Lossless Coding of Correlated Sources With Actions

Oron Sabag, *Student Member, IEEE*, Haim H. Permuter, *Senior Member, IEEE*, and Asaf Cohen, *Member, IEEE*

***Abstract*— This paper studies the problem of the distributed compression of correlated sources with an action-dependent joint distribution. This class of problems is, in fact, an extension of the Slepian–Wolf model, but where cost-constrained actions taken by the encoder or the decoder affect the generation of one of the sources. The purpose of this paper is to study the impact of actions on the achievable rates. In particular, two cases where transmission occurs over a rate-limited link are studied; case A for actions taken at the decoder and case B where actions are taken at the encoder. A complete single-letter characterization of the set of achievable rates is given in both cases. Furthermore, a network coding setup for the case where actions are taken at the encoder is investigated. The sources are generated at different nodes of the network and are required at a set of terminal nodes, yet transmission occurs over a general, acyclic, directed network. For this setup, generalized cut-set bounds are derived, and a full characterization of the set of achievable rates using single-letter expressions is provided. For this scenario, random linear network coding is proved to be optimal, even though this is not a classical multicast problem. In addition, two binary examples are investigated and demonstrate how actions taken at different nodes of the system have a significant effect on the achievable rate region, when compared with a naive time-sharing strategy.**

***Index Terms*— Actions, correlated sources, distributed compression, network coding, random linear network coding, Slepian-wolf source coding.**

I. Introduction

The field of distributed encoding and joint decoding of correlated information sources is fundamental in information theory. In their seminal work, Slepian and Wolf (SW) [1] showed that the total rate used by a system which distributively compresses correlated sources is equal to the rate that is used by a system that performs joint compression. An extension of this model for general networks was studied by Ho *et al.* [2], who showed that this property is maintained, using a novel coding scheme called Random Linear Network Coding (RLNC).

Manuscript received April 16, 2014; revised September 16, 2015; accepted August 19, 2016. Date of publication September 16, 2016; date of current version January 18, 2017. O. Sabag and H. H. Permuter were supported in part by the Israel Science Foundation and in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC under Grant 337752. A. Cohen was supported by the European Union Horizon 2020 Research and Innovation Programme SUPERFLUIDITY under Grant 671566. This paper was presented at the 2014 IEEE International Symposium on Information Theory.

O. Sabag and H. H. Permuter are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beersheba 8410501, Israel (e-mail: oronsa@post.bgu.ac.il; haimp@bgu.ac.il).

A. Cohen is with the Department of Communication Systems Engineering, Ben-Gurion University of the Negev, Beersheba 8410501, Israel (e-mail: coasaf@bgu.ac.il).

Communicated by S. S. Pradhan, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2610432

In past studies, the joint distribution of sources has been perceived as given by nature; however, we pose the question, what if the system can take actions that affect the generation of sources?

We model the system describing this scenario as *correlated sources with actions* with the following source distribution: the source X is a memoryless source that is distributed according to P_X, while the other source, Y, has a memoryless conditional distribution, $P_{Y|X,A}$, that is conditioned on the source X and an action A. The actions are constrained according to a given cost function that determines the nature of actions in the system.

Two possible motivating scenarios for the suggested setting are the following:

1) Consider a sensor network where two measurements of temperature are required at a set of terminal nodes. The first measurement determines whether the temperature is above or below a coarse threshold value and is generated by nature; the second measurement may be either a repeated coarse measurement or a new independent measurement that corresponds to a finer resolution of temperature within the correct coarse region. The system is constrained such that a minimum percentage of the second sensor measurements are allocated to fine measurements.
2) A backup system consists of two sub-units that may back up data on a local device or not. The first unit is controlled ahead of time and thus we have no control over its operation. The operation of the second backup unit can be controlled via actions that effect the nature of its operation. The actions that are taken by the system subject to the observed operations of the first unit. Here, the system is constrained such that the actions that are taken by the units are different during a fraction of the operation time. Lastly, details of the operation of the two sub-units should be transmitted to several decentralized nodes, which may need this information in the case of a recovery requirement.

The information-theoretical question that naturally arises from these examples is how the system should choose actions in order to minimize the transmission rates, subject to the given constraints. The given scenarios will serve as a basis for two examples that will be studied in Section IV, where it is shown that actions might affect the rate region in a non-trivial manner.

In this paper, two concepts are covered regarding the investigated model; the first is a classical multi-user setup where transmission occurs over rate-limited links. Here, actions can be performed at different nodes of the system: in case A actions are taken at the decoder (Fig. 1), and in case B actions are taken at the encoder, as described (Fig. 2). In the second approach, we extend the transmission scenario from

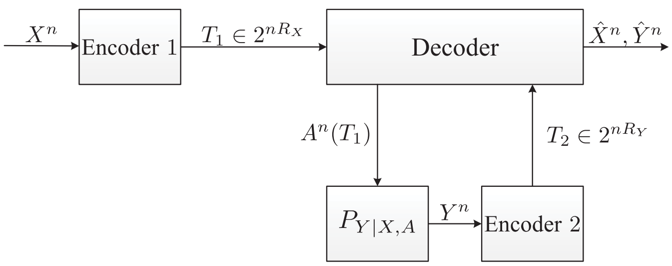

Fig. 1. Case A - Correlated sources with actions taken at the decoder. The actions are based on the index T_1 sent by encoder 1 and affect the generation of the source Y^n.

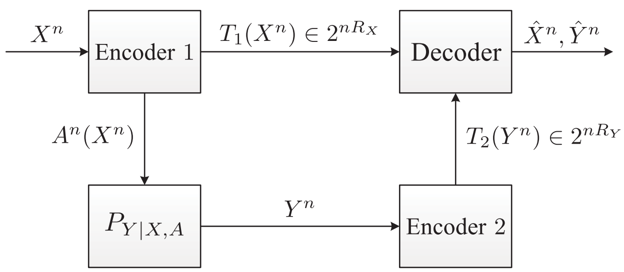

Fig. 2. Case B - Correlated sources with actions taken at the encoder. The actions are based on the source X^n and affect the generation of the source Y^n.

rate-limited links to a *given directed, acyclic network*. This approach is motivated by the needs of large-scale networks that are used to communicate source information to multiple end-nodes.

The first case we consider is depicted in Fig. 1, where actions are taken at the decoder: based on its source observation X^n, which is independent and identically distributed (i.i.d.) according to $\sim P_X$, encoder 1 gives an index $T_1(X^n)$ to the decoder. Having received the index T_1, the decoder chooses the action sequence A^n. Nature then generates the other source sequence, Y^n, which is the output of a discrete memoryless channel $P_{Y|X,A}$, whose input is the pair (X^n, A^n). Based on its observation Y^n, an index $T_2(Y^n)$ is sent to the decoder by encoder 2. The reconstruction sequences $(\hat{X}^n, \hat{Y}^n)$ are then generated at the decoder, based on the indices that were given by the encoders. For this case, a single-letter characterization of the optimal rate region is presented in Theorem 1.

The second case we consider is depicted in Fig. 2, where actions are taken at encoder 1: based on its source observation X^n, which is i.i.d. $\sim P_X$, the first encoder chooses an action sequence A^n. The other source, Y^n, is then generated as in case A and is made available at encoder 2. Each encoder now chooses an index to be given to the decoder, based on its source observation. The reconstruction sequences $(\hat{X}^n, \hat{Y}^n)$ are then generated at the decoder based on the indices that were given by the encoders. This case is found to display better performance than case A, which is is intuitively clear since in case A actions are constrained to be a function of T_1, while in case B actions are a function of the explicit source X^n. In Theorem 2, the optimal rate region for this case is characterized by single-letter expressions. In Section IV, we demonstrate and prove, in two binary examples, that performing actions at the encoder or the decoder might affect the optimal rate region, and also display a significant advantage compared to a naive time-sharing strategy.

The problem of multi-terminal source coding is fundamental in information theory and is still an open problem [3], [4]; among its fully-characterized special cases are the Wyner-Ziv (WZ) [5], Ahlswede-Korner (AK) [6] and SW [1] settings. The connection of this open problem and actions was first introduced in [7] by considering an action-dependent WZ setting. This setting was useful for modelling a scenario where the user can perform actions that affect the quality of the side information (SI) available to the decoder. This model was termed the Vending Machine (VM), and actions could be performed at the encoder or the decoder. In the current paper, the action-dependent version of the SW problem is studied, and we also remark on the difficulty when considering an action-dependent AK problem in Section VII. From the operational point, the extended SW and the AK problems differ from the VM mainly in the multi-user nature of their setting; specifically, their setting consists of two rates which imply a two-dimensional rate region.

In the settings that are investigated here and in the VM, the action-dependent sequence plays a different role; in our current setting, this sequence is a source of information and, thus, we also have a reconstruction constraint on the source sequence at the decoder, while in the VM the affected sequence is a SI that is available at the decoder. Moreover, in both settings the generated source, Y^n, is not distributed i.i.d. since the actions' sequence should not admit the i.i.d. property. Hence, our setting deals with the lossless compression of the source Y^n that might have a memory, whereas in the VM it is made available to the decoder by definition. Despite the different roles of the action-dependent sequence, one can find the following similarity between the models: if we take the link between encoder 2 and the decoder to be of infinite capacity, it follows that the sequence Y^n is available to the decoder and, as a by-product, the reconstruction constraint is satisfied. For this special case, our problem is reduced to the lossless transmission of the source X to the decoder, that is, the VM setting.

The optimal coding of case A is given by a simple and intuitive scheme that reveals the SW multi-user nature of our setting. In the achievability proof of case B, we utilize the VM coding together with a lossless transmission of the source Y^n to achieve one of the corner points and the other corner point is achieved by applying standard lossless coding arguments. For case B, we also provide an alternative achievability proof that is independent of the VM coding. Specifically, as case B is a special case of the investigated network, we obtain an alternative proof directly from the novel coding scheme that is provided for the general network scenario and is based on RLNC.

Various studies on actions in source coding can be found in the literature: in [8], Zhao *et al.* studied a model where an action-dependent source is generated and reconstruction of the source is required at terminal node. In [9], Simeone considered a VM model, but with sources that are not memoryless and with actions that might also be affected by causal

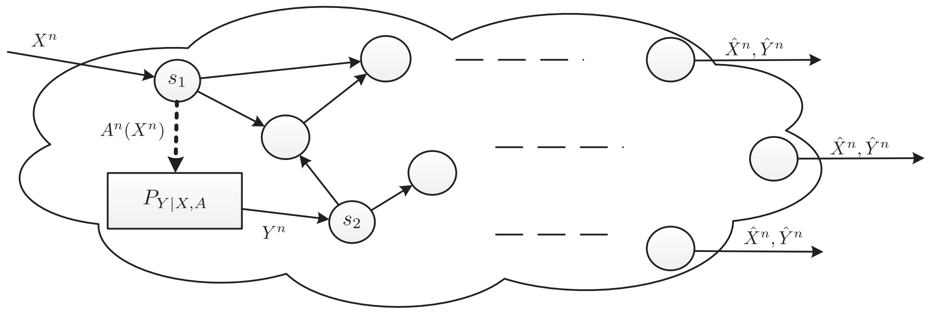

Fig. 3. Correlated sources in general networks with actions. Based on source X^n, node s_1 performs actions that affect the generation of Y^n. Transmission of the encoded sources occurs over an arbitrary acyclic directed network. Both sources are required at a set of terminal nodes. Note, the dashed arrow is the actions' cost-constrained link.

observation of the SI. In [10], Kittichokechai *et al.* considered a model where actions affect the generation of two-sided SI sequences; one is available to the encoder and the other one to the decoder. In [11], Ahmadi *et al.* studied a new role of actions, where an additional decoder observes a function of the actions and should reconstruct the source information. In [12], Chia *et al.* studied a multi-user setup of the VM; two decoders can observe different SI sequences, where both sequences are generated according to the same performed actions. More studies on actions in channel coding can be found in [13]–[16]. In all the cited papers, actions were proved to be efficient, not only for modeling interesting scenarios, but also for improving the transmission rates. However, to the best of our knowledge, actions have not been previously studied in a general network coding setup.

In the network scenario, the case where actions are taken at the encoder is investigated (Fig. 3). The nodes s_1 and s_2 play the role of the encoders, as in case B, and source generation remains the same. However, transmission occurs over a general, acyclic, directed network. Each link in the network has a known capacity, which represents a noiseless link in units of bits per unit time. All intermediate nodes in the network are allowed to perform encoding based on the messages on their input links, and a set of destination nodes, $\mathcal{D}$, is required to reconstruct both sources in a lossless manner. The single-letter characterization for the set of achievable rates for this problem is derived by, first, introducing the *generalized cut-set bounds* for this setting and, then, the tightness of these bounds is completed by providing a novel coding scheme that combines techniques of random coding, random binning and RLNC.

