, it is appropriate to design the system so as to be robust to system failures. As all the computing devices (e.g. sensors or servers) cannot be guaranteed all the time. Therefore, it is appropriate to design the system so as to be robust to system failures. As shown in Figure 5, this aspect can be modelled by assuming that the decoder, unrekonwst to the encoder, may not be able to acquire information sequence $Y^n$. This setting is equivalent to assuming the presence of two decoders, one with the capability to acquire information about $Y^n$ (Node 2) and one without this capability (Node 3). This model is referred to as the Heegard–Berger problem, where Node 2 and Node 3 of Figure 5 are interested in estimating $X^n$ with $X^n$ and $Y^n$, respectively. It is emphasised that $\hat{X}_2^n$ and $\hat{X}_3^n$ are two different descriptions of the source sequence $X^n$ to be reconstructed at Node 2 and Node 3 with distortion levels $D_2$ and $D_3$, respectively.

It was shown in [11] that the rate distortion function is given by

$$R(D_2, D_3) = \min I(X; \hat{X}_3) + I(U; \hat{X}_3, Y)$$

where the minimum is taken over all pmfs such that $E[d(X, \hat{X}_2)] \leq D_2$ and $E[d(X, \hat{X}_3, Y)] \leq D_3$. The optimal strategy achieving (2) is based on successive refinement and binning. The strategy is identical to that used later in the context of the model of Figure 6, and we refer to the discussion later for further details (Figure 7).

The optimal strategy achieving (2) is based on successive refinement (or layered) compression strategy, in which the coarse layer is used to inform the decoder about the actions $A^n$ that are better adapted to the source $X^n$, and the fine layer provides further information that enables the

![Figure 4](image-url) Illustration of binning. The set of all dots represents the codebook of codewords $U^n$, and the subset in gray is a bin. The picture assumes for simplicity $X = U$.

![Figure 5](image-url) Source coding when side information may be absent.

![Figure 6](image-url) Source coding with side information ‘vending machine’.

![Figure 7](image-url) Illustration of the optimal strategy for the set-up of Figure 6. The picture assumes for simplicity that $X = A = U$.

The concept of a side information ‘vending machine’ was introduced in [2] to account for source coding scenarios in which acquiring the side information at the receiver entails some cost and thus should be carried out efficiently. In this class of models, the quality of the side information $Y^n$ can be controlled at the decoder by selecting an action sequence $A^n$, with $A_i \in \mathcal{A}$, that affects the effective channel between the source $X^n$ and the side information $Y^n$ through a conditional memoryless distribution $p_{Y|X,A}(y|x,a)$. Specifically, given $A^n$ and $X^n$, the sequence $Y^n$ is distributed as $p(y^n|a^n,x^n) = \prod_{i=1}^n p_{Y|X,A}(y_i|x_i,a_i)$. The cost of the action sequence is defined by a cost function $\Lambda: \mathcal{A} \rightarrow [0, \Lambda_{\text{max}}]$ with $0 \leq \Lambda_{\text{max}} < \infty$, as $\Lambda(a^n) = \sum_{i=1}^n \Lambda(a_i)$. The estimated sequence $\hat{X}^n$ with $\hat{X}^n \in \hat{X}^n$ is then obtained as a function of $M$ and $Y^n$.

The optimal trade-off between rate $R$ and the average distortion $D$ was introduced in [2] for any given distortion metric $d(x, \hat{x})$, and action cost function $\Lambda(a)$ is given by

$$R(D, \Gamma) = \min I(X; A) + I(X; U|Y, A)$$

where the minimum is taken over all pmfs $p(a, u|x)$ and deterministic function $\hat{x}(u, y)$ such that $E[d(X, \hat{x}(U, Y))] \leq D$ and $E[\Lambda(A)] \leq \Gamma$.

For later reference, it is useful to describe the optimal strategy as derived in [11]. The basic idea is that of using a successive refinement (or layered) compression strategy, in which the coarse layer is used to inform the decoder about the actions $A^n$ that are better adapted to the source $X^n$, and the fine layer provides further information that enables the
decoder to produce the estimate \( \hat{X}^n \). Specifically, as illustrated in Figure 7, the sequence \( X^n \) is quantised using a codebook of action sequences \( A^n \). As discussed, for quantisation to be successful, this quantisation step requires a rate of \( I(X; A) \). Next, a refined description \( U^n \) of \( X^n \) is also obtained and sent to Node 2. From (1), this requires a rate of \( I(X; U|A, Y) \) thanks to binning (a bin is shown in gray in Figure 7).

### 3. SYSTEM MODEL

The two-way source coding problem of interest, sketched in Figure 2, is formally defined by the pmfs \( p_X(x) \) and \( p_{Y|A,X}(y|a,x) \), and by the discrete alphabets \( X, Y, A, X_1, \hat{X}_2 \), along with distortion and cost metrics to be discussed later. The source sequence \( X^n = (X_1, \ldots, X_n) \) consists of \( n \) independent and identically distributed (i.i.d.) entries \( X_i \) for \( i \in [1,n] \) with pmf \( p_X(x) \). Node 1 measures sequence \( X^n \) 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 \( \hat{X}_2^n \), which is distributed as a function of \( X^n \) and the previously observed measurements into an action sequence \( A^n \); a source encoder for Node 2

\[
g_2 : Y^n \times [1,2^nR_1] \rightarrow [1,2^nR_2]
\]

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

\[
h_1 : [1,2^nR_2] \times A^n \rightarrow \hat{X}_1^n
\]

which maps the message \( M_2 \) and the sequence \( X^n \) into the estimated sequence \( \hat{X}_1^n \).

\[
h_2 : [1,2^nR_1] \times Y^n \rightarrow \hat{X}_2^n
\]

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 \( \Gamma \) and distortion constraints \( D_j \) for \( j = 1, 2 \) are satisfied, that is,

\[
\frac{1}{n} \sum_{i=1}^{n} E[A(A_i)] \leq \Gamma
\]

and

\[
\frac{1}{n} \sum_{i=1}^{n} d_j(X_i, Y_i, \hat{X}_2^i) \leq D_j \text{ for } j = 1, 2
\]

**Definition 2.** Given a distortion-cost tuple \( (D_1, D_2, \Gamma) \), 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.

**Definition 3.** The rate distortion-cost region \( R(D_1, D_2, \Gamma) \) is defined as the closure of all rate tuples \( (R_1, R_2) \) that are achievable given the distortion-cost tuple \( (D_1, D_2, \Gamma) \).

