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 $P_{Z|Y}$.

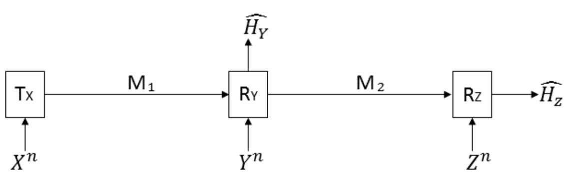

Fig. 1: Cascaded two-hop setup with two decision centers.

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

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

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

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

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 $R_1, R_2 \ge 0$. The exponent pair (θ_1, θ_2) is called (ϵ_1, ϵ_2)*-achievable* if there exists a sequence of encoding and decision functions $\{\phi_1^{(n)}, \phi_2^{(n)}, g_1^{(n)}, g_2^{(n)}\}_{n \ge 1}$ satisfying $\forall i \in \{1,2\}$:

$$\mathbb{E}[\text{len}(\mathsf{M}_i)] \le nR_i, \tag{11}$$

$$\varlimsup_{n\to\infty} \alpha_{i,n} \le \epsilon_i, \tag{12}$$

$$\varliminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{i,n}} \ge \theta_i. \tag{13}$$

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:

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

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

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*:

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

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

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

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

$$\theta_2 \le \theta_{2,\epsilon_2,\text{max}}^*\left(R_1, R_2\right)\}, \tag{18}$$

where

$$\theta_{1,\epsilon_1,\text{max}}^*\left(R_1\right) = \max_{\substack{P_{U_1|X}: \\ R_1 \ge I(U_1;X)}} I\left(U_1; Y\right) \tag{19}$$

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

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 $\mathcal{E}^*_{\max}(R_1, R_2, \epsilon_1, \epsilon_2)$ 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

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

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

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

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

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

$$\mathsf{M}_1 = [0]. \tag{24}$$

<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

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

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

$$\mathsf{M}_2 = \mathrm{string}(m_2) \tag{26}$$

to the receiver.

If no such index m_2 exists, R_Y directly sends string

$$\mathsf{M}_2 = [0]. \tag{27}$$

<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 $m_2 = \mathrm{dec}(\mathsf{m}_2)$, 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

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

for some small number $\mu > 0$.

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

$$\mathsf{M}_1 = \mathsf{M}_2 = [0] \tag{29}$$

and R_Y and R_Z decide on

$$\hat{\mathcal{H}}_Y = \hat{\mathcal{H}}_Z = 1. \tag{30}$$

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

$$R_1 = (1 - \epsilon_1 + \mu)\,(I(U_1; X) + 2\mu) \tag{31}$$

$$R_2 = (1 - \epsilon_2 + \mu)\,(I(U_2; Y) + 2\mu)\,. \tag{32}$$

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).

It can be shown that the presented scheme achieves the error exponents claimed in Equations (48) of Theorem 1 when $n \to \infty$ and $\mu \downarrow 0$. The proof is similar to Appendix A; the details are omitted for brevity.

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

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

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 $\mathcal{D}''_n$

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

such that

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

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

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 $\hat{\mathcal{H}}_Z = 1$.

It can be shown that the presented scheme achieves the error exponents claimed in Equations (49) of Theorem 1 when $n \to \infty$ and $\mu \downarrow 0$. The proof is similar to Appendix A; the details are omitted for brevity.

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

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

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

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

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

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 Appendix A, we prove that the presented scheme achieves the error exponents in Eq. (50) of Theorem 1 when $n \to \infty$ and $\mu \downarrow 0$.

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

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

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

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

$$R_1 \geq (1-\epsilon_1) I(U_1'; X) + (\epsilon_1 - \epsilon_2) I(U_1''; X), \tag{50c}$$

$$R_2 \geq (1-\epsilon_1) I(U_2'; Y) + (\epsilon_1 - \epsilon_2) I(U_2''; Y), \tag{50d}$$

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 Appendix A and [14] for the analyses. For $\epsilon_1 = \epsilon_2$, the converse is proved in [14]. ■

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.

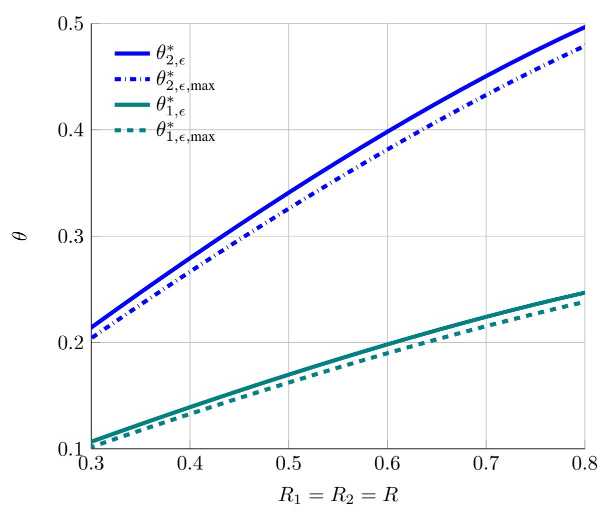

Fig. 2: Optimal error exponents under expected and maximum rate constraints for $\epsilon := \epsilon_1 = \epsilon_2 = 0.05$.

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 sub-optimal 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.

