arXiv:2105.05917v2 [cs.IT] 26 May 2021

Two-Hop Network with Multiple Decision Centers under Expected-Rate Constraints

Mustapha Hamad
LTCI, Telecom Paris, IP Paris
91120 Palaiseau, France
mustapha.hamad@telecom-paris.fr

Michèle Wigger
LTCI, Telecom Paris, IP Paris
91120 Palaiseau, France
michele.wigger@telecom-paris.fr

Mireille Sarkiss
SAMOVAR, Telecom SudParis, IP Paris
91011 Evry, France
mireille.sarkiss@telecom-sudparis.eu

***Abstract*—The paper studies distributed binary hypothesis testing over a two-hop relay network where both the relay and the receiver decide on the hypothesis. Both communication links are subject to *expected* rate constraints, which differs from the classical assumption of maximum rate constraints. We exactly characterize the set of type-II error exponent pairs at the relay and the receiver when both type-I error probabilities are constrained by the same value $\epsilon > 0$. No tradeoff is observed between the two exponents, i.e., one can simultaneously attain maximum type-II error exponents both at the relay and at the receiver. For $\epsilon_1 \neq \epsilon_2$, we present an achievable exponents region, which we obtain with a scheme that applies different versions of a basic two-hop scheme that is optimal under *maximum* rate constraints. We use the basic two-hop scheme with two choices of parameters and rates, depending on the transmitter's observed sequence. For $\epsilon_1 = \epsilon_2$, a single choice is shown to be sufficient. Numerical simulations indicate that extending to three or more parameter choices is never beneficial.**

***Index Terms*—Multi-hop, distributed hypothesis testing, error exponents, expected rate constraints, variable-length coding,**

I. Introduction

In many Internet of things (IoT) and sensor networks, the sensors may not communicate directly with the decision center due to limited resources or environmental effects. This motivates us to consider multi-hop networks where the sensor can communicate to the decision center only via a relay. In certain scenarios, the relays also wish to decide on the hypothesis, for example to faster raise alarms. In such distributed hypothesis testing problems, the relays and the receiver have to decide on a binary hypothesis to determine the joint distributions underlying all terminals' observations including their own. In particular, maximizing the accuracy of any taken decision under imposed *communication rate constraints* is an important concern in many applications related to security, health monitoring, or incident-detection. In these applications, often the error under the alternative hypothesis corresponding to a *missed detection* is more critical than the error under the null hypothesis corresponding to *false alarms*. We thus aim at maximizing the exponential decays of the missed detection probabilities under given thresholds on the false alarm probabilities. As we shall see, a particular challenge arises when the relay and the decision center have different thresholds on the tolerable false-alarm probabilities.

Most information-theoretic works on distributed hypothesis testing focus on *maximum rate constraints* [1]–[6]. *Expected rate constraints* were introduced in [7], [8], which also characterized the maximum error exponents for single-sensor single-decision center setups in the special case of testing-against independence. The optimal coding and decision scheme in [7], [8] chooses an event $\mathcal{S}_n$ of probability close to the permissible type-I error probability ϵ. Under this event, the transmitter sends a single flag bit to the decision center, which then decides on the hypothesis $\mathcal{H} = 1$. Otherwise, the transmitter and the receiver run the optimal scheme under the maximum rate constraints [1], [2]. The described scheme achieves same type-II error exponent as in [1], [2], but with a communication rate reduced by the factor of $(1-\epsilon)$. Similar conclusions also hold for more complicated networks with multiple communication links, as we showed in [9] at hand of the partially-cooperating multi-access network with two sensors.

In this paper, we consider the two-hop network, where the observations at the transmitter X^n, the relay Y^n, and the receiver Z^n form a Markov chain $X^n \to Y^n \to Z^n$. Such a Markov chain often occurs simply because the transmitter is closer to the relay than to the receiver. Under maximum rate-constraints, the optimal exponents at the relay and the receiver were characterized in [10], [11]. We show that when both the transmitter and the relay have same $\epsilon_1 = \epsilon_2$, then under expected rate constraints one can boost both rates by a factor $(1-\epsilon)^{-1}$ as compared to a maximum rate-constraint. The case $\epsilon_1 \neq \epsilon_2$ differs in various ways. Firstly, our set of achievable exponent pairs indicates a tradeoff between the relay's and the receiver's exponents. Secondly, a more complicated coding and decision scheme is required. Specifically, we propose a strategy where the transmitter chooses three events, and depending on the event, applies either a degenerate single-flagbit strategy or the scheme in [10] with one of two different choices of parameters and rates, depending on the transmitter's observation. Extending to more than three events (i.e., to more than two parameter and rate choices for the scheme in [10]) however does not seem to yield further improvements.

Notation: We follow the notation in [12], [8]. In particular, we use sans serif font for bit-strings: e.g., m for a deterministic and M for a random bit-string. We let $\mathrm{string}(m)$ denote the shortest bit-string representation of a positive integer m, and for any bit-string m we let $\mathrm{len}(\mathsf{m})$ and $\mathrm{dec}(\mathsf{m})$ denote its length and its corresponding positive integer. In addition, $\mathcal{T}_{\mu}^{(n)}$ denotes the strongly typical set given by [13, Definition 2.8].

II. System Model

Consider the distributed hypothesis testing problem in Fig. 1 under the Markov chain

$$X^n \to Y^n \to Z^n \tag{1}$$

and in the special case of testing against independence, i.e., depending on the binary hypothesis $\mathcal{H} \in \{0,1\}$, the tuple (X^n, Y^n, Z^n) is distributed as:

$$\text{under } \mathcal{H}=0 : (X^n, Y^n, Z^n) \sim \text{i.i.d. } P_{XY} \cdot P_{Z|Y}; \tag{2a}$$
$$\text{under } \mathcal{H}=1 : (X^n, Y^n, Z^n) \sim \text{i.i.d. } P_X \cdot P_Y \cdot P_Z \tag{2b}$$

for given probability mass functions (pmfs) P_{XY} and $P_{Z|Y}$.

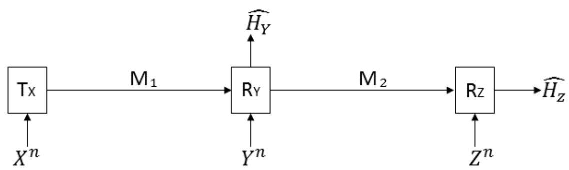

Fig. 1: Cascaded two-hop setup with two decision centers.

The system consists of a transmitter T_X, a relay R_Y, and a receiver R_Z. The transmitter T_X observes the source sequence X^n and sends its bit-string message $\mathsf{M}_1 = \phi_1^{(n)}(X^n)$ to R_Y, where the encoding function is of the form $\phi_1^{(n)} : \mathcal{X}^n \to \{0,1\}^\star$ and satisfies the *expected* rate constraint

$$\mathbb{E}\left[\text{len}\left(\mathsf{M}_1\right)\right] \leq nR_1. \tag{3}$$

The relay R_Y observes the source sequence Y^n and with the message M_1 received from T_X, it produces a guess $\hat{\mathcal{H}}_Y$ of the hypothesis $\mathcal{H}$ using a decision function $g_1^{(n)} : \mathcal{Y}^n \times \{0,1\}^\star \to \{0,1\}$:

$$\hat{\mathcal{H}}_Y = g_1^{(n)}\left(\mathsf{M}_1, Y^n\right) \in \{0,1\}. \tag{4}$$

Relay R_Y also computes a bit-string message $\mathsf{M}_2 = \phi_2^{(n)}\left(Y^n, \mathsf{M}_1\right)$ using some encoding function $\phi_2^{(n)} : \mathcal{Y}^n \times \{0,1\}^\star \to \{0,1\}^\star$ that satisfies the expected rate constraint

$$\mathbb{E}\left[\text{len}\left(\mathsf{M}_2\right)\right] \leq nR_2. \tag{5}$$

Then it sends M_2 to the receiver R_Z, which guesses hypothesis $\mathcal{H}$ using its observation Z^n and the received message M_2, i.e., using a decision function $g_2^{(n)} : \mathcal{Z}^n \times \{0,1\}^\star \to \{0,1\}$, it produces the guess:

$$\hat{\mathcal{H}}_Z = g_2^{(n)}\left(\mathsf{M}_2, Z^n\right) \in \{0,1\}. \tag{6}$$

The goal is to design encoding and decision functions such that their type-I error probabilities

$$\alpha_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_Y = 1 | \mathcal{H} = 0] \tag{7}$$
$$\alpha_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_Z = 1 | \mathcal{H} = 0] \tag{8}$$

stay below given thresholds $\epsilon_1 > 0$, $\epsilon_2 > 0$ and the type-II error probabilities

$$\beta_{1,n} \triangleq \Pr[\hat{\mathcal{H}}_Y = 0 | \mathcal{H} = 1] \tag{9}$$
$$\beta_{2,n} \triangleq \Pr[\hat{\mathcal{H}}_Z = 0 | \mathcal{H} = 1] \tag{10}$$

decay to 0 with largest possible exponential decay.

Definition 1: Fix maximum type-I error probabilities $\epsilon_1, \epsilon_2 \in [0,1]$ and rates $R_1, R_2 \geq 0$. The exponent pair (θ_1, θ_2) is called (ϵ_1, ϵ_2)*-achievable* if there exists a sequence of encoding and decision functions $\{\phi_1^{(n)}, \phi_2^{(n)}, g_1^{(n)}, g_2^{(n)}\}_{n \geq 1}$ satisfying $\forall i \in \{1,2\}$:

$$\mathbb{E}[\text{len}(\mathsf{M}_i)] \leq nR_i, \tag{11}$$
$$\varlimsup_{n \to \infty} \alpha_{i,n} \leq \epsilon_i, \tag{12}$$
$$\varliminf_{n \to \infty} \frac{1}{n} \log \frac{1}{\beta_{i,n}} \geq \theta_i. \tag{13}$$

Definition 2: The closure of the set of all (ϵ_1, ϵ_2)-achievable exponent pairs (θ_1, θ_2) is called the (ϵ_1, ϵ_2)*-exponents region* and is denoted by $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$.

The maximum exponents that are achievable at each of the two decision centers are also of interest:

$$\theta_{1,\epsilon_1}^*(R_1) := \max\{\theta_1 \colon (\theta_1, \theta_2) \in \mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2) \text{ for some } \epsilon_2 > 0, \theta_2 \geq 0\} \tag{14}$$
$$\theta_{2,\epsilon_2}^*(R_1, R_2) := \max\{\theta_2 \colon (\theta_1, \theta_2) \in \mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2) \text{ for some } \epsilon_1 > 0, \theta_1 \geq 0\}. \tag{15}$$

Remark 1: The multi-hop hypothesis testing setup of Fig. 1 and Equations (2) was also considered in [10] and [11], but under *maximum rate constraints*:

$$\text{len}(\mathsf{M}_i) \leq nR_i, \qquad i \in \{1,2\}, \tag{16}$$

instead of the *expected rate constraints* (3) and (5).

