Covert Distributed Detection over Discrete Memoryless Channels

Abdelaziz Bounhar*, Mireille Sarkiss§, Michèle Wigger*
*LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France
{abdelaziz.bounhar, michele.wigger}@telecom-paris.fr
§SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France
{mireille.sarkiss}@telecom-sudparis.eu

***Abstract*—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. This paper studies the problem of distributed detection (binary hypothesis testing) over a discrete memoryless channel (DMC) under the constraint that an eavesdropping adversary should not be able to determine whether communication is ongoing or not, i.e., communication over the DMC has to remain covert. The main contribution of the paper is an upper bound on the largest possible Stein exponent, showing that it cannot exceed the largest exponent achievable under zero-rate communication over a noise-free link. In interesting special cases, the upper bound is achieved by a local test at the decision center that completely ignores the communication. In these cases, the covertness constraint thus renders communication useless for improving the Stein exponent.**

***Index Terms*—Hypothesis testing, covert communication, error exponents.**

I. Introduction

Distributed binary hypothesis testing problems have been widely investigated in the information-theory literature, focusing on the single-sensor and single-decision center setup (see Figure 1 without the external warden). In this setup, both terminals observe correlated sources, whose underlying joint distribution depends on a binary hypothesis. The sensor communicates to the decision center over a perfect link or a memoryless channel, and the decision center guesses the underlying hypothesis based on the communicated symbols and its local observations.

For certain classes of source distributions, Ahlswede and Csiszàr [1], and later Rahman and Wagner [2], derived the optimal Stein exponent when communication is over a noise-free but rate-limited link. The Stein exponent refers to the largest possible exponential decay-rate of the probability of error under the alternative hypothesis given a bound on the probability of error under the null hypothesis. Such asymmetric constraints arise in alert systems for which it often suffices to keep the false-alert error probability below a given threshold but the miss-detection error probability has to be as small as possible. Despite these early advances on the problem, the optimal Stein exponent remains unknown for general source distributions. Lower bounds on the optimal exponent for such general source distributions were presented in [3]–[7] and recently in [8]. Similar results were also reported for the setup when communication is over a discrete memoryless channel (DMC) [9], [10].

The setup in Figure 1 with an external warden was considered in [11]–[15]. In particular, [11], [14], [15] imposed the security constraint that the external warden is not allowed to learn too much information (in a distortion or equivocation sense) about the sensor's source observation. The recent work [15] determined the largest possible Stein exponent for this setup under the Ahlswede-Csiszár source distributions, and showed that with the eavesdropping constraint the Stein exponent depends on the probability of error allowed under the null hypothesis. This was not the case without security constraint, neither for the DMC nor for noise-free links [1], [16], where a strong converse result holds.

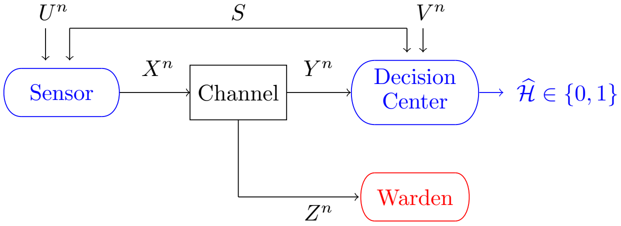

Fig. 1: Distributed binary hypothesis testing system with an external warden.

A somehow separate line of work [4], [17], [18] considered the distributed hypothesis testing problem where the sensor can send only a sublinear (in the observation blocklength) number of bits over a noise-free link to the decision center. In this case, the sensor's optimal strategy [4], [17] is to send a single bit indicating whether its observed source sequence is typical according to the distribution under the null hypothesis. The decision center then declares this null hypothesis if also its own observation is typical according to the distribution under the null hypothesis, and it declares the alternative hypothesis in all other cases.

In this paper, we impose a covertness constraint, i.e., we impose the somewhat stronger security constraint that the warden is not allowed to determine whether communication is ongoing or not. Technically, the output distributions at the warden under the null and the alternative hypotheses are required to be similar to the output distribution induced when the sensor sends the same predefined zero-symbol consistently during the entire transmission, indicating the absence of communication. For standard data transmission such a covertness constraint implies that the capacity is zero and the number of information bits that can reliably be sent over the channel without being detected by the warden only grows as square-root of n, for n denoting the blocklength of communication [19]–[21]. These results were extended for various other data communication scenarios, e.g., for multi-access and broadcast communications [22]–[24], for state-dependent communication [25]–[28], or to mixed networks allowing both for covert and non-covert users [29]. Recently the impact of covertness constraints has also been analyzed for sensing systems [30]–[32] and other communication systems whose task goes beyond pure data communication [33].

Our contributions: In this work, we consider a communication network whose final task is distributed hypothesis testing. Specifically, for this setup we show that the largest Stein exponent that is achievable over a binary DMC under a covertness constraint cannot exceed the Stein exponent of the setup where communication is over a noise-free but zero-rate link. For example, for the testing against independence setup, our converse implies that no positive error exponent is possible under a covertness constraint, while without such a covertness constraint the largest Stein exponent can be large. Our converse result depends only on some mild assumptions on the source distribution (all source pairs have positive probabilities under the alternative hypothesis) and the DMC (the zero-output distribution at the warden cannot be simulated by other inputs) and holds for all sublinear key-lengths. We further compare our results to a trivial lower bound achieved when the sensor does not communicate anything, i.e., always sends the non-detection zero-symbol, and the decision center guesses the binary hypothesis solely based on its local observation. This trivial lower bound matches our upper bound on Stein's exponent in some special cases, thus establishing that the covertness constraint completely eliminates the benefit of the communication for the Stein exponent. Without the covertness constraint, a much better exponent can be achieved when the sensor transmits a quantized version of its local observation to the decision center.

