

Asymptotic Performance of a Censoring Sensor Network

Wee Peng Tay, *Student Member, IEEE*, John N. Tsitsiklis, *Fellow, IEEE*, and Moe Z. Win, *Fellow, IEEE*

Abstract—We consider the problem of decentralized binary detection in a sensor network where the sensors have access to side information that affects the statistics of their measurements, or reflects the quality of the available channel to a fusion center. Sensors can decide whether or not to make a measurement and transmit a message to the fusion center (“censoring”), and also have a choice of the mapping from measurements to messages. We consider the case of a large number of sensors, and an asymptotic criterion involving error exponents. We study both a Neyman–Pearson and a Bayesian formulation, characterize the optimal error exponent, and derive asymptotically optimal strategies for the case where sensor decisions are only allowed to depend on locally available information. Furthermore, we show that for the Neyman–Pearson case, global sharing of side information (“sensor cooperation”) does not improve asymptotic performance, when the Type I error is constrained to be small.

Index Terms—Censoring, cooperation, decentralized detection, error exponent, sensor networks.

I. INTRODUCTION

A sensor network is often constrained by energy limitations, because the sensors are typically low-cost battery-powered devices, and by communication limitations, because of transmission costs or limited available radio resources. When deploying a large-scale sensor network for the purposes of detection, e.g., for environmental monitoring or for detecting anomalies in industrial applications, we are often faced with a tradeoff between energy efficiency and detection reliability. To address this issue, “censoring networks” have been introduced in [1] and later in [2]. These references consider a binary detection problem, assume that the sensors obtain independent measurements X_i , and raise the question of deciding which sensors should transmit their measurements to a fusion center, subject to a constraint on the average number of transmitting sensors. In particular, they assume that the sensors are operating independently from each other, i.e., the censoring decisions do not involve any sensor cooperation or exchange of information. Their main results state that each sensor should base its decision on the likelihood ratio associated with its measurement,

Manuscript June 30, 2006; revised June 2, 2007. This work was supported, in part, by the National Science Foundation under Contracts ECS-0426453 and ANI-0335256, the Charles Stark Draper Laboratory Robust Distributed Sensor Networks Program, and an Office of Naval Research Young Investigator Award N00014-03-1-0489. The material in this paper was presented in part at the Conference on Information Sciences and Systems, Princeton, NJ, March 2006 and IEEE International Symposium on Information Theory, Seattle, WA, July 2006.

The authors are with the Laboratory for Information and Decision Systems, the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: wptay@mit.edu; jnt@mit.edu; moewin@mit.edu).

Communicated by A. Høst-Madsen, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2007.907441

and should transmit X_i only if the likelihood ratio falls outside a “censoring interval.” Subsequently, [3] and [4] consider the asymptotic performance of “constrained networks,” including the case of an overall power constraint and the case of capacity-constrained communications. The question of deciding which sensors should transmit is replaced by the question of choosing the mode of sensor transmissions. There are differences between the problems considered in [1] and the problems studied in [3] and [4], but there are also significant similarities, suggesting that a unified treatment may be possible. Such a unified treatment, at a higher level of generality, is one of the objectives of this paper. However, we will be concerned with sensor networks with an asymptotically large number of nodes, unlike in [1], [2], where the problem of censoring is treated for the case of a fixed number of nodes. In particular, we provide a formulation that involves:

- 1) a general class of “network constraints,”
- 2) a general class of possible “transmission modes” that the sensors can choose from, and
- 3) the possibility of having access to some side information R_i , which can be used to choose between transmission modes.

We use the term “side information” in a very general way to refer to some observable that affects the operation of each sensor. In general, R_i could provide information on the quality of the channel from sensor i to the fusion center, or on the quality of the measurement available at sensor i . The choice of what side information is available depends on the specific problem and its constraints. We illustrate our framework by presenting two motivating examples, which will be revisited in Section VIII.

Example 1 (Fading Channels): Consider a large number n of sensors deployed for the purposes of detection, that is, testing between two hypotheses H_0 and H_1 . Each sensor i obtains an independent measurement \tilde{X}_i (with a different distribution under each hypothesis), which it can encode and transmit to a fusion center through a noisy channel. The message received at the fusion center is of the form

$$Y_i = Q_i \tilde{\gamma}_i(\tilde{X}_i) + W_i$$

where Q_i is a stochastic fading coefficient, and W_i is zero-mean Gaussian noise with known variance σ_i^2 , independent of everything else. In order to conserve power, or to avoid divulging the presence of the sensors, we introduce a constraint that under “normal conditions” (that is, under the null hypothesis H_0), the expected number of transmitting sensors is bounded by nc , where $c \in (0, 1]$ is a given constant. Then, the sensor network

is faced with the problem of choosing which sensors should transmit their measurements to the fusion center. Suppose that the network has knowledge of the channel state information Q_i and σ_i , $i = 1, \dots, n$. Obviously, we would like to choose only those sensors that have a favorable channel to the fusion center, so the choice should be based on $R_i = (Q_i, \sigma_i)$. (In some cases, σ_i is a known constant, then the choice is made based only on $R_i = Q_i$.) Furthermore, we would like to examine and compare a cooperative scheme (the decision to transmit or not by each sensor depends on the channel parameters of *all* sensors) and a distributed scheme (the decision of each sensor depends only on the local channel parameters). Finally, we may want to optimize the choice of the “transmission function” $\tilde{\gamma}_i$ from within a class of possible such functions. \square

Example 2 (Spatial Signal): Consider the problem of detecting a spatial signal on the domain $[-1, 1]$ (or more generally on a bounded subset of \mathbb{R}^d). The sensors are placed randomly and uniformly in the set $[-1, 1]$, with the fusion center at the origin. Let R_i be the location of sensor i . This serves as the side information that is available. There are two possible spatial signals $s_0(\cdot)$ and $s_1(\cdot)$, and we wish to detect which of the two is present. Each sensor i makes a noisy measurement of the local signal. We assume that the power required for a sensor to transmit its measurement depends on the distance from the sensor to the fusion center. Given a constraint on the total power used by the sensors, which ones should be chosen to transmit? It is not necessarily the case that the sensors closest to the fusion center should be the ones transmitting; for instance, if the spatial signals $s_0(r)$ and $s_1(r)$ are equal when r is in the vicinity of the fusion center, the sensors close to the fusion center do not have any information worth transmitting. In Section VIII-B, we will give an example where the transmitting sensors should be the ones furthest away from the fusion center. \square

In this paper, we provide a framework that is general enough to address the questions posed in the above examples. In particular, we generalize the idea of a censoring network presented in [1], and introduce network constraints in the spirit of [3] and [4]. To be specific, we consider a Neyman–Pearson decentralized binary detection problem where the sensors utilize side information R_i when deciding to transmit their measurements to the fusion center. We introduce a cost function that depends only on the side information and the transmission policy used, and require that the expected average cost per sensor be below a given threshold. Finding an optimal solution to such a problem is rather intractable. For this reason, we focus on an asymptotic analysis, involving a large number of sensors and the minimization of the error exponent, in the spirit of [5], which introduced and studied such a formulation in the context of unconstrained sensor networks.

In our formulation, we allow the sensors to cooperate, in the sense that the sensors’ censoring decisions can be made by the fusion center, on the basis of the entire vector (R_1, \dots, R_n) of side-information values at each sensor. This can arise, for example, when the local pieces of side information are some low-resolution data that can be transmitted to the fusion center inexpensively, or when the fusion center is able to monitor the state of the channels from the sensors to itself. Nevertheless, we

will establish that when the Type I error probability is asymptotically small, optimal performance can be achieved even in the absence of such cooperation, by having each sensor make its censoring and transmission decisions only on the basis of the locally available side information. Furthermore, all the sensors can use the same policy, which shows that a simple distributed scheme is asymptotically optimal. The case where there is no cooperation is the asymptotic counterpart of the censoring problem considered in [1] (cf. Section V-C).

We then proceed to consider the Bayesian counterpart of the above formulation, except that for reasons described in Section IX, the cooperation among sensors is explicitly ruled out. We characterize the asymptotically optimal performance and the strategies that achieve it. We show that an optimal scheme is to divide the sensors into two groups, each group using the same policy. We also show how some of the results in [3] and [4] can be derived by converting the problems studied therein to our framework. Finally, we provide a generalization of some of the results in [4].

The rest of the paper is organized as follows. In Section II, we present our model in detail. In Section III, we state the Neyman–Pearson version of the problem and our assumptions. In Section IV, we consider the case where the side information variables R_i are independent and identically distributed (i.i.d.) under either hypothesis, and prove the main results for the Neyman–Pearson case. In Section V, we discuss some variations of the results, and in Section VI, we provide an extension to the case where the variables R_i are stationary and ergodic. In Section VII, we characterize optimal policies, and in Section VIII, we revisit the two examples given in this Introduction. In Section IX, we analyze the Bayesian formulation. In Section X, we rederive some of the results in [3], and extend some of the results in [4]. Finally, in Section XI, we summarize and offer some closing comments.

II. PROBLEM FORMULATION

In this section, we introduce our model. With a few exceptions, we will be using upper case letters to denote random variables, and lower case letters to denote values of these random variables. We will also be using the notation $z^{(n)}$ to denote a vector (z_1, \dots, z_n) , where the components of the vector may be numbers, random variables, or functions.

A. The Basic Elements of the Model

We consider a hypothesis testing problem involving two hypotheses, H_0 and H_1 . There are n sensors and a fusion center. Each sensor i observes some side information R_i , which is a random variable taking values in a set \mathcal{R} , and a measurement X_i taking values in a set \mathcal{X} . In addition, there is an auxiliary random variable V , taking values in a set \mathcal{V} of our own choosing, which will be used as the “seed” whenever a randomized decision is to be made. These are all the basic random variables in our model. We assume a suitably large measurable space (Ω, \mathcal{F}) so that all random variables can be defined on that space, for any number n of sensors. To avoid technical distractions, we will not delve into measurability issues.

Under each hypothesis H_j , $j = 0, 1$, we assume that we have a measure \mathbb{P}_j on (Ω, \mathcal{F}) , and a corresponding expectation operator \mathbb{E}_j , with the following properties.

- 1) The random variable R_i is distributed according to a given marginal probability law μ_j , for every i .
- 2) Conditioned on $R_1 = r_1, R_2 = r_2, \dots, R_n = r_n$, the measurements X_i are (conditionally) independent, and each X_i is distributed according to a given regular conditional distribution $\nu_j(\cdot | r_i)$.
- 3) The random variable V is independent of the random variables R_i and X_i , with a distribution that is the same under both hypotheses, and which will be of our choosing.

Note that we have only specified the marginal distributions of the variables R_i . Regarding their joint distribution, we will be making in the sequel one of the following alternative assumptions.

- 1) Under either hypothesis, the random variables R_i are independent (and therefore i.i.d.).
- 2) Under either hypothesis, the sequence (R_1, R_2, \dots) is stationary and ergodic. In this case, we also assume $\mu_0 = \mu_1$, so that the variables R_i provide no information about the true hypothesis.

B. Sensor Policies and Strategies

There are two types of decisions to be made at each sensor: deciding whether to make a measurement (not censoring), and if a measurement is made, deciding what to transmit to the fusion center. These decisions are to be made based on available information, according to a set of rules (policies). We describe here the types of policies to be considered.

We assume that $R^{(n)}$ is known at the fusion center (in a mathematically equivalent scenario, we could have each sensor communicate its side information to every other sensor) and that the same is true for the auxiliary random variable V . Based on $R^{(n)}$ and V , we let the fusion center decide which of the sensors should make a measurement X_i . (This is what we term as *cooperation*: the decision depends on the side information of all sensors.) Subsequently, each uncensored sensor is to generate a message to the fusion center.

Formally, we define a pure censoring policy for sensor i as a function $\xi_i : \mathcal{R}^n \mapsto \{0, 1\}$. Let the set of pure censoring policies be Ξ . A pure transmission policy for sensor i is a function $\gamma_i : \mathcal{X} \times \mathcal{R} \mapsto \mathcal{Y}$, where \mathcal{Y} is a (possibly infinite) transmission alphabet. These policies are called *pure* because they do not make use of the randomization variable V . We restrict pure transmission policies to belong to a given set Γ . The pair (ξ_i, γ_i) is called a *pure policy* for sensor i .