As shown in [11], for any rates $R_1, R_2 \geq 0$ and permissible type-I error probabilities $\epsilon_1, \epsilon_2 \in [0, 1/2]$, the exponents region under the maximum-rate constraints (16) is:

$$\mathcal{E}_{\text{max}}^*(R_1, R_2, \epsilon_1, \epsilon_2) = \{(\theta_1, \theta_2) : \theta_1 \leq \theta_{1,\epsilon_1,\text{max}}^*\left(R_1\right), \tag{17}$$
$$\theta_2 \leq \theta_{2,\epsilon_2,\text{max}}^*\left(R_1, R_2\right)\}, \tag{18}$$

where

$$\theta_{1,\epsilon_1,\text{max}}^*\left(R_1\right) = \max_{\substack{P_{U_1|X}: \\ R_1 \geq I(U_1;X)}} I\left(U_1; Y\right) \tag{19}$$
$$\theta_{2,\epsilon_2,\text{max}}^*\left(R_1, R_2\right) = \theta_{1,\epsilon_1,\text{max}}^*\left(R_1\right) + \max_{\substack{P_{U_2|Y}: \\ R_2 \geq I(U_2;Y)}} I\left(U_2; Z\right) \tag{20}$$

and the mutual information quantities are calculated using the joint pmfs $P_{U_1XY} \triangleq P_{U_1|X} P_{XY}$ and $P_{U_2YZ} \triangleq P_{U_2|Y} P_{YZ}$.

In the following subsection III-A we present a coding and decision scheme that achieves $\mathcal{E}_{\text{max}}^*(R_1, R_2, \epsilon_1, \epsilon_2)$. It is a simplification of the scheme in [10].

III. Coding and Decision Schemes

In Subsection III-A, we present a basic two-hop hypothesis testing scheme, which we obtain by simplifying the general scheme in [10] and which suffices to achieve the exponents region $\mathcal{E}_{\text{max}}^*$ under maximum rate constraints.

For the setup with *expected* rate constraints studied in this paper, in Subsections III-B–III-D we propose to use different versions of this two-hop scheme (with different parameters and different communication rates) depending on the transmitter's observation x^n, where for certain sequences x^n we even apply degenerate versions of the scheme where only zero-rate flag-bits are sent over one or both communication links. Notice that in principle, we could apply a different set of parameters for each observation $x^n \in \mathcal{X}^n$. Our numerical examples however indicate that without loss in optimality one can restrict to only one or two parameter choices and an additional degenerate version of the scheme with zero communication rates on both links. As proved by the scheme in Subsection III-B and Theorem 1 a single parameter choice suffices when $\epsilon_1 = \epsilon_2$. For $\epsilon_1 \neq \epsilon_2$ two parameter choices are strictly better as we show in our numerical simulations in Section IV-A. More choices seem unnecessary.

A. A basic two-hop coding and decision scheme [10]

We revisit a simplified version of the scheme in [10], which achieves the exponents region under maximum rate constraints $\mathcal{E}^*_{\text{max}}(R_1, R_2, \epsilon_1, \epsilon_2)$ for any ϵ_1, ϵ_2.

Fix a blocklength n and choose the following parameters: a small positive number $\mu > 0$, conditional pmfs $P_{U_1|X}$ and $P_{U_2|Y}$. In the following, all mutual informations will be evaluated according to the joint pmf $P_{XYZU_1U_2} := P_X P_{Y|X} P_{Z|Y} P_{U_1|X} P_{U_2|Y}$.

Randomly generate the codebooks

$$\mathcal{C}_{U_1} \triangleq \left\{ u_1^n(m_1) : m_1 \in \left\{1, \cdots, 2^{n(I(U_1;X)+\mu)}\right\}\right\} \quad (21)$$

$$\mathcal{C}_{U_2} \triangleq \left\{ u_2^n(m_2) : m_2 \in \left\{1, \cdots, 2^{n(I(U_2;Y)+\mu)}\right\}\right\}, \quad (22)$$

by drawing all entries i.i.d. according to the marginal pmfs P_{U_1} and P_{U_2}.

<u>T_X:</u> Assume it observes $X^n = x^n$. If $x^n \in \mathcal{T}_\mu^{(n)}(P_X)$, it looks for indices m_1 satisfying $(u_1^n(m_1), x^n) \in \mathcal{T}_\mu^{(n)}(P_{U_1X})$, randomly picks one of these indices, and sends its corresponding bit-string

$$\mathsf{M}_1 = [\text{string}(m_1)]. \quad (23)$$

If no such index exists or if $x^n \notin \mathcal{T}_\mu^{(n)}(P_X)$, then T_X sends

$$\mathsf{M}_1 = [0]. \quad (24)$$

<u>R_Y:</u> Assume it observes $Y^n = y^n$ and receives the bit-string message $\mathsf{M}_1 = \mathsf{m}_1$.

If $\mathsf{m}_1 = [0]$, then

$$\hat{\mathcal{H}}_Y = 1 \qquad \text{and} \qquad \mathsf{M}_2 = [0]. \quad (25)$$

Else it checks if $(u_1^n(\mathsf{m}_1), y^n) \in \mathcal{T}_\mu^{(n)}(P_{U_1Y})$. If the check is successful R_Y declares $\hat{\mathcal{H}}_Y = 0$; otherwise it declares $\hat{\mathcal{H}}_Y = 1$ and sends $\mathsf{M}_2 = [0]$.

If $\hat{\mathcal{H}}_Y = 0$, R_Y next looks for indices m_2 satisfying $(u_2^n(m_2), y^n) \in \mathcal{T}_\mu^{(n)}(P_{U_2Y})$, randomly picks one of them and sends

$$\mathsf{M}_2 = \text{string}(m_2) \quad (26)$$

to the receiver.

If no such index m_2 exists, R_Y directly sends string

$$\mathsf{M}_2 = [0]. \quad (27)$$

<u>R_Z:</u> Assume it observes the sequence $Z^n = z^n$ and receives message $\mathsf{M}_2 = \mathsf{m}_2$.

If $\mathsf{m}_2 = [0]$, it declares $\hat{\mathcal{H}}_Z = 1$.

Else it sets $m_2 = \text{dec}(\mathsf{m}_2)$, and checks if $(u_2^n(m_2), z^n) \in \mathcal{T}_\mu^{(n)}(P_{U_2Z})$. It declares $\hat{\mathcal{H}}_Z = 0$ if the check succeeds, and $\hat{\mathcal{H}}_Z = 1$ otherwise.

In the following subsections, we explain how to employ this basic scheme in a variable-length coding framework.

B. Variable-length coding for $\epsilon_1 = \epsilon_2$

We employ only a single version of the two-hop scheme, and combine it with a degenerate scheme that has zero communication rates over both links. Specifically, as for the point-to-point setup in [8], we choose a subset $\mathcal{S}_n \subseteq \mathcal{T}_\mu^{(n)}(P_X)$ of probability

$$\Pr[X^n \in \mathcal{S}_n] = \epsilon_2 - \mu = \epsilon_1 - \mu, \quad (28)$$

for some small number $\mu > 0$.

Whenever $X^n \in \mathcal{S}_n$, T_X and R_Y both send

$$\mathsf{M}_1 = \mathsf{M}_2 = [0] \quad (29)$$

and R_Y and R_Z decide on

$$\hat{\mathcal{H}}_Y = \hat{\mathcal{H}}_Z = 1. \quad (30)$$

Whenever $X^n \notin \mathcal{S}_n$, the terminals T_X, R_Y, R_Z all follow the basic two-hop scheme in Subsection III-A for parameters $\mu, P_{U_1|X}, P_{U_2|Y}$ satisfying

$$R_1 = (1 - \epsilon_1 + \mu)\left(I(U_1; X) + 2\mu\right) \quad (31)$$

$$R_2 = (1 - \epsilon_2 + \mu)\left(I(U_2; Y) + 2\mu\right). \quad (32)$$

The factors $(1-\epsilon_1+\mu)$ and $(1-\epsilon_2+\mu)$ in front of the mutual information terms represent the gain obtained by *expected* rate constraints, because with probability $\epsilon_1 - \mu = \epsilon_2 - \mu$ in our scheme both messages M_1 and M_2 are of zero rate, see (29).

In Appendix A, we prove that the presented scheme achieves the error exponents in Eq. (48) of Theorem 1 when $n \to \infty$ and $\mu \downarrow 0$.

C. Variable-length coding for $\epsilon_2 > \epsilon_1$

We employ two versions of the basic two-hop scheme as we will explain shortly. Moreover, we again choose a subset $\mathcal{S}_n \subseteq \mathcal{T}_\mu^{(n)}(P_X)$ of probability

$$\Pr[X^n \in \mathcal{S}_n] = \epsilon_1 - \mu, \quad (33)$$

and all terminals T_X, R_Y, R_Z apply the degenerate scheme in (29)–(30) whenever $X^n \in \mathcal{S}_n$.

We now partition the remaining set $\mathcal{X}^n \backslash \mathcal{S}_n$ into two disjoint sets $\mathcal{D}_n'$ and $\mathcal{D}_n''$

$$\mathcal{D}_n' \cup \mathcal{D}_n'' = \mathcal{X}^n \backslash \mathcal{S}_n \quad \text{and} \quad \mathcal{D}_n' \cap \mathcal{D}_n'' = \emptyset \quad (34)$$

such that

$$\Pr[X^n \in \mathcal{D}_n'] = 1 - \epsilon_2 + \mu \quad (35)$$

$$\Pr\left[X^n \in \mathcal{D}_n''\right] = \epsilon_2 - \epsilon_1. \tag{36}$$

We further split $R_1 = R_1' + R_1''$ for $R_1', R_1'' > 0$.

Then, whenever $x^n \in \mathcal{D}_n'$, all terminals T_X, R_Y, R_Z follow the basic two-hop scheme for a set of parameters $\mu, P_{U_1'|X}, P_{U_2'|Y}$ satisfying

$$R_1' = (1 - \epsilon_2 + \mu)(I(U_1'; X) + 2\mu) \tag{37}$$

$$R_2 = (1 - \epsilon_2 + \mu)(I(U_2'; Y) + 2\mu). \tag{38}$$

To inform the relay and the receiver about the event $x^n \in \mathcal{D}_n'$, both T_X and R_Y add $[1,0]$-flag bits at the beginning of their communication to R_Y and R_Z, respectively. (Notice that two additional bits do not change the rate of communication.)