While the converse result is rather intuitive, the mathematical proof is somehow technical and requires combining tools from converses on covert communication with techniques to establish converse results for hypothesis testing. Additional new proof steps are required to show that one can limit attention to inputs with Hamming weight slightly less than linear in the blocklength, and to account for the secret key in the hypothesis testing converse steps.

A. Notation

We mostly follow standard notation. In particular, random variables are denoted by upper case letters (e.g., X), while their realizations are denoted by lowercase (e.g. x). We abbreviate $(x_1,\ldots,x_n)$ by x^n and $(x_{t+1},\ldots,x_n)$ by x_{t+1}^n. To indicate the Hamming weight, we use $w_{\mathrm{H}}(\cdot)$ and for the Hamming distance we use $d_{\mathrm{H}}(\cdot,\cdot)$. We further abbreviate *independent and identically distributed* as *i.i.d.* and *probability mass function* as *pmf.* Also, we denote by $\pi_{x^ny^n}$ the joint type of the sequences (x^n,y^n):

$$\pi_{x^ny^n}(a,b) \triangleq \frac{n_{x^n,y^n}(a,b)}{n}, \tag{1}$$

and we use $\mathcal{T}_{\mu}^{(n)}(P_{XY})$ to denote the jointly strongly-typical set as in [34, Definition 2.9], and accordingly $\mathcal{T}_{0}^{(n)}(P_{XY})$ the set of all sequence of constant type P_{XY}. We use Landau notation $o(1)$ to indicate any function that tends to 0 for blocklengths $n\to\infty$.

Throughout this manuscript, $\{\mu_n\}_{n=1}^{\infty}$ is a sequence of small positive numbers satisfying[1],

$$\lim_{n\to\infty}\mu_n = 0 \tag{2a}$$

$$\lim_{n\to\infty} n\cdot\mu_n^2 = \infty. \tag{2b}$$

II. Problem Setup

Consider the distributed hypothesis testing problem in Figure 1 where for a given blocklength n, a sensor observes a sequence U^n and a secret-key S, and communicates to a decision center, which also knows the secret key S in addition to its local observations V^n. The secret-key S is uniform over a given finite set $\mathcal{K}$, which grows sub-exponentially in the blocklength:[2]

$$\lim_{n\to\infty}\frac{1}{n}\log|\mathcal{K}| = 0. \tag{3}$$

The distribution of the observations (U^n,V^n) depends on a binary hypothesis $\mathcal{H}\in\{0,1\}$:

$$\text{if } \mathcal{H}=0: \quad (U^n,V^n) \text{ i.i.d. } \sim P_{UV}; \tag{4a}$$

$$\text{if } \mathcal{H}=1: \quad (U^n,V^n) \text{ i.i.d. } \sim Q_{UV}, \tag{4b}$$

for given pmfs P_{UV} and Q_{UV} over the product alphabet $\mathcal{U}\times\mathcal{V}$, where we assume that $Q_{UV}(u,v)>0$ for all $(u,v)\in\mathcal{U}\times\mathcal{V}$. Let P_U and P_V denote the marginal pmfs of P_{UV}.

The sensor communicates with a decision center over n uses of a discrete memoryless channel DMC, and we assume that communication needs to remain undetected to an external warden. We thus have a third hypothesis $\mathcal{H}=-1$ which models the absence of communication, and the goal of the warden is to distinguish whether $\mathcal{H}=-1$ or $\mathcal{H}\in\{0,1\}$. In contrast, the sensor and the decision center are aware of when hypothesis $\mathcal{H}=-1$ occurs. However, in case $\mathcal{H}=-1$ is not valid, they do not know whether $\mathcal{H}=0$ or $\mathcal{H}=1$. In fact, the decision center's goal is exactly to distinguish between these two hypotheses.

The DMC is described by the finite input and output alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and a transition law $\Gamma_{YZ|X}$, where Y denotes the output at the legitimate receiver (the decision center) and Z the output at the warden. It is assumed that the warden's output distribution induced by the zero-input cannot be simulated by any linear combination of the other symbols, i.e., there are no convex weights $\{\lambda_x\}_{x\in\mathcal{X}}$ such that:

$$\sum_{x\in\mathcal{X}}\lambda_x\Gamma_{Z|X}(z|x) = \Gamma_{Z|X}(z|0), \quad \forall z\in\mathcal{Z}, \tag{5}$$

where $\Gamma_{Z|X}$ is the conditional marginal obtained from $\Gamma_{YZ|X}$.

Under the hypothesis $\mathcal{H}=-1$, the sensor sends the all-zero sequence

$$X^n = 0^n, \tag{6}$$

where we assume that $0\in\mathcal{X}$. In this case, the decision center is aware of hypothesis $\mathcal{H}=-1$ and does not produce any guess. The warden observes an output sequence Z^n that follows the product distribution

$$P_{Z^n|\mathcal{H}=-1} = \Gamma_{Z|X}^{\otimes n}(\cdot|0^n). \tag{7}$$

[1] Condition (2b) ensures that the probability of the strongly typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{XY})$ under $P_{XY}^{\otimes n}$ tends to 1 as $n\to\infty$ [34, Remark to Lemma 2.12].

[2] Notice that sub-exponential key sizes have been proved to be sufficient [21] for standard covert communication.