The optimal coding scheme for this setting is based on the random generation of the actions' codebook together with RLNC for the transmission over the network. Throughout the direct proof we differentiate between two cases; specifically, the cases are determined by the sign of the expression $I(X; A) - I(Y; A)$. The construction of the actions at the encoder will be similar to a WZ coding when the sign is positive, and to the Gel'fand and Pinsker [17] scheme for a negative sign. When both sources are generated, RLNC in a finite field is used for transmission in the network as in [2]. For our case, the inputs to the network are sources together with the chosen actions' sequence.

The decoding procedure that was applied in [2] was based on min-entropy or maximum a posteriori procedures that exploit the i.i.d. nature of the sources. In our setting, the triplet (X^n, A^n, Y^n) is not i.i.d. since actions are a function of the complete source sequence X^n and, therefore, we adopt a strong typicality decoding procedure. Moreover, the analysis of the probability of error for the proposed coding required a derivation of an upper bound on the probability that two different inputs to a randomized linear network induce the same output at a receiver node. Based on the result in [2, Appendix A], we derive an upper bound on the probability of this event that can fit a broader class of network coding problems, including our problem. The generalized upper bound appears in Lemma 1, followed by an alternative proof and, finally, it is demonstrated through an example how this upper bound is an efficient tool when solving network coding problems.

It is known that linear network coding is optimal in multicast problems, as shown by Ahlswede *et al.* [18]. Following this result, the RLNC technique was introduced by Ho *et al.* in [2] for a model of the compression of correlated sources' over an arbitrary network. Our model *does not fall into the class of multicast problems* since no requirement for actions reconstruction is defined, yet it is very clear that the actions taken affect the rate region. Moreover, our set of achievable rates includes terms of mutual information, which are not typical in multicast problems. Nevertheless, we prove that RLNC achieves optimality in our network model as well.

The remainder of the paper is organized as follows. In Section II, we formulate the problem for all communication models. Section III summarizes our main results regarding the optimal rate regions for case A, case B and the set of achievable rates for the general network scenario. Section IV presents two binary examples. Section V outlines the proofs of case A and case B. A detailed proof for the network coding scenario is provided in Section VI. Finally, Section VII summarizes the main achievements and insights presented in this work along with suggestions for possible future work.

II. Notation and Problem Definition

Let $\mathcal{X}$ be a finite set, and let $\mathcal{X}^n$ denote the set of all n-tuples of elements from $\mathcal{X}$. An element from $\mathcal{X}^n$ is denoted by $x^n = (x_1, x_2, \dots, x_n)$. Random variables are denoted by uppercase letters, X, and the previously mentioned notation holds also here, e.g. $X^n = (X_1, X_2, \dots, X_n)$. The probability mass function of X, the joint distribution function of X and Y, and the conditional distribution of X given Y will be denoted by P_X, $P_{X,Y}$ and $P_{X|Y}$, respectively. The notation $\lceil x \rceil$ stands for the smallest integer greater than or equal to x and, lastly, $\bar{\alpha}$ stands for $1 - \alpha$, with $\alpha \in [0, 1]$.

We consider a system of correlated sources with actions. Let us refer to the case where the decoder is allowed to perform actions as case A and to the case where encoder 1 performs actions as case B. We provide here a definition for the setting of case A, while the definition for the setting of case B is straightforward. The source sequence X^n is such that $X_i \in \mathcal{X}$ for $i \in [1, n]$ and is distributed i.i.d. with a pmf P_X. The first encoder measures a sequence X^n and encodes it in a message $T_1 \in \{1, \dots, 2^{nR_X}\}$, which is transmitted to the decoder. The decoder receives the index T_1 and selects an action sequence, where $A^n \in \mathcal{A}^n$. The action sequence affects the generation of the other source sequence Y^n, which is the output of a discrete memoryless channel without feedback $P_{Y|X,A}$ with inputs of (X^n, A^n). Specifically, given $X^n = x^n$ and $A^n = a^n$, the source sequence Y^n is distributed as

$$p(y^n|x^n, a^n) = \prod_{i=1}^{n} p(y_i|x_i, a_i). \tag{1}$$

Encoder 2 receives the observation y^n and encodes it in a message $T_2 \in \{1, \dots, 2^{nR_Y}\}$. The estimated sequences $(\hat{X}^n, \hat{Y}^n)$ are then obtained at the decoder as a function of the messages T_1 and T_2.

For the settings described above, a $(2^{nR_X}, 2^{nR_Y}, n)$ *code* for a block of length n and rate pairs (R_X, R_Y) consists of encoding functions:

$$T_1 : \mathcal{X}^n \to [1 : 2^{nR_X}],$$
$$T_2 : \mathcal{Y}^n \to [1 : 2^{nR_Y}],$$

strategy functions:

$$h_d : [1 : 2^{nR_X}] \to \mathcal{A}^n \text{ for case A,}$$
$$h_e : \mathcal{X}^n \to \mathcal{A}^n \text{ for case B,} \tag{2}$$

and a decoding function:

$$g : [1 : 2^{nR_X}] \times [1 : 2^{nR_Y}] \to \widehat{\mathcal{X}}^n \times \widehat{\mathcal{Y}}^n.$$

Actions taken are subject to a cost constraint Γ, that is,

$$E\left[\frac{1}{n}\sum_{i=1}^{n} \Lambda(A_i)\right] \le \Gamma. \tag{3}$$

The *probability of error* for a code $(2^{nR_X}, 2^{nR_Y}, n)$ is defined as $P_e^{(n)} = \Pr((X^n, Y^n) \neq g(T_1, T_2))$. For a given cost constraint Γ, a rate pair (R_X, R_Y) is said to be *achievable* if there exists a sequence of codes $(2^{nR_X}, 2^{nR_Y}, n)$ such that $P_e^{(n)} \to 0$ as $n \to \infty$ and the cost constraint, (3), is satisfied. The *optimal rate region* is the convex closure of the set of achievable rate pairs. Let us denote the optimal rate regions as $\mathcal{R}_A$ and $\mathcal{R}_B$ for case A and case B, respectively.

A. Network Model

A network is represented as a directed, acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of network nodes and $\mathcal{E}$ is the set of links, such that information can be sent noiselessly from node i to node j if $(i, j) \in \mathcal{E}$. Each edge $l \in \mathcal{E}$ is associated with a nonnegative real number c_l, which represents its capacity in bits per unit time. We also denote the origin node of the link l as $\sigma(l)$ and the destination of a link l as $\tau(l)$.

We specify a *network of correlated sources with actions* $(\mathcal{V}, \mathcal{E}, s_1, s_2, \mathcal{D})$ as follows. The source sequence X^n is such that $X_i \in \mathcal{X}$ for $i \in [1, n]$ is i.i.d. with a pmf P_X. Based on its source observation X^n, node $s_1 \in \mathcal{V}$ selects an action sequence $A^n \in \mathcal{A}^n$. The action sequence affects the generation of the other source sequence Y^n as specified in (1). The source sequence Y^n is available at node $s_2 \in \mathcal{V} \setminus \{s_1\}$. The source sequences (X^n, Y^n) are demanded at a set of terminal nodes denoted as $\mathcal{D} \subseteq \mathcal{V} \setminus \{s_1, s_2\}$. We assume that the source nodes s_1, s_2 have no incoming links and that each node $t \in \mathcal{D}$ has no outgoing links.

For any vector of rates $(R_l)_{l \in \mathcal{E}}$, a $\left(\left(2^{nR_l}\right)_{l \in \mathcal{E}}, n\right)$ *source code* consists of a strategy function:

$$h : \mathcal{X}^n \to \mathcal{A}^n,$$

encoding functions:

$$g_l : \mathcal{X}^n \to [1 : 2^{nR_l}] \forall l \in \mathcal{E}, \quad \sigma(l) = s_1,$$
$$g_l : \mathcal{Y}^n \to [1 : 2^{nR_l}] \forall l \in \mathcal{E}, \quad \sigma(l) = s_2,$$
$$g_l : \textstyle\prod_{l' : \tau(l') = \sigma(l)} [1 : 2^{nR_{l'}}]$$
$$\to [1 : 2^{nR_l}] \forall l \in \mathcal{E}, \quad \sigma(l) \notin \{s_1, s_2\},$$

and decoding functions, for each $t \in \mathcal{D}$:

$$\phi_t : \textstyle\prod_{l : \tau(l) = t} \{1, \dots, 2^{nR_l}\} \to \hat{\mathcal{X}}^n \times \hat{\mathcal{Y}}^n.$$

We are interested in the set of possible values $(c_l)_{l \in \mathcal{E}}$, such that for any $\epsilon > 0$ there exists a sufficiently large n and a $\left(\left(2^{nR_l}\right)_{l \in \mathcal{E}}, n\right)$ code with $R_l \le c_l$ for all $l \in \mathcal{E}$, such that $\max_{t \in \mathcal{D}} \Pr((\hat{X}_t^n, \hat{Y}_t^n) \neq (X^n, Y^n)) \ge 1 - \epsilon$ and $E\left[\frac{1}{n}\sum_{i=1}^{n} \Lambda(A_i)\right] \le \Gamma$. We call the closure of this set the *set of the achievable rates*, which we denote by $\mathcal{R}_N$.

Given any set $A \subset \mathcal{V}$ and a node $t \in \mathcal{V} \setminus A$, a *cut* $\mathcal{V}_{A;t}$ is a subset of vertices that includes A but is disjoint from t, that is, $A \subseteq \mathcal{V}_{A;t}$ and $\mathcal{V}_{A;t} \cap t = \emptyset$. Given a cut $\mathcal{V}_{A;t}$, the *capacity of a cut* $\mathcal{C}(\mathcal{V}_{A;t})$ is the sum over all capacities of edges $l \in \mathcal{E}$ such that $\sigma(l) \in \mathcal{V}_{A;t}$ and $\tau(l) \notin \mathcal{V}_{A;t}$; that is,

$$\mathcal{C}(\mathcal{V}_{A;t}) = \sum_{l \in \mathcal{E} : \sigma(l) \in \mathcal{V}_{A;t}, \tau(l) \notin \mathcal{V}_{A;t}} c_l.$$

For a given set A and a node t, let $\mathcal{V}_{A;t}^*$ be the *minimum cut*, which is the cut minimizes the capacity of a cut among all cuts $\mathcal{V}_{A;t}$. Finally, for given non-intersecting sets $A, \mathcal{D}$ we define $\mathcal{C}(\mathcal{V}_{A;\mathcal{D}}^*) = \min_{t \in \mathcal{D}} \mathcal{C}(\mathcal{V}_{A;t}^*)$.

III. Main Results

This section concerns with the main results of this paper. Specifically, the next two theorems state the optimal rate regions of case A and case B.

Theorem 1: The optimal rate region $\mathcal{R}_A$ for case A (See Fig. 1), i.e. correlated sources with actions taken at the decoder, is the closure of the set of triplets (R_X, R_Y, Γ) such that

$$R_X \geq H(X|Y,A) + I(X;A), \tag{4a}$$

$$R_Y \geq H(Y|X,A), \tag{4b}$$

$$R_X + R_Y \geq H(X,Y|A) + I(X;A), \tag{4c}$$

where the joint distribution of (X, A, Y) is of the form:

$$P_{X,A,Y} = P_X P_{A|X} P_{Y|A,X}, \tag{5}$$

under which $E[\Lambda(A)] \leq \Gamma$.

Theorem 2: The optimal rate region $\mathcal{R}_B$ for case B (See Fig. 2), i.e. correlated sources with actions taken at the encoder, is the closure of the set of triplets (R_X, R_Y, Γ) such that

$$R_X \geq H(X|Y,A) + I(X;A) - I(Y;A), \tag{6a}$$

$$R_Y \geq H(Y|X,A), \tag{6b}$$

$$R_X + R_Y \geq H(X,Y|A) + I(X;A), \tag{6c}$$

where the joint distribution of (X, A, Y) is of the form (5), under which $E[\Lambda(A)] \leq \Gamma$.

The proofs of Theorem 1 and Theorem 2 appear in Section V.

Remark 1: Given a fixed distribution of the form (5), the regions satisfy $\mathcal{R}_A \subseteq \mathcal{R}_B$; specifically, when minimizing R_Y, the regions share a common corner point which be represented as $(R_X, R_Y) = (H(X), H(Y|X,A))$. For the other corner point, one can note that R_X has a looser constraint in $\mathcal{R}_B$, while the sum rate is equal in both regions. The difference between the rates in R_X is by a non-negative factor of $I(Y;A)$ and can be justified by the operative definitions in (2).

Remark 2: In the special case where R_Y is unlimited, both regions $\mathcal{R}_A$ and $\mathcal{R}_B$ reduce to those investigated in [7]. In this case, (4b),(4c) and (6b),(6c) are redundant constraints and, thus, the only constraint is on the rate R_X. From here it follows that Theorem 1 coincides with [7, Th. 1] source coding with SI where actions are taken at the decoder with lossless reconstruction, and Theorem 2 coincides with [7, Th. 5] source coding with SI where actions are taken at the encoder.