For $x^n \in \mathcal{D}_n''$, the transmitter and the relay still follow the basic two-hop scheme in Subsection III-A but now for a different parameter choice $\mu, P_{U_1''|X}$ satisfying

$$R_1'' = (\epsilon_2 - \epsilon_1)(I(U_1''; X) + 2\mu), \tag{39}$$

and where T_X additionally sends the $[1,1]$-flag as part of M_1 to R_Y, which simply relays this flag $\mathsf{M}_2 = [1,1]$ without adding additional information. Upon observing $\mathsf{M}_2 = [1,1]$, R_Z immediately declares $\hat{\mathcal{H}}_Z = 1$.

In Appendix B, we prove that the presented scheme achieves the error exponents in Eq. (49) of Theorem 1 when $n \to \infty$ and $\mu \downarrow 0$.

D. Variable-length coding for $\epsilon_1 > \epsilon_2$

In this case, we employ two full versions of the basic two-hop scheme. Moreover, we again choose a subset $\mathcal{S}_n \subseteq \mathcal{T}_\mu^{(n)}(P_X)$ of probability

$$\Pr\left[X^n \in \mathcal{S}_n\right] = \epsilon_2 - \mu, \tag{40}$$

and partition the remaining subset of $\mathcal{X}^n$ into two disjoint sets $\mathcal{D}_n'$ and $\mathcal{D}_n''$

$$\mathcal{D}_n' \cup \mathcal{D}_n'' = \mathcal{X}^n \backslash \mathcal{S}_n \quad \text{and} \quad \mathcal{D}_n' \cap \mathcal{D}_n'' = \emptyset \tag{41}$$

such that

$$\Pr\left[X^n \in \mathcal{D}_n'\right] = 1 - \epsilon_1 + \mu \tag{42}$$

$$\Pr\left[X^n \in \mathcal{D}_n''\right] = \epsilon_1 - \epsilon_2. \tag{43}$$

We further split $R_1 = R_1' + R_1''$ and $R_2 = R_2' + R_2''$ for $R_1', R_1'', R_2', R_2'' > 0$.

Whenever $X^n \in \mathcal{S}_n$, T_X, R_Y, and R_Z, all apply the degenerate scheme in (29)–(30).

Whenever $X^n \in \mathcal{D}_n'$, all terminals T_X, R_Y, and R_Z follow the basic two-hop scheme for a choice of parameters $\mu, P_{U_1'|X}, P_{U_2'|Y}$ satisfying

$$R_1' = (1 - \epsilon_1 + \mu)(I(U_1'; X) + 2\mu) \tag{44}$$

$$R_2' = (1 - \epsilon_1 + \mu)I(U_2'; Y) + 2\mu). \tag{45}$$

Additionally, T_X and R_Y add $[1,0]$-flag bits to their messages M_1 and M_2 to indicate to R_Y and R_Z that $X^n \in \mathcal{D}_n'$.

Whenever $X^n \in \mathcal{D}_n''$, all terminals T_X, R_Y, and R_Z mostly follow the basic two-hop scheme but now for parameters $\mu, P_{U_1''|X}, P_{U_2''|Y}$ satisfying

$$R_1'' = (\epsilon_1 - \epsilon_2)(I(U_1''; X) + 2\mu) \tag{46}$$

$$R_2'' = (\epsilon_1 - \epsilon_2)(I(U_2''; Y) + 2\mu). \tag{47}$$

The only exceptions are that T_X and R_Y add a $[1,1]$-flag to their messages M_1 and M_2 to indicate to R_Y and R_Z that $X^n \in \mathcal{D}_n''$, and that R_Y *always* declares $\hat{\mathcal{H}}_Y = 1$ upon observing this $[1,1]$-flag in M_1, irrespective of the remaining bits of M_1 or its observation Y^n. Besides this decision, R_Y however follows the protocol of the basic two-hop scheme which forces it to compute a tentative decision $\hat{\mathcal{H}}_Y''$, which determines its communication to R_Z. (In particular, if $\hat{\mathcal{H}}_Y'' = 1$, R_Y sends only the $[1,1]$-flag to R_Z so that R_Z immediately declares $\hat{\mathcal{H}}_Z = 1$.) Notice that while R_Y can ignore the tentative decision $\hat{\mathcal{H}}_Y''$ because of its larger permissible type-I error probability $\epsilon_1 > \epsilon_2$, this decision is important for R_Z so that this latter can satisfy its constraint on the type-I probability ϵ_2.

In a similar way to the previous schemes, it can be shown that this scheme achieves the error exponents in Eq. (50) of Theorem 1 when $n \to \infty$ and $\mu \downarrow 0$. Details are presented in Appendix C.

IV. Results on the Exponents Region

Our main result provides inner bounds to the exponent region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ achieved by the schemes presented in the preceding Section III. The theorem further provides an exact characterization of exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ when $\epsilon_1 = \epsilon_2$.

Theorem 1: If $\epsilon_1 = \epsilon_2$, the (ϵ_1, ϵ_2)-exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ *is the set of* all (θ_1, θ_2) pairs satisfying

$$\theta_1 \leq I(U_1; Y), \tag{48a}$$

$$\theta_2 \leq I(U_1; Y) + I(U_2; Z), \tag{48b}$$

for some conditional pmfs $P_{U_1|X}, P_{U_2|Y}$ so that

$$R_1 \geq (1 - \epsilon_1) I(U_1; X), \tag{48c}$$

$$R_2 \geq (1 - \epsilon_2) I(U_2; Y), \tag{48d}$$

and where the mutual information quantities are calculated using the joint pmfs $P_{U_1XY} \triangleq P_{U_1|X} P_{XY}$ and $P_{U_2YZ} \triangleq P_{U_2|Y} P_{YZ}$.

If $\epsilon_1 < \epsilon_2$, the (ϵ_1, ϵ_2)-exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ *contains* all (θ_1, θ_2) pairs that satisfy

$$\theta_1 \leq \min\{I(U_1'; Y), I(U_1''; Y)\}, \tag{49a}$$

$$\theta_2 \leq I(U_1'; Y) + I(U_2'; Z), \tag{49b}$$

for some conditional pmfs $P_{U_1'|X}, P_{U_1''|X}, P_{U_2'|Y}$ so that

$$R_1 \geq (1 - \epsilon_2) I(U_1'; X) + (\epsilon_2 - \epsilon_1) I(U_1''; X), \tag{49c}$$

$$R_2 \geq (1 - \epsilon_2) I(U_2'; Y), \tag{49d}$$

and where the mutual information quantities are calculated using the joint pmfs $P_{U_1'XY} \triangleq P_{U_1'|X} P_{XY}$, $P_{U_1''XY} \triangleq P_{U_1''|X} P_{XY}$, and $P_{U_2'YZ} \triangleq P_{U_2'|Y} P_{YZ}$.

If $\epsilon_1 > \epsilon_2$, the (ϵ_1, ϵ_2)-exponents region $\mathcal{E}^*(R_1, R_2, \epsilon_1, \epsilon_2)$ *contains* all (θ_1, θ_2) pairs that satisfy

$$\theta_1 \leq I(U_1'; Y), \tag{50a}$$

$$\theta_2 \leq \min\{I(U_1';Y) + I(U_2';Z), I(U_1'';Y) + I(U_2'';Z)\}, \quad (50b)$$

for some conditional pmfs $P_{U_1'|X}, P_{U_1''|X}, P_{U_2'|Y}, P_{U_2''|Y}$ so that

$$R_1 \geq (1-\epsilon_1)I(U_1';X) + (\epsilon_1 - \epsilon_2)I(U_1'';X), \quad (50c)$$
$$R_2 \geq (1-\epsilon_1)I(U_2';Y) + (\epsilon_1 - \epsilon_2)I(U_2'';Y), \quad (50d)$$

and where the mutual information quantities are calculated using the joint pmfs $P_{U_1'XY} \triangleq P_{U_1'|X}P_{XY}$, $P_{U_1''XY} \triangleq P_{U_1''|X}P_{XY}$, $P_{U_2'YZ} \triangleq P_{U_2'|Y}P_{YZ}$, and $P_{U_2''YZ} \triangleq P_{U_2''|Y}P_{YZ}$.

Proof: Achievability results are based on the schemes in Section III, see Appendices A, B, and C for the analyses. For $\epsilon_1 = \epsilon_2$ the converse is proved in Appendix D. ■

A. Numerical Simulations

In this section, we illustrate the benefits of variable-length coding as opposed to fixed-length coding (or the benefits of having the relaxed expected rate constraints in (3) and (5) instead of the more stringent maximum rate-constraints (16)). We also show for $\epsilon_2 \neq \epsilon_1$ the benefits of having two auxiliary random variables U_1' and U_1'' in (49)–(50) instead of only a single random variable, which is equivalent to applying the basic two-hop scheme for two parameter choices (depending on X^n) and not just one. And finally, for $\epsilon_2 < \epsilon_1$, we illustrate the benefits of having both U_2' and U_2'' in (50), which stems from applying two full versions of the basic two-hop scheme in Subsection III-A.

Throughout this section we consider the following example. Let X, S, T be independent Bernoulli random variables of parameters $p_X = 0.4, p_S = 0.8, p_T = 0.8$ and set $Y = X \oplus T$ and $Z = Y \oplus S$.

We first consider the case of equal permissible type-I error exponents $\epsilon_1 = \epsilon_2$. By Theorem 1, in this case the optimal exponents region $\mathcal{E}^*$ is given by the rectangle determined by $\theta^*_{1,\epsilon_1}(R_1)$ and $\theta^*_{2,\epsilon_2}(R_1, R_2)$. Under maximum rate-constraints, the optimal exponents region $\mathcal{E}_{\text{max}}$ is also a rectangle, but now determined by $\theta^*_{1,\epsilon_1,\text{max}}(R_1)$ and $\theta^*_{2,\epsilon_2,\text{max}}(R_1, R_2)$. Fig. 2 plots these optimal error exponents for $\epsilon_1 = \epsilon_2 = 0.05$ and in function of $R_1 = R_2$. It thus illustrates the gain of having expected rate constraints instead of maximum rate-constraints.