Under hypotheses $\mathcal{H} = 0$ or $\mathcal{H} = 1$, the sensor sends an input sequence

$$X^n = f^{(n)}(U^n, S) \in \mathcal{X}^n \tag{8}$$

over the channel, where $f^{(n)}(\cdot,\cdot)$ is an encoding function on appropriate domains. The decision center observes the corresponding outputs Y^n of the DMC $\Gamma_{YZ|X}$ and the warden the outputs Z^n, where depending on the hypothesis this sequence follows the distribution

$$P_{Z^n|\mathcal{H}=0} = \frac{1}{|\mathcal{K}|} \sum_{s\in\mathcal{K}} \sum_{u^n\in\mathcal{U}^n} P_U^{\otimes n}(u^n) \Gamma_{Z|X}^{\otimes n}(\cdot|f^{(n)}(u^n,s)), \tag{9}$$

or

$$P_{Z^n|\mathcal{H}=1} = \frac{1}{|\mathcal{K}|} \sum_{s\in\mathcal{K}} \sum_{u^n\in\mathcal{U}^n} Q_U^{\otimes n}(u^n) \Gamma_{Z|X}^{\otimes n}(\cdot|f^{(n)}(u^n,s)). \tag{10}$$

Based on the received sequence V^n, its observations Y^n, and the shared secret key S, the decision center produces a guess of the hypothesis:

$$\hat{\mathcal{H}} = g^{(n)}(V^n, Y^n, S) \in \{0,1\} \tag{11}$$

for an appropriate guessing function $g^{(n)}$.

Covertness under hypothesis $\mathcal{H} = \mathsf{H} \in \{0,1\}$ is measured by the Kullback-Leibler divergence:

$$\delta_{n,\mathsf{H}} := D(P_{Z^n|\mathcal{H}=\mathsf{H}} \| P_{Z^n|\mathcal{H}=-1}), \quad \mathsf{H} \in \{0,1\}. \tag{12}$$

Definition 1: Given $\epsilon \in [0,1)$, a miss-detection error exponent $\theta > 0$ is called ϵ-achievable *under a covertness constraint* if there exists a sequence of encoding and decision functions $\{(f^{(n)}, g^{(n)})\}_{n=1}^{\infty}$ satisfying

$$\varlimsup_{n\to\infty} \Pr\left[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\right] \leq \epsilon \tag{13a}$$

$$\varliminf_{n\to\infty} -\frac{1}{n}\log \Pr\left[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\right] \geq \theta. \tag{13b}$$

and under both hypotheses $\mathsf{H} \in \{0,1\}$:

$$\lim_{n\to\infty} \delta_{n,\mathsf{H}} = 0. \tag{13c}$$

Theorem 1 (Converse): Given $\epsilon \in [0,1)$ and any key set satisfying (3), a miss-detection error exponent θ is not ϵ-achievable if it satisfies

$$\theta > \min_{\substack{\pi_{UV}:\\ \pi_U = P_U\\ \pi_V = P_V}} D(\pi_{UV} \| Q_{UV}). \tag{14}$$

Remark 1: Our converse result Theorem 1 remains valid also when covertness, i.e., constraint (13c), is imposed only under $\mathcal{H} = 0$ but not necessarily under $\mathcal{H} = 1$. This can model a practically relevant situation where one wishes to keep the distributed detection system invisible for an adversary under normal circumstances. In contrast, under alert situations one only cares about the alert itself and not about keeping the distributed detection system invisible. (In certain situations one even wishes the opposite: alerts should be accompanied by loud or highly visible alert signals.)

Theorem 2 (Achievability): Given $\epsilon \in [0,1)$, a miss-detection error exponent θ is ϵ-achievable if it satisfies

$$\theta < D(P_V \| Q_V). \tag{15}$$

The performance in (15) can even be achieved with $|\mathcal{K}| = 1$.

Proof: The exponent is achieved by the following scheme. The sensor always sends $X^n = 0$ and thus communication is covert. The decision center ignores the channel outputs and produces $\hat{\mathcal{H}} = 0$ if V^n lies in $\mathcal{T}_{\mu_n}^{(n)}(P_V)$, and it produces $\hat{\mathcal{H}} = 1$ otherwise. ■

In certain cases our converse and achievability results in Theorems 1 and 2 do coincide.

Corollary 3 (Exact Exponent): Consider pmfs P_{UV} and Q_{UV} satisfying

$$\sum_v P_V(v) Q_{U|V}(u|v) = P_U(u), \quad \forall u \in \mathcal{U}. \tag{16}$$

Then, for any $\epsilon \in [0,1)$, a miss-detection error exponent θ is ϵ-achievable if, and only if, it satisfies

$$\theta \leq D(P_V \| Q_V). \tag{17}$$

In particular, for testing against independence ($Q_{UV} = P_U \cdot P_V$) no positive exponent is possible.

The performance in (17) can even be achieved with $|\mathcal{K}| = 1$.

Proof: Achievability follows directly from Theorem 2. To see that the converse follows from Theorem 1 notice that

$$\begin{aligned} D(\pi_{UV} \| Q_{UV}) &= D(\pi_V \| Q_V) + \mathbb{E}_{\pi_V}[D(\pi_{U|V} \| Q_{U|V})] &(18)\\ &\geq D(\pi_V \| Q_V), &(19) \end{aligned}$$

where the inequality holds with equality if, and only if $\pi_{U|V} = Q_{U|V}$. Notice that this choice is admissible if, and only if, Condition (16) holds. ■

III. Proof of Theorem 1

We shall show a converse result under the stronger setup where the decision center directly observes the DMC inputs X^n instead of the outputs Y^n. Based on the inputs, it can itself (if it wishes) generate outputs $\tilde{Y}^n$ that are equivalent to the outputs Y^n observed in the original setup. The new setup is thus stronger than the original setup and a converse result for the new setup applies also a converse for the old setup.