**Remark 1.** For the special case in which the side information \( Y \) is independent of the action \( A \), the rate region \( R(D_1, D_2, \Gamma) \) has been derived in [15]. Instead, if \( D_1 = D_{1,\max} \), the set of all achievable rates \( R_1 \) was characterised in [2].

**Remark 2.** The definition (6) of an action encoder allows for adaptation of the actions to the previously observed values of the side information \( Y \). This possibility was studied in [18] for the point-to-point one-way model, which is obtained by setting \( R_2 = 0 \) in the setting of Figure 2.

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.
4. RATE DISTORTION-COST REGION

In this section, a single-letter characterisation of the rate distortion-cost region is derived.

**Proposition 1.** The rate distortion-cost region $\mathcal{R}(D_1, D_2, \Gamma)$ for the two-way source coding problem illustrated in Figure 2 is given by the union of all rate pairs $(R_1, R_2)$ that satisfy the conditions

\[
R_1 \geq I(X; A) + I(X; U|A, Y) \quad (12a)
\]
and

\[
R_2 \geq I(Y; V|A, X, U) \quad (12b)
\]

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

\[
p(x, y, a, u, v) = p(x)p(a, u|x)p(y|a, x)p(v|a, u, y)
\]

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

\[
E[d_1(X, Y, f_1(V, X))] \leq D_1 \quad (14a)
\]
\[
E[d_2(X, Y, f_2(U, Y))] \leq D_2 \quad (14b)
\]

and

\[
E[A(A)] \leq \Gamma, \quad (14c)
\]

are satisfied for some functions $f_1 : V \times X \to \bar{X}_1$ and $f_2 : U \times Y \to \bar{X}_2$. Finally, $U$ and $V$ are auxiliary random variables whose alphabet cardinality can be constrained as $|U| \leq |X||A| + 4$ and $|V| \leq |U||Y||A| + 1$ without loss of optimality.

**Remark 3.** For the special case in which the side information $Y$ is independent of the action $A$ given $X$, that is, 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 [15,16]. Instead, if $D_1 = D_{1,\max}$, the result reduces to that in [2].

The proof of the converse is provided in Appendix A. The achievability follows as a combination of the techniques proposed in [2] and [15], 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 [2] (Section 2 and Figure 7), 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 $X^n$, thus allowing Node 2 to adapt its actions for information acquisition to the current realisation of the source $X^n$. The refinement layer of the code used by Node 1 is leveraged, instead, to provide additional information to Node 2 to meet Node 2’s distortion requirement. Node 2 then employs standard Wyner–Ziv coding (i.e. binning) [1] for the backward link to satisfy Node 1’s distortion requirement.

We now briefly outline the main technical aspects of the achievability proof, because the details follow from standard arguments and do not require further elaboration here. To be more precise and with reference to Figure 7, Node 1 first maps sequence $X^n$ into the action sequence $A^n$ using the standard joint typicality criterion. This mapping requires a codebook of rate $I(X; A)$ (see, e.g. [1, pp. 62-63]). Given the sequence $A^n$, the description of sequence $X^n$ is further refined through mapping to a sequence $U^n$. This requires a codebook of size $I(X; U|A, Y)$ for each action sequence $A^n$ using Wyner–Ziv binning with respect to side information $Y^n$ [1, pp. 62-63]. In the reverse link, Node 2 employs Wyner–Ziv coding for the sequence $Y^n$ by leveraging the side information $X^n$ 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 (12b). Finally, Node 1 and Node 2 produce the estimates $\hat{X}_1^n$ and $\hat{X}_2^n$ as the symbol-by-symbol functions $\hat{X}_{1i} = f_1(V_i, X_i)$ and $\hat{X}_{2i} = f_2(U_i, Y_i)$ for $i \in [1, n]$, respectively.

**Remark 4.** The achievability scheme, discussed previously, uses actions that do not adapt to the previous values of the side information $Y$. The fact that this scheme attains the optimal performance characterised in Proposition 1 shows that, as demonstrated in [18] for the one-way model with $R_2 = 0$, adaptive actions do not improve the rate distortion performance.

4.1. Indirect rate distortion-cost region

In this section, we consider a more general model in which Node 1 observes only a noisy version of the source $X^n$, as depicted in Figure 8. Following [19], we refer to this setting as posing an indirect source coding problem. The example studied in Section 5 illustrates the relevance of this generalisation. The system model is as defined in Section 3 with the following differences. The source encoder for Node 1

\[
g_1 : \mathcal{Z}^n \to \{1, 2^n R_1\}
\]

is

\[
\begin{array}{cccc}
\text{Node 1} & \rightarrow & \text{Node 2} \\
\downarrow & & \downarrow \\
\hat{X}_1^n & & \hat{X}_2^n \\
\uparrow & & \uparrow \\
X^n & & Y^n \\
\end{array}
\]

Figure 8. Indirect two-way source coding with a side information vending machine at Node 2.
maps the sequence $Z^n$ into a message $M_1$: the decoder for Node 1

$$h_1 : [1, 2^{nR_2}] \times Z^n \rightarrow \hat{X}_1^n$$

(16)

maps the message $M_2$ and the sequence $Z^n$ into the estimated sequence $X_1^n$: given $(X^n, A^n, Z^n)$, the side information $Y^n$ is distributed as $p(y^n|a^n, x^n, z^n) = \prod_{i=1}^n p(y_i|a_i, x_i, z_i)$ and the distortion constraints are given as

$$\frac{1}{n} \sum_{i=1}^n E[d_j(x_i, y_i, z_i, \hat{x}_j)] \leq D_j \quad \text{for } j = 1, 2$$

(17)

for some distortion metrics $d_j(x, y, z, \hat{x}_j) : X \times Y \times Z \times \hat{X}_j \rightarrow R_+ \cup \{\infty\}$, for $j = 1, 2$. The next proposition derives a single-letter characterization of the rate distortion-cost region.

**Proposition 2.** The rate distortion-cost region $R(D_1, D_2, \Gamma)$ for the indirect two-way source coding problem illustrated in Figure 8 is given by the union of all rate pairs $(R_1, R_2)$ that satisfy the conditions

$$R_1 \geq I(Z; A) + I(Z; U|A, Y)$$

(18a)

and

$$R_2 \geq I(Y; V|A, Z, U)$$

(18b)

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

$$p(x, y, z, a, u, v) = p(x, z) p(a, u|z) p(y|a, x, z)$$

$$\times p(v|a, u, y)$$

(19)

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

$$E[d_1(X, Y, Z, f_1(V, Z))] \leq D_1$$

(20a)

$$E[d_2(X, Y, Z, f_2(U, Y))] \leq D_2$$

(20b)

and

$$E[\Lambda(A)] \leq \Gamma$$

(20c)

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

The proof of the achievability and converse follows with slight modifications from that of Proposition 1. Specifically, in the achievability the sequence $X^n$ is replaced by its noisy version, that is, the sequence $Z^n$, and the rest of the proof remains essentially unchanged. The proof of the converse is provided in Appendix A.