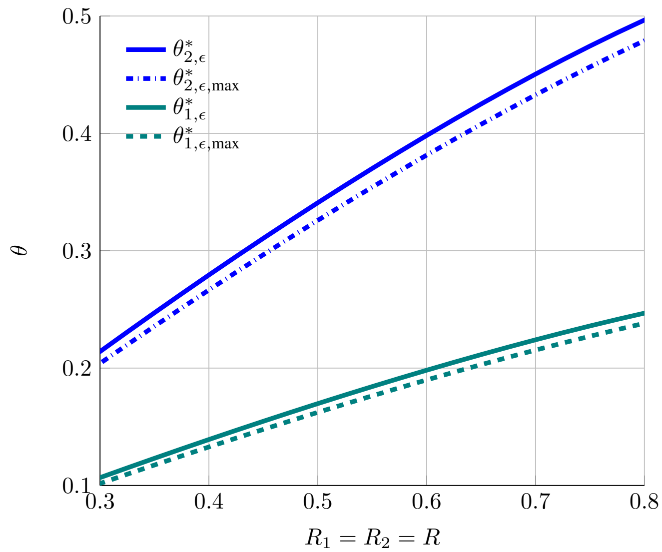

Fig. 2: Optimal error exponents under expected and maximum rate constraints for $\epsilon := \epsilon_1 = \epsilon_2 = 0.05$.

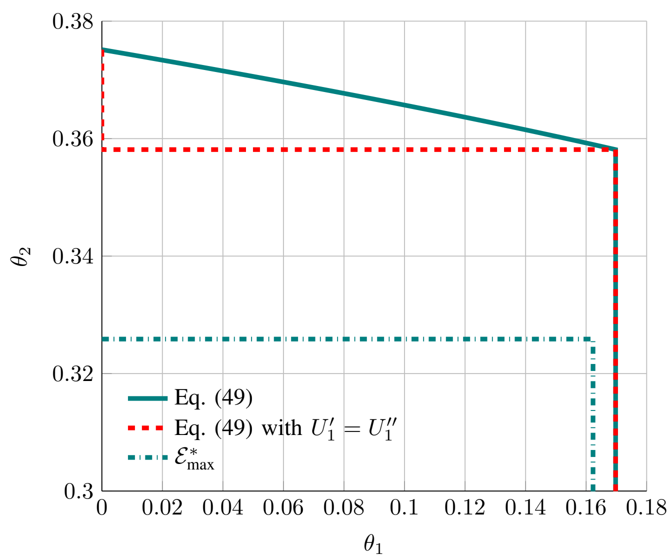

Fig. 3: Exponents regions for $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$ and $R_1 = R_2 = 0.5$.

We now consider the case $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$, and plot our inner bound to $\mathcal{E}^*$ in Fig. 3 for rates $R_1 = R_2 = 0.5$. We note a tradeoff between the two exponents θ_1, θ_2, which was not present for $\epsilon_1 = \epsilon_2$. (This tradeoff occurs because both exponents have to be optimized over the same choices of random variables U_1', U_1''.) The figure also shows a suboptimal version of the inner bound in Theorem 1, where we set $U_1' = U_1''$ but still optimize over all choices of U_1'. We observe that using two different auxiliary random variables U_1' and U_1'' (i.e., two different versions of the basic two-hop scheme) allows to obtain a better tradeoff between the two exponents. Finally, for comparison, Fig. 3 also shows the exponents region $\mathcal{E}^*$ under maximum rate-constraints, so as to illustrate the gain provided by having the weaker expected rate constraints instead of a maximum rate constraint.

We finally consider the case $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$. Fig. 4 shows our inner bound in Theorem 1 together with sub-optimal versions of this inner bound where we either set $U_2' = U_2''$ or $U_1' = U_1''$. Similarly to the previous figure we observe that having multiple auxiliary random variables (i.e., two versions of the basic two-hop scheme) allows to improve the tradeoff between the two exponents.

V. Conclusion

In this work, distributed hypothesis testing over a two-hop network with two decision centers is studied under *expected* rate constraints. Different coding and decision schemes are

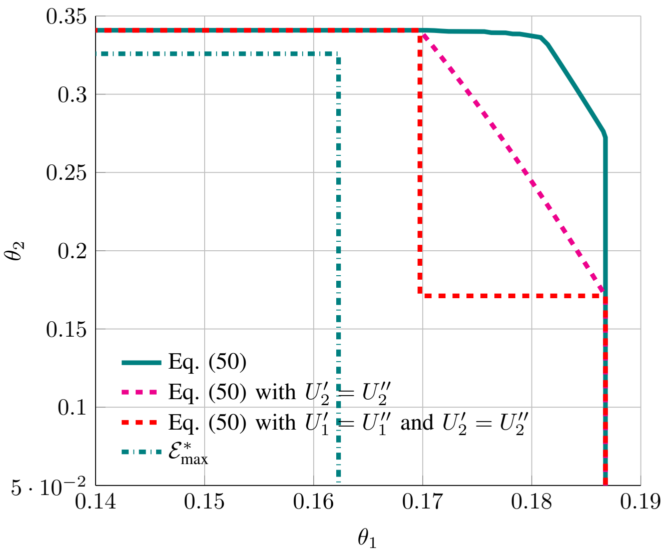

Fig. 4: Exponents regions under expected and maximum rate constraints for $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R_1 = R_2 = 0.5$.

proposed for different cases of permissible type-I error probabilities. These schemes are designed to choose different set of parameters and rates based on the transmitter's observation, aiming to maximize the achievable type-II error exponents at both decision centers. Optimal error exponents are obtained when the decision centers share equal type-I error constraints. Otherwise, a tradeoff between the exponents at the two decision centers occur. Supported by numerical simulations, the benefits of the proposed schemes are shown in this work, where the gain induced by expected rate constraints instead of maximum rate constraints is highlighted too.

Acknowledgment

M. Wigger and M. Hamad have been supported by the European Union's Horizon 2020 Research And Innovation Programme under grant agreement no. 715111.

References

[1] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.
[2] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, pp. 759–772, Nov. 1987.
[3] H. Shimokawa, T. Han, and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. ISIT*, p. 114, Jul. 1994.
[4] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6282–6303, Oct. 2012.
[5] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detection," in *Proc. ISIT*, Jun. 2018.
[6] W. Zhao and L. Lai, "Distributed testing with cascaded encoders," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7339–7348, 2018.
[7] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," in *18th Int. Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 2020.
[8] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 681–694, 2020.
[9] M. Hamad, M. Wigger, and M. Sarkiss, "Cooperative multi-sensor detection under variable-length coding." [Online]. Available: https://arxiv.org/pdf/2010.09616.pdf, 2020.
[10] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Trans. Inf. Theory*, vol. 65, pp. 4411–4433, Jul. 2019.
[11] D. Cao, L. Zhou, and V. Y. F. Tan, "Strong converse for hypothesis testing against independence over a two-hop network," *Entropy (Special Issue on Multiuser Information Theory II)*, vol. 21, Nov. 2019.
[12] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[13] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* Wiley, 2006.
[15] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 689–703, 2019.
[16] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.

Appendix A
Analysis of the Coding Scheme in Subsection III-B for $\epsilon_1 = \epsilon_2$

Denote by $\tilde{\mathcal{H}}_Y$ and $\tilde{\mathcal{H}}_Z$ the guesses produced by the basic two-hop scheme in Subsection III-A for the chosen parameters $\mu, P_{U_1|X}, P_{U_2|Y}$. We can then write for the type-I error probabilities:

$$\begin{aligned}
\alpha_{1,n} &= \Pr[\hat{\mathcal{H}}_Y = 1|\mathcal{H} = 0] && (51)\\
&= \Pr[\hat{\mathcal{H}}_Y = 1, X^n \in \mathcal{S}_n|\mathcal{H} = 0]\\
&\quad + \Pr[\hat{\mathcal{H}}_Y = 1, X^n \notin \mathcal{S}_n|\mathcal{H} = 0] && (52)\\
&= \Pr[X^n \in \mathcal{S}_n|\mathcal{H} = 0]\\
&\quad + \Pr[\tilde{\mathcal{H}}_Y = 1, X^n \notin \mathcal{S}_n|\mathcal{H} = 0] && (53)\\
&\le \epsilon_1 - \mu + \Pr[\tilde{\mathcal{H}}_Y = 1|\mathcal{H} = 0], && (54)
\end{aligned}$$

and

$$\begin{aligned}
\alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_Z = 1|\mathcal{H} = 0] && (55)\\
&= \Pr[\hat{\mathcal{H}}_Z = 1, X^n \in \mathcal{S}_n|\mathcal{H} = 0]\\
&\quad + \Pr[\hat{\mathcal{H}}_Z = 1, X^n \notin \mathcal{S}_n|\mathcal{H} = 0] && (56)\\
&= \Pr[X^n \in \mathcal{S}_n|\mathcal{H} = 0]\\
&\quad + \Pr[\tilde{\mathcal{H}}_Z = 1, X^n \notin \mathcal{S}_n|\mathcal{H} = 0] && (57)\\
&\le \epsilon_2 - \mu + \Pr[\tilde{\mathcal{H}}_Z = 1|\mathcal{H} = 0]. && (58)
\end{aligned}$$

Since by [10], $\Pr[\tilde{\mathcal{H}}_Y = 1|\mathcal{H} = 0]$ and $\Pr[\tilde{\mathcal{H}}_Z = 1|\mathcal{H} = 0]$ both tend to 0 as $n \to \infty$, we conclude that $\overline{\lim}_{n\to\infty} \alpha_{1,n} \le \epsilon_1$, and $\overline{\lim}_{n\to\infty} \alpha_{2,n} \le \epsilon_2$.

We notice that when $X^n \in \mathcal{S}_n$, then $\hat{\mathcal{H}}_Y = \hat{\mathcal{H}}_Z = 1$. The type-II error probabilities of the scheme can therefore be bounded as:

$$\begin{aligned}
\beta_{1,n} &= \Pr[\hat{\mathcal{H}}_Y = 0|\mathcal{H} = 1] && (59)\\
&= \Pr[\tilde{\mathcal{H}}_Y = 0, X^n \notin \mathcal{S}_n|\mathcal{H} = 1] && (60)\\
&\le \Pr[\tilde{\mathcal{H}}_Y = 0|\mathcal{H} = 1] && (61)\\
&\le 2^{-n(I(U_1;Y)+\delta(\mu))} && (62)
\end{aligned}$$

and

$$\begin{aligned}
\beta_{2,n} &= \Pr[\hat{\mathcal{H}}_Z = 0|\mathcal{H} = 1] && (63)\\
&= \Pr[\tilde{\mathcal{H}}_Z = 0, X^n \in \mathcal{S}_n|\mathcal{H} = 1] && (64)\\
&\le \Pr[\tilde{\mathcal{H}}_Z = 0|\mathcal{H} = 1] && (65)\\
&\le 2^{-n(I(U_1;Y)+I(U_2;Z)+\delta(\mu))} && (66)
\end{aligned}$$

where (62) and (66) are proved in [10], and $\delta(\mu) \to 0$ as $\mu \downarrow 0$.