Another special case is when considering deterministic actions, that is, $A = a$; let us write the original optimal rate region of SW as $\mathcal{R}_{SW}(P_X, P_{Y|X})$, with the explicit dependence on P_X and $P_{Y|X}$. For this setting, both $\mathcal{R}_A$ and $\mathcal{R}_B$ reduce to $\mathcal{R}_{SW}(P_X, P_{Y|X,A=a})$.

We proceed to the next theorem which concerns with our main result on the general networks scenario.

Theorem 3: Given a correlated sources with action network $(\mathcal{V}, \mathcal{E}, s_1, s_2, \mathcal{D}, \Gamma)$ (See Fig. 3), the set of achievable rates $\mathcal{R}_N$ is such that

$$\mathcal{C}(\mathcal{V}^*_{s_1;\mathcal{D}}) \geq I(X;A) - I(Y;A) + H(X|Y,A), \tag{7a}$$

$$\mathcal{C}(\mathcal{V}^*_{s_2;\mathcal{D}}) \geq H(Y|X,A), \tag{7b}$$

$$\mathcal{C}(\mathcal{V}^*_{s_1,s_2;\mathcal{D}}) \geq I(X;A) + H(X,Y|A), \tag{7c}$$

where the joint distribution of (X, A, Y) is of the form (5), under which $E[\Lambda(A)] \leq \Gamma$.

The network investigated here is an extension of case B, and thus there are solid similarities between both results. Specifically, the right hand side of (7) coincides with the information measurements in Theorem 2. Moreover, the region of Theorem 3 is reduced to the optimal rate region given in 2 by taking the set of nodes as $\mathcal{V} = \{s_1, s_2, t\}$, and the set of edges as $\mathcal{E} = \{(s_1, t), (s_2, t)\}$.

IV. Examples

In this section, we study two binary examples and calculate their corresponding optimal rate regions $\mathcal{R}_A$, $\mathcal{R}_B$. We also an additional special scenario where actions are taken before the first source X^n is known, and thus actions are independent of this source and play the role of time-sharing of the possible actions. This special scenario may seem a degenerate setup, but can lead to some insights when considering an implementation of such a system with actions. The first example illustrates a scenario where actions taken at different nodes of the system cannot affect the set of achievable rates, while the second example demonstrated how taking actions at different nodes of the system improve significantly the optimal rate region under a cost regime.

Example 1: This binary example is a continuation of the first motivating scenario in Section I: two sensors' measurements, X and Y, are known at different nodes of the system and are required at a terminal node. The measurement X is a coarse measurement which is binary and distributed uniformly, while the measurement Y corresponds to fine or coarse measurement that depends on the taken actions. A low-cost actions correspond to a fine measurement within the measured range, and high-cost actions correspond to a coarse measurement identical to the X measurement. This cost implies that the number of fine measurements needs to be above some threshold. Our goal is to characterize the transmission rates that are required to receive both measurements at the decoder, subject to the given constraint.

The example is illustrated in Fig. 4; all alphabets are binary, i.e. $\mathcal{X} = \mathcal{Y} = \mathcal{A} = \{0, 1\}$, and $X \sim \text{Bern}(0.5)$. The source Y is an output of a clean channel if $A = 0$, and the output of a noisy-channel with crossover probability 0.5 if $A = 1$. Actions can be taken at the decoder (switch 2 is closed), at the encoder (switch 1 is closed) or in the special case of actions taken before the source X is known (switch 1 and switch 2 are open). We consider a cost function $\Lambda(A) = A$ that implies the constraint $P(A = 1) \leq \Gamma/$

<u>Case A</u> - actions are taken at the decoder; the setup is depicted in Fig. 4, with switch 2 closed. A general conditional distribution connecting X and A is considered, with $P_{A|X}(1|0) = \alpha$ and $P_{A|X}(0|1) = \beta$. The optimal rate region,

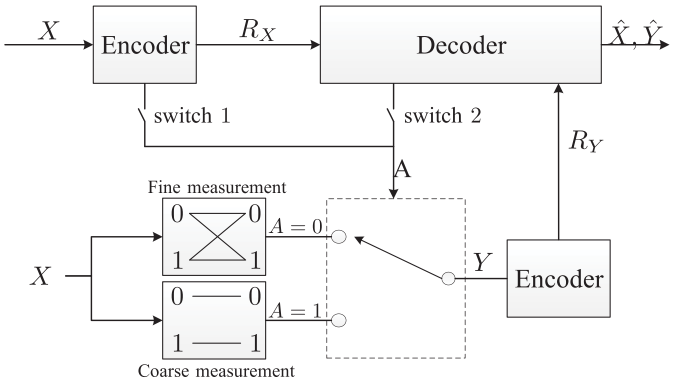

Fig. 4. The setup for example 1. Actions can be performed by the decoder (switch 2 is closed), by the encoder (switch 1 is closed) or before X is known (switch 1 and switch 2 are open). The switch in the dashed box corresponds to actions' performance.

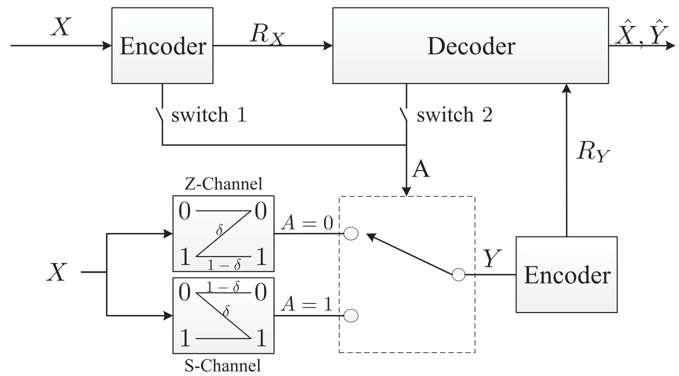

Fig. 5. The setup for example 2. Actions can be taken at the decoder (switch 2 is closed), at the encoder (switch 1 is closed) or before X is known (switch 1 and switch 2 are open). The switch in the dashed box corresponds to actions' performance.

$\mathcal{R}_A$, is as follows:

$$\begin{aligned} R_X &\geq 1 - 0.5(\alpha + \bar{\beta}) H_b\left(\frac{\alpha}{\alpha+\bar{\beta}}\right), \\ R_Y &\geq 0.5(\bar{\alpha} + \beta), \\ R_X + R_Y &\geq 1 + 0.5(\bar{\alpha} + \beta), \end{aligned} \tag{8}$$

for some $\alpha, \beta \in [0, 1]$ such that $0.5(\alpha + \bar{\beta}) \leq \Gamma$.

Case B - actions are taken at the encoder; the setup is depicted in Fig. 4, with switch 1 closed. Calculating $\mathcal{R}_B$ with the same pmf as in the previous case yields:

$$\begin{aligned} R_X &\geq 1 - H_b(0.5\alpha + 0.25[\beta + \bar{\alpha}]) + 0.5(\bar{\alpha} + \beta), \\ R_Y &\geq 0.5(\bar{\alpha} + \beta), \\ R_X + R_Y &\geq 1 + 0.5(\bar{\alpha} + \beta), \end{aligned} \tag{9}$$

for some $\alpha, \beta \in [0, 1]$ such that $0.5(\alpha + \bar{\beta}) \leq \Gamma$.

Case C - Actions are taken before the source X^n is known - for this case, actions contain no information of the source X^n and play the role of a time-sharing random variable that is available to all nodes in the system. Definitions of the probability of error, achievable rates and the optimal rate region, that is denoted by $\mathcal{R}_{A \perp X}$, are defined in a straightforward manner as in the previous cases. For this scenario, it can be shown that the optimal rate region $\mathcal{R}_{A \perp X}$ is the set of (R_X, R_Y, Γ) that satisfy:

$$\begin{aligned} R_X &\geq H(X|Y, A), \\ R_Y &\geq H(Y|X, A), \\ R_X + R_Y &\geq H(X, Y|A), \end{aligned}$$

for some joint distribution $P_{X,A,Y} = P_X P_A P_{Y|A,X}$, under which $E[\Lambda(A)] \leq \Gamma$.

The setup is depicted in Fig. 4 where both switches are open; we assume that $X \sim \text{Bern}(\bar{\alpha})$ and the optimal rate region, $\mathcal{R}_{A \perp X}$, is as follows:

$$\begin{aligned} R_X &\geq \bar{\alpha}, \\ R_Y &\geq \bar{\alpha}, \\ R_X + R_Y &\geq 1 + \bar{\alpha}, \end{aligned} \tag{10}$$

for some $\alpha \leq \Gamma$.

Remarkably, the unions over the input distribution in the three resulted regions coincide for any value of Γ. The input distribution that maximizes the regions is as follows: substitute $\alpha = \Gamma$ and $\bar{\beta} = \Gamma$ (which satisfies the cost constraint on its boundary), and it then follows that optimal rate regions given in (8)-(9) and (10) are:

$$\begin{aligned} R_X &\geq 1 - \Gamma, \\ R_Y &\geq 1 - \Gamma, \\ R_X + R_Y &\geq 2 - \Gamma, \end{aligned}$$

where Γ is the cost constraint.

This equivalence of the optimal rate regions can happen in systems for which greedy policy is optimal. A greedy policy is associated with a system for which different observations of X lead to the same actions strategy that minimize the transmission rates. For instance, in example 1, the greedy policy is to maximize the appearances of $A = 1$ that yield more correlation between X and Y, and thus a greater achievable region. Note that this policy has no dependence on the source X, and the only constraint is given by the cost Γ.

Example 2: This example is based on the second scenario described in the introduction and depicted in Fig. 5. Let us specify the technical details of this scenario; the operation of the first unit is modeled as the source X that is distributed according to Bern(0.5). We refer to $X = 1$ as the backup operation of the first unit, and $X = 0$ to no backup operation. The other backup unit is modeled as the other source Y, which is governed by actions. The scenario under consideration is where actions determine the operation that will be repeated in the second unit with a probability of 1. The above is modeled as S and Z channels; the S channel assures that the backup operation will be repeated in the second unit, while Z-channel assures that no backup operation will be repeated in the second unit. This choice also verifies that for a certain percentage of the time both operators are different, independent of the actions taken.

The cost of actions is taken as $\Lambda(A) = A$, due to power consumptions; specifically, we allocate higher cost for repeating the backup operation than for repeating the no backup operation. Also here, actions can be taken at the decoder (switch 2 is closed), at the encoder (switch 1 is closed) or before the source X is known (switch 1 and switch 2 are open).

Case A - actions are taken at the decoder; the setup is depicted in Fig. 5 for the case that switch 2 is closed. A general conditional distribution connecting X and A is considered, with $P_{A|X}(1|0) = \alpha$ and $P_{A|X}(0|1) = \beta$ that imply the constraint $0.5(\alpha + \bar{\beta}) \leq \Gamma$. Straightforward calculation of (4), gives that the optimal rate region, $\mathcal{R}_A$, is:

$$\begin{aligned} R_X \geq 1 - 0.5 \Bigg[&(\alpha + \bar{\beta}) H_b\left(\frac{\bar{\beta}}{\alpha + \bar{\beta}}\right) \\ &- (\beta + \bar{\alpha}) H_b\left(\frac{\bar{\alpha}}{\beta + \bar{\alpha}}\right)\Bigg] \\ + 0.5 \Bigg[&(\bar{\alpha} + \beta\delta) H_b\left(\frac{\bar{\alpha}}{\bar{\alpha} + \beta\delta}\right) \\ &+ (\bar{\beta} + \alpha\delta) H_b\left(\frac{\bar{\beta}}{\bar{\beta} + \alpha\delta}\right)\Bigg] \\ R_Y &\geq 0.5(\alpha + \beta) H_b(\delta) \\ R_X + R_Y &\geq 1 + 0.5(\alpha + \beta) H_b(\delta), \end{aligned}$$

for some $\alpha, \beta \in [0, 1]$ such that $0.5(\alpha + \bar{\beta}) \leq \Gamma$.

Case B - actions are taken at the encoder; the setup is depicted in Fig. 5 for the case that switch 1 is closed. A conditional distribution is assumed as in case A. The optimal rate region ,$\mathcal{R}_B$, for this case is as follows:

$$\begin{aligned} R_X &\geq 1 + 0.5(\alpha + \beta) H_b(\delta) - H_b(0.5[1 + \alpha\delta - \beta\delta]), \\ R_Y &\geq 0.5(\alpha + \beta) H_b(\delta), \\ R_X + R_Y &\geq 1 + 0.5(\alpha + \beta) H_b(\delta), \end{aligned}$$

for some $\alpha, \beta \in [0, 1]$ such that $0.5(\alpha + \bar{\beta}) \leq \Gamma$.

Note, the optimal rate region $\mathcal{R}_B$ is minimized by taking $A = X$ for the case of $\Gamma \geq 0.5$.