## 5. CASE STUDY AND NUMERICAL RESULTS

In this section, we consider an example for the set-up in Figure 8 to illustrate the main aspects of the problem and the relevance of the theoretical results derived previously. 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 $X \sim \text{Bern}(0.5)$ (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 $\epsilon$ (referred to as ‘erasure’) with probability $\epsilon$. This implies that we have $Z_i = e$ with probability $\epsilon$, 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 modelled by assuming that the vending machine at Node 2 operates as follows:

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

(21)

with cost constraint $\Lambda(a) = a$, for $a \in \{0, 1\}$, where $\phi$ 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 $\Gamma$ limits the average number of samples of the sequence $Y$ that can be measured by Node 2 to around $n\Gamma$ given the constraint (10).

Node 1 wishes to reconstruct the source $X^n$, whereas Node 2 is interested in recovering $Z^n$ 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 $d_2(z, \hat{z}_2) = 1_{z \neq \hat{z}_2}$. Therefore, the maximum distortions (4) are easily seen to be given by $D_1, \text{max} = 0.5$ and $D_2, \text{max} = 1 - \max(\epsilon, (1 - \epsilon)/2)$.

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

### 5.1. $D_1 = D_1, \text{max}$ and $D_2 = 0$

Consider the distortion requirements $D_1 = D_1, \text{max}$ and $D_2 = 0$. As a result, Node 1 requires no backward communication from Node 2, whereas 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 2 can be evaluated as

$$R_1 \geq H_2(\epsilon) + (1 - \epsilon - \Gamma)^+ \quad (22a)$$
and
$$R_2 \geq 0 \quad (22b)$$

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 B. The rate region (22) shows that, as the cost budget \( \Gamma \) 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 \( E_i = 1 \) for all \( i = 1, \ldots, n \), losslessly to Node 2. This can be explained by noting that the time sharing strategy discussed next achieves region (22) and is thus optimal.

In the time sharing strategy, Node 1 describes the process \( E^n \) losslessly to Node 2 with \( H_2(\epsilon) \) bits per symbol. In addition to \( E^n \), 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 \( Z_1 \). Node 1 describes \( Z_i = X_i \) for \( n(1 - \epsilon - \Gamma)^+ \) of these samples, whereas the remaining \( n_{\min}(\Gamma, 1 - \epsilon) \) are measured by Node 2 through the vending machine. Note that, in the strategy just described, sequence \( Z^n \) can be interpreted as control data that is used by Node 2 to adapt its information acquisition process. An alternative strategy based directly on Proposition 2 can be found in Appendix B.

Figure 9 illustrates the rate \( R_1 \) in (22a) versus the cost budget \( \Gamma \) for \( \epsilon = 0.2 \) (line with circles). The first observation is that for \( \Gamma = 0 \), because 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 \( \Gamma \) 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 because Node 2 only needs to measure a fraction \( (1 - \epsilon) \) of values of \( X^n \) to recover \( Z^n \) losslessly.

5.2. \( D_1 = 0 \) and \( D_2 = D_{2,\text{max}} \)

Here, we consider the dual case in which Node 1 wishes to reconstruct sequence \( X^n \) losslessly (\( D_1 = 0 \)), whereas Node 2 does not have any distortion requirements (\( D_2 = D_{2,\text{max}} \)). 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 B, if \( \Gamma \geq \epsilon \), the rate-cost region is given by the union of all rate pairs \((R_1, R_2)\) such that

$$R_1 \geq H_2(\epsilon) - \Gamma H\left(\frac{\epsilon}{\Gamma}\right) \quad (23a)$$
and
$$R_2 \geq \epsilon \quad (23b)$$

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 2 can be found in Appendix B. In the following, we argue that a natural time sharing strategy, akin to that used for the previous case \( D_1 = D_{1,\text{max}}, D_2 = 0 \), would be suboptimal, implying that the optimal strategy requires a more sophisticated approach based on the successive refinement code presented in Section 4.

A natural time sharing strategy would be the following. Node 1 describes \( n\eta \) samples of the erasure process \( E^n \), for some \( 0 < \eta < 1 \), 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 \( A_j = 1 \) if \( E_j = 1 \), thus observing around \( n\eta \) samples \( Y_j = X_j \) 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 \( A_j = 0 \), thus acquiring all the side information samples. Again, this is necessary given Node 1’s requirements. Node 2 conveys losslessly the \( n\eta \) samples \( Y_j = X_j \) obtained when \( E_j = 1 \), which requires \( n\eta \) bits per sample, along with the \( n(1 - \eta) \) samples \( Y_j \) 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 [1, Chapter 10], because Node 1 has side information \( Z_j \) for the second set of samples. Overall, we have that \( R_2 = n\eta + (1 - \eta)\epsilon = \epsilon \) bits/source symbol. This entails a cost budget of \( \Gamma = n\eta + 1 - \eta \), and thus \( \eta = (1 - \Gamma)/(1 - \epsilon) \).

Figure 9 compares the rate \( R_1 \) as in (23a) (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 previously, achieving \( D_1 = 0 \) is impossible for \( \Gamma \leq \epsilon \), because Node 2 must obtain the fraction \( \epsilon \) of samples of \( X^n \) that Node 1 fails to measure.
to guarantee lossless reconstruction at Node 1. Finally, for \( \Gamma = 1 \), there is no need for Node 1 to send any information to Node 2, as Node 2 is able to acquire the sequence \( X^n \) and send back to Node 1 on the backward link (using rate \( R_2 = \epsilon \) by the Slepian–Wolf theorem).

### 5.3. \( D_1 = D_2 = 0 \)

We now consider the case in which both nodes wish to achieve lossless reconstruction, that is, \( D_1 = D_2 = 0 \). 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

\[
R_1 \geq H_2(\epsilon) + (1 - \Gamma) \tag{24a}
\]

and

\[
R_2 \geq \epsilon \tag{24b}
\]

as shown in Appendix B.