Thus, we assume in the following that (11) can be replaced by

$$\hat{\mathcal{H}} = g^{(n)}(V^n, X^n, S) \in \{0,1\}. \tag{20}$$

Covertness Constraint Implies Low-Weight Inputs: In this part, we shall define a set $\tilde{\mathcal{X}}^n \subseteq \mathcal{X}$ with only low-weight vectors, and show that the covertness constraint implies the following two properties

$$\lim_{n\to\infty} \Pr\left[X^n \notin \tilde{\mathcal{X}}^n | \mathcal{H} = 0\right] = 1 \tag{21}$$

$$\lim_{n\to\infty} \frac{1}{n}\log\left|\tilde{\mathcal{X}}^n\right| = 0. \tag{22}$$

In the remaining of the proof we can then restrict to channel inputs $x^n \in \tilde{\mathcal{X}}^n$.

Start by noting that for the random tuple $Z^n \sim P_{Z^n|\mathcal{H}=0}$ in (9), we have:

$$\begin{aligned} &D(P_{Z^n|\mathcal{H}=0} \| P_{Z^n|\mathcal{H}=-1})\\ &\quad = -H(Z^n) + \mathbb{E}_{Z^n}\left[\log\left(\frac{1}{\Gamma_{Z|X}^{\otimes n}(Z^n|0^n)}\right)\right] \end{aligned} \tag{23}$$

$$\overset{(a)}{\geq} -\sum_{i=1}^{n} H(Z_i) + \mathbb{E}_{Z_i}\left[\log\left(\frac{1}{\Gamma_{Z|X}(Z_i|0)}\right)\right] \tag{24}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} D\Big(\sum_{x\in\mathcal{X}\setminus\{0\}} \beta_x \alpha_{n,i}^{0} \Gamma_{Z|X}(\cdot \mid x) + (1-\alpha_{n,i}^{0})\Gamma_{Z|X}(\cdot \mid 0) \| \, \Gamma_{Z|X}(\cdot \mid 0)\Big) \tag{25}$$

where (a) holds because conditioning reduces entropy; and (b) by defining

$$\alpha_{n,i}^{0} := \Pr[X_i \neq 0 | \mathcal{H}=0], \quad i \in \{1,\dots,n\}, \tag{26}$$

and

$$\beta_x := \Pr[X_i = x | X_i \neq 0, \mathcal{H}=0]. \tag{27}$$

By the covertness constraint (13c) for $\mathcal{H}=0$ and the non-negativity of Kullback-Leibler divergence, we deduce that each summand on the right-hand side of (25) must vanish, and thus

$$\lim_{n\to\infty} \alpha_{n,i}^{0} = 0, \quad i\in\{1,\dots,n\}, \tag{28}$$

and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \alpha_{n,i}^{0} = 0. \tag{29}$$

Define next for each n:

$$a_n := \sqrt{\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,i}^{0}}, \tag{30}$$

so that a_n vanishes but slowlier than $\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,i}^{0}$. Define the set of all low-weight inputs

$$\tilde{\mathcal{X}}^n := \{x^n \in \mathcal{X}^n \colon w_{\mathrm{H}}(x^n) < a_n \cdot n\}, \tag{31}$$

and notice that

$$\sum_{i=1}^{n}\alpha_{n,i}^{0} = \mathbb{E}[w_{\mathrm{H}}(X^n)|\mathcal{H}=0] \tag{32}$$

$$\geq a_n \cdot n \cdot \Pr[X^n \notin \tilde{\mathcal{X}}^n | \mathcal{H}=0], \tag{33}$$

which by (30) implies that

$$\Pr\left[X^n \notin \tilde{\mathcal{X}}^n|\mathcal{H}=0\right] \leq \frac{\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,i}^{0}}{a_n} \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\alpha_{n,i}^{0}}. \tag{34}$$

By (29) this proves (21).

To see that the size of $\tilde{\mathcal{X}}^n$ does not grow exponentially, we notice that this set can be described as the union over all type-classes (i.e., sets of sequences with same type) for types that assign frequency larger or equal to $1-a_n$ to the 0 symbol. Since the type-class for type $\boldsymbol{\pi}$ is of size at most $2^{nH(\boldsymbol{\pi})}$ and since the number of type-classes is bounded by $(n+1)^{|\mathcal{X}|}$, we have:

$$|\tilde{\mathcal{X}}^n| \leq (n+1)^{|\mathcal{X}|} 2^{n \max_{\boldsymbol{\pi}} H(\boldsymbol{\pi})}, \tag{35}$$

where the maximum is over all types $\boldsymbol{\pi}$ with $\boldsymbol{\pi}(0) \geq 1-a_n$. Since a_n vanishes as $n\to\infty$ and by the continuity of the entropy functional, we obtain

$$\lim_{n\to\infty}\frac{1}{n}\log\left|\tilde{\mathcal{X}}^n\right| \leq \lim_{n\to\infty}\left[\frac{|\mathcal{X}|}{n}\log(n+1) + \max_{\substack{\boldsymbol{\pi}:\\ \boldsymbol{\pi}(0)\geq 1-a_n}} H(\boldsymbol{\pi})\right]$$

$$= 0. \tag{36}$$

Simplified Acceptance Region under Low-Weight Inputs: In this part, we determine a square region over $\mathcal{U}^n$ and $\mathcal{V}^n$ and show that each of the two components has large probability under P_U and P_V respectively, while the error exponent under the alternative hypothesis is only slightly decreased compared to the original decision regions.