The described scheme satisfies the rate constraints for all blocklengths n that are sufficiently large so that $(1-\epsilon_1+\mu)n\mu \geq (\epsilon_1-\mu) \Leftrightarrow (1-\epsilon_2+\mu)n\mu \geq (\epsilon_2-\mu)$ hold:

$$\begin{aligned}\mathbb{E}(\mathrm{len}(\mathsf{M}_1)) &\leq (\epsilon_1-\mu)+(1-\epsilon_1+\mu)\cdot n(I(U_1;X)+\mu) && (67)\\ &\leq (1-\epsilon_1+\mu)\cdot n(I(U_1;X)+2\mu) && (68)\\ &= nR_1 && (69)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(\mathrm{len}(\mathsf{M}_2)) &\leq (\epsilon_2-\mu)+(1-\epsilon_2+\mu)\cdot n(I(U_2;Y)+\mu) && (70)\\ &\leq (1-\epsilon_2+\mu)\cdot n(I(U_2;Y)+2\mu) && (71)\\ &= nR_2. && (72)\end{aligned}$$

Letting first $n \to \infty$ and then $\mu \downarrow 0$, establishes the desired achievability result in (48).

Appendix B
Analysis of the coding scheme in Subsection III-C for $\epsilon_2 > \epsilon_1$

Let $\tilde{\mathcal{H}}'_Y$ and $\tilde{\mathcal{H}}'_Z$ denote the hypotheses guessed by R_Y and R_Z for the basic two-hop scheme with the first parameter choices $\mu, P_{U'_1|X}, P_{U'_2|Y}$. Similarly, let $\tilde{\mathcal{H}}''_Y$ be the hypothesis produced by R_Y for the basic two-hop scheme with the parameter choice $\mu, P_{U''_1|X}$. We then obtain for the type-I error probabilities:

$$\begin{aligned}\alpha_{1,n} &= \Pr[\hat{\mathcal{H}}_Y=1, X^n\in\mathcal{S}_n|\mathcal{H}=0]\\ &\quad+\Pr[\hat{\mathcal{H}}_Y=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0]\\ &\quad+\Pr[\hat{\mathcal{H}}_Y=1, X^n\in\mathcal{D}''_n|\mathcal{H}=0] && (73)\\ &= \Pr[X^n\in\mathcal{S}_n|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}'_Y=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}''_Y=1, X^n\in\mathcal{D}''_n|\mathcal{H}=0] && (74)\\ &\leq \epsilon_1-\mu+\Pr[\tilde{\mathcal{H}}'_Y=1|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}''_Y=1|\mathcal{H}=0] && (75)\end{aligned}$$

and

$$\begin{aligned}\alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_Z=1, X^n\in(\mathcal{S}_n\cup\mathcal{D}''_n)|\mathcal{H}=0]\\ &\quad+\Pr[\hat{\mathcal{H}}_Z=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0] && (76)\\ &= \Pr[X^n\in(\mathcal{S}_n\cup\mathcal{D}''_n)|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}'_Z=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0] && (77)\\ &\leq \epsilon_2-\mu+\Pr[\tilde{\mathcal{H}}'_Z=1|\mathcal{H}=0]. && (78)\end{aligned}$$

Since by [10], $\Pr[\tilde{\mathcal{H}}'_Y=1|\mathcal{H}=0]$ and $\Pr[\tilde{\mathcal{H}}''_Y=1|\mathcal{H}=0]$, and $\Pr[\tilde{\mathcal{H}}'_Z=1|\mathcal{H}=0]\downarrow 0$ as $n\to\infty$, we conclude that for the scheme in Subsection III-C $\overline{\lim}_{n\to\infty}\alpha_{1,n}\leq\epsilon_1$ and $\overline{\lim}_{n\to\infty}\alpha_{2,n}\leq\epsilon_2$.

For the type-II error probabilities we obtain

$$\begin{aligned}\beta_{1,n} &= \Pr[\tilde{\mathcal{H}}'_Y=0, X^n\in\mathcal{D}'_n|\mathcal{H}=1]\\ &\quad+\Pr[\tilde{\mathcal{H}}''_Y=0, X^n\in\mathcal{D}''_n|\mathcal{H}=1] && (79)\\ &\leq \Pr[\tilde{\mathcal{H}}'_Y=0|\mathcal{H}=1]+\Pr[\tilde{\mathcal{H}}''_Y=0|\mathcal{H}=1] && (80)\\ &\leq 2^{-n(I(U'_1;Y)+\delta(\mu))}+2^{-n(I(U''_1;Y)+\delta(\mu))} && (81)\end{aligned}$$

and

$$\begin{aligned}\beta_{2,n} &= \Pr[\tilde{\mathcal{H}}'_Z=0, X^n\in\mathcal{D}'_n|\mathcal{H}=1] && (82)\\ &\leq \Pr[\tilde{\mathcal{H}}'_Z=0|\mathcal{H}=1] && (83)\\ &\leq 2^{-n\left(I(U'_1;Y)+I(U'_2;Z)+\delta(\mu)\right)}, && (84)\end{aligned}$$

where (81) and (84) are proved in [10], and $\delta(\mu)\downarrow 0$ as $\mu\downarrow 0$.

The described scheme satisfies the rate constraints for all blocklengths n that are sufficiently large so that both $(1-\epsilon_1+\mu)n\mu \geq (2-\epsilon_1+\mu)$ and $(1-\epsilon_2+\mu)n\mu \geq (2-\epsilon_1+\mu)$ hold:

$$\begin{aligned}\mathbb{E}[\mathrm{len}(\mathsf{M}_1)] &\leq (\epsilon_1-\mu)\\ &\qquad+(1-\epsilon_2+\mu)\cdot(n(I(U'_1;X)+\mu)+2)\\ &\qquad+(\epsilon_2-\epsilon_1)\cdot(n(I(U''_1;X)+\mu)+2) && (85)\\ &\leq n(R'_1+R''_1)=nR_1 && (86)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\mathrm{len}(\mathsf{M}_2)] &\leq (\epsilon_1-\mu)+(\epsilon_2-\epsilon_1)\cdot 2\\ &\qquad+(1-\epsilon_2+\mu)\cdot(n(I(U'_2;Y)+\mu)+2) && (87)\\ &\leq nR_2. && (88)\end{aligned}$$

Letting first $n\to\infty$ and then $\mu\downarrow 0$, establishes the desired result in (49).

Appendix C
Analysis of the coding scheme in Subsection III-D for $\epsilon_1 > \epsilon_2$

Let $\tilde{\mathcal{H}}'_Y$ and $\tilde{\mathcal{H}}'_Z$ denote the hypotheses guessed by R_Y and R_Z for the basic two-hop scheme with the first parameter choices $\mu, P_{U'_1|X}, P_{U'_2|Y}$. Similarly, let $\tilde{\mathcal{H}}''_Z$ be the hypothesis produced by R_Z for the basic two-hop scheme with the parameter choices $\mu, P_{U''_1|X}, P_{U''_2|Y}$. We then obtain for the type-I error probabilities:

$$\begin{aligned}\alpha_{1,n} &= \Pr[\hat{\mathcal{H}}_Y=1, X^n\in(\mathcal{S}_n\cup\mathcal{D}''_n)|\mathcal{H}=0]\\ &\quad+\Pr[\hat{\mathcal{H}}_Y=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0] && (89)\\ &= \Pr[X^n\in(\mathcal{S}_n\cup\mathcal{D}''_n)|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}'_Y=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0] && (90)\\ &\leq \epsilon_1-\mu+\Pr[\tilde{\mathcal{H}}'_Y=1|\mathcal{H}=0] && (91)\end{aligned}$$

and

$$\begin{aligned}\alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_Z=1, X^n\in\mathcal{S}_n|\mathcal{H}=0]\\ &\quad+\Pr[\hat{\mathcal{H}}_Z=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0]\\ &\quad+\Pr[\hat{\mathcal{H}}_Z=1, X^n\in\mathcal{D}''_n|\mathcal{H}=0] && (92)\\ &= \Pr[X^n\in\mathcal{S}_n|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}'_Z=1, X^n\in\mathcal{D}'_n|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}''_Z=1, X^n\in\mathcal{D}''_n|\mathcal{H}=0] && (93)\\ &\leq \epsilon_2-\mu+\Pr[\tilde{\mathcal{H}}'_Z=1|\mathcal{H}=0]\\ &\quad+\Pr[\tilde{\mathcal{H}}''_Z=1|\mathcal{H}=0]. && (94)\end{aligned}$$

Since by [10], $\Pr[\tilde{\mathcal{H}}'_Y = 1|\mathcal{H} = 0]$, $\Pr[\tilde{\mathcal{H}}'_Z = 1|\mathcal{H} = 0]$, and $\Pr[\tilde{\mathcal{H}}''_Z = 1|\mathcal{H} = 0]$ all tend to 0 as $n \to \infty$, we conclude that for the scheme in Subsection III-D $\overline{\lim}_{n\to\infty} \alpha_{1,n} \leq \epsilon_1$ and $\overline{\lim}_{n\to\infty} \alpha_{2,n} \leq \epsilon_2$.

For the type-II error probabilities we obtain

$$\begin{aligned}\beta_{1,n} &= \Pr[\tilde{\mathcal{H}}'_Y = 0, X^n \in \mathcal{D}'_n|\mathcal{H} = 1] && (95)\\ &\leq \Pr[\tilde{\mathcal{H}}'_Y = 0|\mathcal{H} = 1] && (96)\\ &\leq 2^{-n\left(I(U_1';Y)+\delta(\mu)\right)}, && (97)\end{aligned}$$

and

$$\begin{aligned}\beta_{2,n} &= \Pr[\tilde{\mathcal{H}}'_Z = 0, X^n \in \mathcal{D}'_n|\mathcal{H} = 1]\\ &\quad + \Pr[\tilde{\mathcal{H}}''_Z = 0, X^n \in \mathcal{D}''_n|\mathcal{H} = 1] && (98)\\ &\leq \Pr[\tilde{\mathcal{H}}'_Z = 0|\mathcal{H} = 1] + \Pr[\tilde{\mathcal{H}}''_Z = 0|\mathcal{H} = 1] && (99)\\ &\leq 2^{-n\left(I(U_1';Y)+I(U_2';Z)+\delta(\mu)\right)}\\ &\quad + 2^{-n\left(I(U_1'';Y)+I(U_2'';Z)+\delta(\mu)\right)}. && (100)\end{aligned}$$

where (100) and (97) are proved in [10], and $\delta(\mu) \downarrow 0$ as $\mu \downarrow 0$.