We allow censoring and transmission policies to be randomized. Let $\pi_i = (\xi_{i,v}, \gamma_{i,v})_{v \in \mathcal{V}}$ be a collection of pure policies indexed by \mathcal{V} . We call π_i a *policy* for sensor i . We envisage the following sequence of events. A realization v of the randomization variable V is generated (this can be done at the fusion center, with the result communicated to all sensors, or at each sensor using a common seed). Sensor i then uses the pure policy $(\xi_{i,v}, \gamma_{i,v})$. It is censored (no measurement is made) if and only if $\xi_{i,v}(R^{(n)}) = 0$. If, on the other hand, $\xi_{i,v}(R^{(n)}) = 1$, a message $Y_i = \gamma_{i,v}(X_i, R_i)$ is transmitted to the fusion center. Although we say that the message Y_i is transmitted to the fusion center, our formulation allows for the inclusion of channel

noise in the transmission function $\gamma_{i,v}$. More specifically, suppose that the message $\tilde{Y}_i = \tilde{\gamma}_{i,v}(\tilde{X}_i, R_i)$ is transmitted over a noisy channel so that $Y_i = f(\tilde{Y}_i, R_i, W_i)$ is received at the fusion center. Here, f is the channel transfer function and W_i is a random variable conditionally independent of \tilde{X}_i , given R_i . Then, we can define $X_i = (\tilde{X}_i, W_i)$, and the transmission function as $\gamma_{i,v}(X_i, R_i) = f(\tilde{\gamma}_{i,v}(\tilde{X}_i, R_i), R_i, W_i)$. As an example, consider Example 1 of Section I. In our present notation, we have $X_i = (\tilde{X}_i, W_i)$, $R_i = (Q_i, \sigma_i)$, and $\gamma_{i,v}(X_i, R_i) = Q_i \tilde{\gamma}_{i,v}(\tilde{X}_i) + W_i$. Therefore, in the sequel, we will assume that the message received at the fusion center is the same as Y_i . For convenience, we also assume that X_i and $\gamma_{i,v}(X_i, R_i)$ are always defined, even if sensor i is censored and nothing gets transmitted.

A collection $\pi^{(n)}$ of policies, one for each sensor, all of which involve the same set \mathcal{V} and the same randomization variable V , together with the distribution of V , will be called a *strategy*. We will often abuse terminology, however, and will be referring to $\pi^{(n)}$ as a strategy.

C. Resource Constraints

We assume that when sensor i makes a measurement X_i and transmits Y_i to the fusion center, certain resources are consumed, and therefore a cost is incurred, possibly depending on the side information at that sensor. To model such costs, we introduce a function $\rho : \mathcal{R} \times \Gamma \mapsto [0, \infty)$, and interpret $\rho(r, \gamma)$ as the cost incurred by a sensor i that uses a pure policy $\pi = (\xi, \gamma)$, if the side information at that sensor takes on the value r , and the sensor is not censored, i.e., $\xi(R^{(n)}) = 1$. When the sensor is censored, we assume that no cost is incurred, so that the resulting expected cost at sensor i (under H_0) equals $\rho(\pi) = \mathbb{E}_0 [\xi(R^{(n)})\rho(R_i, \gamma)]$. For a more general (randomized) policy $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}}$, $\rho(\pi)$ is defined to be equal to $\mathbb{E}_0 [\xi_V(R^{(n)})\rho(R_i, \gamma_V)]$, where the expectation is taken with respect to both $R^{(n)}$ and V . We will say that a strategy $\pi^{(n)} = (\pi_1, \dots, \pi_n)$ is *admissible* if

$$\frac{1}{n} \sum_{i=1}^n \rho(\pi_i) \leq c \quad (1)$$

where c is a given constant.

Note that the resource constraint is in place only under H_0 . The presumption here is that H_0 (the ‘‘null hypothesis’’) corresponds to a ‘‘normal’’ situation. Thus, we are constraining the resource utilization to be low under normal circumstances, but allow higher resource utilization under exceptional circumstances. However, in a Bayesian formulation, we will define $\mathbb{E} = q_0 \mathbb{E}_0 + q_1 \mathbb{E}_1$, where q_j is the prior probability of hypothesis H_j , and will replace \mathbb{E}_0 with \mathbb{E} in the definition of $\rho(\pi_i)$.

The following are two examples of resource constraints. Many other choices are possible, to reflect particular constraints of interest to a system designer.

Example 3 (Proportional Censoring): If $\rho(r, \gamma) = 1$ for all $r \in \mathcal{R}$ and all $\gamma \in \Gamma$, then (1) becomes a constraint on the average proportion of sensors that make a measurement. \square

Example 4 (Power Constraints): Suppose that $\rho(r, \gamma) = \mathbb{E}_0 [|\gamma(X_1, r)|^2 | R_1 = r]$. In this case, (1) becomes a constraint on the average transmission power. \square

D. The Fusion Center

The fusion center receives the messages Y_i from each sensor. Based on this information, together with the side information $R^{(n)}$ and the random variable V , it decides between the two hypotheses. Recall that in classical (centralized) Neyman–Pearson hypothesis testing, randomization can reduce the Type II error probability. Accordingly, we assume that the fusion center has access to another random variable V' which is uniformly distributed in $[0, 1]$, and independent of everything else. We then let the fusion center use a randomized fusion rule $\phi : \mathcal{Y}^n \times \mathcal{R}^n \times \mathcal{V} \times [0, 1] \rightarrow \{0, 1\}$ to select one of the two hypotheses. Let $\hat{H}_n = \phi(Y^{(n)}, R^{(n)}, V, V')$, which is a binary random variable indicating the selected hypothesis. In the above expression, and in order to keep notation simple, we assume that whenever sensor i is censored, Y_i is set to a special symbol y^* .

We summarize the elements of our model in the following definition.

Definition 1: An overall strategy consists of the following.

- 1) A set \mathcal{V} , and the distribution of a \mathcal{V} -valued random variable V .
- 2) An admissible strategy $\pi^{(n)}$ (i.e., one that satisfies the resource constraints).
- 3) A fusion rule ϕ .

For given n and c , and a given overall strategy, the Type I error and the Type II error probabilities $\mathbb{P}_0(\hat{H}_n = 1)$ and $\mathbb{P}_1(\hat{H}_n = 0)$ are well defined. In a Neyman–Pearson formulation (Section III), we will aim at minimizing the probability of the Type II error (more precisely, its error exponent), subject to a constraint on the Type I error probability. In a Bayesian formulation (Section IX), we will aim at minimizing a weighted average of these two error probabilities.

E. Independent Randomization

Our model allows for randomization based on a globally known randomization variable V , whose distribution is subject to our choice. Such a V can be generated at each sensor using a common seed, or it can be generated at the fusion center and communicated to the sensors. As discussed in [6], the above model of *dependent randomization* includes the special case of *independent randomization*, where the sensors rely on locally generated independent random variables.

Formally, we will say that we have *independent randomization* if the set \mathcal{V} is a Cartesian product of n copies of another set \mathcal{V}_0 , i.e., $\mathcal{V} = \mathcal{V}_0^n$, and is endowed with a product measure, so that V is of the form $V = (V_1, \dots, V_n) = V^{(n)}$, where the V_i are independent.

F. Local and Homogeneous Strategies

Loosely speaking, in a local strategy every sensor i has access only to an independent, locally generated random variable V_i and its own side information R_i , thus allowing for distributed implementation. Furthermore, in a homogeneous local strategy, every sensor responds in the same way to its local variables. In the definition below, $v = (v_1, \dots, v_n)$.

Definition 2:

- 1) A policy $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}}$ is said to be *local* (for sensor i), if i) independent randomization is used; ii) $\xi_v(r^{(n)})$ can be expressed as a function of only r_i and v_i ; and iii) $\gamma_v(x, r)$ can be expressed as a function of only x , r , and v_i .
- 2) A strategy $\pi = (\pi_1, \dots, \pi_n)$ is said to be *local* if each π_i is a local policy for sensor i .

A local policy for sensor i is denoted as $\pi_i = (\xi_{i,v_i}, \gamma_{i,v_i})_{v_i \in \mathcal{V}_0}$, where the functions ξ_{i,v_i} and γ_{i,v_i} are now functions whose arguments are the local random variables X_i and R_i .

Definition 3: A local strategy is said to be *homogeneous* if the independent random variables V_i are identically distributed, and if the policy of every sensor is identified with the same local policy.

Let us remark that for a homogeneous local strategy associated with a common local policy π , the resource constraint (1) simplifies to $\rho(\pi) \leq c$. We let $\Pi(c)$ be the set of local policies that satisfy this constraint.

For the reader's convenience, we summarize the notation introduced so far:

X_i	Measurement of sensor i .
R_i	Side information of sensor i .
V	Randomization variable.
ξ_i	Pure censoring policy for sensor i .
γ_i	Pure transmission policy for sensor i .
$\pi_i = (\xi_{i,v}, \gamma_{i,v})_{v \in \mathcal{V}}$	A policy for sensor i . Given V , $\xi_{i,V}$ is a pure censoring policy, and $\gamma_{i,V}$ is a pure transmission policy.
$\pi^{(n)}$	A strategy (π_1, \dots, π_n) .
$\rho(\pi)$	Expected cost of policy π .
$\Pi(c)$	Set of local policies π satisfying the resource constraint $\rho(\pi) \leq c$.

III. THE NEYMAN–PEARSON PROBLEM

Given an overall strategy for the n -sensor problem, we will use β_n to denote the resulting Type II error probability $\mathbb{P}_1(\hat{H}_n = 0)$. For any given n , c , and α , we define

$$\beta_n^*(c, \alpha) = \inf \beta_n$$

where the infimum is taken over all overall strategies that satisfy the resource constraint (1), as well as the constraint $\mathbb{P}_0(\hat{H}_n = 1) \leq \alpha$.

The above optimization problem is intractable, even in the absence of censoring. Even if it were tractable, implementing an optimal cooperative censoring strategy would involve complicated feedback from the fusion center to the sensors. We will see, however, that the problem becomes tractable if n is large and α is small, and under an asymptotic optimality criterion.

Furthermore, the resulting strategy will take a simple form that allows for distributed implementation.

For any reasonable overall strategy, the resulting Type II error probability β_n falls exponentially with n . For this reason, as in [5], we are interested in the optimal error exponent. Furthermore, we will focus on the case of an asymptotically small Type I error probability and the associated optimal error exponent

$$\lim_{\alpha \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(c, \alpha).$$

A. Assumptions and Notation

Recall that under H_j , the measure μ_j describes the distribution of R_i , and $\nu_j(\cdot | \cdot)$ describes the conditional distribution of X_i given R_i . We use the notation $\mu \ll \mu'$ to indicate that a measure μ is absolutely continuous with respect to another measure μ' .

Assumption 1: We have $\mu_0 \ll \mu_1$, and for every $r \in \mathcal{R}$, $\nu_0(\cdot | r) \ll \nu_1(\cdot | r)$.

This assumption results in little loss of generality. For example, suppose that for some measurable set $A \subset \mathcal{R}$, we have $\mu_0(A) > 0$ and $\mu_1(A) = 0$. Consider then the fusion rule that decides in favor of H_0 if and only if there exists some i for which $R_i \in A$. Clearly, the Type I error probability equals $\mathbb{P}_0(R_i \notin A, \text{ for all } i)$, and converges to zero, so that the constraint $\mathbb{P}_0(\hat{H}_n = 1) \leq \alpha$ is satisfied for large enough n . In addition, the Type II error probability β_n is equal to zero. Thus, the optimal error exponent is equal to $-\infty$ and is achieved by a very simple overall strategy.

Let $d\mu_0/d\mu_1$ be the Radon–Nikodym derivative (likelihood ratio) of the measures μ_0 and μ_1 . Similarly, we define a function $\ell_{01} : \mathcal{X} \times \mathcal{R} \mapsto [0, \infty)$, so that for every $r \in \mathcal{R}$, $\ell_{01}(X_i | r)$ is the likelihood ratio between the two hypotheses, when X_i is observed, given that $R_i = r$. Formally, this is the Radon–Nikodym derivative of the measures $\nu_0(\cdot | r)$ and $\nu_1(\cdot | r)$ on the set \mathcal{X} .

In the same vein, for any pure transmission policy $\gamma \in \Gamma$, we define a function $\ell'_{01} : \mathcal{Y} \times \mathcal{R} \mapsto [0, \infty)$, so that for every $r \in \mathcal{R}$, $\ell'_{01}(Y_i | r)$ is the likelihood ratio between the two hypotheses, when $Y_i = \gamma(X_i, R_i)$ is received at the fusion center, given that $R_i = r$. Formally, this is the Radon–Nikodym derivative of the measures $\nu'_0(\cdot | r)$ and $\nu'_1(\cdot | r)$ on the set \mathcal{Y} , where $\nu'_j(\cdot | r)$, is the measure $\nu_j(\cdot | r)$ restricted to the σ -algebra generated by $Y_i = \gamma(X_i, r)$.

Let us fix a strategy $((\xi_{i,v}, \gamma_{i,v})_{v \in \mathcal{V}})_{1 \leq i \leq n}$, and recall that $(\xi_{i,V}, \gamma_{i,V})$ is the resulting pure policy of sensor i , as determined by V . With the above introduced notation, the likelihood ratio calculated at the fusion center, on the basis of the available information $(R^{(n)}, Y^{(n)}, V)$, is

$$\prod_{i: \xi_{i,V}(R^{(n)})=1} \ell_{01}^{\gamma_{i,V}}(Y_i | R_i) \prod_{i=1}^n \frac{d\mu_0}{d\mu_1}(R_i). \quad (2)$$

For convenience, we define the random variables Z_i and S_n by

$$Z_i = -\xi_{i,V}(R^{(n)}) \log \ell_{01}^{\gamma_{i,V}}(Y_i | R_i) - \log \frac{d\mu_0}{d\mu_1}(R_i) \quad (3)$$

and

$$S_n = \sum_{i=1}^n Z_i \quad (4)$$

so that S_n is the negative of the log-likelihood ratio at the fusion center.