A time sharing strategy that achieves (24) 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. 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, whereas 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 9 illustrates the rate \( R_1 \) in (24a) versus the cost budget \( \Gamma \) for \( \epsilon = 0.2 \) (line with triangular markers). As discussed previously, achieving \( D_1 = 0 \) is impossible for \( \Gamma \leq \epsilon \). Moreover, for \( \Gamma = 1 \), the performance of system is identical to that with \( D_1 = D_{1,\text{max}} \) and \( D_2 = 0 \), because 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 \( R_1 = H_2(\epsilon) = 0.72 \)).

### 6. WHEN THE SIDE INFORMATION MAY BE ABSENT

In this section, we generalise the results of the previous section to the scenario in Figure 10 in which, unbeknownst to Node 1, Node 2 may be unable to perform information acquisition due, for example, to energy shortage or malfunctioning. The set-up of the general model is illustrated in Figure 10.

#### 6.1. System model

The formal description of a \((n, R_1, R_2, D_1, D_2, D_3, \Gamma, \epsilon)\) code for the set-up of Figure 10 is given in Section 4.1

![Figure 10. Indirect two-way source coding when the side information vending machine may be absent at the recipient of the message from Node 1.](image)

(which generalises the model in Section 3) with the addition of Node 3. This added node, which has no access to side information, models the case in which the receiver is not able to acquire the side information due to, for example, malfunctioning. Note that the same message \( M_1 \) from Node 1 is received by both Node 2 and Node 3. This captures the fact that the information about whether or not the recipient is able to access the side information is not available to Node 1. The model in Figure 10 is a generalisation of the so-called Heegard–Berger problem [11, 12].

Formally, Node 3 is defined by the decoding function

\[
h_3 : [1, 2^n R_1] \rightarrow \hat{X}_3^n \tag{25}
\]

which maps the message \( M_1 \) into the estimated sequence \( \hat{X}_3^n \), and the additional distortion constraint

\[
\frac{1}{n} \sum_{i=1}^{n} E \left[ d_3(X_i, Y_i, Z_i, \hat{X}_{3i}) \right] \leq D_3 \tag{26}
\]

We remark that adding a link between Node 3 and Node 1 cannot improve the system performance because Node 3 does not have access to any additional information. Therefore, there is no advantage in having a backward link from Node 3 to Node 1 because the information available at Node 3 is a subset of the information available at Node 1. Therefore, this link is not included in the model.

#### 6.2. Rate distortion-cost region

In this section, a single-letter characterisation of the rate distortion-cost region is derived for the set-up in Figure 10.

**Proposition 3.** The rate distortion-cost region \( R(D_1, D_2, D_3, \Gamma) \) for the two-way source coding problem illustrated in Figure 10 is given by the union of all rate pairs \((R_1, R_2)\) that satisfy the conditions

\[
R_1 \geq I(Z; A) + I(Z; \hat{X}_3 | A)
\]

\[
+ I(Z; U | A, Y, \hat{X}_3) \tag{27a}
\]

and

\[
R_2 \geq I(Y; V | A, Z, U, \hat{X}_3) \tag{27b}
\]
where the mutual information terms are evaluated with respect to the joint pmf

\[ p(x, y, z, a, u, v) = p(x, z)p(a, u, \hat{x}_3|z)p(y|a, x, z) \]
\[ p(v|a, u, y, \hat{x}_3) \]

(28)

for some pmfs \( p(a, u, \hat{x}_3|z) \) and \( p(v|a, u, y) \) such that the inequalities

\[ E[d_1(X, Y, Z, f_1(V, Z))] \leq D_1 \]
\[ E[d_2(X, Y, Z, f_2(U, Y))] \leq D_2 \]
\[ E[d_3(X, Y, Z, \hat{X}_3)] \leq D_3 \]

\[ \text{and } E[A(A)] \leq \Gamma \]

are satisfied for some functions \( f_1: \mathcal{V} \times \mathcal{Z} \to \hat{X}_1 \) and \( f_2: \mathcal{U} \times \mathcal{Y} \to \hat{X}_2 \). Finally, \( \mathcal{U} \) and \( \mathcal{V} \) 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.

The proof of the converse is provided in Appendix A. The achievable rate (27a) can be interpreted as follows. Node 1 uses a successive refinement code with three layers. The first layer is defined as for Section 4 and carries query and control information. The second and third layers are designed as in the optimal Heegard–Berger scheme [11]. Specifically, the second layer is destined to both Node 2 and Node 3, whereas the third layer targets only Node 2, which has enhanced decoding capabilities due to the availability of side information.

To provide further details, as for Proposition 1, the encoder first maps the input sequence \( Z^n \) into an action sequence \( A^n \) so that the two sequences are jointly typical, which requires \( I(Z, A) \) bits/source sample. Then, it maps \( Z^n \) into the estimate \( \hat{X}_3^n \) for Node 3 using a conditional codebook with rate \( I(Z; \hat{X}_3|A) \). Finally, it maps \( Z^n \) into another sequence \( U^n \) using the fact that Node 2 has the action sequence \( A^n \), the estimate \( \hat{X}_2^n \) and the measurement \( Y^n \). Using conditional codebooks (with respect to \( \hat{X}_2^n \) and \( A^n \)) and from the Wyner–Ziv theorem, this requires \( I(Z; U|A, Y, \hat{X}_3) \) bit/source code (Section 2 and Figure 7). As for the rate (27b), 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 \( U^n, A^n, \hat{X}_2^n, \) which are known to both Node 1 and Node 2 as a result of the forward communication. This requires a rate equal to the right-hand side of (27b) (Section 2 and Figure 7). Finally, Node 1 and Node 2 produce the estimates \( \hat{X}_1^n \) and \( \hat{X}_2^n \) as the symbol-by-symbol functions \( \hat{X}_{1i} = f_1(V_i, Z_i) \) and \( \hat{X}_{2i} = f_2(U_i, Y_i) \) for \( i \in [1, n] \), respectively.

### Table I. Erasure distortion for reconstruction at Node 3.

| E     | \( \hat{X}_3 \) | 0 | 1 | * |
|-------|-----------------|---|---|---|
| 0     | 0               | ∞ | 1 |
| 1     | ∞               | 0 | 1 |

### 6.3. Case study and numerical results

In this section, we extend the binary example of Section 5 to the set-up in Figure 10. Specifically, we consider the same setting as in Section 5, with the addition of Node 3. For the latter, we assume a ternary reconstruction alphabet \( \hat{X}_3 = \{0, 1, *\} \) and the distortion metric \( d_3(x, \hat{x}_3) = d_3(1_{\hat{x}_3=0}, \hat{x}_3) \) in Table 1, where we recall that \( E_j = I(Z, e_j) \) is the erasure process. Accordingly, Node 3 is interested in recovering the erasure process \( E^n \) under an erasure distortion metric (see, e.g. [20]), where \( \ast \) represents the ‘don’t care’ or erasure reproduction symbol.