Start by noting that since

$$\sum_{x^n\in\tilde{\mathcal{X}}^n} \Pr\left[\hat{\mathcal{H}}=0, X^n = x^n|\mathcal{H}=0\right]$$

$$= \Pr\left[\hat{\mathcal{H}}=0|\mathcal{H}=0\right] - \sum_{x^n\notin\mathcal{X}^n}\Pr\left[\hat{\mathcal{H}}=0, X^n=x^n|\mathcal{H}=0\right] \tag{37}$$

$$\geq 1-\epsilon-\Pr[X^n\notin\tilde{\mathcal{X}}^n|\mathcal{H}=0], \tag{38}$$

for any $\eta\in(0,\epsilon)$ and sufficiently large n we have

$$\sum_{x^n\in\tilde{\mathcal{X}}^n}\Pr\left[\hat{\mathcal{H}}=0, X^n=x^n|\mathcal{H}=0\right] \geq 1-\epsilon-\eta. \tag{39}$$

In particular, there must be a special sequence $\bar{x}^n\in\tilde{\mathcal{X}}^n$ and a key $\bar{s}\in\mathcal{K}$ so that:

$$\Pr\left[\hat{\mathcal{H}}=0, X^n=\bar{x}^n\middle|\mathcal{H}=0, S=\bar{s}\right] \geq \frac{1-\epsilon-\eta}{|\tilde{\mathcal{X}}^n|}. \tag{40}$$

Pick now an arbitrary $\tilde{P}_{UV}$ with marginals $\tilde{P}_U = P_U$ and $\tilde{P}_V = P_V$ and define the sets

$$\bar{\mathcal{C}} := \{u^n \in \mathcal{U}^n \colon f^{(n)}(u^n,\bar{s}) = \bar{x}^n\} \tag{41}$$

$$\bar{\mathcal{D}} := \{v^n \in \mathcal{V}^n \colon g^{(n)}(v^n,\bar{x}^n,\bar{s}) = 0\}. \tag{42}$$

Note that (40) is equivalent to

$$P_{UV}^{\otimes n}(\bar{\mathcal{C}}\times\bar{\mathcal{D}}) \geq \frac{1-\epsilon-\eta}{|\tilde{\mathcal{X}}^n|}, \tag{43}$$

and thus, because $\tilde{P}_U = P_U$ and $\tilde{P}_V = P_V$, we have

$$\tilde{P}_U^{\otimes n}(\bar{\mathcal{C}}) = P_U^{\otimes n}(\bar{\mathcal{C}}) \geq \frac{1-\epsilon-\eta}{|\tilde{\mathcal{X}}^n|} \tag{44a}$$

and

$$\tilde{P}_V^{\otimes n}(\bar{\mathcal{D}}) = P_V^{\otimes n}(\bar{\mathcal{D}}) \geq \frac{1-\epsilon-\eta}{|\tilde{\mathcal{X}}^n|}. \tag{44b}$$

In the following we shall slightly blow up (enlarge) the sets $\bar{\mathcal{C}}$ and $\bar{\mathcal{D}}$ to obtain $\hat{\mathcal{C}}$ and $\hat{\mathcal{D}}$, and then show that these enlarged sets contain a large portion of the typical set $\mathcal{T}_{\mu_n}^{(n)}(\tilde{P}_{UV})$. To this end, let $\{\ell_n\}_{n\geq 1}$ be a sequence satisfying $\lim_{n\to\infty}\ell_n/\sqrt{n} = \infty$ and $\lim_{n\to\infty}\ell_n/n = 0$, and define the blown up regions

$$\hat{\mathcal{C}} := \left\{\tilde{u}^n \colon \exists u^n \in \bar{\mathcal{C}} \;\; s.t. \;\; d_{\mathrm{H}}(\tilde{u}^n, u^n) \leq \ell_n\right\} \tag{45}$$

$$\hat{\mathcal{D}} := \left\{\tilde{v}^n \colon \exists v^n \in \bar{\mathcal{C}} \;\; s.t. \;\; d_{\mathrm{H}}(\tilde{v}^n, v^n) \leq \ell_n\right\}. \tag{46}$$

By (44) and the blowing-up lemma [35, remark on p. 446]:

$$\tilde{P}_U^{\otimes n}(\hat{\mathcal{C}}) \geq 1-\lambda_n, \tag{47a}$$

$$\tilde{P}_V^{\otimes n}(\hat{\mathcal{D}}) \geq 1-\lambda_n, \tag{47b}$$

for some sequence λ_n that tends to 0 as $n\to\infty$.

The product $\hat{\mathcal{C}} \times \hat{\mathcal{D}}$ is the desired set because we have:

$$\begin{aligned}
\Pr[\hat{\mathcal{H}}=0|\mathcal{H}=1] &\geq \Pr[\hat{\mathcal{H}}=0|\mathcal{H}=1, S=s]\cdot \frac{1}{|\mathcal{K}|} \\
&\geq Q_{UV}^{\otimes n}\left(\bar{\mathcal{C}}\times\bar{\mathcal{D}}\right)\cdot\frac{1}{|\mathcal{K}|} \qquad (48)\\
&\geq Q_{UV}^{\otimes n}\left(\hat{\mathcal{C}}\times\hat{\mathcal{D}}\right)\cdot 2^{-n\xi_n}\cdot\frac{1}{|\mathcal{K}|}, \qquad (49)
\end{aligned}$$

where

$$\xi_n := H_{\mathrm{b}}(\ell_n/n) + \frac{\ell_n}{n}\log(|\mathcal{U}||\mathcal{V}|) - \frac{\ell_n}{n}\log\min_{(u,v)}\underbrace{Q_{UV}(u,v)}_{>0}, \tag{50}$$

and the last inequality is obtained by simple counting arguments and because $2^{nH_{\mathrm{b}}(\ell_n/n)}$ upper bounds the set of all binary vectors with Hamming weight ℓ_n/n. Notice that the terms $2^{-n\xi_n}$ and $\frac{1}{|\mathcal{K}|}$ both do not grow exponentially in n and thus will not affect the error exponent.