The amount of relevant information contained in Y_i , given that $R_i = r$ and that sensor i employs a pure transmission policy $\gamma \in \Gamma$, is quantified by the Kullback–Leibler (KL) divergence, defined by

$$I(r, \gamma) = \mathbb{E}_0 [\log \ell'_{01}(Y_i | r) | R_i = r].$$

Our last assumption is introduced for the same technical reasons as in [5].

Assumption 2: We have

$$\begin{aligned} \mathbb{E}_0 \left[\log^2 \frac{d\mu_0}{d\mu_1}(R_1) \right] &< \infty \\ \mathbb{E}_0 [\log^2 \ell_{01}(X_1 | R_1)] &< \infty. \end{aligned}$$

We record a consequence of Assumption 2.

Lemma 1: We have $\mathbb{E}_0 [\log(d\mu_0/d\mu_1)] < \infty$. Furthermore, for every $\gamma \in \Gamma$, we have $0 \leq I(R_1, \gamma) < \infty$, \mathbb{P}_0 -a.s., and

$$\mathbb{E}_0 [\log^2 \ell'_{01}(Y_1 | R_1) | R_1] \leq a(R_1), \quad \mathbb{P}_0\text{-a.s.},$$

for some function $a(\cdot)$ that satisfies $\mathbb{E}_0[a(R_1)] < \infty$.

Proof: This follows from Jensen's inequality, as in the proof of Proposition 3 in [5].¹ \square

IV. THE I.I.D. CASE

In this section, we characterize the optimal exponent for the Neyman–Pearson problem. Furthermore, we show that the optimal exponent does not change when we restrict to homogeneous local strategies. Throughout this section, we assume that under either hypothesis, the random variables R_i are i.i.d.

According to Stein's lemma [7], in the absence of censoring or side information, and if all sensors use the same pure transmission policy, the error exponent is the negative of the associated KL divergence. By a similar argument, if the sensors use a common local policy $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0} \in \Pi(c)$, we expect to obtain an error exponent equal to

$$\lambda(\pi) = -\mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] - \mathbb{E}_0 [\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})].$$

It is then natural to optimize over all admissible local policies $\pi \in \Pi(c)$, and define

$$\lambda^*(c) = -\mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] - \sup_{\pi \in \Pi(c)} \mathbb{E}_0 [\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})]$$

where the optimization includes the choice of the local randomization variable V_1 and its distribution.

¹In reference to that proof, the argument needs to be carried out by using the function $t \mapsto (t \log^2 t) \mathbf{1}(t \geq 1)$, which is convex for $t \geq 0$, together with the fact that $t \log^2 t < 1$ when $t \in [0, 1)$.

We will show that $\lambda^*(c)$ is the optimal error exponent, even if we remove the restriction to homogeneous local strategies, in the limit as α goes to zero. In deriving the required lower bound, we will not be able to invoke standard results from Large Deviations Theory, because the summands in the log-likelihood ratio are all affected by the overall side information $R^{(n)}$, and are not independent. For this reason, the proof of the lower bound will proceed from first principles.

Our main result is as follows.

Theorem 1: Suppose that Assumptions 1 and 2 hold, and that the random variables R_i are i.i.d., under either hypothesis.

- i) For every $\alpha \in (0, 1)$ and $c > 0$, the optimal error exponent $\liminf_{n \rightarrow \infty} (1/n) \log \beta_n^*(c, \alpha)$ is bounded below by $\lambda^*(c)/(1 - \alpha)$.
- ii) For every $\alpha \in (0, 1)$ and $c > 0$, there exists a sequence of admissible homogeneous local strategies (one for each n) that satisfy the Type I error constraint, and such that the corresponding Type II error probabilities β_n satisfy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n = \lambda^*(c).$$

- iii) For every $c > 0$,

$$\lim_{\alpha \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(c, \alpha) = \lambda^*(c).$$

Furthermore, if the random variables R_i are stationary and ergodic, and $\mu_0 = \mu_1$, then iii) still holds.

We record an elementary fact that will be used later.

Lemma 2: The function $\lambda^* : [0, \infty) \mapsto (-\infty, 0]$ is convex, and in particular, continuous on $(0, \infty)$.

Proof: Suppose that $c = \delta c_1 + (1 - \delta)c_2$ for some $\delta \in [0, 1]$. Fix some $\epsilon > 0$ and consider two local policies $\pi^k \in \Pi(c_k)$, $k = 1, 2$ (so that $\rho(\pi^k) \leq c_k$), which satisfy

$$\lambda(\pi^k) \leq \lambda^*(c_k) + \epsilon.$$

Consider a new local policy π^0 that uses π^1 with probability δ , and π^2 with probability $1 - \delta$. We then have

$$\rho(\pi^0) = \delta \rho(\pi^1) + (1 - \delta) \rho(\pi^2) \leq \delta c_1 + (1 - \delta)c_2 = c$$

so that $\pi^0 \in \Pi(c)$. Furthermore

$$\begin{aligned} \lambda^*(c) &\leq \lambda(\pi^0) = \delta \lambda(\pi^1) + (1 - \delta) \lambda(\pi^2) \\ &\leq \delta \lambda^*(c_1) + (1 - \delta) \lambda^*(c_2) + \epsilon. \end{aligned}$$

The result follows by letting ϵ decrease to zero. \square

A. Proof of the Lower Bound

In this subsection, we prove the first part of Theorem 1. Throughout this subsection, α is held at a fixed value.

Suppose that a strategy $\pi^{(n)}$ has been fixed. Since we are interested in a lower bound on the resulting error exponent, we assume that the fusion center uses the best possible fusion rule. As the fusion center is faced with a classical Neyman–Pearson problem, where the information available is

$(Y^{(n)}, R^{(n)}, V, V')$, a corresponding optimal fusion rule is a likelihood ratio test of the form

$$\hat{H}_n = 0 \text{ if and only if } \frac{S_n}{n} \leq T_n$$

where T_n is a possibly randomized threshold (determined by V'). It is convenient to consider the expected value of the log-likelihood ratio, given that the side information has been revealed and the randomization variable has been realized. We thus define

$$\begin{aligned} \Lambda_n &= \frac{1}{n} \mathbb{E}_0 \left[S_n \mid R^{(n)}, V \right] \\ &= -\frac{1}{n} \sum_{i=1}^n \log \frac{d\mu_0}{d\mu_1}(R_i) - \frac{1}{n} \sum_{i=1}^n \xi_{i,V}(R^{(n)}) I(R_i, \gamma_{i,V}) \end{aligned}$$

where the second equality follows from (3)–(4) and the definition of $I(r, \gamma)$. We start by showing that S_n/n is asymptotically close (in probability) to Λ_n .

Lemma 3: For every $\eta > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}_0 \left(\left| \frac{S_n}{n} - \Lambda_n \right| < \eta \right) = 1$.

Proof: We condition on $R^{(n)}$ and V . Then $S_n - n\Lambda_n$ becomes a sum of (conditionally) independent random variables, each having (conditional) variance bounded by $a(R_i)$ (cf. Lemma 1). Chebychev's inequality yields

$$\begin{aligned} &\mathbb{P}_0 \left(\left| \frac{S_n}{n} - \Lambda_n \right| \geq \eta \mid R^{(n)}, V \right) \\ &\leq \frac{1}{n^2 \eta^2} \mathbb{E}_0 \left[(S_n - n\Lambda_n)^2 \mid R^{(n)}, V \right] \\ &\leq \frac{1}{n^2 \eta^2} \mathbb{E}_0 \left[\sum_{i=1}^n \log^2 \ell_{01}^{i,V}(Y_i \mid R_i) \mid R^{(n)}, V \right] \\ &\leq \frac{1}{n^2 \eta^2} \sum_{i=1}^n a(R_i). \end{aligned} \quad (5)$$

Taking unconditional expectations of both sides, we obtain

$$\mathbb{P}_0 \left(\left| \frac{S_n}{n} - \Lambda_n \right| \geq \eta \right) \leq \frac{1}{n \eta^2} \mathbb{E}_0[a(R_1)]$$

which converges to zero because $\mathbb{E}_0[a(R_1)] < \infty$. \square

The next lemma is crucial in that it relates the amount of information provided by an admissible strategy to the best possible exponent $\lambda^*(c)$ under local admissible strategies. The key idea is that as far as sensor i is concerned, the side information at the other sensors has the same effect as using additional local randomization variables.

Lemma 4: For any sequence of admissible strategies (one for each n), and for every sequence of measurable subsets A_n of Ω , we have

$$\liminf_{n \rightarrow \infty} \mathbb{E}_0[\Lambda_n \mathbf{1}_{A_n}] \geq \lambda^*(c).$$

Proof: Suppose that a sequence of admissible strategies $\pi^{(n)}$ has been fixed, and let $(\xi_{i,v}, \gamma_{i,v})_{v \in \mathcal{V}}$ be the policy of sensor i . We have

$$\begin{aligned} \mathbb{E}_0[\Lambda_n \mathbf{1}_{A_n}] &\geq -\mathbb{E}_0 \left[\frac{1}{n} \sum_{i=1}^n \log \frac{d\mu_0}{d\mu_1}(R_i) \mathbf{1}_{A_n} \right] \\ &\quad - \frac{1}{n} \sum_{i=1}^n \mathbb{E}_0 \left[\xi_{i,V}(R^{(n)}) I(R_i, \gamma_{i,V}) \right]. \end{aligned} \quad (6)$$

To bound the first term, we have

$$\begin{aligned}
& \mathbb{E}_0 \left[\frac{1}{n} \sum_{i=1}^n \log \frac{d\mu_0}{d\mu_1}(R_i) \mathbf{1}_{A_n} \right] \\
&= \mathbb{E}_0 \left[\left(\frac{1}{n} \sum_{i=1}^n \log \frac{d\mu_0}{d\mu_1}(R_i) - \mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] \right) \mathbf{1}_{A_n} \right] \\
&\quad + \mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] \mathbb{P}_0(A_n) \\
&\leq \mathbb{E}_0 \left[\left| \frac{1}{n} \sum_{i=1}^n \log \frac{d\mu_0}{d\mu_1}(R_i) - \mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] \right| \right] \\
&\quad + \mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right]. \tag{7}
\end{aligned}$$

In the limit as $n \rightarrow \infty$, the first term in the right-hand side (RHS) of (7) converges to 0 (because the R_i are i.i.d. and the L^1 ergodic theorem applies), which leaves the term $\mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right]$.

For any $v \in \mathcal{V}$, let $\tilde{\xi}_{i,v}(R_i) = \mathbb{E}_0[\xi_{i,v}(R^{(n)}) | R_i]$. Then

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \mathbb{E}_0 \left[\xi_{i,v}(R^{(n)}) I(R_i, \gamma_{i,v}) \right] \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_0 \left[\tilde{\xi}_{i,v}(R_i) I(R_i, \gamma_{i,v}) \right] \\
&= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_0 \left[\tilde{\xi}_{i,v}(R_1) I(R_1, \gamma_{i,v}) \right] \tag{8}
\end{aligned}$$

$$\leq \sup_{\pi \in \Pi(c)} \mathbb{E}_0 \left[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1}) \right]. \tag{9}$$

The equality in (8) follows from the stationarity of $R^{(n)}$. The inequality in (9) is obtained by considering a local policy $(\tau_{\bar{v}}, \delta_{\bar{v}})_{\bar{v} \in \bar{\mathcal{V}}}$, where $\bar{\mathcal{V}} = \{1, \dots, n\} \times [0, 1] \times \mathcal{V}$, defined as follows. Let $\bar{V} = (J, U, V)$, where J is uniform on $\{1, 2, \dots, n\}$, U is uniform on $[0, 1]$, and J, U , and V are independent. Let $\delta_{\bar{v}} = \gamma_{J,V}$, and for every $r \in \mathcal{R}$, let

$$\tau_{\bar{v}}(r) = \begin{cases} 1, & \text{if } U \leq \tilde{\xi}_{J,v}(r) \\ 0, & \text{otherwise.} \end{cases}$$

In particular, if $J = i$, $U = u$, and $V = v$, the new local policy $(\tau_{\bar{v}}, \delta_{\bar{v}})_{\bar{v} \in \bar{\mathcal{V}}}$ applies the pure transmission policy $\gamma_{i,v}$, and censors if $u > \tilde{\xi}_{i,v}(r)$, when the local side information is r . Then, $(\tau_{\bar{v}}, \delta_{\bar{v}})_{\bar{v} \in \bar{\mathcal{V}}} \in \Pi(c)$ and the RHS of (8) is equal to $\mathbb{E}_0[\tau_{\bar{v}}(R_1) I(R_1, \delta_{\bar{v}})]$. Hence, (9) follows.