We first observe that for cases 1) and 3) in Section 5 the distortion requirements of Node 3 do not change the rate distortion function. This is because, as discussed in Section 5, the requirement that \( D_2 \) be equal to zero entails that the erasure process \( E^n \) be communicated losslessly to Node 2 without leveraging the side information from the vending machine (which cannot provide information about the erasure process). It follows that one can achieve \( D_3 = 0 \) at no additional rate cost. We thus now focus on the case 2) in Section 5, namely \( D_1 = 0 \) and \( D_2 = D_{2, max} = 1 - \max\{\epsilon, (1-\epsilon)/2\} \).

In the case at hand, Node 1 wishes to recover \( X^n \) losslessly, Node 2 has no distortion requirements and Node 3 wants to recover \( E^n \) with distortion \( D_3 \). As explained in Section 5.2, to reconstruct \( X^n \) losslessly at Node 1, we must have \( \Gamma \geq \epsilon \) and \( \Pr(A = 1|Z = e) = 1 \). Moreover, due to symmetry of the problem with respect to \( Z = 0 \) and \( Z = 1 \), we can set \( \Pr(A = 1|Z = 0) = \Pr(A = 1|Z = 1) = \frac{\epsilon}{1 - \epsilon} \), for some \( 0 \leq \epsilon \leq \Gamma \). To evaluate the rate distortion-cost region (27), we then define \( \Pr(\hat{X}_3 = *|A = 1, Z = e) \triangleq p_1, \Pr(\hat{X}_3 = *|A = 0, Z = 0) \triangleq p_2 \) and \( \Pr(\hat{X}_3 = *|A = 1, Z = 0) \triangleq p_3 \). We thus obtain the rate distortion-cost region as given by

\[
R_1 \geq H_2(\epsilon) + 1 - \epsilon - (1 - \Gamma)(1 - p_2) - (\Gamma - \epsilon) \\
(1 - p_3) - (1 - \Gamma)p_2 - (\epsilon p_1 + (\Gamma - \epsilon)p_3) \\
\left( H_2\left(\frac{\epsilon p_1}{\epsilon p_1 + (\Gamma - \epsilon)p_3}\right) + \frac{(\Gamma - \epsilon)p_3}{\epsilon p_1 + (\Gamma - \epsilon)p_3}\right)
\]

(30a)

\( \ast \)Infinity in Table 1 means that the corresponding reconstruction is unacceptable at the decoder and is thus measured to be infinity. This is a standard approach to account for errors that are not allowed by design.
As it can be seen, for parameters and Node 3 is able to recover at no additional rate cost. Thus increasing the cost budget is dominated by the distortion requirement of Node 3 and $\frac{p}{c_{128}}$.

Figure 11 illustrates the rate $R_1$ versus cost $\Gamma$ for the examples in Section 6.3 with $\epsilon = 0.2$, $D_1 = 0$ and $D_2 = D_{2,\text{max}}$.

and $R_2 \geq \epsilon$  \hspace{1cm} (30b)

where parameters $p_1, p_2, p_3 \in [0, 1]$ must be selected so as to satisfy the distortion constraint of Node 3, namely $D_3 \geq \epsilon p_1 + (1 - \Gamma) p_2 + (\Gamma - \epsilon) p_3$.

Figure 11 illustrates the rate $R_1$ in (30a), minimised over $p_1, p_2$ and $p_3$ under the constraints mentioned previously versus the cost budget $\Gamma$ for $\epsilon = 0.2$ and different values of $D_3$, namely $D_3 = 0.4, 0.6, 0.8$ and $D_3 = D_{2,\text{max}} = 1$. Note that for $D_3 = D_{2,\text{max}} = 1$ we obtain the rate in (23a). As it can be seen, for $\Gamma \leq D_3$, the rate decreases with increasing cost $\Gamma$, but for $\Gamma \geq D_3$, the rate remains constant while increasing $\Gamma$. The reason is that for the latter region, that is, $\Gamma \geq D_3$, the performance of the system is dominated by the distortion requirement of Node 3 and thus increasing the cost budget $\Gamma$ does not improve the rate. Instead, for $\Gamma \leq D_3$, it is sufficient to cater only to Node 2, and Node 3 is able to recover $E$ with distortion $D_3 = \Gamma$ at no additional rate cost.

7. 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, set-ups in which there are multiple sources communicating to a single receiver in an interactive fashion, or, dually, multiple receivers, each connected to its own vending machine.

APPENDIX A: CONVERSE PROOF FOR PROPOSITION 1 AND PROPOSITION 2

Here, we prove the converse part of Proposition 1. For any $(n, R_1, R_2, D_1 + \epsilon, D_2 + \epsilon, \Gamma + \epsilon)$ code, we have the series of inequalities

$$n R_1 \geq H(M_1) \hspace{1cm} (a)$$

$$H(X^n) + H(Y^n | X^n) - H(Y^n | M_1) = I(M_1; X^n, Y^n) \hspace{1cm} (b)$$

$$\sum_{i=1}^{n} H(Y_i | X^{i-1}, X^n, M_1, A_i) + H(Y_i | Y^{i-1}, X^n, M_1, A_i) - H(Y_i | Y^{i-1}, M_1, A_i) \hspace{1cm} (c)$$

where $(a)$ follows because $M_1$ is a function of $X^n$ and because conditioning reduces entropy; $(b)$ follows because $A_i$ is a function of $(M_1, Y^{i-1})$ and $M_1$ is a function of $X^n$ and $(c)$ follows because conditioning decreases entropy, by defining $U_i = (M_1, X^{i+1}, A_i, Y^{i-1})$ and using the fact that the vending machine is memoryless. We also have the series of inequalities

$$n R_2 \geq H(M_2) \hspace{1cm} (a)$$

$$H(M_2; X^n, M_1) \hspace{1cm} (b)$$

$$\sum_{i=1}^{n} H(Y_i | X^{i-1}, X^n, M_1, A_i) - H(Y_i | X^{i-1}, X^n, M_1, A_i) \hspace{1cm} (c)$$

where $(a)$ follows because $M_2$ is a function of $(M_1, X^n)$, $(b)$ follows because $A_i$ is a function of $(M_i, Y^{i-1})$ and $(c)$ follows because the Markov chain $Y_i - (X_i, U_i, A_i) - X^{i-1}$ holds by the problem definition (the validity of the Markov chain can be verified using d-separation on the Bayesian network representation of the joint distribution of the variables at hand as induced by the system model in Figure A.1, see, for example, [21, Sec. A.9]), because conditioning reduces entropy and by defining $V_i = M_2$. We recall that d-separation [21, Sec. A.9] is a standard procedure that allows to test whether a set of variables $X$ is independent of another
set $Y$, when conditioning on a third set $Z$. The procedure operates on the Bayesian network that describes the joint distribution of all the variables.