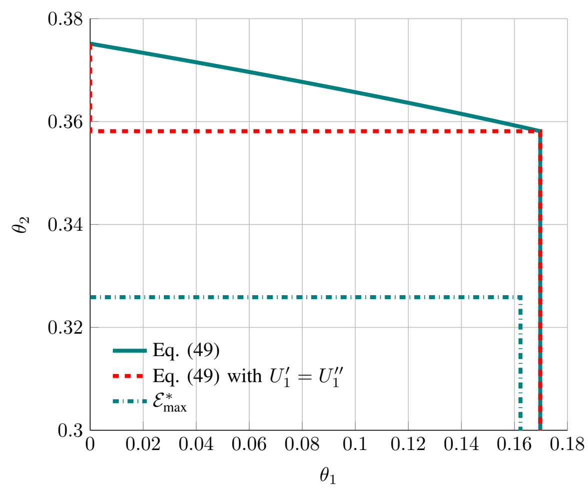

Fig. 3: Exponents regions for $\epsilon_1 = 0.05 < \epsilon_2 = 0.15$ and $R_1 = R_2 = 0.5$.

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 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.

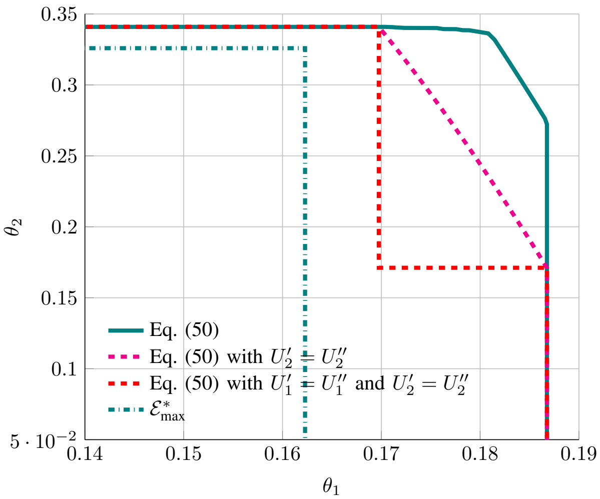

Fig. 4: Exponents regions under expected and maximum rate constraints for $\epsilon_1 = 0.15 > \epsilon_2 = 0.05$ and $R_1 = R_2 = 0.5$.

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] M. Hamad, M. Wigger, and M. Sarkiss, "Two-hop network with multiple decision centers under expected-rate constraints." [Online]. https://arxiv.org/pdf/2105.05917.pdf, 2021.

Appendix A
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:

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

and

$$\begin{aligned}\alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_Z = 1, X^n \in \mathcal{S}_n|\mathcal{H}=0] \\ &\quad + \Pr[\hat{\mathcal{H}}_Z = 1, X^n \in \mathcal{D}'_n|\mathcal{H}=0] \\ &\quad + \Pr[\hat{\mathcal{H}}_Z = 1, X^n \in \mathcal{D}''_n|\mathcal{H}=0] \qquad (54)\\ &= \Pr[X^n \in \mathcal{S}_n|\mathcal{H}=0] \\ &\quad + \Pr[\tilde{\mathcal{H}}'_Z = 1, X^n \in \mathcal{D}'_n|\mathcal{H}=0] \\ &\quad + \Pr[\tilde{\mathcal{H}}''_Z = 1, X^n \in \mathcal{D}''_n|\mathcal{H}=0] \qquad (55)\\ &\leq \epsilon_2 - \mu + \Pr[\tilde{\mathcal{H}}'_Z = 1|\mathcal{H}=0] + \Pr[\tilde{\mathcal{H}}''_Z = 1|\mathcal{H}=0] \qquad (56)\end{aligned}$$

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

$$\begin{aligned}\beta_{1,n} &= \Pr[\tilde{\mathcal{H}}'_Y = 0, X^n \in \mathcal{D}'_n|\mathcal{H}=1] \qquad (57)\\ &\leq \Pr[\tilde{\mathcal{H}}'_Y = 0|\mathcal{H}=1] \qquad (58)\\ &\leq 2^{-n\left(I(U_1';Y)+\delta(\mu)\right)}, \qquad (59)\end{aligned}$$

and

$$\begin{aligned}\beta_{2,n} &= \Pr[\tilde{\mathcal{H}}'_Z = 0, X^n \in \mathcal{D}'_n|\mathcal{H}=1] \\ &\quad + \Pr[\tilde{\mathcal{H}}''_Z = 0, X^n \in \mathcal{D}''_n|\mathcal{H}=1] \qquad (60)\\ &\leq \Pr[\tilde{\mathcal{H}}'_Z = 0|\mathcal{H}=1] + \Pr[\tilde{\mathcal{H}}''_Z = 0|\mathcal{H}=1] \qquad (61)\\ &\leq 2^{-n\left(I(U_1';Y)+I(U_2';Z)+\delta(\mu)\right)} \\ &\quad + 2^{-n\left(I(U_1'';Y)+I(U_2'';Z)+\delta(\mu)\right)}. \qquad (62)\end{aligned}$$

where (62) and (59) 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:

$$\begin{aligned}\mathbb{E}[\mathrm{len}(\mathsf{M}_1)] &\leq (\epsilon_2 - \mu) \\ &\quad + (1-\epsilon_1+\mu)\cdot(n(I(U_1';X)+\mu)+2) \\ &\quad + (\epsilon_1-\epsilon_2)\cdot(n(I(U_1'';X)+\mu)+2) \qquad (63)\\ &\leq n(R_1' + R_1'') = nR_1 \qquad (64)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\mathrm{len}(\mathsf{M}_2)] &\leq (\epsilon_2 - \mu) \\ &\quad + (1-\epsilon_1+\mu)\cdot(n(I(U_2';Y)+\mu)+2) \\ &\quad + (\epsilon_1-\epsilon_2)\cdot(n(I(U_2'';Y)+\mu)+2) \qquad (65)\\ &\leq n(R_2' + R_2'') = nR_2. \qquad (66)\end{aligned}$$

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