Combining the above with (7), we obtain

$$\begin{aligned}
\liminf_{n \rightarrow \infty} \mathbb{E}_0[\Lambda_n \mathbf{1}_{A_n}] &\geq - \mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] \\
&\quad - \sup_{\pi \in \Pi(c)} \mathbb{E}_0[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})] \\
&= \lambda^*(c)
\end{aligned}$$

and the lemma is proved. \square

Lemma 5: For all $\alpha \in (0, 1)$ and $c \geq 0$, we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(c, \alpha) \geq \frac{\lambda^*(c)}{1 - \alpha}.$$

Proof: Fix some $\eta > 0$. For every n , consider an admissible strategy $\pi^{(n)}$ and a fusion rule with $\mathbb{P}_0(\hat{H}_n = 0) \geq 1 - \alpha$. We use a change of measure argument, similar to that in the proof of Stein's lemma in [7], to get

$$\begin{aligned}
\beta_n &= \mathbb{P}_1 \left(\frac{S_n}{n} \leq T_n \right) \\
&= \mathbb{E}_0 \left[e^{S_n} \mathbf{1} \left(\frac{S_n}{n} \leq T_n \right) \right] \\
&\geq \mathbb{E}_0 \left[e^{S_n} \mathbf{1} \left(\Lambda_n - \eta < \frac{S_n}{n} \leq T_n \right) \right] \\
&\geq \mathbb{E}_0 \left[e^{n(\Lambda_n - \eta)} \mathbf{1} \left(\Lambda_n - \eta < \frac{S_n}{n} \leq T_n \right) \right]. \tag{10}
\end{aligned}$$

Let $A_n = \{\omega : \Lambda_n - \eta < \frac{S_n}{n} \leq T_n\}$. Then

$$\begin{aligned}
& \liminf_{n \rightarrow \infty} \mathbb{P}_0(A_n) \\
&\geq \liminf_{n \rightarrow \infty} \left(\mathbb{P}_0 \left(\frac{S_n}{n} \leq T_n \right) + \mathbb{P}_0 \left(\frac{S_n}{n} > \Lambda_n - \eta \right) - 1 \right) \\
&\geq 1 - \alpha > 0
\end{aligned}$$

where we have made use of Lemma 3. Hence, for sufficiently large n , we can condition on A_n and obtain

$$\begin{aligned}
\frac{1}{n} \log \beta_n &\geq \frac{1}{n} \log \mathbb{E}_0 \left[e^{n(\Lambda_n - \eta)} \mathbf{1}_{A_n} \right] \\
&= \frac{1}{n} \log \mathbb{P}_0(A_n) + \frac{1}{n} \log \mathbb{E}_0 \left[e^{n(\Lambda_n - \eta)} \mid A_n \right] \\
&\geq \frac{1}{n} \log \mathbb{P}_0(A_n) + \mathbb{E}_0[\Lambda_n | A_n] - \eta \tag{11}
\end{aligned}$$

where the last step follows from Jensen's inequality. Applying Lemma 4, we have

$$\begin{aligned}
\liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^* &\geq \liminf_{n \rightarrow \infty} \frac{\mathbb{E}_0[\Lambda_n \mathbf{1}_{A_n}]}{\mathbb{P}_0(A_n)} - \eta \\
&\geq \frac{\lambda^*(c)}{1 - \alpha} - \eta.
\end{aligned}$$

The result follows by letting $\eta \rightarrow 0$. \square

Lemma 5 concludes the proof of the lower bound [part i) of Theorem 1].

B. Proof of the Upper Bound

In this subsection, we construct a sequence of admissible homogeneous local strategies (one for each n), under which the lower bound is asymptotically attained.

For each n , consider a strategy involving a common local policy $\pi^n = (\xi_v^n, \gamma_v^n)_{v \in \mathcal{V}_0} \in \Pi(c)$, such that

$$\lim_{n \rightarrow \infty} \lambda(\pi^n) = \lambda^*(c). \tag{12}$$

Consider the fusion rule $\hat{H}_n = \phi_n(Y^{(n)}, R^{(n)}, V)$ that selects H_0 if and only if

$$\frac{S_n}{n} \leq \lambda(\pi^n) + n^{-\frac{1}{4}}.$$

Let

$$\mathcal{S}(R^{(n)}, V) = \left\{ Y^{(n)} \mid \phi_n(Y^{(n)}, R^{(n)}, V) = 0 \right\}. \tag{13}$$

Since the random variables Z_i are i.i.d., with mean $\lambda(\pi^n)$ and variance bounded by some constant a (cf. Assumption 2), we have

$$\frac{1}{n^2} \mathbb{E}_0 \left[(S_n - n\lambda(\pi^n))^2 \right] \leq \frac{a}{n}$$

and hence

$$\mathbb{P}_0 \left(\left| \frac{S_n}{n} - \lambda(\pi^n) \right| \leq n^{-\frac{1}{4}} \right) \rightarrow 1. \quad (14)$$

This implies that $\mathbb{P}_0(\hat{H}_n = 1) \rightarrow 0$ as $n \rightarrow \infty$. It follows that for any given $\alpha > 0$, the constraint $\mathbb{P}_0(\hat{H}_n = 1) \leq \alpha$ will be satisfied for large n . We let $\tilde{\beta}_n$ be the minimum possible Type II error (over all fusion rules), when we use the common local policy π^n at all sensors. In particular, $\tilde{\beta}_n \leq \mathbb{P}_1(\hat{H}_n = 0)$.

The next lemma is a modification of Stein's lemma. The proof is almost the same as the standard proof (see [8]) but we include it for completeness, and because we want to check that it remains valid in the case where $\mu_0 = \mu_1$ and the variables R_i are stationary and ergodic (as opposed to i.i.d.), as will be discussed in Section VI.

Lemma 6: For the above defined common local policy π^n , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\beta}_n = \lambda^*(c).$$

Proof: We have

$$\begin{aligned} & \mathbb{P}_1(\hat{H}_n = 0 \mid R^{(n)}, V) \\ &= \int_{\mathcal{S}(R^{(n)}, V)} \prod_{i: \xi_{V_i}^n(R_i)=1} \nu_1^{\gamma_{V_i}^n}(dy_i \mid R_i) \\ &= \int_{\mathcal{S}(R^{(n)}, V)} \left\{ e^{-\sum_{i=1}^n \xi_{V_i}^n(R_i) \log \ell_{01}^{\gamma_{V_i}^n}(y_i \mid R_i)} \right. \\ & \quad \left. \cdot \prod_{i: \xi_{V_i}^n(R_i)=1} \nu_0^{\gamma_{V_i}^n}(dy_i \mid R_i) \right\}. \end{aligned}$$

Hence

$$\begin{aligned} \tilde{\beta}_n &\leq \mathbb{P}_1(\hat{H}_n = 0) \\ &= \mathbb{E}_1 \left[\mathbb{P}_1(\hat{H}_n = 0 \mid R^{(n)}, V) \right] \\ &= \mathbb{E}_0 \left[\mathbb{P}_1(\hat{H}_n = 0 \mid R^{(n)}, V) \exp \left(-\sum_{i=1}^n \log \frac{d\mu_0}{d\mu_1}(R_i) \right) \right] \\ &= \mathbb{E}_0 \left[\int_{\mathcal{S}(R^{(n)}, V)} e^{S_n} \prod_{i: \xi_{V_i}^n(R_i)=1} \nu_0^{\gamma_{V_i}^n}(dy_i \mid R_i) \right]. \end{aligned}$$

Recall that on the set $\mathcal{S}(R^{(n)}, V)$, we have $S_n \leq n(\lambda(\pi^n) + n^{-1/4})$, which yields

$$\begin{aligned} \tilde{\beta}_n &\leq \exp \left(n(\lambda(\pi^n) + n^{-\frac{1}{4}}) \right) \mathbb{E}_0[\mathbb{P}_0(\hat{H}_n = 0 \mid R^{(n)}, V)] \\ &\leq \exp \left(n(\lambda(\pi^n) + n^{-\frac{1}{4}}) \right). \end{aligned}$$

Thus

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\beta}_n &\leq \lim_{n \rightarrow \infty} \left(\lambda(\pi^n) + n^{-\frac{1}{4}} \right) \\ &= \lambda^*(c). \end{aligned} \quad (15)$$

To show the lower bound, we mimic the proof of Lemma 5, with $\lambda(\pi^n)$ replacing Λ_n in that proof. Then from (10), with T_n as the threshold for the optimal Neyman–Pearson test, we obtain

$$\begin{aligned} \tilde{\beta}_n &\geq e^{n(\lambda(\pi^n) - n^{-1/4})} \mathbb{P}_0 \left(\lambda(\pi^n) - n^{-\frac{1}{4}} < \frac{S_n}{n} \leq T_n \right) \\ &\geq e^{n(\lambda(\pi^n) - n^{-1/4})} \\ &\quad \times \left(1 - \alpha + \mathbb{P}_0 \left(\frac{S_n}{n} > \lambda(\pi^n) - n^{-\frac{1}{4}} \right) - 1 \right). \end{aligned}$$

Taking logarithms, dividing by n , and then letting $n \rightarrow \infty$, and using (14)

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\beta}_n \geq \lambda^*(c). \quad (16)$$

The lemma is proved. \square

This concludes the proof of the upper bound [part ii) of Theorem 1]. Part iii) is an immediate consequence of parts i) and ii). The last part, involving a stationary and ergodic sequence of random variables R_i will be discussed in Section VI.

C. The Role of Randomization

Suppose that the cost function is independent of the transmission function, i.e., $\rho(r, \gamma) = \tilde{\rho}(r)$ for some nonnegative function $\tilde{\rho}$. Suppose also that the set of pure transmission policies Γ is of the form $\Gamma = \prod_{r \in \mathcal{R}} \Gamma(r)$, where $\Gamma(r)$ is a set of allowed pure transmission policies $\gamma(\cdot, r)$, when the side information takes the value r . We interpret this as each sensor i being able to choose its own transmission policy separately for each possible value of the side information R_i . Then, finding an asymptotically optimal strategy is simplified because

$$\begin{aligned} \lambda^*(c) &= -\mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] \\ &\quad - \sup \mathbb{E}_0 \left[\xi_{V_1}(R_1) \sup_{\gamma \in \Gamma(R_1)} I(R_1, \gamma) \right] \end{aligned}$$

where the first supremum is taken over all local censoring policies $(\xi_v)_{v \in \mathcal{V}_0}$ that satisfy

$$\mathbb{E}_0[\xi_{V_1}(R_1) \tilde{\rho}(R_1)] \leq c.$$

In particular, a pure transmission policy γ can be used. In achieving $\lambda^*(c)$, we choose transmission policies that maximize the KL divergence between the distributions of the messages Y_i , separately for each possible value of R_i . It is intuitively clear that this has to be the case, since the KL divergence quantifies the discrimination between two distributions.

Randomization goes a long way to simplify the form of an optimal policy, as can be seen in the proof of Lemma 4. If we are restricted to pure strategies, then homogeneous local strategies need not be asymptotically optimal. Indeed, suppose that \mathcal{R} and Γ are finite, so that we have to choose from a finite set of pure

local policies $\{(\xi^k, \gamma^k) : k = 1, \dots, K\}$. We are then faced with the optimization problem

$$\begin{aligned} \max_{x_1, \dots, x_K} \quad & \sum_{k=1}^K x_k \mathbb{E}_0 [\xi^k(R_1) I(R_1, \gamma^k)] \\ \text{s.t.} \quad & \sum_{k=1}^K x_k \mathbb{E}_0 [\xi^k(R_1) \rho(R_1, \gamma^k)] \leq c \\ & \sum_{k=1}^K x_k = 1 \\ & x_k \geq 0, \quad \forall i \end{aligned}$$

where x_k is the proportion of sensors that use the pure policy (ξ^k, γ^k) . This is a linear program with two constraints, and generically, an optimal solution will have two nonzero variables. Let x^* and $1 - x^*$ be the optimal values of these two variables, and assume (to simplify the discussion) that x^*n is integer. As long as we are restricted to pure local strategies, then we have to divide the sensors into two groups, one group consisting of x^*n sensors that use one local policy and another group consisting of $(1 - x^*)n$ sensors that use another local policy. Thus, in the Neyman–Pearson problem randomization and nonhomogeneity are alternatives to each other.

V. DISCUSSION OF THE NEYMAN–PEARSON RESULTS

In this section, we discuss some variations, extensions and generalizations of the results in Section IV.

A. Cooperation Can Improve Performance

We observe that if $\alpha \in (0, 1)$ is held fixed, there is a gap between the upper and lower bounds in Theorem 1. We discuss here the extent to which cooperation among the sensors improves detection performance.

If we do not allow cooperation among the sensors, i.e., if $\pi^{(n)}$ is local, we can use an argument similar to the proof of Lemma 6 to show that $\lambda^*(c)$ is a lower bound, for any fixed $\alpha \in (0, 1)$. In particular, homogeneous local strategies are asymptotically optimal.

If cooperation is allowed, and α is fixed, the optimal exponent can be less than $\lambda^*(c)$, as shown in Proposition 1 below. In that case, asymptotically optimal strategies are difficult to find, and we do not believe that a simple closed-form expression for the optimal error exponent is possible.