Change of Measure: We will restrict the acceptance region found in the previous paragraph to ensure that the source sequences in the new region have joint type close to $\tilde{P}_{UV}$. That means, we define the new set:

$$\hat{\mathcal{E}}_n \triangleq \{(u^n,v^n)\in\mathcal{T}_{\mu_n}^{(n)}(\tilde{P}_{UV})\colon u^n\in\hat{\mathcal{C}},\ v^n\in\hat{\mathcal{D}}\}. \tag{51}$$

Denote the probability of this set under $P_{UV}^{\otimes}$ by Δ_n,

$$\Delta_n := P_{UV}^{\otimes n}\left(\hat{\mathcal{E}}_n\right), \tag{52}$$

and define the tuple $(\tilde{U}^n,\tilde{V}^n)$ to be of joint pmf

$$P_{\tilde{U}^n\tilde{V}^n}(u^n,v^n) = \frac{P_{UV}^{\otimes n}(u^n,v^n)}{\Delta_n}\cdot\mathbb{1}\{(u^n,v^n)\in\hat{\mathcal{E}}_n\}. \tag{53}$$

Inequalities (47) and typicality arguments imply the following bound on the normalization factor in (53).

Lemma 1: It holds that

$$\Delta_n = \left(1-2\lambda_n-\frac{|\mathcal{U}||\mathcal{V}|}{4\mu_n^2 n}\right)\cdot 2^{-n(D(\tilde{P}_{UV}\|P_{UV})+o(1))}, \tag{54}$$

and thus

$$\lim_{n\to\infty}\frac{1}{n}\log\Delta_n := -D(\tilde{P}_{UV}\|P_{UV}). \tag{55}$$

Proof: See Appendix A. ■

Bound on the Exponent: Since $\hat{\mathcal{E}}$ is a subset of $\hat{\mathcal{C}}\times\hat{\mathcal{D}}$, we continue from (49) to obtain:

$$\begin{aligned}
&-\frac{1}{n}\log\Pr[\hat{\mathcal{H}}=0|\mathcal{H}=1] \\
&\leq -\frac{1}{n}\log Q_{UV}^{\otimes n}\left(\mathcal{E}\right)+\xi_n+\frac{1}{n}\log|\mathcal{K}| \qquad (56)\\
&\overset{(a)}{=} \frac{1}{n}P_{\tilde{U}^n\tilde{V}^n}\left(\mathcal{E}\right)\log\frac{P_{\tilde{U}^n\tilde{V}^n}\left(\mathcal{E}\right)}{Q_{UV}^{\otimes n}\left(\mathcal{E}\right)}+\xi_n+\frac{1}{n}\log|\mathcal{K}| \qquad (57)\\
&\overset{(b)}{\leq} \frac{1}{n}D\left(P_{\tilde{U}^n\tilde{V}^n}\|Q_{UV}^{\otimes n}\right)+\xi_n+\frac{1}{n}\log|\mathcal{K}| \qquad (58)\\
&\overset{(c)}{\leq} \frac{1}{n}\sum_{(u^n,v^n)\in\mathcal{E}} P_{\tilde{U}^n\tilde{V}^n}(u^n,v^n)\log\frac{P_{UV}^{\otimes n}(u^n,v^n)}{Q_{UV}^{\otimes n}(u^n,v^n)} \\
&\quad -\frac{1}{n}\log(\Delta_n\cdot|\mathcal{K}|)+\xi_n \qquad (59)\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{(u^n,v^n)\in\mathcal{E}} P_{\tilde{U}^n\tilde{V}^n}(u^n,v^n)\log\frac{P_{UV}(u_i,v_i)}{Q_{UV}(u_i,v_i)} \\
&\quad -\frac{1}{n}\log(\Delta_n|\mathcal{K}|)+\xi_n \qquad (60)\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{u_i,v_i} P_{\tilde{U}_i\tilde{V}_i}(u_i,v_i)\log\frac{P_{UV}(u_i,v_i)}{Q_{UV}(u_i,v_i)} \\
&\quad -\frac{1}{n}\log(\Delta_n\cdot|\mathcal{K}|)+\xi_n \qquad (61)\\
&= \sum_{u,v} P_{\tilde{U}_T\tilde{V}_T}(u,v)\log\frac{P_{UV}(u,v)}{Q_{UV}(u,v)} \\
&\quad -\frac{1}{n}\log(\Delta_n\cdot|\mathcal{K}|)+\xi_n, \qquad (62)
\end{aligned}$$

where T is a uniform random variable over $\{1,\ldots,n\}$ independent of $(\tilde{U}^n,\tilde{V}^n)$. In above inequalities, (a) holds because $P_{\tilde{U}^n\tilde{V}^n}$ is only defined on $\hat{\mathcal{E}}$ and thus $P_{\tilde{U}^n\tilde{V}^n}(\hat{\mathcal{E}})=1$; (b) holds by the data processing inequality and because (57) is the KL-divergence between the two binary distributions obtained by applying $P_{\tilde{U}^n\tilde{V}^n}$ and $Q_{UV}^{\otimes n}$ to the indicator function $\mathbb{1}\{(u^n,v^n)\in\mathcal{E}\}$ and its complement event; and (c) holds by the definition of $P_{\tilde{U}^n\tilde{V}^n}$.

Using Lemma 1, the subexponential behaviour of the key size (3), the fact that $\xi_{\to}0$ as $n\to\infty$, and finally also equality $P_{\tilde{U}_T\tilde{V}_T}(u,v)=\tilde{P}_{UV}(u,v)+o(1)$, by taking $n\to\infty$, we obtain:

$$\begin{aligned}
&\varlimsup_{n\to\infty} -\frac{1}{n}\log\Pr\left[\hat{\mathcal{H}}=1|\mathcal{H}=0\right] \qquad (63)\\
&\leq \sum_{u,v}\tilde{P}_{UV}(u,v)\log\frac{P_{UV}(u,v)}{Q_{UV}(u,v)}-D(\tilde{P}_{UV}\|P_{UV}) \qquad (64)\\
&= D(\tilde{P}_{UV}\|Q_{UV}). \qquad (65)
\end{aligned}$$

Since the inequality holds for any choice of $\tilde{P}_{UV}$ with marginals P_U and P_V, we have proved the desired converse result.