Case C - actions are taken before the source X^n is known; the setup is depicted in Fig. 5 where both switches are open. The optimal rate region, $\mathcal{R}_{A \perp X}$, is:

$$\begin{aligned} R_X &\geq 0.5(1 + \delta) H_b\left(\frac{1}{1 + \delta}\right), \\ R_Y &\geq 0.5 H_b(\delta), \\ R_X + R_Y &\geq 1 + 0.5 H_b(\delta). \end{aligned}$$

Note that the region is independent of α and no union is needed here. This fact implies that $\mathcal{R}_{A \perp X}$ is also independent of the cost Γ and only depends on the value of δ.

To gain some intuition regarding the optimal rate regions, we draw the results for $\Gamma = 0.3$ and $\delta = 0.5$ in Fig. 6. Let us examine the curved dashed blue line, which corresponds to case A; its corner point coincides with the black line (squared-marker) and tends to the red line (triangled-marker) for decreasing R_X. The corner point that coincides with the black line is the common corner point that is discussed in Remark 1. As long as R_X is decreased, this implies that $P_{A|X}$ is constrained to have less correlation between A and X. Although it seems that the blue line tends to the red line, it is notable that the blue plot achieves better performance in R_X than the red plot, that is, independent A and X is not optimal

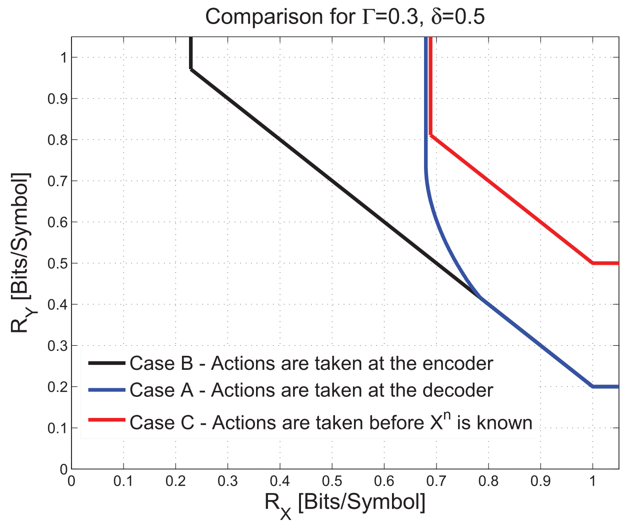

Fig. 6. The optimal rate regions for the three cases of Example 2.

also when minimizing R_X. Clearly, case A and case B have greater optimal region than the region of case C, thus time-sharing is not optimal in general when considering an action-dependent system.

V. Proofs of Case A and Case B

In this section we present the proof of Theorem 1 and the achievability of Theorem 2. Since the network model that is studied in Theorem 3 is a generalized version of case B, we omit the converse of Theorem 2, and it follows directly from the cut-set bounds in Section VI. Although the achievability of case B also follows from the network case, we provide an alternative achievability proof, which is less complicated than the direct method of Theorem 3.

A. Proof of Theorem 1

In this section, we provide an optimal code construction for case A and then provide a converse that completes the proof of Theorem 1. The coding scheme combines both the idea of actions and the SW machinery: at first, we will use random coding arguments for the actions' generation and explicitly transmit the sequence that is typical with x^n to the decoder. The first step has also been done in [7] for the case of source coding with SI where actions are taken at the decoder. However, the second stage in our proof is different from [7]; we use the SW machinery of random binning in the decentralized encoders to compress both sources. Later, at the decoding stage, we will exploit the actions' codeword that is available at the decoder as SI, when looking for jointly typical codewords in the pair of assigned bins.

1) Sketch of Achievability: At the first stage, the identity of the action sequence is transmitted from encoder 1 to the decoder as follows: generate a codebook of actions containing $2^{nI(X;A)}$ independent codewords, where each codeword is generated according to the marginal distribution P_A. Encoder 1 looks in the codebook for a codeword which is jointly typical with the source observation x^n, and transmits this codeword to the decoder using a rate of $I(X; A)$. The existence of a typical actions' codeword with the source observation follows directly from the covering lemma [19, Ch. 3].

We proceed and note that the optimal rate region in (4), $\mathcal{R}_A$, can be written as:

$$\begin{aligned} R_X - I(X;A) &\geq H(X|Y,A), \\ R_Y &\geq H(Y|X,A), \\ [R_X - I(X;A)] + R_Y &\geq H(X,Y|A). \end{aligned} \tag{11}$$

The right hand side of (11) reveals the SW nature of this region; the information measures correspond to a SW mechanism, but with the actions' sequence A^n available to the decoder. Before proceeding to the last step of the proof, note that the triplet (A^n, X^n, Y^n) is jointly typical with high probability. Specifically, the source Y^n is an output of a memoryless channel that is conditioned on the pair (X^n, A^n) which is also jointly typical with high probability.

Finally, the proof is completed by implementing a SW coding scheme with the actions treated as SI available at the decoder. A straightforward calculation gives that the right hand side of (11) is achieved.

2) Converse: For a sequence $(2^{nR_X}, 2^{nR_Y}, n)$ of codes with corresponding achievable rates, consider the rate that is used by encoder 1:

$$\begin{aligned} nR_X &\geq H(T_1) \\ &\overset{(a)}{=} H(T_1, A^n) + H(X^n|Y^n, T_1) - H(X^n|Y^n, T_1) \\ &\overset{(b)}{\geq} H(A^n) + H(T_1|A^n) + H(X^n|Y^n, T_1) - n\epsilon_n \\ &\overset{(c)}{\geq} H(A^n) + H(T_1|A^n, Y^n) + H(X^n|Y^n, T_1) - n\epsilon_n \\ &\overset{(d)}{=} H(A^n) + H(X^n, T_1|A^n, Y^n) - n\epsilon_n \\ &\overset{(e)}{=} H(A^n) + H(X^n|A^n, Y^n) - n\epsilon_n \\ &\overset{(f)}{=} H(A^n) - H(A^n|X^n) + H(X^n|A^n, Y^n) - n\epsilon_n \\ &\overset{(g)}{=} H(X^n) - H(Y^n|A^n) + H(Y^n|A^n, X^n) - n\epsilon_n \\ &\overset{(h)}{\geq} \sum_{i=1}^{n} [H(X_i) - H(Y_i|A_i) + H(Y_i|A_i, X_i)] - n\epsilon_n \\ &\overset{(i)}{=} \sum_{i=1}^{n} [I(X_i; A_i) + H(X_i|A_i, Y_i)] - n\epsilon_n, \end{aligned}$$

where:

(a) follows from the fact that A^n is a deterministic function of the index T_1;
(b) follows from Fano's inequality and properties of joint entropy;
(c) follows from the fact that conditioning reduces entropy;
(d) follows from the fact that A^n is a deterministic function of the index T_1;
(e) follows from the fact that T_1 is a deterministic function of X^n;
(f) follows from the fact that A^n is a deterministic function of T_1 and, therefore, also a deterministic function of X^n;
(g) follows from the properties of mutual information;
(h) follows from the facts that X^n is i.i.d., conditioning reduces entropy and the memoryless property (1);
(i) follows from the properties of mutual information.

In the derivation above, steps $(f) - (j)$ are similar to those taken in [7, eq. (18)] and presented here for completeness.

For the second rate that is used by encoder 2, consider

$$\begin{aligned} nR_Y &\geq H(T_2) \\ &\overset{(a)}{\geq} H(T_2, Y^n|X^n) - H(Y^n|T_2, X^n) \\ &\overset{(b)}{\geq} H(T_2, Y^n|X^n) - n\epsilon_n \\ &\overset{(c)}{=} H(Y^n|X^n) - n\epsilon_n \\ &\overset{(d)}{=} H(Y^n|X^n, A^n) - n\epsilon_n \\ &\overset{(e)}{=} \sum_{i=1}^{n} [H(Y_i|A_i, X_i)] - n\epsilon_n, \end{aligned}$$

where:

(a) follows from the fact that conditioning reduces entropy;
(b) follows from Fano's inequality;
(c) follows from the fact that T_2 is a deterministic functions of Y^n;
(d) follows from the fact that A^n is a deterministic function of T_1 and, therefore, also a deterministic function of X^n;
(e) follows from the memoryless property (1).

The last lower bound is for the achievable sum-rate of the two encoders:

$$\begin{aligned} n(R_X + R_Y) &\geq H(T_1, T_2) \\ &= H(T_1, T_2, X^n, Y^n) - H(X^n, Y^n|T_1, T_2) \\ &\overset{(a)}{\geq} H(X^n, Y^n) + H(T_1, T_2|X^n, Y^n) - n\epsilon_n \\ &\overset{(b)}{=} H(X^n, Y^n) - n\epsilon_n \\ &\overset{(c)}{=} H(X^n) + H(Y^n|X^n, A^n) - n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^{n} [H(X_i) + H(Y_i|X_i, A_i)] - n\epsilon_n \\ &= \sum_{i=1}^{n} [I(X_i; A_i) + H(X_i, Y_i|A_i)] - n\epsilon_n \end{aligned}$$

where:

(a) follows from Fano's inequality and the properties of joint entropy;
(b) follows from the fact that T_1 and T_2 are deterministic functions of X^n and Y^n, respectively;
(c) follows from the fact that A^n is a deterministic function of T_1 and, therefore, also a deterministic function of X^n;
(d) follows from the fact that X^n is memoryless and the memoryless property (1).

Derivation of the single letter terms is by using a standard time-sharing technique. Thus, we have shown that the pair of achievable rates (R_X, R_Y) satisfy the next set of inequalities:

$$\begin{aligned} R_X &\geq I(X;A) + H(X|A,Y) - \epsilon_n, \\ R_Y &\geq H(Y|A,X) - \epsilon_n, \\ R_X + R_y &\geq I(X;A) + H(X,Y|A) - \epsilon_n. \end{aligned} \tag{12}$$

The proof is completed by taking $n \to \infty$ in (12), which implies $\epsilon_n \to 0$ since (R_X, R_Y) are achievable.

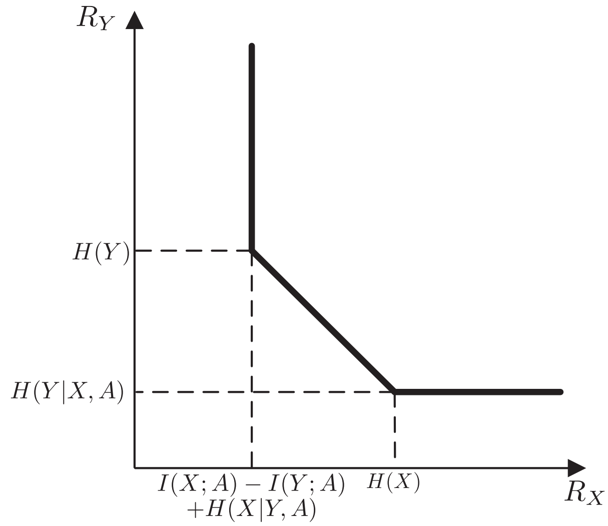

Fig. 7. The optimal rate region, $\mathcal{R}_B$, for case B.

B. Achievability of Theorem 2

For the achievability proof that is presented in this section we use time sharing arguments; specifically, we prove that the corner points of $\mathcal{R}_B$ are achievable and conclude that the convex region is also achievable. The corner points of $\mathcal{R}_B$ are illustrated in Fig. 7 for a fixed joint distribution, and can be written as:

$$(R_X, R_Y) = (I(X;A) - I(Y;A) + H(X|Y,A), H(Y)) \quad (13)$$

$$(R_X, R_Y) = (H(X), H(Y|X,A)). \quad (14)$$

The corner point in (13) can be achieved as follows: actions' generation is performed according to the method proposed in [7, Sec. III]. Note that actions are taken at the encoder and no transmission rates are required for this step. The source sequence Y^n is then generated and transmitted in a lossless manner at a rate of $H(Y)$ to the decoder. It then follows that the current problem is reduced to that of [7, Sec. III]-source coding with SI where actions are taken at the encoder. The proof for the rate $R_X = I(X;A) - I(Y;A) + H(X|Y,A)$ is omitted here, and can be found in [7, Sec. III].

The corner point in (14) is, indeed, the common corner point for case A and case B as mentioned in Section III. The rate R_X in (14) can be written also as $H(X)$; thus, this rate is used to transmit the source X^n in a lossless manner to the decoder. Having received the source X^n, the decoder obtains A^n, which is a deterministic function of X^n. Later, a trivial source coding scheme for the source Y^n is used at a rate of $H(Y|X,A)$, where (X^n, A^n) are considered as SI available to the decoder.

VI. Proof of Theorem 3

In this section, the converse for Theorem 3 is given in subsection VI-A and the code construction, encoding and the decoding procedures are presented in subsection VI-B. In subsection VI-C, Lemma 1 states an upper bound on the probability of error that two different inputs to a randomized network yield the same output, followed by a multicast example and the proof of the lemma. Finally, the analysis of the probability of error will be given in subsection VI-D.