Proposition 1: For a fixed $\alpha \in (0, 1)$, we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(c, \alpha) \leq \lambda^* \left(\frac{c}{1 - \alpha} \right).$$

Proof: Let us fix n and α , and some $\epsilon \in (0, \alpha)$. Consider the following strategy. With probability $p = \alpha - \epsilon$, we use the censoring policy $\xi_i \equiv 0$ for all i , and always declare H_1 . In this case, we satisfy the resource constraint (1) with c replaced by $c_1 = 0$, and have a Type I error probability of $\alpha_1 = 1$. Let

$$\alpha_2 = \frac{\epsilon}{1 - \alpha + \epsilon}, \quad c_2 = \frac{c}{1 - \alpha + \epsilon}.$$

With probability $1 - p$, we use a homogeneous local strategy involving a common local policy $\pi \in \Pi(c_2)$ that satisfies

$$\lambda(\pi) \leq \lambda^*(c_2) + \frac{1}{n},$$

and a fusion rule that achieves a Type I error probability of α_2 . Note that $\alpha = p\alpha_1 + (1 - p)\alpha_2$ and $c = pc_1 + (1 - p)c_2$, so that the composite strategy we have constructed is admissible. Using Lemma 6, its Type II error probability β_n satisfies

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(c, \alpha) & \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n \\ & = \lambda^*(c_2) = \lambda^* \left(\frac{c}{1 - \alpha + \epsilon} \right). \end{aligned}$$

Taking $\epsilon \rightarrow 0$ and using the continuity of $\lambda^*(\cdot)$, we obtain the desired result. \square

From Proposition 1, the improvement in the error exponent when using cooperation instead of using homogeneous, local strategies has a magnitude of at least

$$\lambda^*(c) - \lambda^* \left(\frac{c}{1 - \alpha} \right),$$

which is strictly positive when c is small enough, and is upper-bounded by (cf. Lemma 5)

$$\left(1 - \frac{1}{1 - \alpha} \right) \lambda^*(c) = \frac{\alpha}{1 - \alpha} |\lambda^*(c)|.$$

We see that in a severely constrained network (small c), the price paid for not cooperating is positive but not very large. Thus, the communications overhead and resulting complexity may not justify the use of cooperative censoring.

B. Generalized Sensor Policies

In this subsection, we provide a generalization of our framework, by allowing a more general class of policies. In the preceding, each sensor could choose separately a censoring policy and a transmission policy. Here, these two choices will be subsumed under a single choice of a “generalized policy” $\bar{\zeta}_i$. We will see that when specialized to our earlier setting, the generalized formulation will also allow the choice of transmission functions to be made cooperatively, on the basis of the global side-information vector $R^{(n)}$.

Formally, we define a (*generalized*) *policy* as a function $\bar{\zeta} : \mathcal{X} \times \mathcal{R}^{(n)} \times \mathcal{V} \mapsto \mathcal{Y}$. A sensor that uses policy $\bar{\zeta}$, transmits $Y_i = \bar{\zeta}(X_i, R^{(n)}, V)$ to the fusion center. Assuming independent randomization, the notion of a local policy for sensor i is defined as before, namely, the dependence on $(R^{(n)}, V^{(n)})$ is only through (R_i, V_i) . Once more, a local strategy is called homogeneous if every sensor uses the same mapping from (X_i, R_i, V_i) to Y_i .

Let $r^{(n,i)}$ be the vector $r^{(n)}$ after removing the i th component. As before, for every $r^{(n,i)} \in \mathcal{R}^{n-1}$ and for every $v \in \mathcal{V}$, we require that the function $\zeta_{r^{(n,i)},v}$, defined by $\zeta_{r^{(n,i)},v}(x, r_i) = \bar{\zeta}(x, (r^{(n,i)}, r_i), v)$ be in a given set Γ of functions from $\mathcal{X} \times \mathcal{R}$ to \mathcal{Y} . The function $\zeta_{r^{(n,i)},v}$ is called a *pure* local policy. A generalized policy $\bar{\zeta}$ can be viewed as a random choice of a *pure* local policy $\zeta_{R^{(n,i)},V}$, based on the value of $R^{(n,i)}$ and V . We assume that every pure local policy ζ consumes an amount $\tilde{\rho}(\zeta)$ of a certain resource. Given a generalized policy $\bar{\zeta}$ for sensor i , the policy of that sensor is chosen to be $\zeta_{R^{(n,i)},V}$. The cost is defined to be $\mathbb{E}_0 [\tilde{\rho}(\zeta_{R^{(n,i)},V})]$, where the expectation

is taken over $R^{(n,i)}$ and V . Similar to (1), we are interested in *admissible* strategies $(\bar{\zeta}^1, \dots, \bar{\zeta}^n)$, that satisfy the constraint $\sum_{i=1}^n \mathbb{E}_0 \left[\tilde{\rho}(\zeta_{R^{(n,i)}, V}^i) \right] \leq nc$.

Example 5 (Censoring): Consider the setting of Section II, and assume without loss of generality, that there is a special element y^* of \mathcal{Y} that is never used, i.e., $\gamma(x, r) \neq y^*$, for every $x \in \mathcal{X}$, $r \in \mathcal{R}$, and $\gamma \in \Gamma$. Given a pure local policy (ξ, γ) , we will represent a sensor i that decides to censor ($\xi_i(R_i) = 0$) as one that transmits a “dummy” message equal to y^* . Such a dummy message carries the same information to the fusion center as censoring (the absence of a message). We let $\zeta(x, r) = y^*$ whenever $\xi(r) = 0$, and $\zeta(x, r) = \gamma(x, r)$ otherwise. If $\rho(\cdot, \cdot)$ is the resource function used in our earlier formulation, it is natural to define $\tilde{\rho}(\zeta) = \mathbb{E}_0[\xi(R_1)\rho(R_1, \gamma)]$. \square

Example 6 (Power Constraints): Suppose that $\mathcal{Y} = \mathbb{R}$, and that the cost of a pure local policy ζ at sensor i is $\tilde{\rho}(\zeta) = \mathbb{E}_0[\zeta(X_i, R_i)]^2$. Then, $\tilde{\rho}(\zeta)$ corresponds to the expected power consumed by ζ . In this setting, a message with value equal to zero can also be viewed as a censoring decision. \square

When specialized to the censoring problem of earlier sections, the main difference introduced by the current framework is the following: the transmission function γ_i used by sensor i can now be chosen on the basis of not only the randomization variable V , but also the side information $R^{(n,i)}$ at the other sensors.

For any pure local policy ζ for sensor 1, let $I(r, \zeta)$ be the KL divergence associated with the measurement $Y_1 = \zeta(X_1, R_1)$. For a randomized local policy $\bar{\zeta} : \mathcal{X} \times \mathcal{R} \times \mathcal{V}_0 \mapsto \mathcal{Y}$ for sensor 1, let ζ_{V_1} be a random variable whose realized value is the pure local policy $\zeta_v(\cdot, \cdot) = \bar{\zeta}(\cdot, \cdot, v)$ whenever $V_1 = v$. We have the following generalization of Theorem 1, where $\beta_n^*(c, \alpha)$ is, as before, the optimal error probability.

Theorem 2: Suppose that Assumptions 1 and 2 hold, and that the random variables R_i are i.i.d. under either hypothesis. Then, for every $c > 0$

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(c, \alpha) \\ = -\mathbb{E}_0 \left[\log \frac{d\mu_0}{d\mu_1}(R_1) \right] - \sup \mathbb{E}_0 [I(R_1, \zeta_{V_1})] \end{aligned}$$

where the supremum is taken over all local policies $\bar{\zeta}$ for sensor 1 that satisfy $\mathbb{E}_0[\tilde{\rho}(\bar{\zeta})] \leq c$. Furthermore, there exists a sequence of homogeneous local strategies that asymptotically achieves the optimal error exponent.

The proof of Theorem 2 is similar to the proof of Theorem 1. The main difference is that we need to replace the transmission policies $\gamma_{i,V}$ with generalized policies $\bar{\zeta}_i$, and eliminate the censoring policies. However, with generalized policies, an extension to the case where the R_i are stationary and ergodic is not apparent (in contrast to the results of Section VI).

C. Unknown Side Information

So far, we have been assuming that even in the case of no cooperation (local strategies), the fusion center has access to the entire vector $R^{(n)}$. We will now consider the case of no

cooperation when the fusion center does not have access to $R^{(n)}$. Thus, the only information available at the fusion center is $V^{(n)} = (V_1, \dots, V_n)$, the identity of the sensors that are censoring, and the messages of the sensors that do not censor. (Note that just the act of censoring provides some information to the fusion center.) Reference [1] considers a setting in which (when translated to our framework) we have R_i equal to the local-likelihood ratio $l_i(X_i)$ of the measurement $\rho(r, \gamma) \equiv 1$, and the transmission policy is $\gamma_i(X_i, R_i) = X_i$. Reference [1] shows that for any fixed n , it is optimal to choose the censoring regions to be intervals, i.e., if $l_i(X_i)$ falls within some interval $(t_{i,1}, t_{i,2})$, then the sensor does not send its measurement to the fusion center. Note that [1] assumes only that the measurements X_i are independent, but even when they have identical distributions, each sensor uses a different censoring interval. Optimizing over $(t_{i,1}, t_{i,2})$, for all i , can be a daunting task even if the number of sensors n is moderate [1]. Hence, it is of interest to examine whether the problem simplifies when the variables R_i are i.i.d. and n is large.

From our discussion in Section V-A, we expect that homogeneous local censoring strategies are asymptotically optimal. This is indeed the case if we assume that the fusion center knows each sensor's policy. For example, $V^{(n)}$ can be determined beforehand, and made known at every sensor and the fusion center, while the fusion center has a table of all the censoring policies employed by the sensors. For $j = 0, 1$, let v_j^γ be the distribution of $Y_1 = \gamma(X_1, R_1)$, under hypothesis H_j . Let

$$\tilde{I}(\gamma) = \mathbb{E}_0 \left[\log \frac{dv_0^\gamma}{dv_1^\gamma} \right].$$

We have the following result.

Proposition 2: Suppose Assumptions 1 and 2 hold and that the random variables R_i are i.i.d., under either hypothesis. Suppose that the fusion center knows V but not $R^{(n)}$, and that we are restricted to local strategies. Then, homogeneous strategies are asymptotically optimal as $n \rightarrow \infty$, for every $\alpha \in (0, 1)$. Furthermore, the optimal exponent is equal to $\hat{\lambda}(c)$, defined by

$$\begin{aligned} \hat{\lambda}(c) = - \sup_{\pi \in \Pi(c)} \mathbb{E}_0 \left[\xi_{V_1}(R_1) \tilde{I}(\gamma_{V_1}) \right. \\ \left. + (1 - \xi_{V_1}(R_1)) \log \frac{\mathbb{P}_0(\xi_{V_1}(R_1) = 0 | V_1)}{\mathbb{P}_1(\xi_{V_1}(R_1) = 0 | V_1)} \right]. \end{aligned}$$

Proof: (Outline) We first note that using Assumption 2, and an argument as in Lemma 1, there exists $b \in (0, \infty)$, such that for all $\gamma \in \Gamma$ we have $\mathbb{E}_0 \left[\log^2 \frac{dv_0^\gamma}{dv_1^\gamma} \right] \leq b$.

In the current setting, a censoring decision can be viewed as a transmission of a special symbol to the fusion center. We redefine Z_i so that

$$\begin{aligned} Z_i = -\xi_{i,V_i}(R_i) \log \frac{dv_0^\gamma}{dv_1^\gamma}(Y_i) \\ - (1 - \xi_{i,V_i}(R_i)) \log \frac{\mathbb{P}_0(\xi_{i,V_i}(R_i) = 0 | V_i)}{\mathbb{P}_1(\xi_{i,V_i}(R_i) = 0 | V_i)}. \end{aligned}$$

We first check the inequality $\liminf_{n \rightarrow \infty} \mathbb{E}_0[\Lambda_n] \geq \hat{\lambda}(c)$, which is obtained as in Lemma 4. The rest of the proof proceeds as in Section IV. \square

VI. THE ERGODIC CASE

We now consider the case where $(R_i)_{i=1}^{\infty}$ is a stationary and ergodic sequence, and each R_i has the same distribution under either hypothesis. This case is of interest, because in many situations, the side information at the sensors is correlated. For example, in the sensor network described in Example 1 of Section I, if the sensors are geographically densely colocated, then we would expect the fading channels from the sensors to the fusion center to have correlated characteristics. Note also that in that example, the side information does not provide any information on the true hypothesis.

We now assume that $\mu_0 = \mu_1$.² We have the following result, which shows that cooperation is unnecessary in the asymptotic regime of large n and small α .

Theorem 3: Suppose $(R_i)_{i=1}^{\infty}$ is a stationary and ergodic sequence $\mu_0 = \mu_1$, and Assumptions 1 and 2 hold. Then

$$\lim_{\alpha \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^*(c, \alpha) = - \sup_{\pi \in \Pi(c)} \mathbb{E}_0[\xi_{V_1}(R_1)I(R_1, \gamma_{V_1})].$$

The proof of the above theorem is similar to that in Section IV. The proof of the lower bound in Section IV-A still holds. For the upper bound, we require the following result.

Lemma 7: Suppose that $(R_i)_{i=1}^{\infty}$ is a stationary ergodic sequence, and that Assumptions 1 and 2 hold. Then under hypothesis H_0 , for any homogeneous local strategy involving a common local policy π , we have $S_n/n \rightarrow \lambda(\pi)$ in probability.