Defining $Q$ to be a random variable uniformly distributed over $[1, n]$ and independent of all the other random variables and with $X \triangleq X_Q$, $Y \triangleq Y_Q$, $A \triangleq A_Q$, $\hat{X}_1 \triangleq \hat{X}_1Q$, $\hat{X}_2 \triangleq \hat{X}_2Q$, $V \triangleq (V_Q, Q)$ and $U \triangleq (U_Q, Q)$, from (31), we have

$$nR_1 \geq H(X|Q) - H(X|A, Y, U, Q) + H(Y|X, A, Q) - H(Y|A, Q) \geq H(X) - H(X|A, Y, U) + H(Y|X, A) - H(Y|A) = I(X; A) + I(X; U|A, Y).$$ \hfill (33)

where (a) follows by the fact that source $X^n$ and side information vending machine are memoryless and because conditioning decreases entropy. Next, from (32), we have

$$nR_2 \geq H(Y|X, A, U) - H(Y|X, A, U, V) = I(Y; V|X, A, U).$$ \hfill (34)

Moreover, from Figure A.1 and using d-separation, it can be seen that Markov chains $U_i-(X_i, A_i)-Y_i$ and $V_i-(A_i, U_i, Y_i)-X_i$ hold. This implies that the random variables $(X, Y, A, U, V)$ factorise as in (13).

We now need to show that the estimates $\hat{X}_1$ and $\hat{X}_2$ can be taken to be functions of $(V, X)$ and $(U, Y)$, respectively. To this end, recall that, by the problem definition, the reconstruction $\hat{X}_{1i}$ is a function of $(M_2, X^n)$ and thus of $(X_i, U_i, V_i, X^{i-1})$. Moreover, we can take $\hat{X}_{1i}$ to be a function of $(X_i, U_i, V_i)$ only without loss of optimality, due to the Markov chain relationship $Y_i-(X_i, U_i, V_i)-X^{i-1}$, which can be again proved by d-separation using Figure A.1. This implies that the distortion $d_1(X_i, Y_i, \hat{X}^{i-1})$ cannot be reduced by including also $X^{i-1}$ in the functional dependence of $X^i$. Similarly, the reconstruction $\hat{X}_{2j}$ is a function of $(M_1, Y^n)$ by the problem definition, and can be taken to be a function of $(U_i, Y_i)$ only without loss of optimality, because the Markov chain relationship $X_i-(Y_i, A_i, U_i)-Y^n$ holds. These arguments and the fact that the definition of $V$ and $U$ includes the time sharing variable $Q$ allow us to conclude that we can take $\hat{X}_1$ to be a function of $(U, V, X)$ and $\hat{X}_2$ of $(U, Y)$. We finally observe that $V$ is arbitrarily correlated with $U$ as per (13), and thus, we can without loss of generality set $\hat{X}_1$ to be a function of $(V, X)$ only. The bounds (14) follow immediately from the previous discussion and the constraints (10)–(11).

To bound the cardinality of auxiliary random variable $U$, we observe that (13) factorises as

$$p(x, y, a, u, v) = p(u) p(a, x|u) p(y|a, x) p(v|a, u, y)$$ \hfill (35)

Therefore, for fixed $p(y|a, x)$, $p(a, u|y)$ and $p(v|a, u, y)$ the characterisation in Proposition 1 can be expressed in terms of integrals $\int g_f(x) d\mathcal{F}(u)$, for $f = 1, \ldots, [A] \times [A]+1$, of functions $g_f(.)$ of the given pmfs. Specifically, we have $g_f$ for $f = 1, \ldots, [A] \times [A]+1$, given by $p(a, x|u)$ for all values of $x \in X$ and $a \in A$, and by the characterisation in Proposition 1. We finally observe that $V$ is arbitrarily correlated with $U$ as per (13), and thus, we can without loss of generality set $\hat{X}_1$ to be a function of $(V, X)$ only. The bounds (14) follow immediately from the previous discussion and the constraints (10)–(11).

To bound the cardinality of auxiliary random variable $V$, we note that (13) can be factorised as

$$p(x, y, a, u, v) = p(v) p(a, u, y) p(y|a, x) p(x|a, u, y)$$ \hfill (36)

so that, for fixed $p(x|a, u, y)$, the characterisation in Proposition 1 can be expressed in terms of integrals $\int g_f(x) p(a, u|y) d\mathcal{F}(v)$, for $f = 1, \ldots, [A] \times [U]+1$, of functions $g_f(.)$ that are continuous on the space of probabilities over alphabet $[A] \times [U] \times [\mathcal{Y}]$. Specifically, we have $g_f$ for $f = 1, \ldots, [A] \times [U]+1$, given by $p(a, u, y)$ for all values of $a \in A$, $u \in U$ and $y \in \mathcal{Y}$ (except one); $g_{[A] \times [U]+1} = H(Y|X, A, U); g_{[A] \times [2]+1} = I(Y; V|A, X, U); g_{[A] \times [3]+1} = E[d_1(X, Y, f_1(V, X))|U = u]$ and $g_{[A] \times [4]+1} = E[d_2(X, Y, f_2(U, Y))|U = u]$. The proof is concluded by invoking Fenchel–Eggleston–Caratheodory Theorem [1, Appendix C].

The converse for Proposition 2 follows similar steps as mentioned earlier with the only difference that here we have

$$nR_1 \geq \sum^n_{i=1} H(Z_i) - H(Z_i|Z_{i+1}^n, Y^n, M_1, A_i') + H(Y_i|Y^{i-1}, Z^n, M_1, A_i') - H(Y_i|Y^{i-1}, M_1, A_i)$$

$$\geq \sum^n_{i=1} H(Z_i) - H(Z_i|A_i, Y_i, U_i) + H(Y_i|Z_i, A_i') - H(Y_i|A_i)$$ \hfill (37)
where (a) follows as in (a)–(b) of (31); and (b) follows because Markov chain relationship $Y_i - (Z_1, A_i) - (Y_{i-1}, Z^n, y, M_1)$ holds. The rest of the proof is mentioned earlier.