A. Generalized Cut-Set Bounds (Converse)

In this subsection, we derive an outer bound on the set of achievable rates for our model. The outer bound is indeed a generalization of the known cut-set bound, this method of generalized cut-set bounds was adopted also in [20].

For the converse of Theorem 3, given an achievable $\left((2^{nR_l})_{l\in\mathcal{E}}, n\right)$ source code we need to show that there exists a joint distribution, $P_{X,A,Y} = P_X P_{A|X} P_{Y|X,A}$, such that the inequalities in Theorem 3 hold.

For any set of messages denoted by $\mathcal{M}_1$, across a cut $\mathcal{V}_{s_1;t}$, we have

$$\begin{aligned}
n\mathcal{C}(\mathcal{V}_{s_1;t}) &\geq H(\mathcal{M}_1)\\
&= H(\mathcal{M}_1) + H(X^n|Y^n,\mathcal{M}_1) - H(X^n|Y^n,\mathcal{M}_1)\\
&\overset{(a)}{\geq} H(\mathcal{M}_1) + H(X^n|Y^n,\mathcal{M}_1) - n\epsilon_n\\
&\overset{(b)}{\geq} I(\mathcal{M}_1;X^n,Y^n) + H(X^n|Y^n,\mathcal{M}_1) - n\epsilon_n\\
&= H(X^n,Y^n) - H(Y^n|\mathcal{M}_1) - n\epsilon_n\\
&\overset{(c)}{\geq} H(X^n) + H(Y^n|X^n,A^n) - H(Y^n) - n\epsilon_n\\
&\overset{(d)}{\geq} \sum_{i=1}^{n}[H(X_i) - H(Y_i) + H(Y_i|A_i,X_i)] - n\epsilon_n\\
&= \sum_{i=1}^{n}[I(X_i;A_i) + H(Y_i,X_i|A_i) - H(Y_i)] - n\epsilon_n\\
&= \sum_{i=1}^{n}[I(X_i;A_i) - I(Y_i;A_i) + H(X_i|A_i,Y_i)] - n\epsilon_n,
\end{aligned}$$

where:

(a) follows from Fano's inequality;
(b) follows from the fact that $\mathcal{M}_1$ is a deterministic function of X^n, Y^n;
(c) follows from the fact that A^n is a deterministic function of X^n;
(d) follows from the X^n is memoryless, conditioning reduces entropy and the memoryless property (1).

For the second inequality in (7), we have

$$\begin{aligned}
n\mathcal{C}(\mathcal{V}_{s_2;t}) &\geq H(\mathcal{M}_2)\\
&\geq H(\mathcal{M}_2, Y^n|X^n) - H(Y^n|X^n,\mathcal{M}_2)\\
&\overset{(a)}{\geq} H(\mathcal{M}_2, Y^n|X^n) - n\epsilon_n\\
&\overset{(b)}{=} H(Y^n|X^n,A^n) + H(\mathcal{M}_2|X^n,Y^n) - n\epsilon_n\\
&\overset{(c)}{=} H(Y^n|X^n,A^n) - n\epsilon_n\\
&= \sum_{i=1}^{n} H(Y_i|A_i,X_i) - n\epsilon_n,
\end{aligned}$$

where:

(a) follows from Fano's inequality;
(b) follows from the fact that A^n is a deterministic function of X^n;
(c) follows from the fact that $\mathcal{M}_2$ is a deterministic function of X^n, Y^n.

For the sum-rate, we have

$$
\begin{aligned}
n\mathcal{C}(\mathcal{V}_{s_1,s_2;t}) &\geq H(\mathcal{M}_3) \\
&= H(X^n, Y^n, \mathcal{M}_3) - H(X^n, Y^n|\mathcal{M}_3) \\
&\overset{(a)}{\geq} H(X^n, Y^n, \mathcal{M}_3) - n\epsilon_n \\
&\overset{(b)}{=} H(X^n, Y^n) - n\epsilon_n \\
&\overset{(c)}{=} H(X^n) + H(Y^n|X^n, A^n) - n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} [H(X_i) + H(Y_i|X_i, A_i)] - n\epsilon_n \\
&= \sum_{i=1}^{n} [I(X_i; A_i) + H(Y_i, X_i|A_i)] - n\epsilon_n,
\end{aligned}
$$

where:

(a) follows from Fano's inequality;

(b) follows from the fact that $\mathcal{M}_3$ is a deterministic function of X^n, Y^n;

(c) follows from the fact that A^n is a deterministic function of X^n;

(d) follows from the fact the X^n is memoryless and the memoryless property (1).

Let us summarize the lower bounds we have characterized:

$$
\begin{aligned}
\mathcal{C}(\mathcal{V}_{s_1;t}) &\geq 1 \sum_{i=1}^{n} \frac{1}{n} [I(X_i; A_i) - I(Y_i; A_i) \\
&\qquad\qquad + H(X_i|A_i, Y_i)] - \epsilon_n, \\
\mathcal{C}(\mathcal{V}_{s_2;t}) &\geq \sum_{i=1}^{n} \frac{1}{n} H(Y_i|A_i, X_i) - \epsilon_n, \\
\mathcal{C}(\mathcal{V}_{s_1,s_2;t}) &\geq \sum_{i=1}^{n} \frac{1}{n} [I(X_i; A_i) + H(Y_i, X_i|A_i)] - \epsilon_n,
\end{aligned}
\tag{15}
$$

for some cuts $\mathcal{V}_{s_1;t}, \mathcal{V}_{s_2;t}, \mathcal{V}_{s_1,s_2;t}$.

To complete the proof, we minimize the left hand side of (15) by taking the cuts to be $\mathcal{C}(\mathcal{V}^*_{s_1;t}), \mathcal{C}(\mathcal{V}^*_{s_2;t})$, and $\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})$, respectively. Derivation of the single-letter characterization in (15) is done by the common time-sharing technique as in [21].

B. Direct

As the direct proof involves linear network coding over a finite filed, throughout this proof, without loss of generality, we assume that each edge has unit capacity. Namely, each edge can transmit one element in the field $\mathbb{F}_2$ per unit time.[1] This model is common in the linear network coding literature. For instance, see [2] or [22]. It is justified by allowing parallel edges to achieve any integer capacity and, further, by normalizing over multiple transmissions to achieve any rational capacity. Thus, we will use this notation throughout the direct. The converse proof, however, holds for the more general definition of network coding, as was given in subsection VI-A.

The construction of the code is based on a random codebook generation of the actions codewords that is followed by transmission of the network inputs using RLNC in a finite field. The rate of the actions' codebook generation will be depend on the sign of the expression $I(X; A) - I(Y; A)$, and throughout the proof, we differentiate between these two cases. In each case, it will be shown how actions can be exploited such that the optimal region is achieved. Furthermore, the source sequences X^n and Y^n are binned randomly, and these bins together with the chosen action codewords will be the inputs to the network. Recall that we have assumed that each link in the network has a unit capacity (i.e. transmit one bit in each time instance) and, therefore, inputs to the network for n channel uses are represented as vectors consisting of elements from $\mathbb{F}_{2^n}$. The transmission is based on RLNC in $\mathbb{F}_{2^n}$, using the algebraic approach formulation introduced by Koetter and Medard [22].

1) The Case $I(X; A) - I(Y; A) \geq 0$: Fix a joint distribution of $P_{X,A,Y} = P_X P_{A|X} P_{Y|A,X}$, where source distributions P_X and $P_{Y|A,X}$ are given.

Code construction:

- The X^n sequences are randomly binned into 2^{nr_1} bins, where $r_1 \triangleq H(X) + \epsilon$, for some $\epsilon > 0$. Each bin can be represented as nr_1 bits, or alternatively as a vector of $\lceil r_1 \rceil$ elements from the finite field $\mathbb{F}_{2^n}$. The bin vector of X^n will be denoted as $\underline{X}^n$, consisting of $\lceil r_1 \rceil$ elements. The Y^n sequences are randomly binned into 2^{nr_2} bins, where $r_2 \triangleq H(Y) + \epsilon$. Again, the bin vector of the sequence Y^n will be denoted by $\underline{Y}^n$, consisting of $\lceil r_2 \rceil$ elements from $\mathbb{F}_{2^n}$. The bin vectors $\underline{X}^n$ and $\underline{Y}^n$ will be part of the input to the network.
- A codebook $\mathcal{C}$ of actions codewords is generated, consisting of 2^{nr_A} independent codewords, $A^n(i)$, $i \in \{1, 2, \ldots, 2^{nr_A}\}$, where each codeword is distributed i.i.d. according to $\sim \prod_{j=1}^{n} P_A(a_j)$. Each codeword A^n is represented by a vector of elements from $\mathbb{F}_{2^n}$, denoted by $\underline{A}^n$ and consisting of $\lceil r_A \rceil$ elements.
- The inputs to the network will be the source bins $\underline{X}^n$, $\underline{Y}^n$, and actions codewords $\underline{A}^n$, each consisting of elements in $\mathbb{F}_{2^n}$. Each element in the input vectors $\underline{X}^n$, $\underline{Y}^n$, and $\underline{A}^n$ is denoted by U_i, where $i \in \{1, \ldots, \lceil r_1 \rceil + \lceil r_2 \rceil + \lceil r_A \rceil\}$. Let $\sigma(U_i)$ be equal to s_1 if U_i is an element in the vector $\underline{X}^n$ or $\underline{A}^n$, and $\sigma(U_i) = s_2$ if U_i is an element in the vector $\underline{Y}^n$.

The information process V_j transmitted on a link $j \in \mathcal{E}$ is formed as a linear combination, in $\mathbb{F}_{2^n}$, of link j's inputs, i.e. source elements, U_i, for which $\sigma(U_i) = \sigma(j)$ and input processes V_l for which $\tau(l) = \sigma(j)$. This can be represented by the equation

$$V_j = \sum_{i:\sigma(U_i)=\sigma(j)} b_{i,j} U_i + \sum_{l:\tau(l)=\sigma(j)} f_{l,j} V_l.$$

The coefficients $\{b_{i,j}, f_{l,j}\}$ are generated uniformly from the finite field $\mathbb{F}_{2^n}$ and collected into matrices $\mathbf{B} = \{b_{i,j}\}$ and $\mathbf{F} = \{f_{l,j}\}$; note the dimensions $|\mathbf{B}| = (\lceil r_1 \rceil + \lceil r_2 \rceil + \lceil r_A \rceil) \times |\mathcal{E}|$, and $|\mathbf{F}| = |\mathcal{E}| \times |\mathcal{E}|$. For acyclic graphs, we can assume that there exists an ancestral indexing of the links in $\mathcal{E}$. It then follows that the matrix $\mathbf{F}$ is upper triangular with zeros on the diagonal and there exists the inverse of $(\mathbf{I} - \mathbf{F})$, denoted by $\mathbf{G} \triangleq (\mathbf{I} - \mathbf{F})^{-1}$. Let $\mathbf{G}_v$ denote the sub-matrix consisting of

[1]The finite field $\mathbb{F}_2$ is chosen to ease the representation of the code. The techniques herein are nor restricted to finite fields with characteristic 2.

only the columns of $\mathbf{G}$ corresponding to the input links of node v. Now, we can write the complete mapping from the input vector of the network, e.g. $\underline{U} = [\underline{X}^n, \underline{A}^n, \underline{Y}^n]$, to the input processes of some terminal node t as:

$$\underline{Z}_t = [\underline{X}^n, \underline{A}^n, \underline{Y}^n]\mathbf{BG}_t,$$

where $\underline{Z}_t$ is a vector consisting of the processes V_j satisfying $d(j) = \{t\}$.

Encoding: Given the source realization x^n, node s_1 looks in the codebook for an index i such that $A^n(i)$ is jointly typical with x^n; if there is none it outputs $i = 1$. If there is more than one index, i is set to the smallest among them. The source Y^n is then generated and available at node s_2. The input to the network will then be the vector $[\underline{x}^n, \underline{a}^n, \underline{y}^n]$, where $\underline{x}^n, \underline{y}^n$ are the bins' sources, and $\underline{a}^n$ is the chosen actions codeword.

Decoding: Having received the vector $\underline{Z}_t$, each node $t \in \mathcal{D}$ looks for a unique triplet $(X^n, A^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X, A, Y)$ satisfying $[\underline{X}^n, \underline{A}^n, \underline{Y}^n]\mathbf{BG}_t = \underline{Z}_t$.

2) The Case $I(X; A) - I(Y; A) \leq 0$: Fix a joint distribution of $P_{X,A,Y} = P_X P_{A|X} P_{Y|A,X}$, where the source distributions P_X and $P_{Y|A,X}$ are given.