The described scheme satisfies the rate constraints for all blocklengths n that are sufficiently large so that $(1 - \epsilon_2 + \mu)n\mu \geq (2 - \epsilon_2 + \mu)$ holds:

$$\begin{aligned}\mathbb{E}[\mathsf{len}(\mathsf{M}_1)] &\leq (\epsilon_2 - \mu)\\ &\quad + (1 - \epsilon_1 + \mu) \cdot (n(I(U_1';X) + \mu) + 2)\\ &\quad + (\epsilon_1 - \epsilon_2) \cdot (n(I(U_1'';X) + \mu) + 2) && (101)\\ &\leq n(R_1' + R_1'') = nR_1 && (102)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\mathsf{len}(\mathsf{M}_2)] &\leq (\epsilon_2 - \mu)\\ &\quad + (1 - \epsilon_1 + \mu) \cdot (n(I(U_2';Y) + \mu) + 2)\\ &\quad + (\epsilon_1 - \epsilon_2) \cdot (n(I(U_2'';Y) + \mu) + 2) && (103)\\ &\leq n(R_2' + R_2'') = nR_2. && (104)\end{aligned}$$

Letting first $n \to \infty$ and then $\mu \downarrow 0$, establishes the desired result in (50).

Appendix D
Converse Proof to Theorem 1 when $\epsilon_1 = \epsilon_2$

Throughout this section, let $h_b(\cdot)$ denote the binary entropy function, and $D(P\|Q)$ denote the Kullback-Leibler divergence between two probability mass functions on the same alphabet.

Define $\epsilon \triangleq \epsilon_1 = \epsilon_2$ and fix $\theta_1 < \theta^*_{1,\epsilon}(R_1)$ and $\theta_2 < \theta^*_{2,\epsilon}(R_1, R_2)$. The proof consists of three parts. In the first two parts (Subsections D-A-D-B) we establish constraints based on the decisions at R_Y and at R_Z, respectively, and in the third part we combine the constraints.

A. Constraints based on R_Y's decision

Considering only the decision at R_Y but not at R_Z, by [8] we conclude that there exists an auxiliary random variable U_1' jointly distributed with the pair $(X, Y) \sim P_{XY}$ so that the following conditions hold:

$$\begin{aligned}\theta_1 &\leq I(U_1';Y), && (105)\\ R_1 &\geq (1-\epsilon)I(U_1';X), && (106)\\ U_1' &\to X \to Y. && (107)\end{aligned}$$

B. Constraints based on R_Z's decision

In what follows we establish similar constraints but based on the decision at R_Z. The proof is basically an extension of the proof in [8] but to a multi-hop network. Consider a sequence of encoding and decision functions $\{(\phi_1^{(n)}, \phi_2^{(n)}, g_2^{(n)})\}_{n\geq 1}$ satisfying the type-I and type-II error constraints (12)–(13) for $i = 2$. Then, fix a blocklength n and a small number $\eta \geq 0$ and define

$$\begin{aligned}&\mu_n \triangleq n^{-\frac{1}{3}}, && (108)\\ &\mathcal{B}_n(\eta) \triangleq \{(x^n, y^n) :\\ &\qquad \Pr[\hat{\mathcal{H}}_Z = 0|X^n = x^n, Y^n = y^n, \mathcal{H} = 0] \geq \eta\}, && (109)\\ &\mathcal{D}_n(\eta) \triangleq \mathcal{B}_n(\eta) \cap \mathcal{T}^n_{\mu_n}(P_{XY}). && (110)\end{aligned}$$

By constraint (12) on the type-I error probability, we have:

$$\begin{aligned}1 - \epsilon &\leq \Pr[\hat{\mathcal{H}}_Z = 0|\mathcal{H} = 0] && (111)\\ &= \sum_{(x^n,y^n)\in\mathcal{B}_n} \underbrace{\Pr[\hat{\mathcal{H}}_Z = 0|X^n = x^n, Y^n = y^n, \mathcal{H} = 0]}_{\leq 1}\\ &\qquad \cdot P_{X^nY^n}(x^n, y^n)\\ &\quad + \sum_{(x^n,y^n)\notin\mathcal{B}_n} \underbrace{\Pr[\hat{\mathcal{H}}_Z = 0|X^n = x^n, Y^n = y^n, \mathcal{H} = 0]}_{\leq \eta}\\ &\qquad \cdot P_{X^nY^n}(x^n, y^n) && (112)\\ &\leq P_{X^nY^n}(\mathcal{B}_n(\eta)) + \eta(1 - P_{X^nY^n}(\mathcal{B}_n(\eta))), && (113)\end{aligned}$$

and thus

$$P_{X^nY^n}(\mathcal{B}_n(\eta)) \geq \frac{1-\epsilon-\eta}{1-\eta}. \qquad (114)$$

Moreover, by [13, Remark to Lemma 2.12], the probability that the pair (X^n, Y^n) lies in the jointly strong typical set $\mathcal{T}^{(n)}_{\mu_n}(P_{XY})$ satisfies

$$P^n_{XY}\left(\mathcal{T}^{(n)}_{\mu_n}(P_{XY})\right) \geq 1 - \frac{|\mathcal{X}|\ |\mathcal{Y}|}{4\mu_n^2 n}, \qquad (115)$$

and thus by (110) and (114),

$$P^n_{XY}(\mathcal{D}_n) \geq \frac{1-\epsilon-\eta}{1-\eta} - \frac{|\mathcal{X}|\ |\mathcal{Y}|}{4\mu_n^2 n} \triangleq \Delta_n. \qquad (116)$$

We define the random variables $\left(\tilde{\mathsf{M}}_1, \tilde{\mathsf{M}}_2, \tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n\right)$ as the restriction of the random variables $(\mathsf{M}_1, \mathsf{M}_2, X^n, Y^n, Z^n)$ to $(X^n, Y^n) \in D_n(\eta)$ with their probability distribution given by:

$$\begin{aligned}&P_{\tilde{\mathsf{M}}_1\tilde{\mathsf{M}}_2\tilde{X}^n\tilde{Y}^n\tilde{Z}^n}(\mathsf{m}_1, \mathsf{m}_2, x^n, y^n, z^n) \triangleq\\ &\quad P_{X^nY^nZ^n}(x^n, y^n, z^n) \cdot \frac{\mathbb{1}\{(x^n, y^n) \in D_n(\eta)\}}{P_{X^nY^n}(D_n(\eta))}\\ &\qquad \cdot \mathbb{1}\{\phi_1(x^n) = \mathsf{m}_1\} \cdot \mathbb{1}\{\phi_2(y^n, \phi_1(x^n)) = \mathsf{m}_2\}, && (117)\end{aligned}$$

leading to the following inequalities:

$$\begin{aligned}P_{\tilde{\mathsf{M}}_1\tilde{\mathsf{M}}_2}(\mathsf{m}_1, \mathsf{m}_2) &\leq P_{\mathsf{M}_1\mathsf{M}_2}(\mathsf{m}_1, \mathsf{m}_2)\Delta_n^{-1}, && (118)\\ P_{\tilde{Z}^n}(z^n) &\leq P^n_Z(z^n)\Delta_n^{-1}, && (119)\\ D(P_{\tilde{X}^n\tilde{Y}^n}\|P^n_{XY}) &\leq \log \Delta_n^{-1}. && (120)\end{aligned}$$

1) Single-Letter Characterization of Rate Constraints: Define the following random variables:

$$\tilde{L}_i \triangleq \text{len}(\tilde{\mathsf{M}}_i), \qquad i=1,2. \tag{121}$$

By the rate constraints (3) and (5), we get under $\mathcal{H}=0$:

$$\begin{aligned} nR_i &\geq \mathbb{E}[L_i] & (122)\\ &\geq \mathbb{E}[L_i|(X^n,Y^n)\in\mathcal{D}_n(\eta)]P_{X^nY^n}(\mathcal{D}_n(\eta)) & (123)\\ &= \mathbb{E}[\tilde{L}_i]P_{X^nY^n}(\mathcal{D}_n(\eta)) & (124)\\ &\geq \mathbb{E}[\tilde{L}_i]\Delta_n, & (125)\end{aligned}$$

where the last inequality holds by (116). Moreover, by definition, $\tilde{L}_i$ is a function of $\tilde{\mathsf{M}}_i$, for $i=1,2$, so we can upper bound the entropy of $\tilde{\mathsf{M}}_i$ as follows:

$$\begin{aligned} H(\tilde{\mathsf{M}}_i) &= H(\tilde{\mathsf{M}}_i,\tilde{L}_i) & (126)\\ &= \sum_{l_i}\Pr[\tilde{L}_i=l_i]H(\tilde{\mathsf{M}}_i|\tilde{L}_i=l_i)+H(\tilde{L}_i) & (127)\\ &\leq \sum_{l_i}\Pr[\tilde{L}_i=l_i]l_i+H(\tilde{L}_i) & (128)\\ &= \mathbb{E}[\tilde{L}_i]+H(\tilde{L}_i) & (129)\\ &\leq \frac{nR_i}{\Delta_n}+\frac{nR_i}{\Delta_n}h_b\left(\frac{\Delta_n}{nR_i}\right) & (130)\\ &= \frac{nR_i}{\Delta_n}\left(1+h_b\left(\frac{\Delta_n}{nR_i}\right)\right), & (131)\end{aligned}$$

where (130) holds by (125) and since the maximum possible entropy of $\tilde{L}_i$ is obtained by a geometric distribution of mean $\mathbb{E}[\tilde{L}_i]$, which is further bounded by $\frac{nR_i}{\Delta_n}$ [14, Theorem 12.1.1].