Proof: We have

$$\begin{aligned} \frac{S_n}{n} &= \frac{1}{n} \sum_{i=1}^n Z_i \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_0[Z_i | R_i] + \frac{1}{n} \sum_{i=1}^n (Z_i - \mathbb{E}_0[Z_i | R_i]). \end{aligned} \quad (17)$$

Since the sequence $(R_i)_{i \geq 1}$ is stationary and ergodic, the first term $\frac{1}{n} \sum_{i=1}^n \mathbb{E}_0[Z_i | R_i]$ on the RHS of (17) converges in probability to $\lambda(\pi)$ (cf. Birkhoff's Ergodic Theorem [10]). For the second term, we have for each $\epsilon > 0$

$$\begin{aligned} &\mathbb{P}_0 \left(\left| \sum_{i=1}^n (Z_i - \mathbb{E}_0[Z_i | R_i]) \right| > n\epsilon \right) \\ &\leq \frac{1}{n^2 \epsilon^2} \text{var} \left(\sum_{i=1}^n (Z_i - \mathbb{E}_0[Z_i | R_i]) \right) \\ &= \frac{1}{n^2 \epsilon^2} \mathbb{E}_0 \left[\text{var} \left(\sum_{i=1}^n (Z_i - \mathbb{E}_0[Z_i | R_i]) \middle| R^{(n)} \right) \right] \end{aligned} \quad (18)$$

$$= \frac{1}{n^2 \epsilon^2} \mathbb{E}_0 \left[\sum_{i=1}^n \text{var} \left(Z_i - \mathbb{E}_0[Z_i | R_i] \middle| R^{(n)} \right) \right] \quad (19)$$

$$\begin{aligned} &\leq \frac{1}{n^2 \epsilon^2} \mathbb{E}_0 \left[\sum_{i=1}^n a(R_i) \right] \\ &= \frac{1}{n \epsilon^2} \mathbb{E}_0[a(R_1)] \end{aligned} \quad (20)$$

²One of the reasons for this assumption is that the asymptotic KL rate of the stochastic process $(R_i)_{i \geq 1}$ may not exist [9].

where (18) follows because $\mathbb{E}_0[Z_i | R_i] = \mathbb{E}_0[Z_i | R^{(n)}]$, and (19) follows because given $R^{(n)}$, the Z_i are independent. The last inequality (20) follows from Lemma 1. Therefore, as $n \rightarrow \infty$, the second term on the RHS of (17) converges in probability to 0, and the lemma is proved. \square

To complete the proof of Theorem 3, we proceed as in Section IV-B, except that we fix an $\epsilon > 0$ and consider a homogeneous local strategy involving a common local policy $\pi^\epsilon \in \Pi(c)$ that satisfies $\lambda(\pi^\epsilon) \leq \lambda^*(c) + \epsilon$. With this strategy, from Lemma 7, $S_n/n \rightarrow \lambda(\pi^\epsilon)$ in probability under hypothesis H_0 . Hence, we have the same result as (14) with π^ϵ replacing π^n , and η replacing $n^{-1/4}$, for some fixed $\eta > 0$. Corresponding changes are made in Lemma 6. The proof of Theorem 3 is now complete.

VII. OPTIMAL CENSORING

To find an optimal common local policy π , we need to maximize $f(\pi) = \mathbb{E}_0[\xi_{V_1}(R_1)I(R_1, \gamma_{V_1})]$ over all $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0} \in \Pi(c)$, i.e., over all π that satisfy $\rho(\pi) = \mathbb{E}_0[\xi_{V_1}(R_1)\rho(R_1, \gamma_{V_1})] \leq c$. We now show that it is sufficient to consider local policies that randomize between only two pure local policies. In particular, each sensor need only use an extra bit to communicate to the fusion center which policy it has chosen.

Suppose that a common local policy π has been fixed, including the range \mathcal{V}_0 of the randomization variable V_1 , except that the distribution μ of V_1 is left unspecified. Let π_v be the pure local policy obtained when $V_1 = v$. To optimize the distribution of V_1 , we have to maximize $\int f(\pi_v)\mu(dv)$, subject to $\int \rho(\pi_v)\mu(dv) \leq c$, over all measures μ on \mathcal{V}_0 . If \mathcal{V}_0 were finite, this would be a linear programming problem over the unit simplex, together with one additional constraint. As is well known, the optimum would be attained on an edge of the feasible set, that is, there would exist an optimal μ whose support consists of at most two points [11]. The lemma that follows states that the optimality of two-point distributions remains valid even when \mathcal{V}_0 is infinite (except that the optimum need not be attained), and establishes our claim that we only need to consider local policies that randomize between two pure local policies. (As in the rest of the paper, we omit the standard measurability conditions that are needed in this lemma.) The proof is provided in the Appendix for completeness.

Lemma 8: Let \mathcal{M} be the set of probability measures on a set \mathcal{V}_0 , and let f, g be given nonnegative functions from \mathcal{V}_0 into $[0, \infty)$. Then

$$\sup \left\{ \int f(v)\mu(dv) : \mu \in \mathcal{M}, \int g(v)\mu(dv) \leq c \right\} \quad (21)$$

$$= \sup \{ u f(v_1) + (1-u) f(v_2) : u \in [0, 1], v_1, v_2 \in \mathcal{V}_0, u g(v_1) + (1-u) g(v_2) \leq c \} \quad (22)$$

Furthermore, if the supremum in (21) is finite and is attained, then the supremum in (22) is also attained.

We close with a characterization of an optimal local censoring policy $(\xi_v)_{v \in \mathcal{V}_0}$, given that a local transmission policy $(\gamma_v)_{v \in \mathcal{V}_0}$ and the distribution of V_1 have been fixed. Let $\tilde{\xi}(r, v) = \xi_v(r)$, $\tilde{I}(r, v) = I(r, \gamma_v)$, and $\tilde{\rho}(r, v) = \rho(r, \gamma_v)$.

We then need to optimize $\mathbb{E}_0 [\bar{\xi}(R_1, V_1) \tilde{I}(R_1, V_1)]$ over all $\bar{\xi} : \mathcal{R} \times \mathcal{V}_0 \mapsto \{0, 1\}$ that satisfy $\mathbb{E}_0 [\bar{\xi}(R_1, V_1) \tilde{\rho}(R_1, V_1)] \leq c$. It is an easy exercise (whose proof is omitted) to show that there exists an optimal censoring policy of the following form. There is a threshold t such that $\bar{\xi}(r, v) = 1$ if $\tilde{I}(r, v)/\tilde{\rho}(r, v) > t$ and $\bar{\xi}(r, v) = 0$ if $\tilde{I}(r, v)/\tilde{\rho}(r, v) < t$. Randomization is only used to make a censoring decision when $\tilde{I}(r, v)/\tilde{\rho}(r, v) = t$, and a binary randomization variable at each sensor suffices. This is a solution of the “water-filling” type, whereby the uncensored “states” (r, v) are chosen starting with those with a higher value of $\tilde{I}(r, v)/\tilde{\rho}(r, v)$, and continuing until the resource constraint is met. Note also that for a pure transmission policy γ , the relevant ratio is $I(r, \gamma)/\rho(r, \gamma)$, which has the intuitive interpretation of information content per unit resource consumed.

VIII. APPLICATIONS

In this section, we revisit Examples 1 and 2 from Section I, and illustrate the form of an optimal censoring policy. Given our focus on the censoring policy, we will assume that all sensors send their observations “in the clear” to the fusion center, i.e., the pure local transmission policy $\gamma(x, r) = x$ is employed at all the sensors. Accordingly, γ will be suppressed in our notation below.

A. Fading Channels

We will focus on a special case of the problem posed in Example 1. We consider a wireless sensor network transmitting measurements to a fusion center over slowly fading Gaussian channels. We assume that $\rho(r) = 1$ for all r , so that we are only concerned with restricting the number of sensors transmitting. Depending on the condition of the channel, we will naturally want to allow sensors to transmit over good channels and allow sensors that have bad channels to censor. This raises the issue of identifying the key parameters of the channel on the basis of which censoring decisions should be made.

Suppose that

$$\begin{aligned} H_0 : X_i &\sim N(-m, \sigma^2) \\ H_1 : X_i &\sim N(m, \sigma^2) \end{aligned}$$

and that the fusion center receives

$$Y_i = Q_i X_i + W_i$$

where Q_i is the fading coefficient, with a known density $g(\cdot)$, and $W_i \sim N(0, \sigma_i^2)$. Assume that the channel characteristics $R_i = (Q_i, \sigma_i)$ are stationary and ergodic, with the same stationary distribution under either hypotheses. This can be used to model the case where sensors are placed in a line so that an ergodic assumption on the distribution of the variables R_i is reasonable. (Random (i.i.d.) placement of sensors is another example.) Since this is a slow-fading channel, each sensor can measure R_i . From Theorem 3, the important design parameter is

$$\begin{aligned} I(Q_i, \sigma_i) &= \mathbb{E}_0 \left[-\frac{(Y_i + Q_i m)^2}{2(Q_i^2 \sigma^2 + \sigma_i^2)} + \frac{(Y_i - Q_i m)^2}{2(Q_i^2 \sigma^2 + \sigma_i^2)} \right] \\ &= \frac{2Q_i^2 m^2}{Q_i^2 \sigma^2 + \sigma_i^2}. \end{aligned}$$

According to Theorem 3 and the discussion in the previous section, we want to censor when $|Q_i/\sigma_i|$ is small. Thus, an asymptotically optimal censoring policy (where censoring is based on the channel characteristics) is of the form

$$\xi(R_i) = \begin{cases} 1, & \text{if } \left| \frac{Q_i}{\sigma_i} \right| > \eta \\ 0, & \text{otherwise} \end{cases}$$

where η depends on the value of c and the density $g(\cdot)$. Note that randomization when $|Q_i/\sigma_i| = \eta$ is unnecessary, because this event happens with zero probability.

B. Detection of Spatial Signals

Consider n sensor nodes, placed uniformly and independently in $[-1, 1]$, with the fusion center at the origin, for the purpose of detecting a spatial signal. Consider the hypotheses

$$\begin{aligned} H_0 : X_i &= s_0(R_i) + W_i \quad \forall i \\ H_1 : X_i &= s_1(R_i) + W_i \quad \forall i \end{aligned}$$

where each $s_j(\cdot)$ is a known spatial signal, and $W_i \sim N(0, \sigma^2)$ is Gaussian noise. When sensor i sends its measurement X_i to the fusion center, it consumes power $\rho(R_i)$ (assumed positive), which depends on its relative position to the fusion center. We constrain the overall average power to be less than a given positive constant c . From Theorem 1, each sensor should use a common local censoring policy $(\xi_v)_{v \in \mathcal{V}_0}$, obtained by maximizing $\mathbb{E}_0[\xi_{V_1}(R_1)I(R_1)]$ subject to $\mathbb{E}_0[\xi_{V_1}(R_1)\rho(R_1)] \leq c$. According to the discussion in Section VII, a sensor i should be censored when $I(R_i)/\rho(R_i)$ is below a threshold. As a specific illustration, let $s_0(r) = 1 - r$ and $s_1(r) = 1 + r$. Then

$$\begin{aligned} I(r) &= \mathbb{E}_0 \left[-\frac{1}{2\sigma^2} ((X_1 - 1 + r)^2 - (X_1 - 1 - r)^2) \right] \\ &= \frac{2r^2}{\sigma^2}. \end{aligned}$$

Suppose $\rho(r) = 1 + a|r|^d$, where $2 \leq d \leq 4$. (This is in line with standard models used for power decay in a wireless network, see [12]. The unit cost is due to the power used to make the measurement X_i .) Then, we have

$$\frac{I(r)}{\rho(r)} = \frac{2r^2}{\sigma^2(1 + a|r|^d)}.$$

A specific case is shown in Fig. 1. We have taken $a = 1$, $d = 2$, $\sigma = 1$, and a constraint of $c = 19/12$. As shown, only sensors at a large enough distance from the fusion center should transmit their measurements X_i .