Code construction:

- Generate a codebook $\mathcal{C}$, consisting of $2^{n(I(Y;A)-\epsilon)}$ independent codewords, $A^n(i)$, $i \in \{1, \ldots, 2^{n(I(Y;A)-\epsilon)}\}$, where each element is i.i.d. $\sim \prod_{j=1}^n P_A(a_j)$, for some $\epsilon > 0$. Randomly bin the codewords in $\mathcal{C}$ into $2^{n\Delta}$ bins, where $\Delta = (I(Y; A) - I(X; A) - 2\epsilon)$, such that in each bin there are $2^{n(I(X;A)+\epsilon)}$ codewords. For each $A^n \in \mathcal{C}$, the bin that contains A^n will be denoted as $\mathcal{B}_{A^n}$. Each bin can be represented by a message of $n\Delta$ bits, which is the rate that is reduced from the source X^n. Let $\underline{A}^n$ denote the representation of each codeword by $\lceil I(Y; A) - \epsilon \rceil$ elements from $\mathbb{F}_{2^n}$.
- The X^n sequences are randomly binned into 2^{nr_1} bins, where $r_1 \triangleq H(X|Y, A) + \epsilon$. The notation $\mathcal{B}_{X^n}$ stands for the first $n\Delta$ bits of the bin index where X^n falls. Additionally, each bin index is denoted by $\underline{X}^n(j)$, $j \in \{1, \ldots, 2^{nr_1}\}$, consisting of $\lceil r_1 \rceil$ elements from the finite field $\mathbb{F}_{2^n}$.
- The Y^n sequences are randomly binned into 2^{nr_2} bins, where $r_2 \triangleq H(Y)+\epsilon$. Each bin is represented by a vector consisting of $\lceil r_2 \rceil$ elements from $\mathbb{F}_{2^n}$, and denoted by $\underline{Y}^n(k)$ $k \in \{1, \ldots, 2^{nr_2}\}$.
- The process of network coefficients generation is the same as for the case $I(X; A)-I(Y; A) \geq 0$, and therefore omitted here.

Encoding: Given the source realization x^n, node s_1 looks in the actions' bin satisfying $\mathcal{B}_{A^n} = \mathcal{B}_{x^n}$ for a codeword A^n which is jointly typical with x^n. The source y^n is then generated and available at node s_2. The input to the network will then be the vector $[\underline{x}^n, \underline{a}^n, \underline{y}^n]$ corresponding to the bins where the source sequences fall and the chosen actions codeword.

Decoding: Having received the vector $\underline{Z}_t$, each node $t \in \mathcal{D}$ looks for a unique triplet $(X^n, A^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X, A, Y)$ satisfying $[X^n, A^n, Y^n]\mathbf{BG}_t = \underline{Z}_t$ and $\mathcal{B}_{A^n} = \mathbf{I}_{\mathbf{X^n}}$.

C. An Upper Bound in Randomized Networks

Following the result in [2, Appendix I], the next lemma provides an upper bound on the probability of the event that two different inputs to a randomized linear network yield the same output at a terminal node t. Due to the fact that the network is linear, this event is equivalent to the event that the difference between two inputs yields the zero processes at the terminal node. The next lemma will be at the assist of our direct proof. Moreover, as we will see in Example. 3, it has implications beyond the scoop of our proof as well.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed, acyclic graph where each link has a unit-capacity. The matrix $\mathbf{BG}_t$ represents the complete mapping of the network from inputs to some node t as presented in the code construction of subsection VI-B. Now, consider a set of sources with no incoming links, denoted by $\mathcal{S} \subseteq \mathcal{V}$, such that $\mathcal{S} = \{1, \ldots, k\}$. Each node $i \in \mathcal{S}$ consists of a vector, $\underline{u}_i$, which comprises elements from $\mathbb{F}_{2^n}$. For any two different inputs to the network, denoted by $\underline{u} = [\underline{u}_1 \underline{u}_2 \ldots \underline{u}_k]$ and $\underline{v} = [\underline{v}_1 \underline{v}_2 \ldots \underline{v}_k]$, let $\mathcal{W}$ be a subset of $\mathcal{S}$, such that if $\underline{u}_i \neq \underline{v}_i$ then $i \in \mathcal{W}$.

Lemma 1: For any pair of different inputs $\underline{u}$ and $\underline{v}$, the probability that these inputs induce the same output in node t is bounded by:

$$\Pr\left([\underline{u} - \underline{v}]\mathbf{BG}_t = \mathbf{0}\right) \leq \left(\frac{L}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{\mathcal{W};t})},$$

where L denotes the maximum source-receiver path length, and $\mathcal{C}(\mathcal{V}^*_{\mathcal{W};t})$ is the minimum cut-set between $\mathcal{W}$ and t.

Note that the upper bound is independent of the number of elements in the vector $\underline{u}_i, \forall i$. This remarkable fact allows us to think of $\underline{A}^n$ and $\underline{X}^n$ as the same input in our network; thus, for the following two useful probabilities:

$$\Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \underline{\tilde{a}}^n - \underline{a}^n, \mathbf{0}]\mathbf{BG}_t = \mathbf{0} | \underline{\tilde{x}}^n \neq \underline{x}^n, \underline{\tilde{a}}^n \neq \underline{a}^n\right)$$
$$\Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \mathbf{0}, \mathbf{0}]\mathbf{BG}_t = \mathbf{0} | \underline{\tilde{x}}^n \neq \underline{x}^n\right),$$

we will have the same upper bound, $\left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})}$, with L_1 stands for the maximum length of a path between s_1 and t.

We now show how the lemma above can serve as an easy and elegant proof to the capacity of multicast networks: a sender wishes to transmit a message to a set of terminal nodes through a directed, acyclic network. The sender transmits a message from the set $\mathcal{M} = \{1, \ldots, 2^{nR}\}$, and each receiver $t \in \mathcal{D}$ is required to decode the correct message in a lossless manner. We want to characterize the maximal rate R that can be used for a reliable communication in a given network.

Example 3 (Multicast Network): Consider a directed, acyclic network, where sender denoted as node 1 is required to transmit a message from $\mathcal{M} = \{1, \ldots, 2^{nR}\}$ to a set of terminal nodes denoted as $\mathcal{D}$. The sender can choose any message, $m \in \mathcal{M}$, and each receiver $t \in \mathcal{D}$ is required to decode the correct message in a lossless manner. We provide here a simple n block-length coding scheme follows by an analysis of the probability of error.

To encode the message, we rely on the scalar algebraic approach we have shown earlier in the code construction of the proof for Theorem 3. The input to the network is $\underline{m}$, where $\underline{m}$ is a vector representing m by elements from $\mathbb{F}_{2^n}$. Each terminal node, $t \in \mathcal{D}$, having received $\underline{z}_t$ looks for $m \in \mathcal{M}$ satisfying $\underline{m}\mathbf{BG}_t = \underline{z}_t$.

Now, assume without loss of generality that the message m was sent. An error occurs only if there exists $m' \neq m$ satisfying $\underline{m}'\mathbf{BG}_t = \underline{z}_t$ for some $t \in \mathcal{D}$.

Upper bounding the probability of error for some receiver $t \in \mathcal{D}$ yields:

$$\begin{aligned}\Pr(\text{error}) &= \Pr(\exists \tilde{m} \neq m : [\underline{\tilde{m}} - \underline{m}]\mathbf{BG}_t = \mathbf{0}) \\ &= \sum_{\tilde{m}\in\mathcal{M}} \Pr([\underline{\tilde{m}} - \underline{m}]\mathbf{BG}_t = \mathbf{0}) \\ &\leq 2^{nR}\left(\frac{L}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{1;t})} \\ &= L^{\mathcal{C}(\mathcal{V}^*_{1;t})} 2^{n(R-\mathcal{C}(\mathcal{V}^*_{1;t}))}. \end{aligned} \tag{16}$$

Note that if $R \leq \mathcal{C}(\mathcal{V}^*_{1;t})$, the term (16) tends to zero for sufficiently large n. Our requirement is to decode the message correctly at all the receivers; thus, using the union-bound we achieve that the overall probability tends to zero for large n if,

$$R < \min_{t\in\mathcal{D}} \mathcal{C}(\mathcal{V}^*_{1;t}),$$

which is the known multicast result.

Proof of Lemma 1: Let $\mathcal{G}_1$ be a subgraph of $\mathcal{G}$ consisting of all links downstream of $\mathcal{W}$, where a link l is considered downstream if $\sigma(l) \in \mathcal{W}$, or if there is a directed path from some source $s \in \mathcal{W}$ to $\sigma(l)$. Since information sources can differ only in source nodes satisfying $i \in \mathcal{W}$, this fact induces that only links in $\mathcal{G}_1$ will affect the bound on probability.

Note that in a random linear network code with block length n, any link l which has at least one nonzero input transmits the zero process with probability 2^{-n}. This is the same as the probability that a pair of distinct values for the inputs of l are mapped to the same output value on l.

For a given pair of distinct input values, let E_l be the event where the corresponding inputs to link l are distinct, but the corresponding values on l are the same. Let $E(\mathcal{G}_1)$ be the event that E_l occurs for some link l on every source-terminal path in graph $\mathcal{G}_1$. Note, the probability of the event $E(\mathcal{G}_1)$ is equal to the probability that two inputs induce the same output at the terminal node, i.e. $\Pr([\underline{u} - \underline{v}]\mathbf{BG}_t = \mathbf{0})$.

We proceed and look at the set of source-terminal paths in the graph $\mathcal{G}_1$. By Menger's Theorem [23], there exist $C(\mathcal{V}^*_{\mathcal{W};t})$ disjoint paths, since each link in the network has unit capacity. We denote each disjoint path as $\mathcal{P}_{\mathcal{G}_1 i}$ with its corresponding length L_i, where $i \in \{1, \ldots, C(\mathcal{V}^*_{\mathcal{W};t})\}$. Furthermore, we denote $E(\mathcal{P}_{\mathcal{G}_1 i})$ as the event that E_l occurs for some link on $\mathcal{P}_{\mathcal{G}_1 i}$.

$$\begin{aligned}\Pr(E(\mathcal{G}_1)) &= \Pr\left(\bigcap_{i=1}^{C(\mathcal{V}^*_{\mathcal{W};t})} E(\mathcal{P}_{\mathcal{G}_1 i})\right) \\ &\overset{(a)}{=} \prod_{i=1}^{C(\mathcal{V}^*_{\mathcal{W};t})} \Pr(E(\mathcal{P}_{\mathcal{G}_1 i})) \\ &\overset{(b)}{=} \prod_{i=1}^{C(\mathcal{V}^*_{\mathcal{W};t})} 1 - \left(1 - \frac{1}{2^n}\right)^{L_i} \\ &\overset{(c)}{\leq} \prod_{i=1}^{C(\mathcal{V}^*_{\mathcal{W};t})} 1 - \left(1 - \frac{1}{2^n}\right)^{L} \\ &= \left(1 - \left(1 - \frac{1}{2^n}\right)^{L}\right)^{C(\mathcal{V}^*_{\mathcal{W};t})} \\ &\overset{(d)}{\leq} \left(\frac{L}{2^n}\right)^{C(\mathcal{V}^*_{\mathcal{W};t})},\end{aligned}$$

where:

(a) follows from the fact that the coefficients are generated independently on each path;

(b) follows from the fact that the complement event of $E(\mathcal{P}_{\mathcal{G}_1 i})$ is the event that E_l does not occur on each link in the path $\mathcal{P}_{\mathcal{G}_1 i}$;

(c) follows from the notation $L = \max_i L_i$;

(d) follows from applying Bernoulli's inequality, $(1 + x)^r \geq 1 + rx$, with $x = -\frac{1}{2^n}$ and $r = L$.

D. Analysis of the Probability of Error

Following the direct method in Section VI, the probability of error is analyzed for both cases: a negative and positive sign of the term $I(X; A) - I(Y; A)$.