On the other hand, we lower bound the entropy of $\tilde{\mathsf{M}}_i$ as:

$$\begin{aligned} H(\tilde{\mathsf{M}}_i) &\geq I(\tilde{\mathsf{M}}_i;\tilde{X}^n\tilde{Y}^n)+D(P_{\tilde{X}^n\tilde{Y}^n}||P_{XY}^n)+\log\Delta_n & (132)\\ &= H(\tilde{X}^n\tilde{Y}^n)+D(P_{\tilde{X}^n\tilde{Y}^n}||P_{XY}^n)\\ &\quad -H(\tilde{X}^n\tilde{Y}^n|\tilde{\mathsf{M}}_i)+\log\Delta_n & (133)\\ &\geq n[H(\tilde{X}_T\tilde{Y}_T)+D(P_{\tilde{X}_T\tilde{Y}_T}||P_{XY})]\\ &\quad -\sum_{t=1}^n H(\tilde{X}_t\tilde{Y}_t|\tilde{U}_{i,t})+\log\Delta_n & (134)\\ &= n[H(\tilde{X}_T\tilde{Y}_T)+D(P_{\tilde{X}_T\tilde{Y}_T}||P_{XY})]\\ &\quad -nH(\tilde{X}_T\tilde{Y}_T|\tilde{U}_{i,T},T)+\log\Delta_n & (135)\end{aligned}$$

Here, (132) holds by (120); (134) holds by the super-additivity property in [15, Proposition 1], by the chain rule, and by defining

$$\tilde{U}_{i,t} \triangleq (\tilde{\mathsf{M}}_i\tilde{X}^{t-1}\tilde{Y}^{t-1}), \quad i\in\{1,2\}; \tag{136}$$

and (135) holds by defining T uniform over $\{1,\ldots,n\}$ independent of all other random variables. Finally, defining

$$\begin{aligned} \tilde{U}_i &\triangleq (\tilde{U}_{i,T},T), \quad i\in\{1,2\}, & (137a)\\ \tilde{X} &\triangleq \tilde{X}_T & (137b)\\ \tilde{Y} &\triangleq \tilde{Y}_T & (137c)\end{aligned}$$

results in:

$$H(\tilde{\mathsf{M}}_1) \geq n\left[I(\tilde{X};\tilde{U}_1)+\frac{1}{n}\log\Delta_n\right], \tag{138}$$

$$H(\tilde{\mathsf{M}}_2) \geq n\left[I(\tilde{Y};\tilde{U}_2)+\frac{1}{n}\log\Delta_n\right], \tag{139}$$

thus following (131), (138), and (139), we deduce that:

$$R_1 \geq \frac{I(\tilde{X};\tilde{U}_1)+\frac{1}{n}\log\Delta_n}{1+h_b\left(\frac{\Delta_n}{nR_1}\right)}\cdot\Delta_n, \tag{140}$$

$$R_2 \geq \frac{I(\tilde{Y};\tilde{U}_2)+\frac{1}{n}\log\Delta_n}{1+h_b\left(\frac{\Delta_n}{nR_2}\right)}\cdot\Delta_n. \tag{141}$$

2) Upper Bounding the Type-II Error Exponent θ_2: Define for each m_2 the set

$$\mathcal{A}_{Z,n}(\mathsf{m}_2) \triangleq \{z^n\colon g_2(\mathsf{m}_2,z^n)=0\}, \tag{142}$$

and its *Hamming neighborhood*

$$\hat{\mathcal{A}}^{\ell_n}_{Z,n}(\mathsf{m}_2) \triangleq \{\tilde{z}^n : \exists\, z^n \in \mathcal{A}_{Z,n}(\mathsf{m}_2) \text{ s.t. } d_H(z^n,\tilde{z}^n)\leq \ell_n\} \tag{143}$$

for some real number ℓ_n satisfying $\lim_{n\to\infty}\ell_n/n=0$ and $\lim_{n\to\infty}\ell_n/\sqrt{n}=\infty$. Note that:

$$\mathcal{A}_{Z,n} = \bigcup_{\mathsf{m}_2\in\mathcal{M}_2}\{\mathsf{m}_2\}\times\mathcal{A}_{Z,n}(\mathsf{m}_2). \tag{144}$$

Since by definitions (108) and (110), for all $(x^n,y^n)\in\mathcal{D}_n(\eta)$, and where $\mathsf{m}_2=\phi_2(\phi_1(x^n),y^n)$:

$$P_{\tilde{Z}^n|\tilde{X}^n\tilde{Y}^n}(\mathcal{A}_{Z,n}(\mathsf{m}_2)|x^n,y^n)\geq\eta, \tag{145}$$

then by the blowing-up lemma [16]:

$$P_{\tilde{Z}^n|\tilde{X}^n\tilde{Y}^n}(\hat{\mathcal{A}}^{\ell_n}_{Z,n}(\mathsf{m}_2)|x^n,y^n)\geq 1-\zeta_n, \tag{146}$$

for a real number $\zeta_n>0$ such that $\lim_{n\to\infty}\zeta_n=0$. Therefore:

$$\begin{aligned} P_{\tilde{\mathsf{M}}_2\tilde{Z}^n}(\hat{\mathcal{A}}^{\ell_n}_{Z,n}) &= \sum_{\substack{(x^n,y^n)\in\mathcal{D}_n\\ \mathsf{m}_2\in\mathcal{M}_2}} P_{\tilde{Z}^n|\tilde{X}^n\tilde{Y}^n}(\hat{\mathcal{A}}^{\ell_n}_{Z,n}(\mathsf{m}_2)|x^n,y^n)\\ &\qquad\qquad \cdot P_{\tilde{X}^n\tilde{Y}^n\tilde{\mathsf{M}}_2}(x^n,y^n,\mathsf{m}_2) & (147)\\ &\geq (1-\zeta_n). & (148)\end{aligned}$$

Now define:

$$Q_{\tilde{\mathsf{M}}_2}(\mathsf{m}_2) \triangleq \sum_{y^n,\mathsf{m}_1} P_{\tilde{\mathsf{M}}_1}(\mathsf{m}_1)P_{\tilde{Y}^n}(y^n)\cdot\mathbb{1}\{\phi_2(\mathsf{m}_1,y^n)=\mathsf{m}_2\}, \tag{149}$$

and

$$\begin{aligned} Q_{\mathsf{M}_2}(\mathsf{m}_2) &= \sum_{x^n,y^n,z^n,\mathsf{m}_1} P_X^n(x^n)P_Y^n(y^n)P_Z^n(z^n)\\ &\qquad \cdot\mathbb{1}\{\phi_1(x^n)=\mathsf{m}_1,\phi_2(\mathsf{m}_1,y^n)=\mathsf{m}_2\} & (150)\\ &= \sum_{x^n,y^n,\mathsf{m}_1} P_{X^n\mathsf{M}_1}(x^n,\mathsf{m}_1)P_Y^n(y^n)\\ &\qquad \cdot\mathbb{1}\{\phi_2(\mathsf{m}_1,y^n)=\mathsf{m}_2\} & (151)\\ &= \sum_{y^n,\mathsf{m}_1} P_{\mathsf{M}_1}(\mathsf{m}_1)P_Y^n(y^n)\cdot\mathbb{1}\{\phi_2(\mathsf{m}_1,y^n)=\mathsf{m}_2\} & (152)\end{aligned}$$

Then

$$Q_{\tilde{\mathsf{M}}_2}(\mathsf{m}_2) \le Q_{\mathsf{M}_2}(\mathsf{m}_2)\Delta_n^{-2}, \tag{153}$$

and

$$\begin{aligned} Q_{\tilde{\mathsf{M}}_2} & P_{\tilde{Z}^n}\left(\hat{\mathcal{A}}_{Z,n}^{\ell_n}\right) \\ & \le Q_{\mathsf{M}_2}P_Z^n\left(\hat{\mathcal{A}}_{Z,n}^{\ell_n}\right)\Delta_n^{-3} & (154)\\ & \le \underbrace{Q_{\mathsf{M}_2}P_Z^n\left(\mathcal{A}_{Z,n}\right)}_{\beta_{2,n}} e^{nh_b(\ell_n/n)}|\mathcal{Z}|^{\ell_n}k_n^{\ell_n}\Delta_n^{-3} & (155)\\ & = \beta_{2,n}F_n^{\ell_n}\Delta_n^{-3}, & (156)\end{aligned}$$

where $k_n \triangleq \min_{z,z':P_Z(z')>0} \frac{P_Z(z)}{P_Z(z')}$ and $F_n^{\ell_n} \triangleq e^{nh_b(\ell_n/n)} \cdot k^{\ell_n} \cdot |\mathcal{Z}|^{\ell_n}$. Here, (155) holds by [13, Proof of Lemma 5.1]. Then by standard inequalities (see [8, Lemma 1]), we can obtain the following expression:

$$\frac{1}{n}\log\frac{1}{\beta_{2,n}} \le \frac{1}{n(1-\zeta_n)}\left(D(P_{\tilde{\mathsf{M}}_2\tilde{Z}^n}||Q_{\tilde{\mathsf{M}}_2}P_{\tilde{Z}^n})+1\right)+\delta_n \tag{157}$$

where δ_n tends to 0 as $n\to\infty$.

We further upper bound the divergence terms as follows:

$$\begin{aligned} D(P_{\tilde{\mathsf{M}}_2\tilde{Z}^n}&||Q_{\tilde{\mathsf{M}}_2}P_{\tilde{Z}^n}) \\ &= I(\tilde{\mathsf{M}}_2;\tilde{Z}^n)+D(P_{\tilde{\mathsf{M}}_2}||Q_{\tilde{\mathsf{M}}_2}) & (158)\\ &\le I(\tilde{\mathsf{M}}_2;\tilde{Z}^n)+D(P_{\tilde{Y}^n\tilde{\mathsf{M}}_1}||P_{\tilde{Y}^n}P_{\tilde{\mathsf{M}}_1}) & (159)\\ &= I(\tilde{\mathsf{M}}_2;\tilde{Z}^n)+I(\tilde{\mathsf{M}}_1;\tilde{Y}^n) & (160)\\ &= \sum_{t=1}^n I(\tilde{\mathsf{M}}_2;\tilde{Z}_t|\tilde{Z}^{t-1})+I(\tilde{\mathsf{M}}_1;\tilde{Y}_t|\tilde{Y}^{t-1}) & (161)\\ &\le \sum_{t=1}^n I(\tilde{\mathsf{M}}_2\tilde{X}^{t-1}\tilde{Y}^{t-1}\tilde{Z}^{t-1};\tilde{Z}_t) \\ &\qquad +I(\tilde{\mathsf{M}}_1\tilde{X}^{t-1}\tilde{Y}^{t-1};\tilde{Y}_t) & (162)\\ &= \sum_{t=1}^n I(\tilde{\mathsf{M}}_2\tilde{X}^{t-1}\tilde{Y}^{t-1};\tilde{Z}_t)+I(\tilde{\mathsf{M}}_1\tilde{X}^{t-1}\tilde{Y}^{t-1};\tilde{Y}_t) & (163)\\ &= \sum_{t=1}^n I(\tilde{U}_{2,t};\tilde{Z}_t)+I(\tilde{U}_{1,t};\tilde{Y}_t) & (164)\\ &= n[I(\tilde{U}_{2,T};\tilde{Z}_T|T)+I(\tilde{U}_{1,T};\tilde{Y}_T|T)] & (165)\\ &\le n[I(\tilde{U}_{2,T}T;\tilde{Z}_T)+I(\tilde{U}_{1,T}T;\tilde{Y}_T)] & (166)\\ &= n[I(\tilde{U}_2;\tilde{Z})+I(\tilde{U}_1;\tilde{Y})]. & (167)\end{aligned}$$

Here (159) is obtained by the data processing inequality for Kullback-Leibler divergence; (161) by the chain rule; (163) by the Markov chain $\tilde{Z}^{t-1}\to(\tilde{Y}^{t-1},\tilde{X}^{t-1})\to\tilde{Z}_t$; and (164)–(167) by the definitions of $\tilde{U}_{1,t},\tilde{U}_{2,t},\tilde{U}_1,\tilde{U}_2,\tilde{Y}$ in (136) and (137) and by defining $\tilde{Z}=\tilde{Z}_T$ where T is uniform over $\{1,\dots,n\}$ independent of all other random variables.