**APPENDIX B: PROOFS FOR THE EXAMPLE IN SEC. 5**

(1) $D_1 = D_{1, \text{max}}$ and $D_2 = 0$:

Here, we prove that the rate-cost region in Proposition 2 is given by (22) for $D_1 = D_{1, \text{max}}$ and $D_2 = 0$. We begin with the converse part. Starting from (18a), we have

$$R_1 \overset{(a)}{=} I(A; Z) + H(Z | A, Y)$$
$$= H(Z) - I(Z; Y | A)$$
$$\overset{(b)}{=} H(Z) - \Gamma I(Z; X | A = 1)$$
$$\overset{(c)}{=} H(Z) - \Gamma H(X | A = 1)$$
$$\overset{(d)}{=} H(Z) - \Gamma$$
$$\overset{(e)}{=} H_2(\epsilon) + 1 - \epsilon - \Gamma$$

where (a) follows from (18a) and because $Z$ has to be recovered losslessly at Node 2; (b) follows because $\Pr[A = 1] = \mathbb{E}[A | A] \leq \Gamma$; (c) follows because entropy is non-negative; (d) follows because $H(X | A = 1) \leq 1$; and (e) follows 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 $\Pr(A = 0 | Z = 1) = 1$ in (18).

(2) $D_1 = 0$ and $D_2 = D_{2, \text{max}}$:

Here, we turn to the case $D_1 = 0$ and $D_2 = D_{2, \text{max}}$. We start with the converse. Because $X$ is to be reconstructed losslessly at Node 1, we have the requirement $H(X | V, Z) = 0$ from (20a). It is easy to see that this requires that the equalities $A = 1$ and $V = Y = X$ are met if $Z = e$. 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 $Z = e$ requires that the pmf $p(a | z)$ be such that $\Pr(A = 1 | Z = e) = 1$, which entails $\gamma = \Pr[A = 1] \geq \Pr[Z = e] \geq \epsilon$. 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 (18a), we can thus write

$$R_1 \overset{(a)}{=} I(Z; A)$$
$$= H(Z) - H(Z | A)$$
$$= H_2(\epsilon) + 1 - \epsilon - \eta H(Z | A = 1)$$
$$- (1 - \gamma) H(Z | A = 0)$$
$$\overset{(b)}{=} H_2(\epsilon) + 1 - \epsilon - \eta H\left(\frac{\epsilon}{\gamma}, \frac{\gamma - \epsilon}{2\gamma}, \frac{\gamma - \epsilon}{2\gamma}\right)$$
$$- (1 - \gamma)$$
$$= H_2(\epsilon) - \eta H_2\left(\frac{\epsilon}{\gamma}\right)$$
$$\overset{(c)}{=} H_2(\epsilon) - \Gamma H_2\left(\frac{\epsilon}{\Gamma}\right)$$

where (a) follows from (18a) and because there is no distortion requirement at Node 2; (b) follows by direct calculation; and (c) follows since $H_2(\epsilon) - \gamma H_2\left(\frac{\epsilon}{\gamma}\right)$ is minimised at $\gamma = \Gamma$ over all $0 \leq \gamma \leq \Gamma$.

The bound (38) follows immediately by providing Node 2 with the sequence $X^n$ and then using the bound $R_1 \geq H(X | Z) = \epsilon$.

Achievability follows by setting $U = Z$, $V = \emptyset$, and $\Pr(A = 1 | Z = 0) = \Pr(A = 1 | Z = 1) = \Gamma / (1 - \epsilon)$ and $\Pr(A = 0 | Z = 1) = 1$ in (18).

(3) $D_1 = D_2 = 0$:

Here, we prove the rate-cost region (24) for the case $D_1 = D_2 = 0$. Starting from (18a), we have

$$R_1 \overset{(a)}{=} I(Z; A)$$
$$= H(Z) - H(Z | A)$$
$$= H_2(\epsilon) + 1 - \epsilon - \eta H(Z | A = 1)$$
$$- (1 - \gamma) H(Z | A = 0)$$
$$\overset{(b)}{=} H_2(\epsilon) + 1 - \epsilon - \eta H\left(\frac{\epsilon}{\gamma}, \frac{\gamma - \epsilon}{2\gamma}, \frac{\gamma - \epsilon}{2\gamma}\right)$$
$$- (1 - \gamma)$$
$$= H_2(\epsilon) - \eta H_2\left(\frac{\epsilon}{\gamma}\right)$$
$$\overset{(c)}{=} H_2(\epsilon) - \Gamma H_2\left(\frac{\epsilon}{\Gamma}\right)$$

where (a) follows as in (38); (b) follows because $H(X | A = 1, Z = 0) = H(X | A = 1, Z = 1) = 0$; (c) follows because $H(X | A = 1) \leq 1$, $H(X | A = 1, Z = e) = 1$ and because $p(Z = e | A = 1) = \frac{\epsilon}{\Gamma}$, where latter follows from the requirement $H(X | V, Z) = 0$ as per discussion provided in the previous section.

For the achievability, let $U = Z$, $\Pr(A = 1 | Z = e) = 1$ and $\Pr(A = 1 | Z = 0) = \Pr(A = 1 | Z = 1) = \frac{\Gamma - \epsilon}{\Gamma}$, where

Moreover, let $V = Y = X$ if $Z = e$ and $V = Y = 0$ otherwise. Evaluating (18) with these choices leads to (24).

---

This is due to the fact that, by the problem definition, the events $Z = 0$ and $Z = 1$ are statistically equivalent, and hence, there is no advantage in mapping $Z = 0$ to $A = 1$ with higher probability than mapping $Z = 1$ to $A = 1$ and vice versa.
APPENDIX C: CONVERSE PROOF FOR PROPOSITION 3