1) For the Case $I(X; A) - I(Y; A) \geq 0$: The events corresponding to possible encoding and decoding errors are as follows: An encoding error occurs if:

$$\mathcal{E}_1 = \{\nexists i : (x^n, A^n(i)) \in \mathcal{T}_\epsilon^{(n)}(X, A)\}.$$

For the events of decoding errors, we derive upper bounds for some terminal node $t \in \mathcal{D}$. Later on, we conclude the complete achievable region by a union bound on all $t \in \mathcal{D}$. For a terminal node $t \in \mathcal{D}$, a decoding error will occur for any of the next events:

$$\begin{aligned}\mathcal{E}_2 &= \{(X^n, A^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}(X, A, Y)\} \\ \mathcal{E}_3 &= \{\exists \tilde{X}^n \neq X^n, \tilde{A}^n \neq A^n : [\underline{\tilde{X}}^n, \underline{\tilde{A}}^n, \underline{Y}^n]\mathbf{BG}_t = \underline{Z}_t, \\ &\quad (\tilde{X}^n, \tilde{A}^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X, A, Y)\} \\ \mathcal{E}_4 &= \{\exists \tilde{X}^n \neq X^n : [\underline{\tilde{X}}^n, \underline{A}^n, \underline{Y}^n]\mathbf{BG}_t = \underline{Z}_t, \\ &\quad (\tilde{X}^n, A^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X, A, Y)\} \\ \mathcal{E}_5 &= \{\exists \tilde{Y}^n \neq Y^n : [\underline{X}^n, \underline{A}^n, \underline{\tilde{Y}}^n]\mathbf{BG}_t = \underline{Z}_t, \\ &\quad (X^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X, A, Y)\} \\ \mathcal{E}_6 &= \{\exists \tilde{X}^n \neq X^n, \tilde{Y}^n \neq Y^n : [\underline{\tilde{X}}^n, \underline{A}^n, \underline{\tilde{Y}}^n]\mathbf{BG}_t = \underline{Z}_t, \\ &\quad (\tilde{X}^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X, A, Y)\} \\ \mathcal{E}_7 &= \{\exists \tilde{X}^n \neq X^n, \tilde{A}^n \neq A^n, \tilde{Y}^n \neq Y^n : \\ &\quad [\underline{\tilde{X}}^n, \underline{\tilde{A}}^n, \underline{\tilde{Y}}^n]\mathbf{BG}_t = \underline{Z}_t, \\ &\quad (\tilde{X}^n, \tilde{A}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X, A, Y)\}.\end{aligned}$$

The total probability of an error can be bounded as:

$$\begin{aligned}P_e^{(n)} &= \Pr(\bigcup_{i=1}^{7} \mathcal{E}_i) \\ &\leq \Pr(\mathcal{E}_1 \bigcup \mathcal{E}_2) + \sum_{i=3}^{7} \Pr(\mathcal{E}_i) \\ &\leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2|\mathcal{E}_1^C) + \sum_{i=3}^{7} \Pr(\mathcal{E}_i).\end{aligned}$$

Let us derive an upper bound on each term separately.

For $\mathcal{E}_1$, it is known from the covering lemma [19, Lemma 3.3] that $\Pr(\mathcal{E}_1) \to 0$ for $n \to \infty$ if we fix $r_A = I(X;A) + \epsilon$.

Given the event $\mathcal{E}_1^C$, and the fact that Y^n is generated as the output of a memoryless channel, we use the conditional typicality lemma [19, Ch. 2] to show that $\Pr(\mathcal{E}_2|\mathcal{E}_1^C) \to 0$ as $n \to \infty$.

To upper-bound $\mathcal{E}_3$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_3)\\
&= \Pr(\exists \tilde{X}^n \neq X^n, \tilde{A}^n \neq A^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{X}^n, \underline{\tilde{A}}^n - \underline{A}^n, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, \tilde{A}^n, Y^n) \in \mathcal{T}_\epsilon^{(n)})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \cdot \Pr(\exists \tilde{X}^n \neq x^n, \tilde{A}^n \neq a^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{x}^n, \underline{\tilde{A}}^n - \underline{a}^n, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, \tilde{A}^n, y^n) \in \mathcal{T}_\epsilon^{(n)})\\
&\overset{(a)}{=} \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)\\
&\quad \times \sum_{\tilde{a}^n \in \mathcal{Q}} \Pr(\exists \tilde{X}^n \neq x^n : [\underline{\tilde{X}}^n - \underline{x}^n, \underline{\tilde{a}}^n - \underline{a}^n, \mathbf{0}]\\
&\quad \times \mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, \tilde{a}^n, y^n) \in \mathcal{T}_\epsilon^{(n)} | (\tilde{a}^n, y^n) \in \mathcal{T}_\epsilon^{(n)})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \sum_{\tilde{a}^n \in \mathcal{Q}} \sum_{\substack{\tilde{x}^n \neq x^n:\\ \tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}(X|Y,A)}}\\
&\quad \Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \underline{\tilde{a}}^n - \underline{a}^n, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}\right)\\
&\overset{(b)}{\leq} \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \sum_{\tilde{a}^n \in \mathcal{Q}}\\
&\quad \sum_{\substack{\tilde{x}^n \neq x^n:\\ \tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}(X|Y,A)}} \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})}\\
&\overset{(c)}{\leq} \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)\\
&\quad \times 2^{n(r_A - I(Y;A) + 2\epsilon)} |\mathcal{T}_\epsilon^{(n)}(X|Y,A)| \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})}\\
&\leq 2^{n(I(X;A) - I(Y;A) + H(X|Y,A) + 3\epsilon)} \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})},
\end{aligned}$$

where:

(a) follows from the notation

$$\mathcal{Q} := \{\tilde{a}^n \in \mathcal{C} : \tilde{a}^n \neq a^n, (\tilde{a}^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(A|Y)\};$$

(b) follows from applying Lemma 1. The notation L_1 denotes the maximum path length between s_1 and t. Note that the binning rate r_1 is greater than the source entropy $H(X)$. According to the source-coding theorem [19, Th. 3.4], the probability that a bin contains two typical sequences tends to zero as $n \to \infty$. Hence, we can assume that if $X^n \neq \tilde{X}^n$ are two typical sequences, then $\underline{X}^n \neq \underline{\tilde{X}}^n$;

(c) follows from deriving an upper bound on $|\mathcal{Q}|$. Namely, we are interested in the amount of codewords in $\mathcal{C}$ that are jointly typical with y^n. One may think of it as a random binning of the codebook at a rate of $r_A - I(Y;A) - 2\epsilon$, such that in each bin there are $I(Y;A) - \epsilon$ sequences. Since y^n was generated according to a^n, which is different from $\tilde{a}^n$, then with high probability there will be only one sequence in each bin that is jointly typical with y^n. Therefore, the amount of $\tilde{a}^n$ satisfying $\tilde{a}^n \in \mathcal{Q}$ is bounded by the number of bins, i.e. $2^{n(r_A - I(Y;A) - 2\epsilon)}$.

To upper-bound $\Pr(\mathcal{E}_4)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_4)\\
&= \Pr(\exists \tilde{X}^n \neq X^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{X}^n, \mathbf{0}, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, A^n, Y^n) \in \mathcal{T}_\epsilon^{(n)})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \Pr(\exists \tilde{X}^n \neq x^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{x}^n, \mathbf{0}, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, a^n, y^n) \in \mathcal{T}_\epsilon^{(n)})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \sum_{\substack{\tilde{x}^n \neq x^n:\\ \tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}(X|Y,A)}}\\
&\quad \Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \mathbf{0}, \mathbf{0}]\mathbf{BG}_t = \mathbf{0} | (\tilde{x}^n, a^n, y^n) \in \mathcal{T}_\epsilon^{(n)}\right)\\
&\leq \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) |\mathcal{T}_\epsilon^{(n)}(X|Y,A)| \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})}\\
&\leq \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) 2^{n(H(X|Y,A)+\epsilon)} \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})}\\
&\leq 2^{n(H(X|Y,A)+\epsilon)} \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})}.
\end{aligned}$$

To upper-bound $\Pr(\mathcal{E}_5)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_5)\\
&= \Pr(\exists \tilde{Y}^n \neq Y^n :\\
&\quad [\mathbf{0}, \mathbf{0}, \underline{\tilde{Y}}^n - \underline{Y}^n]\mathbf{BG}_t = \mathbf{0}, (X^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \Pr(\exists \tilde{Y}^n \neq y^n :\\
&\quad [\mathbf{0}, \mathbf{0}, \underline{\tilde{Y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}, (x^n, a^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)\\
&\quad \times \sum_{\substack{\tilde{y}^n \neq y^n:\\ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X,A)}} \Pr\left([\mathbf{0}, \mathbf{0}, \underline{\tilde{y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}\right)\\
&\leq \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) |\mathcal{T}_\epsilon^{(n)}(Y|X,A)| \left(\frac{L_2}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_2;t})}\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) 2^{n(H(Y|X,A)+\epsilon)} \left(\frac{L_2}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_2;t})}\\
&\leq 2^{n(H(Y|X,A)+\epsilon)} \left(\frac{L_2}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_2;t})}.
\end{aligned}$$

To upper-bound $\Pr(\mathcal{E}_6)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_6)\\
&= \Pr(\exists \tilde{X}^n \neq X^n, \tilde{Y}^n \neq Y^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{X}^n, \mathbf{0}, \underline{\tilde{Y}}^n - \underline{Y}^n]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \Pr(\exists \tilde{X}^n \neq x^n, \tilde{Y}^n \neq y^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{x}^n, \mathbf{0}, \underline{\tilde{Y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, a^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)})
\end{aligned}$$

$$= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \times \sum_{\substack{\tilde{x}^n \neq x^n, \tilde{y}^n \neq y^n: \\ (\tilde{x}^n,a^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^{(n)}(X,Y|A)}} \Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \mathbf{0}, \underline{\tilde{y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}\right)$$
$$\leq \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)|\mathcal{T}_\epsilon^{(n)}(X,Y|A)| \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}$$
$$\leq 2^{n(H(X,Y|A)+\epsilon)} \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}.$$

To upper-bound $\Pr(\mathcal{E}_7)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_7) \\
&= \Pr(\exists \tilde{X}^n \neq X^n, \tilde{A}^n \neq A^n, \tilde{Y}^n \neq Y^n : \\
&\quad [\underline{\tilde{X}}^n - \underline{X}^n, \underline{\tilde{A}}^n - \underline{A}^n, \underline{\tilde{Y}}^n - \underline{Y}^n]\mathbf{BG}_t = \mathbf{0}, \\
&\quad (\tilde{X}^n, \tilde{A}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}) \\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \sum_{\tilde{a}^n \neq a^n : \tilde{a}^n \in \mathcal{C}} \Pr(\exists \tilde{X}^n \neq x^n, \tilde{Y}^n \neq y^n : \\
&\quad [\underline{\tilde{X}}^n - \underline{x}^n, \underline{\tilde{a}}^n - \underline{a}^n, \underline{\tilde{Y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}, \\
&\quad (\tilde{X}^n, \tilde{a}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}) \\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \sum_{\tilde{a}^n \neq a^n : \tilde{a}^n \in \mathcal{C}} \\
&\quad \sum_{\substack{\tilde{x}^n \neq x^n, \tilde{y}^n \neq y^n: \\ (\tilde{x}^n,\tilde{a}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^{(n)}(X,Y|A)}} \Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \underline{\tilde{a}}^n - \underline{a}^n, \underline{\tilde{y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}\right) \\
&\leq |\mathcal{C}||\mathcal{T}_\epsilon^{(n)}(X,Y|A)| \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})} \\
&\leq 2^{n(I(X;A)+\epsilon)} 2^{n(H(X,Y|A)+\epsilon)} \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})} \\
&\leq 2^{n(I(X;A)+H(X,Y|A)+2\epsilon)} \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}.
\end{aligned}$$

To conclude the achievable region for this case, note that the events $\mathcal{E}_4$ and $\mathcal{E}_6$ yield redundant constraints; thus, the total probability of error tends to zero for a finite size of network, L_3 and large n only if the inequalities in (7) are satisfied.

2) For the Case $I(X;A) - I(Y;A) \leq 0$: The events corresponding to possible encoding and decoding errors in a terminal node $t \in \mathcal{D}$ are as follows:

$$\begin{aligned}
\mathcal{E}_1 &= \{\nexists A^n : (x^n, A^n) \in \mathcal{T}_\epsilon^{(n)}(X,A), \mathcal{B}_{A^n} = \mathcal{B}_{x^n}\} \\
\mathcal{E}_2 &= \{(X^n, A^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}(X,A,Y)\} \\
\mathcal{E}_3 &= \{\exists \tilde{X}^n \neq X^n : [\underline{\tilde{X}}^n, \underline{A}^n, \underline{Y}^n]\mathbf{BG}_t = \underline{Z}_t, \mathcal{B}_{A^n} = \mathcal{B}_{\tilde{X}^n}, \\
&\quad (\tilde{X}^n, A^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X,A,Y)\} \\
\mathcal{E}_4 &= \{\exists \tilde{Y}^n \neq Y^n : [\underline{X}^n, \underline{A}^n, \underline{\tilde{Y}}^n]\mathbf{BG}_t = \underline{Z}_t, \\
&\quad (X^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X,A,Y)\} \\
\mathcal{E}_5 &= \{\exists \tilde{X}^n \neq X^n, \tilde{Y}^n \neq Y^n : [\underline{\tilde{X}}^n, \underline{A}^n, \underline{\tilde{Y}}^n]\mathbf{BG}_t = \underline{Z}_t, \\
&\quad (\tilde{X}^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X,A,Y)\} \\
\mathcal{E}_6 &= \{\exists \tilde{X}^n \neq X^n, \tilde{A}^n \neq A^n, \tilde{Y}^n \neq Y^n : \\
&\quad [\underline{\tilde{X}}^n, \underline{\tilde{A}}^n, \underline{\tilde{Y}}^n]\mathbf{BG}_t = \underline{Z}_t, \\
&\quad (\tilde{X}^n, \tilde{A}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X,A,Y)\}.
\end{aligned}$$