IV. Conclusion

The paper has introduced the problem of distributed detection over a discrete memoryless channel under a covertness constraint. The main contribution is a converse result that shows that for sublinear key lengths (which are typically employed in covert communication) the largest possible Stein exponent cannot exceed the Stein exponent of the distributed detection problem when communication is over a link that is noise-free but of zero rate. For certain source distributions, this converse result matches with the Stein exponent obtained by a local test at the decision center. For these sources, the covertness constraint renders communication useless in terms of improving the Stein error exponent for hypothesis testing.

The question that remains after our work is whether covert communication can be used to improve Stein exponents (or other hypothesis testing error exponents) for general sources. The fact that no positive decoding error exponents are possible for standard covert data communication [36] hints to a negative answer for our setup.

References

[1] R. Ahlswede and I. Csiszár, “Hypothesis testing with communication constraints,” *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.

[2] 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.

[3] T. S. Han, “Hypothesis testing with multiterminal data compression,” *IEEE Trans. Inf. Theory*, vol. 33, pp. 759–772, Nov. 1987.

[4] H. Shalaby and A. Papamarcou, “Multiterminal detection with zero-rate data compression,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 254–267, 1992.

[5] S. Watanabe, “Neyman-pearson test for zero-rate multiterminal hypothesis testing,” *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4923–4939, 2018.

[6] E. Haim and Y. Kochman, “Binary distributed hypothesis testing via korner-marton coding,” in *Proc. IEEE Info. Theory Work. (ITW)*, 2016.

[7] N. Weinberger, Y. Kochman, and M. Wigger, “Exponent trade-off for hypothesis testing over noisy channels,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1852–1856, 2019.

[8] 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.

[9] S. Sreekumar and D. Gndz, “Distributed hypothesis testing over discrete memoryless channels,” *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2044–2066, 2020.

[10] S. Salehkalaibar and M. Wigger, “Distributed hypothesis testing based on unequal-error protection codes,” *IEEE Trans. Inf. Theory*, vol. 66, pp. 4150–41820, Jul. 2020.

[11] M. Mhanna and P. Piantanida, “On secure distributed hypothesis testing,” in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1605–1609, 2015.

[12] J. Liao, L. Sankar, V. Y. F. Tan, and F. du Pin Calmon, “Hypothesis testing under mutual information privacy constraints in the high privacy regime,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 1058–1071, 2018.

[13] S. B. Amor, A. Gilani, S. Salehkalaibar, and V. Y. F. Tan, “Distributed hypothesis testing with privacy constraints,” in *2018 International Symposium on Information Theory and Its Applications (ISITA)*, pp. 742–746, 2018.

[14] S. Sreekumar and D. Gndz, “Testing against conditional independence under security constraints,” in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 181–185, 2018.

[15] S. Faour, M. Hamad, M. Sarkiss, and M. Wigger, “Testing against independence with an eavesdropper,” in *2023 IEEE Information Theory Workshop (ITW)*, pp. 277–282, 2023.

[16] S. Sreekumar and D. Gndz, “Strong converse for testing against independence over a noisy channel,” in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1283–1288, 2020.

[17] T. Han and K. Kobayashi, “Exponential-type error probabilities for multiterminal hypothesis testing,” *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 2–14, 1989.

[18] P. Escamilla, M. Wigger, and A. Zaidi, “Distributed hypothesis testing: cooperation and concurrent detection,” *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7550–7564, 2020.

[19] B. A. Bash, D. Goeckel, and D. Towsley, “Limits of reliable communication with low probability of detection on awgn channels,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013.

[20] L. Wang, G. W. Wornell, and L. Zheng, “Fundamental limits of communication with low probability of detection,” *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, 2016.

[21] M. R. Bloch, “Covert communication over noisy channels: A resolvability perspective,” *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[22] K. S. K. Arumugam and M. R. Bloch, “Covert communication over a k -user multiple-access channel,” *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7020–7044, 2019.

[23] K. S. Kumar Arumugam and M. R. Bloch, “Embedding covert information in broadcast communications,” *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2787–2801, 2019.

[24] D. Kibloff, S. M. Perlaza, and L. Wang, “Embedding covert information on a given broadcast code,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 2169–2173, 2019.

[25] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, “Covert communication with channel-state information at the transmitter,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2310–2319, 2018.

[26] H. ZivariFard, M. Bloch, and A. Nosratinia, “Keyless covert communication in the presence of non-causal channel state information,” in *2019 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2019.

[27] H. ZivariFard, M. R. Bloch, and A. Nosratinia, “Keyless covert communication via channel state information,” *CoRR*, vol. abs/2003.03308, 2020.

[28] H. ZivariFard, M. R. Bloch, and A. Nosratinia, “Covert communication via non-causal cribbing from a cooperative jammer,” in *2021 IEEE International Symposium on Information Theory (ISIT)*, pp. 202–207, 2021.

[29] A. Bounhar, M. Sarkiss, and M. Wigger, “Mixing a covert and a non-covert user,” *arXiv preprint arXiv:2305.06268*, 2023.

[30] M. Tahmasbi and M. R. Bloch, “Active covert sensing,” in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 840–845, 2020.

[31] B. A. Bash, C. N. Gagatsos, A. Datta, and S. Guha, “Fundamental limits of quantum-secure covert optical sensing,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 3210–3214, 2017.

[32] M.-C. Chang and M. R. Bloch, “Covert sequential hypothesis testing,” in *2021 IEEE Information Theory Workshop (ITW)*, pp. 1–6, 2021.