IX. THE BAYESIAN PROBLEM WITH LOCAL CENSORING

We now consider the decentralized detection problem with censoring in the Bayesian context. Let the prior probability of H_j be $q_j > 0$, for $j = 0, 1$. We define $\mathbb{P} = q_0\mathbb{P}_0 + q_1\mathbb{P}_1$ and let \mathbb{E} be the expectation operator with respect to (w.r.t.) \mathbb{P} . As in the Neyman–Pearson case, we allow sensors to use randomized sensor policies. In contrast to unconstrained Bayesian problems, simple examples show that randomization results in improved performance when the number of sensors n is finite. However,

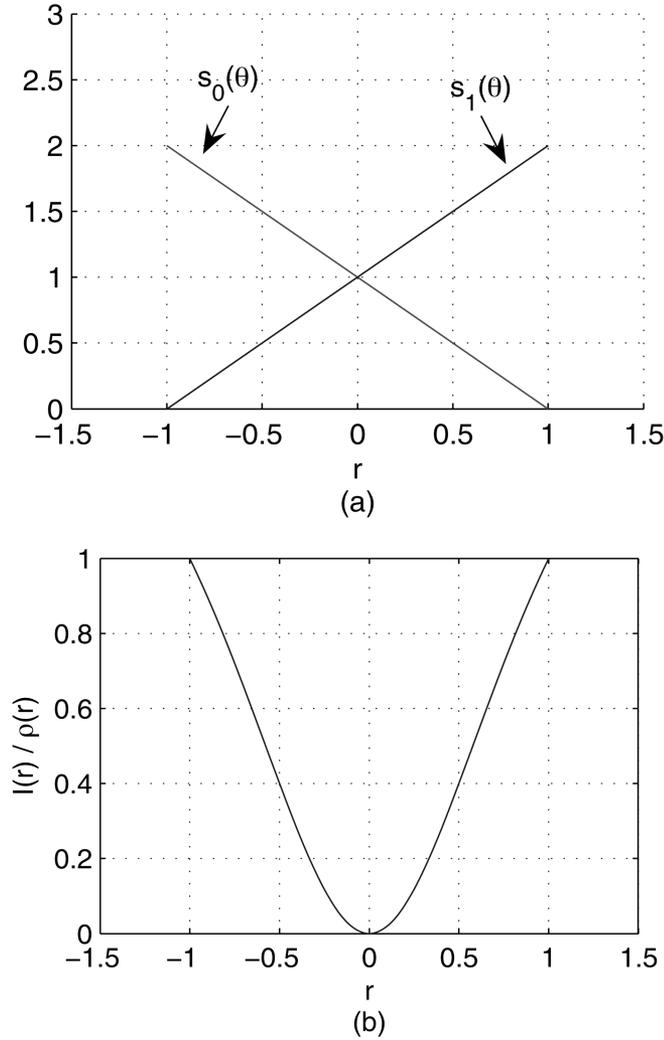


Fig. 1. The top graph shows the spatial signals plotted as a function of sensor location. Let $c = 19/12$. The bottom graph shows a plot of $I(r)/\rho(r)$. A sensor is censored unless its location is in $[-1, -0.5]$ or $[0.5, 1]$.

we will show that for the asymptotic problem considered here (no cooperation), randomization is unnecessary. In the process, we will also characterize the optimal error exponent and associated local policies.

A strategy $\pi^{(n)}$ is admissible if (1) is satisfied, with \mathbb{E} replacing \mathbb{E}_0 . For any admissible strategy $\pi^{(n)}$, let $P_{e,n}(\pi^{(n)})$ denote the resulting probability of error at the fusion center. We will always assume that the fusion center uses an optimal fusion rule, namely the maximum *a posteriori* probability (MAP) rule. Let $P_{e,n}^*$ be the infimum of $P_{e,n}(\pi^{(n)})$ over all admissible strategies. We are interested in finding asymptotically optimal local strategies that achieve

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^*.$$

Before we launch into the analysis, let us consider a simple example that shows that cooperation among the sensors is *strictly* better than using local strategies.

Example 7: Suppose that the random variables R_i belong to $\{0, 1\}$, are i.i.d. under either hypothesis, and that $\mathbb{P}_j(R_1 = 0) = \mathbb{P}_j(R_1 = 1) = 1/2$, for $j = 0, 1$. We assume that all sensors are restricted to using the transmission function $\gamma(x, r) = x$.

We assume that the distribution of X_1 under the two hypotheses is the same when $R_1 = 0$, but different when $R_1 = 1$. Thus, it is only those sensors with $R_1 = 1$ that have useful information to transmit. Under mild conditions (including the special case where X_1 has a finite range), it is a well-known consequence of the Chernoff bound that if exactly m sensors have $R_i = 1$ and transmit to the fusion center, the probability of error is of the form $e^{m\Lambda^*} g(m)$, where Λ^* is a negative constant determined by the distributions of X_1 and where $g(m)$ satisfies $\lim_{m \rightarrow \infty} (\log g(m))/m = 0$. In particular, for every $\epsilon > 0$, we can find some positive a, b , such that $ae^{-\epsilon m} \leq g(m) \leq be^{\epsilon m}$.

Let $\rho(r, \gamma) = 1$ and $c = 1/4$. Thus, the resource constraint (1) becomes $\mathbb{E}[N] \leq n/4$, where N is the (random) number of sensors that are not censored.

Assume for simplicity that $n/4$ is an integer. Consider the following cooperative censoring strategy.

- 1) If $\sum_{i=1}^n R_i \leq n/4$, sensor i transmits if only if $R_i = 1$.
- 2) If $\sum_{i=1}^n R_i > n/4$, among those sensors with $R_i = 1$, arbitrarily choose $n/4$ of them to transmit.

Using the Chernoff bound, we have $\mathbb{P}(N < n/4) \leq e^{-dn}$ for some positive constant d . Let $P_{e,n}$ be the probability of error at the fusion center. We have

$$\begin{aligned} P_{e,n} &\leq \mathbb{P}(N < n/4) + \mathbb{P}(N = n/4)e^{\Lambda^* n/4} g(n/4) \\ &\leq e^{-dn} + be^{(\Lambda^* + \epsilon)n/4}. \end{aligned}$$

Suppose that when $R_1 = 1$, the distribution of X_1 is such that $-d < \Lambda^*$, which is certainly possible. Since $\epsilon > 0$ is arbitrary, we obtain $\liminf_{n \rightarrow \infty} (1/n) \log P_{e,n} \leq \Lambda^*/4$.

Consider now a local and pure censoring strategy. In a best strategy of this kind, every sensor with $R_1 = 0$ is censored, and $\mathbb{E}[N] = n/4$. The only way to achieve this is as follows: $n/2$ sensors are always censored; the remaining sensors are censored if and only if $R_1 = 0$. Thus, N is binomial with parameters $n/2$ and $1/2$. After averaging over all possible values of N , the probability of error satisfies

$$\begin{aligned} P_{e,n} &= \mathbb{E}[e^{\Lambda^* N} g(N)] \geq a \mathbb{E}[e^{(\Lambda^* - \epsilon)N}] \\ &= a \left(\frac{1}{2} + \frac{1}{2} e^{\Lambda^* - \epsilon} \right)^{n/2}. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we obtain

$$\liminf_{n \rightarrow \infty} (1/n) \log P_{e,n} \geq (1/2) \log((1 + e^{\Lambda^*})/2).$$

This is strictly greater than $\Lambda^*/4$, which shows that the cooperative strategy constructed earlier has a better error exponent. Later on, we show that randomization cannot improve performance, bringing the conclusion that cooperative strategies can be strictly better than local ones.

The essence of this example is that in the local case, we have much less control over the tails of the distribution of N ; the possibility of N having a large deviation results in a deterioration in the error exponent. \square

In general, optimal cooperative strategies are difficult to find. As the cooperative strategy may also be practically infeasible, we will focus our attention on finding an optimal *local* strategy. For the remainder of this paper, the words “policy” and “strategy” will always mean “local policy” and “local strategy,” respectively.

A. Notation and Assumptions

Let $\ell_{10}(\cdot | r)$ be the Radon–Nikodym derivative of the measure $\nu_1(\cdot | r)$ w.r.t. $\nu_0(\cdot | r)$. For $\gamma \in \Gamma$, let $\ell_{10}^\gamma(\cdot | r)$ be the Radon–Nikodym derivative of the measure $\nu_1^\gamma(\cdot | r)$ w.r.t. $\nu_0^\gamma(\cdot | r)$. Let $\ell_\mu(R_1) = \frac{d\mu_1}{d\mu_0}(R_1)$, and for any $s \in [0, 1]$, and pure local transmission policy $\gamma \in \Gamma$, let

$$\Lambda(s, r, \gamma) = \log \mathbb{E}_0 \left[e^{s \log \ell_{10}^\gamma(Y_1 | r)} | R_1 = r \right].$$

Finally, for a randomized local policy $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0}$, let

$$\begin{aligned} \Phi(s, \pi) &= \log \mathbb{E}_0 \left[e^{s(\xi_V(R_1) \log \ell_{10}^\gamma(Y_1 | R_1) + \log \ell_\mu(R_1))} \right] \\ &= \log \mathbb{E}_0 \left[\left(\xi_V(R_1) e^{\Lambda(s, R_1, \gamma_V)} + 1 - \xi_V(R_1) \right) \ell_\mu^s(R_1) \right]. \end{aligned} \quad (23)$$

For a policy $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0}$, if $\xi_v(r) = 1$ for all $r \in \mathcal{R}$ and $v \in \mathcal{V}_0$, we will write $\pi = (1, \gamma_v)_{v \in \mathcal{V}_0}$. We will make the following assumptions.

Assumption 3:

- i) Conditioned on either hypothesis, the random variables R_i are i.i.d. Furthermore, μ_0 and μ_1 are equivalent measures.
- ii) The (regular) conditional distributions $\nu_0(\cdot | r)$ and $\nu_1(\cdot | r)$ are equivalent for every $r \in \mathcal{R}$.
- iii) We are restricted to local strategies.
- iv) We have $\rho(r, \gamma) > 0$, for every $r \in \mathcal{R}$ and pure policy $\gamma \in \Gamma$.
- v) There exists an open interval $I = (\tau_1, \tau_2) \subset [0, 1]$, such that for all pure policies $\gamma \in \Gamma$, we have $\arg \min_{s \in [0, 1]} \Phi(s, (1, \gamma)) \in I$. Furthermore, for $k = 1$ and 2, the following holds:

$$\begin{aligned} |\mathbb{E}_0 [(\ell_{10}(X_1 | R_1) \ell_\mu(R_1))^{T_k} \log(\ell_{10}(X_1 | R_1) \ell_\mu(R_1))]| \\ < \infty. \end{aligned}$$

- vi) For the same open interval I as in v) above, there exists a $b \in (0, \infty)$ such that

$$\frac{d^2}{ds^2} \mathbb{E}_0 [\ell_\mu^s(R_1)] \leq b$$

and

$$\frac{d^2}{ds^2} \mathbb{E}_0 [(\ell_{10}^\gamma(Y_1 | R_1) \ell_\mu(R_1))^s] \leq b$$

for all $s \in I$ and all $\gamma \in \Gamma$.

Note that there is little loss of generality in imposing Assumption 3 iv). Indeed, if $\rho(r, \gamma) = 0$ for some r and some pure policy

γ , then we can always transmit Y_i when $R_i = r$, without incurring any cost. So instead of censoring in the state $R_i = r$, the sensor can always choose to transmit using this particular γ .

Assumptions 3 v) and vi) are required for the same technical reasons as in [5], which also gives rather general conditions under which they are satisfied.³ In general, the open interval I can be taken to be $(0, 1)$. Indeed, it can be shown that, under Assumptions 3 i) and ii), and for any pure transmission policy $\gamma \in \Gamma$, the minimizer s^* of $\min_{s \in [0, 1]} \Phi(s, (1, \gamma))$ is in the interior of $[0, 1]$. If we take $I = (0, 1)$, Assumption 3 v) reduces to the condition that the KL divergences $-\mathbb{E}_0[\log(\ell_{10}(X_1 | R_1) \ell_\mu(R_1))]$ and $\mathbb{E}_1[\log(\ell_{10}(X_1 | R_1) \ell_\mu(R_1))]$ are bounded. But we only need the weaker version of Assumptions 3 v) and vi), as stated. This allows us to include cases where Assumptions 3 v) and vi) hold automatically. For example, if Γ is a finite set of transmission policies, the interval I only needs to include certain, finitely many, values of s , and we can choose $I = (a, b)$, where $0 < a < b < 1$. Then, it is easy to show that under Assumptions 3 i) and ii), Assumptions 3 v) and vi) hold automatically. We will make use of this fact in Sections X-C and D. Another sufficient condition for Assumptions 3 v) and vi) are Assumptions 3 i) and ii) together with an assumption similar to Assumption 2 (see [5, Proposition 3]).

The main reason for introducing Assumption 3 is the following lemma, which is proved in the Appendix.

Lemma 9: Suppose that Assumption 3 holds. Then, there exists some $b_1 \in (0, \infty)$ such that for all $s \in I$ and for all π , $|\Phi(s, \pi)| \leq b_1$, $|\frac{d}{ds} \Phi(s, \pi)| \leq b_1$, and $\frac{d^2}{ds^2} \Phi(s, \pi) \leq b_1$.

We record a result from [5], based on the results in [13], which will underlie the rest of our development. This result can also be obtained from [14, Theorem 1.3.13]. The result states that, if the conclusion of Lemma 9 holds, then

$$\begin{aligned} \frac{1}{n} \log P_{e,n}(\pi^{(n)}) &= \min_{s \in I} \frac{1}{n} \sum_{i=1}^n \Phi(s, \pi_i) + o(1) \\ &= \min_{s \in [0, 1]} \frac{1}{n} \sum_{i=1}^n \Phi(s, \pi_i) + o(1) \end{aligned} \quad (24)$$

where $o(1)$ stands for a term that vanishes as $n \rightarrow \infty$, uniformly over all sequences $\pi^{(n)}$. Given this result, we can just focus on the problem of optimizing the RHS of (24), while ignoring the $o(1)$ term.

B. Optimal Strategy

In this subsection, we prove that asymptotic optimality can be obtained by dividing the sensors into two groups with sensors in each group using a common pure policy.

Theorem 4: Under Assumption 3

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^* = \inf_{s \in [0, 1]} \min_{\pi_1, \pi_2} \{u \Phi(s, \pi_1) + (1 - u) \Phi(s, \pi_2)\} \quad (25)$$

where the infimum is taken over all $u \in [0, 1]$, and all pure policies π_1 and π_2 that satisfy $u\rho(\pi_1) + (1 - u)\rho(\pi_2) \leq c$.