$\Pr(\mathcal{E}_1) \to 0$ for $n \to \infty$ from the covering lemma since each bin $\mathcal{B}_{A^n}$ contains $I(X;A)+\epsilon$ codewords. $\Pr(\mathcal{E}_2|\mathcal{E}_1^C) \to 0$ as $n \to \infty$ from the same arguments of the case $I(X;A) - I(Y;A) \geq 0$. To upper-bound $\Pr(\mathcal{E}_3)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_3) \\
&= \Pr(\exists \tilde{X}^n \neq X^n : [\underline{\tilde{X}}^n - \underline{X}^n, \mathbf{0}, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}, \\
&\quad (\tilde{X}^n, A^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}, \mathcal{B}_{A^n} = \mathcal{B}_{\tilde{X}^n}) \\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \Pr(\exists \tilde{X}^n \neq x^n : \\
&\quad [\underline{\tilde{X}}^n - \underline{x}^n, \mathbf{0}, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, a^n, y^n) \in \mathcal{T}_\epsilon^{(n)}, \\
&\quad \mathcal{B}_{a^n} = \mathcal{B}_{\tilde{X}^n}) \\
&\stackrel{(a)}{=} \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \\
&\quad \times \sum_{\tilde{x}^n \in \mathcal{Q}} \Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \mathbf{0}, \mathbf{0}]\mathbf{BG}_t = \mathbf{0}\right), \\
&\quad \mathcal{B}_{a^n} = \mathcal{B}_{\tilde{x}^n}\} \\
&\stackrel{(b)}{\leq} \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)|\mathcal{Q}| \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})} \\
&\stackrel{(c)}{\leq} \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) 2^{n(H(X|Y,A)-\Delta+2\epsilon)} \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})} \\
&\leq 2^{n(I(X;A)-I(Y;A)+H(X|Y,A)+3\epsilon)} \left(\frac{L_1}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1;t})},
\end{aligned}$$

where:

(a) is due to the notation $Q := ...$; See page 5.
(b) follows from applying Lemma 1. The notation L_1 denotes the maximum path length between s_1 and t. Note that $x^n \neq \tilde{x}^n$ implies $\underline{x}^n \neq \underline{\tilde{x}}^n$ from the same arguments in the analysis of the case $I(X;A) - I(Y;A) \geq 0$;
(c) follows from deriving an upper bound on $|\mathcal{Q}|$. Namely, we are interested in the amount of source sequences $\tilde{X}^n$ that are jointly typical with (y^n, a^n), moreover the first $n\Delta$ bits of $\underline{\tilde{x}}^n$ need to be identical to the bin $\mathcal{B}_{a^n}$. The size of this conditional typical set is $2^{n(H(X|Y,A)+2\epsilon)}$, since we know the first $n\Delta$ bits the amount of sequences that fall into this criteria is $2^{n(H(X|Y,A)-\Delta)}$.

To upper-bound $\Pr(\mathcal{E}_4)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_4) \\
&= \Pr(\exists \tilde{Y}^n \neq Y^n : [\underline{X}^n, \underline{A}^n, \underline{\tilde{Y}}^n]\mathbf{BG}_t = \underline{Z}_t, \\
&\quad (X^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X,A,Y)) \\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \cdot \Pr(\exists \tilde{Y}^n \neq y^n : \\
&\quad [\mathbf{0}, \mathbf{0}, \underline{\tilde{Y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}, (x^n, a^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}) \\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \\
&\quad \times \sum_{\tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X,A)} \Pr\left([\mathbf{0}, \mathbf{0}, \underline{\tilde{y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}\right), \\
&\leq \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)|\mathcal{T}_\epsilon^{(n)}(Y|X,A)| \left(\frac{L_2}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_2;t})}
\end{aligned}$$

$$\le \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) 2^{n(H(Y|X,A)+2\epsilon)} \left(\frac{L_2}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_2;t})}$$
$$\le 2^{n(H(Y|X,A)+2\epsilon)} \left(\frac{L_2}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_2;t})},$$

To upper-bound $\Pr(\mathcal{E}_5)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_5)\\
&= \Pr(\exists \tilde{X}^n \neq X^n, \tilde{Y}^n \neq Y^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{X}^n, \mathbf{0}, \underline{\tilde{Y}}^n - \underline{Y}^n]\mathbf{BG}_t = \mathbf{0},\\
&\quad (\tilde{X}^n, A^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}, \mathcal{B}_{A^n} = \mathcal{B}_{\tilde{X}^n})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \cdot \Pr(\exists \tilde{X}^n \neq x^n, \tilde{Y}^n \neq y^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{x}^n, \mathbf{0}, \underline{\tilde{Y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}, (\tilde{X}^n, a^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)})\\
&\le \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)\\
&\quad \times \sum_{\substack{\tilde{x}^n \neq x^n, \tilde{y}^n \neq y^n:\\ (\tilde{x}^n,a^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^{(n)}(X,Y|A)}} \Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \mathbf{0}, \underline{\tilde{y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}\right)\\
&\le \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) |\mathcal{T}_\epsilon^{(n)}(X,Y|A)| \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}\\
&\le 2^{n(H(X,Y|A)+\epsilon)} \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}.
\end{aligned}$$

To upper-bound $\Pr(\mathcal{E}_6)$, we have

$$\begin{aligned}
&\Pr(\mathcal{E}_6)\\
&= \Pr(\exists \tilde{X}^n \neq X^n, \tilde{A}^n \neq A^n, \tilde{Y}^n \neq Y^n : [\underline{\tilde{X}}^n - \underline{X}^n,\\
&\quad \underline{\tilde{A}}^n - \underline{A}^n, \underline{\tilde{Y}}^n - \underline{Y}^n]\mathbf{BG}_t = \mathbf{0},\\
&\quad (\tilde{X}^n, \tilde{A}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}, \mathcal{B}_{\tilde{A}^n} = \mathcal{B}_{\tilde{X}^n})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n)\\
&\quad \times \Pr(\exists \tilde{X}^n \neq x^n, \tilde{A}^n \neq a^n, \tilde{Y}^n \neq y^n :\\
&\quad [\underline{\tilde{X}}^n - \underline{x}^n, \underline{\tilde{A}}^n - \underline{a}^n, \underline{\tilde{Y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0},\\
&\quad (\tilde{X}^n, \tilde{A}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}, \mathcal{B}_{\tilde{A}^n} = \mathcal{B}_{\tilde{X}^n})\\
&= \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \sum_{\tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}(X)} \Pr(\exists \tilde{A}^n \neq a^n, \tilde{Y}^n \neq y^n :\\
&\quad [\underline{\tilde{x}}^n - \underline{x}^n, \underline{\tilde{a}}^n - \underline{a}^n, \underline{\tilde{Y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0},\\
&\quad (\tilde{x}^n, \tilde{A}^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}, \mathcal{B}_{\tilde{A}^n} = \mathcal{B}_{\tilde{x}^n})\\
&\overset{(a)}{=} \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) \sum_{\tilde{x}^n \in \mathcal{T}_\epsilon^{(n)}(X)} \sum_{\substack{\tilde{y}^n \neq y^n:\\ (\tilde{x}^n,\tilde{a}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^{(n)}(Y|X,A)}}\\
&\quad \Pr\left([\underline{\tilde{x}}^n - \underline{x}^n, \underline{\tilde{a}}^n - \underline{a}^n, \underline{\tilde{y}}^n - \underline{y}^n]\mathbf{BG}_t = \mathbf{0}\right)\\
&\le \sum_{(x^n,a^n,y^n)} |\mathcal{T}_\epsilon^{(n)}(X,Y|A)| \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}\\
&\le \sum_{(x^n,a^n,y^n)} P(x^n,a^n,y^n) 2^{n(H(X,Y|,A)+2\epsilon)} \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}\\
&\le 2^{n(H(X)+H(Y|X,A)+2\epsilon)} \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})}\\
&= 2^{n(I(X;A)+H(X,Y|A)+2\epsilon)} \left(\frac{L_3}{2^n}\right)^{\mathcal{C}(\mathcal{V}^*_{s_1,s_2;t})},
\end{aligned}$$

where (a) follows from the fact that for a given $\tilde{x}^n$, there is only one actions codeword denoted by $\tilde{a}^n$ which is jointly typical with $\tilde{x}^n$ and satisfying $\mathcal{B}_{\tilde{a}^n} = \mathcal{B}_{\tilde{x}^n}$.

Note that the constraint induced by the event $\mathcal{E}_5$ is redundant and, therefore, the constraints in (7) are sufficient to show that the total probability of error tends to zero as n tends to infinity.

VII. Conclusions and Future Work

In the current work, we have considered the setup of correlated sources with action-dependent joint distribution. Specifically, the optimal rate regions were characterized for the case where actions taken at the decoder and for the case of actions taken at the encoder. Further, we have presented the set of achievable rates for a scenario where action-dependent sources are known at different nodes of a general network and are required at a set of terminal nodes. Remarkably, RLNC was proved to be optimal also for this scenario, even though this is not a multicast problem. Moreover, the set of achievable rates involved mutual information terms, which are not typical in multicast problems. Two binary examples were studied, and it was shown how actions affect the achievable rate region in a non-trivial manner.

As can be seen from this and additional work [7]–[12], actions have a significant impact on the set of achievable rates in source coding problems and many classical source coding problems can be extended using actions. One particular, as yet unsolved, source coding problem that would be interesting to study is the case of action-dependent source coding with a helper. In this scenario the considered setup is of correlated sources with actions, yet only a reconstruction of X^n is required at the decoder. In the source coding helper problem, the sequence Y^n which is being transmitted on a rate-limited link plays the role of SI and not of an information source as in our model. The main difficulty in proving the converse follows from the fact that Y^n is not distributed i.i.d. as in the original problem of source coding with a helper [6].

Acknowledgment

The authors would like to thank the Associate Editor and anonymous reviewers, for their valuable and constructive comments, which helped to improve this paper.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
[2] T. Ho *et al.*, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
[3] T. Berger, "Multiterminal source coding," in *Information Theory Approach to Communications*, G. Longo, Ed. New York, NY, USA: Springer-Verlag, 1978, pp. 171–231.
[4] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School Elect. Eng., Cornell Univ., New York, NY, USA, 1978.
[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
[6] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, Nov. 1975.

[7] H. H. Permuter and T. Weissman, "Source coding with a side information 'vending machine,'" *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4530–4544, Jul. 2011.

[8] L. Zhao, Y.-K. Chia, and T. Weissman, "Compression with actions," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 796–807, Feb. 2014.

[9] O. Simeone, "Source coding with in-block memory and controllable causal side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 3060–3064.

[10] K. Kittichokechai, T. J. Oechtering, M. Skoglund, and R. Thobaben, "Source and channel coding with action-dependent partially known two-sided state information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2010 pp. 629–633.

[11] B. Ahmadi, H. Asnani, O. Simeone, and H. H. Permuter, "Information embedding on actions," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6902–6916, Nov. 2014.

[12] Y.-K. Chia, H. Asnani, and T. Weissman, "Multiterminal source coding with action-dependent side information," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3653–3667, Jun. 2013.

[13] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.

[14] C. Choudhuri and U. Mitra, "How useful is adaptive action?" in *Proc. Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 2251–2255.

[15] H. H. Permuter and H. Asnani, "Multiple access channel with partial and controlled cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2252–2266, Apr. 2013.

[16] L. Dikstein, H. H. Permuter, and S. Shamai, "MAC with action-dependent state information at one encoder," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 173–188, Jan. 2015.

[17] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.

[18] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[19] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[20] A. Cohen, S. Avestimehr, and M. Effros, "On networks with side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2009, pp. 1343–1347.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[22] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[23] F. Göring, "Short proof of menger's theorem," *Discrete Math.*, vol. 219, nos. 1–3, pp. 295–296, 2000.

Oron Sabag (S'14) received his B.Sc. (cum laude) degree in Electrical and Computer Engineering from the Ben-Gurion University of the Negev, Israel, in 2013. He is currently pursuing his Ph.D in the direct track for honor students in Electrical and Computer Engineering at the same institution.

Oron is a recipient of the Lachish Fellowship for honor students in the direct Ph.D. Program and excellence scholarship from the Electrical and Computer Engineering department, Ben-Gurion University of the Negev, Israel.

Haim H. Permuter (M'08–SM'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY.

Asaf Cohen (S'00–M'07) is a senior lecturer at the Department of Communication Systems Engineering, Ben-Gurion University of the Negev, Israel. Before that, he was a post-doctoral scholar at the California Institute of Technology (Caltech). He received the B.Sc., M.Sc. (both with high honors) and Ph.D. from the Department of Electrical Engineering, Technion, Israel Institute of Technology, in 2001, 2003 and 2007, respectively. From 1998 to 2000 he was with the IBM Research Laboratory in Haifa where he worked on distributed computing. During the 2016 academic year he was at a sabbatical leave at MIT. His areas of interest are information theory, learning and coding. In particular, he is interested in sequential decision making, with applications to detection and estimation; Network security; Network information theory and network coding; Coding theory and performance analysis of codes. Dr. Cohen received several honors and awards, including the Viterbi post-doctoral scholarship, a student paper award at IEEE Israel 2006 and the Dr. Philip Marlin Prize for Computer Engineering, 2000.