[33] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, “Gaussian signalling for covert communications,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3542–3553, 2019.

[34] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[35] K. Marton, “A simple proof of the blowing-up lemma,” *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.

[36] M. Tahmasbi, M. R. Bloch, and V. Y. F. Tan, “Error exponent for covert communications over discrete memoryless channels,” in *2017 IEEE Information Theory Workshop (ITW)*, pp. 304–308, 2017.

[37] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

APPENDIX A
PROOF OF LEMMA 1

The proof is inspired by proof steps in [4].

Let (U^n, V^n) be i.i.d. $\tilde{P}_{UV}$. By the union bound and by (47) and standard bounds on the probability of the typical set [37]

$$\begin{aligned} 1 - \tilde{P}_{UV}^{\otimes n}(\hat{\mathcal{E}}) & \\ &\leq \Pr[U^n \notin \hat{\mathcal{C}} \cup V^n \notin \hat{\mathcal{D}} \cup (U^n, V^n) \notin \mathcal{T}_{\mu_n}^{(n)}(\tilde{P}_{UV})] \qquad (66) \\ &\leq \Pr[U^n \notin \hat{\mathcal{C}}] + \Pr[V^n \notin \hat{\mathcal{D}}] + \Pr[(U^n, V^n) \notin \mathcal{T}_{\mu_n}^{(n)}(\tilde{P}_{UV})] \qquad (67) \\ &\leq 2\lambda_n + \frac{|\mathcal{U}||\mathcal{V}|}{4\mu_n^2 n}, \qquad (68) \end{aligned}$$

and thus

$$\tilde{P}_{UV}^{\otimes n}(\hat{\mathcal{E}}) \geq 1 - 2\lambda_n - \frac{|\mathcal{U}||\mathcal{V}|}{4\mu_n^2 n}. \qquad (69)$$

This probability can be decomposed into the contributions of the various type-classes. In fact, because $\hat{\mathcal{E}} \subseteq \mathcal{T}_{\mu_n}^{(n)}(\tilde{P}_{UV})$ and because within a type-class all sequences are equally-likely:

$$\begin{aligned} \sum_{\substack{\boldsymbol{\pi}_{UV}: \\ |\boldsymbol{\pi}_{UV} - \tilde{P}_{UV}| \leq \mu_n}} & \tilde{P}_{UV}^{\otimes n}\left(\mathcal{T}_0^{(n)}(\boldsymbol{\pi}_{UV})\right) \frac{\left|\hat{\mathcal{C}} \times \hat{\mathcal{D}} \cap \mathcal{T}_0^{(n)}(\boldsymbol{\pi}_{UV})\right|}{\left|\mathcal{T}_0^{(n)}(\boldsymbol{\pi}_{UV})\right|} \\ &= 1 - 2\lambda_n - \frac{|\mathcal{U}||\mathcal{V}|}{4\mu_n^2 n}. \qquad (70) \end{aligned}$$

Since the sum of probabilities

$$\sum_{\substack{\boldsymbol{\pi}_{UV}: \\ |\boldsymbol{\pi}_{UV} - \tilde{P}_{UV}| \leq \mu_n}} \tilde{P}_{UV}^{\otimes n}\left(\mathcal{T}_0^{(n)}(\boldsymbol{\pi}_{UV})\right) \leq 1, \qquad (71)$$

inequality (70) implies the existence of a type $\hat{\boldsymbol{\pi}}_{UV}$ within distance μ_n of $\tilde{P}_{UV}$ so that

$$\frac{\left|\hat{\mathcal{C}} \times \hat{\mathcal{D}} \cap \mathcal{T}_0^{(n)}(\hat{\boldsymbol{\pi}}_{UV})\right|}{\left|\mathcal{T}_0^{(n)}(\hat{\boldsymbol{\pi}}_{UV})\right|} \geq 1 - 2\lambda_n - \frac{|\mathcal{U}||\mathcal{V}|}{4\mu_n^2 n}. \qquad (72)$$

We can use this bound on the cardinalities directly to obtain the desired lower bound on Δ_n. In fact, again using the fact that all sequences in a type-class have same probabilities under i.i.d. distributions:

$$\begin{aligned} \Delta_n &\geq P_{UV}^{\otimes n}\left(\bar{\mathcal{C}} \times \bar{\mathcal{D}} \cap \mathcal{T}_0^{(n)}(\hat{\boldsymbol{\pi}}_{UV})\right) \qquad (73) \\ &= \frac{\left|\hat{\mathcal{C}} \times \hat{\mathcal{D}} \cap \mathcal{T}_0^{(n)}(\hat{\boldsymbol{\pi}}_{UV})\right|}{\left|\mathcal{T}_0^{(n)}(\boldsymbol{\pi}_{UV})\right|} \cdot \tilde{P}_{UV}^{\otimes n}\left(\mathcal{T}_0^{(n)}(\hat{\boldsymbol{\pi}}_{UV})\right) \qquad (74) \\ &\geq \left(1 - 2\lambda_n - \frac{|\mathcal{U}||\mathcal{V}|}{4\mu_n^2 n}\right) 2^{-n(D(\hat{\boldsymbol{\pi}}_{UV} \| P_{UV}) + o(1))} \qquad (75) \end{aligned}$$

Since $\mu_n \to 0$ as $n \to \infty$, the type $\hat{\boldsymbol{\pi}}_{UV}$ tends to $\tilde{P}_{UV}$ and by continuity of the KL divergence, we obtain

$$D(\hat{\boldsymbol{\pi}}_{UV} \| P_{UV}) = D(\tilde{P}_{UV} \| P_{UV}) + o(1), \qquad (76)$$

which proves the lemma.