Observe the Markov chain $\tilde{U}_{2,t}\to\tilde{Y}_t\to\tilde{Z}_t$ for any t, and thus $\tilde{U}_2\to\tilde{Y}\to\tilde{Z}$ holds by construction for any n.

The second desired Markov chain $\tilde{U}_1\to\tilde{X}\to\tilde{Y}$ holds only in the limit as $n\to\infty$. To see this, notice that $\tilde{\mathsf{M}}_1\to\tilde{X}^n\to\tilde{Y}^n$ forms a Markov chain, and thus similar to the analysis in [9, Section V.C]:

$$\begin{aligned} 0 &= I(\tilde{\mathsf{M}}_1;\tilde{Y}^n|\tilde{X}^n) & (168)\\ &\ge H(\tilde{Y}^n|\tilde{X}^n)-H(\tilde{Y}^n|\tilde{X}^n\tilde{\mathsf{M}}_1) \\ &\quad +D(P_{\tilde{X}^n\tilde{Y}^n}||P_{XY}^n)+\log\Delta_n & (169)\\ &\ge n[H(\tilde{Y}_T|\tilde{X}_T)+D(P_{\tilde{X}_T\tilde{Y}_T}||P_{XY})]+\log\Delta_n \\ &\quad -H(\tilde{Y}^n|\tilde{X}^n\tilde{\mathsf{M}}_1) & (170)\\ &\ge n[H(\tilde{Y}_T|\tilde{X}_T)+D(P_{\tilde{X}_T\tilde{Y}_T}||P_{XY})]+\log\Delta_n \\ &\quad -\sum_{t=1}^n H(\tilde{Y}_t|\tilde{X}_t\tilde{X}^{t-1}\tilde{Y}^{t-1}\tilde{\mathsf{M}}_1) & (171)\\ &= n[H(\tilde{Y}_T|\tilde{X}_T)+D(P_{\tilde{X}_T\tilde{Y}_T}||P_{XY})]+\log\Delta_n \\ &\quad -\sum_{t=1}^n H(\tilde{Y}_t|\tilde{X}_t\tilde{U}_{1,t}) & (172)\\ &\ge n[H(\tilde{Y}_T|\tilde{X}_T)-H(\tilde{Y}_T|\tilde{X}_T,\tilde{U}_{1,T},T)]+\log\Delta_n & (173)\\ &\ge nI(\tilde{Y};\tilde{U}_1|\tilde{X})+\log\Delta_n, & (174)\end{aligned}$$

where (170) holds by the super-additivity property in [15, Proposition 1]; (171) by the chain rule and since conditioning reduces entropy; (172) by the definition of $\tilde{U}_{1,t}$ in (136); (173) by the non-negativity of the Kullback-Leibler divergence, and by recalling that T is uniform over $\{1,\dots,n\}$ independent of all other random quantities, and finally (174) holds by the definitions of $\tilde{U}_1,\tilde{X},\tilde{Y}$ in (137).

To sum up, we have proved so far in (140), (141), (157), (167), and (174), that for all $n\ge 1$ there exists a joint pmf $P^{(n)}_{\tilde{X}\tilde{Y}\tilde{Z}\tilde{U}_1\tilde{U}_2}$ (abbreviated as $P^{(n)}$) so that the following conditions hold (where $I_{P^{(n)}}$ indicates that the mutual information should be calculated according to the pmf $P^{(n)}$):

$$\begin{aligned} R_1 &\ge \left(I_{P^{(n)}}(\tilde{U}_1;\tilde{X})+g_1(n)\right)\cdot g_2(n,\eta), & (175a)\\ R_2 &\ge \left(I_{P^{(n)}}(\tilde{U}_2;\tilde{Y})+g_1(n)\right)\cdot g_2'(n,\eta), & (175b)\\ \theta_2 &\le \left(I_{P^{(n)}}(\tilde{U}_2;\tilde{Z})+I_{P^{(n)}}(U_1';\tilde{Y})\right)g_3(n)+g_4(n) & (175c)\\ g_5(n) &\ge I_{P^{(n)}}(\tilde{Y};\tilde{U}_1|\tilde{X}), & (175d)\end{aligned}$$

for some functions $g_1(n),g_2(n,\eta),g_2'(n,\eta),g_3(n),g_4(n),g_5(n)$ with the following asymptotic behaviors:

$$\lim_{n\to\infty}g_1(n)=\lim_{n\to\infty}g_4(n)=\lim_{n\to\infty}g_5(n)=0 \tag{176}$$

$$\lim_{n\to\infty}g_3(n)=1 \tag{177}$$

$$\lim_{n\to\infty}g_2(n,\eta)=\lim_{n\to\infty}g_2'(n,\eta)=\frac{1-\epsilon-\eta}{1-\eta}. \tag{178}$$

By the Markov chains $\tilde{X}\to\tilde{Y}\to\tilde{Z}$ and $\tilde{U}_2\to\tilde{Y}\to\tilde{Z}$ we can further conclude that

$$P^{(n)}_{\tilde{X}\tilde{Y}\tilde{Z}\tilde{U}_1\tilde{U}_2}=P^{(n)}_{\tilde{X}\tilde{Y}\tilde{Z}}\cdot P^{(n)}_{\tilde{U}_1\tilde{U}_2|\tilde{X}\tilde{Y}}, \tag{179}$$

The proof in this section is concluded by letting $n\to\infty$ and $\eta\downarrow 0$, and noting that by (175d) the limiting pmf of the sequence $P^{(n)}$ satisfies the Markov condition $\tilde{U}_1\to\tilde{X}\to\tilde{Y}$. More precisely, we first observe that by Carathéodory's theorem [12, Appendix C] for each n there must exist random variables $\tilde{U}_1$ and $\tilde{U}_2$ satisfying (175) and (179) over alphabets of sizes

$$|\tilde{\mathcal{U}}_1|\le|\mathcal{X}|\cdot|\mathcal{Y}|+2, \tag{180}$$

$$|\tilde{\mathcal{U}}_2|\le|\tilde{\mathcal{U}}_1|\cdot|\mathcal{X}|\cdot|\mathcal{Y}|+1. \tag{181}$$

Then we invoke the Bolzano-Weierstrass theorem and consider a sub-sequence $P^{(n_k)}_{\tilde{X}\tilde{Y}\tilde{Z}\tilde{U}_1\tilde{U}_2}$ that converges to a limiting pmf $P^*_{XYZU_1''U_2}$. For this limiting pmf, which we abbreviate by P^*, we conclude by (175a)–(175c):

$$R_1 \geq (1-\epsilon) I_{P^*}(U_1'';X) \tag{182}$$
$$R_2 \geq (1-\epsilon) I_{P^*}(U_2;Y) \tag{183}$$
$$\theta_2 \leq I_{P^*}(U_2;Z) + I_{P^*}(U_1'';Y). \tag{184}$$

Notice further that since for any k the pair $(\tilde{X}^{n_k}, \tilde{Y}^{n_k})$ lies in the jointly typical set $\mathcal{T}^{(n_k)}_{\mu_{n_k}}(P_{XY})$, we have $|P_{\tilde{X}\tilde{Y}} - P_{XY}| \leq \mu_{n_k}$ and thus the limiting pmf P^* satisfies $P^*_{XY} = P_{XY}$. Moreover, since for each n_k the random variable $\tilde{Z}$ is drawn according to $P_{Z|Y}$ given $\tilde{Y}$, irrespective of $\tilde{X}$, the limiting pmf also satisfies $P^*_{Z|XY} = P_{Z|Y}$. We also notice that under P^* the Markov chain

$$U_2 \to Y \to Z, \tag{185}$$

holds because $\tilde{U}_2 \to \tilde{Y} \to \tilde{Z}$ also forms a Markov chain for any n_k. Finally, by continuity considerations and by (175d), the following Markov chain must hold under P^*:

$$U_1'' \to X \to Y \tag{186}$$

To summarize, we establish the existence of a pmf $P^*_{XYZU_1''U_2}$ with $P^*_{XYZ} = P_{XY}P_{Z|Y}$, and satisfying the Markov chains (185)–(186) and the constraints (182)–(184).

C. *Combining Constraints from Decisions at R_Y and R_Z*

The previous Subsections D-A–D-B established the existence of random variables U_1', U_1'', and U_2 satisfying the three Markov chains $U_1' \to X \to Y$, $U_1'' \to X \to Y$, and $U_2 \to Y \to Z$, and constraints

$$R_1 \geq (1-\epsilon)\max\{I(U_1';X); I(U_1'';X)\}, \tag{187}$$
$$R_2 \geq (1-\epsilon) I(U_2;Y), \tag{188}$$
$$\theta_1 \leq I(U_1';Y), \tag{189}$$
$$\theta_2 \leq I(U_2;Z) + I(U_1'';Y). \tag{190}$$

The proof is concluded by showing that for each choice of U_1', U_1'', U_2, constraints (187)–(190) are relaxed if one replaces both U_1' and U_1'' with the same suitably chosen random variable U_1. In fact, we choose $U_1 = U_1''$ if $I(U_1';Y) \leq I(U_1'';Y)$ and we choose $U_1 = U_1'$ otherwise. For this choice, (187)–(190) imply

$$R_1 \geq (1-\epsilon) I(U_1;X), \tag{191}$$
$$R_2 \geq (1-\epsilon) I(U_2;Y), \tag{192}$$
$$\theta_1 \leq I(U_1;Y), \tag{193}$$
$$\theta_2 \leq I(U_2;Z) + I(U_1;Y). \tag{194}$$

Since the Markov chains $U_1 \to X \to Y$ and $U_2 \to Y \to Z$ hold by definition, this concludes our converse proof for the result in (48).