Here, we prove the converse part of Proposition 3. For any \((n, R_1, R_2, D_1 + \epsilon, D_2 + \epsilon, D_3 + \epsilon, \Gamma + \epsilon)\) code, we have the series of inequalities

\[
n R_1 \geq H(M_1) \geq \sum_{i=1}^{n} H(Z_i) - H(Z_i|Z_{i+1}, Y^n, M_i, A_i, \hat{X}_{3i}) + H(Y_i|Y^{i-1}, X^n, M_i, A_i, \hat{X}_{3i}) - H(Y_i|Y^{i-1}, M_i, A_i, \hat{X}_{3i}) \geq \sum_{i=1}^{n} H(Z_i) - H(Z_i|A_i, Y_i, U_i, \hat{X}_{3i}) + H(Y_i|Z_i, A_i, \hat{X}_{3i}) - H(Y_i|Z_i, A_i, U_i, V_i, \hat{X}_{3i}) \geq H(Z|Q) - H(Z|A, Y, U, \hat{X}_{3}, Q) + H(Y|Z, A, \hat{X}_{3}, Q) - H(Y|A, \hat{X}_{3}, Q) \geq H(Z) - H(Z|A, Y, U, \hat{X}_{3}) + H(Y|Z, A, \hat{X}_{3}) - H(Y|A, \hat{X}_{3}) = I(Z; A) + I(Z; \hat{X}_{3}|A) + I(Z; U|A, Y, \hat{X}_{3})
\]

where \((a)\) follows from \((37)\) by noting that \(\hat{X}_{3i}\) is a function of \(M\) and \((b)\) follows because conditioning decreases entropy, by defining \(U_i = (M_i, X_{i+1}^n, A_i^{i-1}, Y^{i-1})\) and using the Markov chain relationships \(Y_i 
\rightarrow (Z_i, A_i, \hat{X}_{3i}) \rightarrow (Y^{i-1}, X^n, M_i)\). We also have the series of inequalities

\[
n R_2 \geq H(M_2) \geq \sum_{i=1}^{n} H(Y_i|Z_i, A_i, U_i, \hat{X}_{3i}) - H(Y_i|Z_i, A_i, U_i, V_i, \hat{X}_{3i}) \geq H(Z|Q) - H(Z|A, Y, U, \hat{X}_{3}, Q) + H(Y|Z, A, \hat{X}_{3}, Q) - H(Y|A, \hat{X}_{3}, Q) \geq H(Z) - H(Z|A, Y, U, \hat{X}_{3}) + H(Y|Z, A, \hat{X}_{3}) - H(Y|A, \hat{X}_{3}) = I(Z; A) + I(Z; \hat{X}_{3}|A) + I(Z; U|A, Y, \hat{X}_{3})
\]

where \((a)\) follows by the fact that source \(Z^n\) and side information vending machine are memoryless and because conditioning decreases entropy. Next, from \((43)\), we have

\[
n R_1 \geq H(Z) - H(Z|A, Y, U, \hat{X}_{3}) + H(Y|Z, A, \hat{X}_{3}) - H(Y|A, \hat{X}_{3}) = I(Z; A) + I(Z; \hat{X}_{3}|A) + I(Z; U|A, Y, \hat{X}_{3})
\]

Moreover, by just adding \(\hat{X}_{3}\) to the Bayesian graph in Figure A.1 and using d-separation, it can be seen that Markov chains \(U_i \rightarrow (Z_i, A_i) \rightarrow Y_i\) and \(V_i \rightarrow (A_i, U_i, Y_i, \hat{X}_3) \rightarrow Z_i\) hold, which implies that the random variables \((X, Y, Z, A, U, V, \hat{X}_3)\) factorise as in \((28)\). On the basis of the discussion in the converse proof in Appendix A, it is easy to see that the estimates \(\hat{X}_{1}\) and \(\hat{X}_{2}\) are functions of \((V, X)\) and \((U, Y)\), respectively. The bounds \((27)\) follow immediately from the previous discussion and the constraints \((10)–(11)\) and \((26)\).

ACKNOWLEDGEMENT

The work of O. Simeone is supported by the U.S. National Science Foundation under grant CCF-0914899.

REFERENCES

1. El Gamal A, Kim Y. Network Information Theory. Cambridge University Press: New York, NY, 2011.
2. Permuter H, Weissman T. Source coding with a side information “vending machine”. IEEE Transactions on Information Theory 2011; 57: 4530–4544.
3. Chia Y, Asnani H, Weissman T. Multi-terminal source coding with action dependent side information, In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011), July 31–August 5, Saint Petersburg, Russia, 2011; 2035–2039.
4. Ahmadi B, Simeone O. Robust coding for lossy computing with receiver-side observation costs, In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011), July 31–August 5, Saint Petersburg, Russia, 2011; 2939–2943.
5. Ahmadi B, Simeone O. Distributed and cascade lossy source coding with a side information “vending machine”. Available from: http://arxiv.org/abs/1109.6665 [February 2013].
6. Ahmadi B, Choudhuri C, Simeone O, Mitra U. Cascade source coding with a side information “vending machine”. Available from: http://arxiv.org/abs/1207.2793 [July 2012].
7. Wyner A, Ziv J. The rate-distortion function for source coding with side information at the decoder. IEEE Transactions on Information Theory 1976; 22(1): 1–10.
8. Zhao L, Chia YK, Weissman T. Compression with actions, In Proceedings of the Allerton Conference on Communications, Control and Computing, Monticello, Illinois, September 2011; 164–171.
9. Kittichokechai K, Oechtering T, Skoglund M. Coding with action-dependent side information and additional reconstruction requirements. Available from: http://arxiv.org/abs/1202.1484 [February 2012].
10. Kittichokechai K, Oechtering T, Skoglund M. Secure source coding with action-dependent side information, In Proceedings of the IEEE International Symposium on
11. Heegard C, Berger T. Rate distortion when side information may be absent. *IEEE Transactions on Information Theory* 1985; 31(6): 727–734.

12. Kaspi A. Rate-distortion when side-information may be present at the decoder. *IEEE Transactions on Information Theory* 1994; 40(6): 2031–2034.

13. Berger T, Yeung R. Multiterminal source encoding with one distortion criterion. *IEEE Transactions on Information Theory* 1989; 35: 228–236.

14. Steinberg Y. Coding and common reconstruction. *IEEE Transactions on Information Theory* 2009; 55(11): 4995–5010.

15. Kaspi AH. Two-way source coding with a fidelity criterion. *IEEE Transactions on Information Theory* 1985; 31(6): 735–740.

16. Permuter H, Steinberg Y, Weissman T. Two-way source coding with a helper. *IEEE Transactions on Information Theory* 2010; 56(6): 2905–2919.

17. Ma N, Ishwar P. Some results on distributed source coding for interactive function computation. *IEEE Transactions on Information Theory* 2011; 57(9): 6180–6195.

18. Choudhuri C, Mitra U. How useful is adaptive action? In *Proceedings of the IEEE GLOBECOM 2012*, Anaheim, CA, 2012; 2251–2255.

19. Witsenhausen HS. Indirect rate distortion problems. *IEEE Transactions on Information Theory* 1980; IT-26: 518–521.

20. Weissman T, Verdu S. The information lost in erasures. *IEEE Transactions on Information Theory* 2008; 54(11): 5030–5058.

21. Kramer G. *Topics in Multi-User Information Theory*. Now Publishers: Hanover, MA, 2008.