³Although [5] deals with the case of a finite transmission alphabet \mathcal{Y} , the results therein can be easily generalized to the case of infinite alphabets.

Proof: Fix some $s \in [0, 1]$. Let $\bar{\psi}_1, \dots, \bar{\psi}_n$ be some (possibly randomized) policies. Let $\psi_{i,v}$ be the pure policy obtained when $V_i = v$. Using the definition (23) of $\Phi(s, \bar{\psi}_i)$ and Jensen's inequality, we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \Phi(s, \bar{\psi}_i) &= \frac{1}{n} \sum_{i=1}^n \log(\exp(\Phi(s, \bar{\psi}_i))) \\ &= \frac{1}{n} \sum_{i=1}^n \log \mathbb{E}[\exp(\Phi(s, \psi_{i,V_i}))] \\ &\geq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\Phi(s, \psi_{i,V_i})]. \end{aligned} \quad (26)$$

Similarly

$$\frac{1}{n} \sum_{i=1}^n \rho(\bar{\psi}_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\rho(\psi_{i,V_i})].$$

Note that taking the average over all i in the above expressions is equivalent to taking an expectation over a uniformly chosen random i . Let $U = (J, V_J)$, where J is chosen uniformly over $\{1, \dots, n\}$. We minimize the RHS of (26), $\mathbb{E}[\Phi(s, \psi_U)]$, subject to the constraint $\mathbb{E}[\rho(\psi_U)] \leq c$. Applying Lemma 8, with $f(u) = \Phi(s, \psi_u)$ and $g(u) = \rho(\psi_u)$, we obtain

$$\frac{1}{n} \sum_{i=1}^n \Phi(s, \bar{\psi}_i) \geq \inf\{u\Phi(s, \pi_1) + (1-u)\Phi(s, \pi_2)\}$$

where the infimum is taken over all $u \in [0, 1]$, and pure policies π_1, π_2 , satisfying the resource constraint $u\rho(\pi_1) + (1-u)\rho(\pi_2) \leq c$. Hence, from (24),

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^* \geq \inf_{s \in [0,1]} \min\{u\Phi(s, \pi_1) + (1-u)\Phi(s, \pi_2)\}. \quad (27)$$

To achieve the lower bound, suppose that u and $\pi_i, i = 1, 2$, attain the infimum in (27) to within $\epsilon > 0$ and that s^* is a minimizing value of s in (27). We assign $\lfloor un \rfloor$ sensors to use policy π_1 , and $\lfloor (1-u)n \rfloor$ sensors to use policy π_2 . We censor any remaining sensor. Then, from (24)

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^* &\leq u\Phi(s^*, \pi_1) + (1-u)\Phi(s^*, \pi_2) \\ &\leq \inf_{s \in [0,1]} \min\{u\Phi(s, \pi_1) + (1-u)\Phi(s, \pi_2)\} + \epsilon \end{aligned}$$

and taking $\epsilon \rightarrow 0$ completes the proof. \square

Let us remark that similar results are easily obtained for the case where the side information is not transmitted to the fusion center (cf. Section V-C).

C. Characterization of the Optimal Exponent

In this subsection and the next, we will consider the case where \mathcal{R} is finite, for two reasons. First, in many practical cases, because of the limited channel between each sensor and the fusion center, the side information can be assumed to take values from a finite alphabet. Second, when \mathcal{R} is finite, the analysis is

simplified and results in a simple form for the censoring policies. So without loss of generality, we will take $\mathcal{R} = \{1, 2, \dots, |\mathcal{R}|\}$. Let $p_{jr} = \mathbb{P}_j(R_1 = r)$ and $p_r = q_0 p_{0r} + q_1 p_{1r}$.

Let us fix two pure local transmission policies γ_1 and γ_2 . Let

$$K_s(c, \{\gamma_1, \gamma_2\}) = \inf_{\xi_1, \xi_2, u} \{u\Phi(s, (\xi_1, \gamma_1)) + (1-u)\Phi(s, (\xi_2, \gamma_2))\} \quad (28)$$

where the infimum is taken over all $u \in [0, 1]$ and pure censoring policies ξ_1, ξ_2 that satisfy

$$u\mathbb{E}[\xi_1(R_1)\rho(R_1, \gamma_1)] + (1-u)\mathbb{E}[\xi_2(R_2, \gamma_2)\rho(R_2, \gamma_2)] \leq c.$$

From Theorem 4 and under Assumption 3, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^* = \inf_{\gamma_1, \gamma_2 \in \Gamma} \min_{s \in [0,1]} K_s(c, \{\gamma_1, \gamma_2\}). \quad (29)$$

Note that given γ_1 and γ_2 , the minimization in (28) has an optimal solution. (This is because \mathcal{R} is finite, and therefore there are only finitely many possible pure censoring policies.) Let z_1 be the value of $\mathbb{E}[\xi_1(R_1)\rho(R_1, \gamma_1)]$ in such an optimal solution. It follows that ξ_1 must minimize $\Phi(s, (\xi_1, \gamma_1))$ (and, therefore, $\exp(\Phi(s, (\xi_1, \gamma_1)))$ as well), subject to the constraint $\mathbb{E}[\xi_1(R_1)\rho(R_1, \gamma_1)] \leq z_1$. Note that $\exp(\Phi(s, (\xi_1, \gamma_1)))$ is equal to

$$\mathbb{E}_0[\xi(R_1)h_s(R_1, \gamma)] + \mathbb{E}_0[\ell_\mu^s(R_1)]$$

where $h_s(r, \gamma) = (e^{\Lambda(s,r,\gamma)} - 1)\ell_\mu^s(r) \leq 0$.

We can now give a characterization of the optimal ξ_1 , similar to the one at the end of Section VII. For any $r \in \mathcal{R}$, let

$$m_{s,\gamma}(r) = \frac{p_{0r}h_s(r, \gamma)}{p_r\rho(r, \gamma)}.$$

Proposition 3: Suppose Assumption 3 holds. Suppose that \mathcal{R} is finite and that optimal choices of u, s, γ_1, γ_2 have been fixed. Then, there exist thresholds t_1, t_2 such that the corresponding optimal censoring functions ξ_1, ξ_2 satisfy the following: for each $r \in \mathcal{R}$, if $m_{s,\gamma_i}(r) < t_i$, then $\xi_i(r) = 1$, otherwise $\xi_i(r) = 0$.

X. SPECIAL CASES AND EXAMPLES FOR THE BAYESIAN PROBLEM

We now examine some special cases that will lead to simplified versions of Theorem 4.

A. No Side Information

In this subsection, we consider the case of no side information, which is equivalent to having $\mathcal{R} = \{1\}$. Accordingly, r will be suppressed from our notation below. For example, the cost incurred by a sensor making a measurement and transmitting it via a transmission function $\gamma \in \Gamma$ is denoted by $\rho(\gamma)$. We will show that when c in the resource constraint (1) is sufficiently small, we can restrict all uncensored sensors to use the same policy.

Note that there are only two possible pure censoring policies, $\xi = 0$ and $\xi = 1$. In the absence of side information, the likeli-

hood ratio $\ell_\mu(R_1)$ is identically equal to 1. Using the definition of $\Phi(s, (\xi, \gamma))$, for $\xi = 1$ and $\xi = 0$, respectively, we obtain $\Phi(s, (1, \gamma)) = \Lambda(s, \gamma) \leq 0$ and $\Phi(s, (0, \gamma)) = 0$. Let $\Lambda^*(\gamma) = \min_{s \in [0,1]} \Lambda(s, \gamma)$.

Corollary 1: Suppose that Assumption 3 holds, and $0 < c \leq \inf_{\gamma \in \Gamma} \rho(\gamma)$. Then, in the absence of side information, the optimal exponent is equal to $\inf_{\gamma \in \Gamma} c\Lambda^*(\gamma)/\rho(\gamma)$. This remains the optimal error exponent even under the additional restriction that censored sensors are chosen in advance, and all uncensored sensors use the same policy.

Proof: From Theorem 4, we know that at most two different pure policies $\pi_i = (\xi_i, \gamma_i)$, $i = 1, 2$, need be considered. Suppose that one of those policies, say π_2 , involves censoring, namely, $\xi_2 = 0$. Then, $\Phi(s, \pi_2) = 0$ and $\rho(\pi_2) = 0$. Clearly, the other policy should not censor, so that $\pi_1 = (1, \gamma_1)$. For any choice of γ_1 , the optimal choice of u in (25) is to let $u = c/\rho(\gamma_1) \in [0, 1]$, leading to an exponent of $c\Lambda^*(\gamma_1)/\rho(\gamma_1)$. Optimizing over all $\gamma_1 \in \Gamma$, we obtain the claimed error exponent.

Suppose now that neither of the policies π_1 and π_2 involves censoring. Since $c \leq \rho(\gamma)$ for all $\gamma \in \Gamma$, in order to satisfy the resource constraint $u\rho(\gamma_1) + (1-u)\rho(\gamma_2) \leq c$, we must have $\rho(\gamma_1) = \rho(\gamma_2) = c$. In this case, for all $u \in [0, 1]$ and all $s \in [0, 1]$, $u\Phi(s, \pi_1) + (1-u)\Phi(s, \pi_2) \geq \min_{i=1,2} \Lambda^*(\gamma_i) \geq \inf_{\gamma \in \Gamma} c\Lambda^*(\gamma)/\rho(\gamma)$. The corollary is now proven. \square

B. Finite Transmission Policy Sets

In Sections X-C and D, we study problem formulations in which there is no side information, and in which we temporarily restrict transmission policies to belong to a finite subset G of Γ . According to the discussion in Section IX-A, this restriction implies that under Assumptions 3 i) and ii), Assumptions 3 v) and vi) hold automatically. This will allow us to apply Corollary 1 to two problems that have been considered in [3] and [4]. Let $P_{e,n}^*(G)$ be the minimum error probability, when we are restricted to transmission policies in G .

C. Total Power Constraint

In [3], the authors consider the Bayesian problem with no side information and a power constraint of the form

$$\sum_{i=1}^{L_A} \rho(\gamma_i) \leq A \quad (30)$$

where L_A , the number of sensors, is not fixed in advance. The cost $\rho(\gamma)$ is assumed to be positive for all γ .

Let G be a finite subset of Γ , and let \mathcal{G} be the collection of all such subsets G . Let $P_e(G, A)$ be the minimum probability of error when using transmission policies from G that satisfy (30). Recall that we define $\Lambda^*(\gamma) = \min_{s \in [0,1]} \Lambda(s, \gamma)$. Reference [3] shows that

$$\inf_{G \in \mathcal{G}} \liminf_{A \rightarrow \infty} \frac{1}{A} \log P_e(G, A) = \inf_{\gamma \in \Gamma} \frac{\Lambda^*(\gamma)}{\rho(\gamma)} \quad (31)$$

so that it is asymptotically optimal to have all sensors use the same transmission policy. We will rederive (31) from Corollary 1.

To see the connection with our framework, fix a $G \in \mathcal{G}$. Note that for all $\gamma \in G$, there exists a $\delta > 0$ such that $\rho(\gamma) \geq \delta$. Under the constraint (30), the number of sensors that can be used is bounded by $n = \lfloor A/\delta \rfloor$. With n defined in this manner, the constraint (30) is equivalent to the constraint

$$\sum_{i=1}^n \xi_i \rho(\gamma_i) \leq n\delta$$

where $\xi_i \in \{0, 1\}$. Note that the limit $A \rightarrow \infty$, considered in [3], is equivalent to the limit $n \rightarrow \infty$ in our framework.

Therefore, under Assumptions 3 i), ii), and iv), Corollary 1 shows that the optimal error exponent is

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^*(G) = \delta \inf_{\gamma \in G} \frac{\Lambda^*(\gamma)}{\rho(\gamma)}.$$

By taking the infimum over all $G \in \mathcal{G}$, we recover (31). This argument shows that it is asymptotically optimal to use $\lfloor A/\rho(\gamma) \rfloor$ sensors, all of which employ the same pure transmission policy γ , chosen by carrying out the minimization in the RHS of (31).

This discussion elucidates the relationship of a power constraint (in which the number of transmitting sensors is not fixed) to our constrained censoring problem. The decentralized detection problem considered in [5] can be viewed as one where c is so large that censoring is never needed. The problem in this subsection can be viewed as one involving a very small c . In this case, one group of sensors sets $\xi \equiv 0$, and another uses the transmission policy that asymptotically achieves (31). In comparison, the general formulation in this paper also gives the solution for all c , in between these two extremes.

D. Constrained Capacity

Yet another connection can be made to the problem considered in [4], which is summarized as follows. Consider a network of L sensors, z_l , $l = 1, 2, \dots, L$. Each sensor z_l observes a sequence of measurements $\{X_{l,t} : t = 1, 2, \dots, T\}$, and there is no side information. All the measurements are assumed to be conditionally i.i.d. given the hypothesis, over time and across sensors. At each time t , sensor z_l sends $Y_{l,t} = \gamma_l(X_{l,t})$ to the fusion center, where the transmission function is a b_l -bit quantizer, where b_l is a positive integer. Let the set of allowed b -bit transmission functions be Γ_b . We are interested in minimizing the error exponent

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \log P_{e,T}(\gamma^{(L)})$$

where $P_{e,T}(\gamma^{(L)})$ is the probability of error at the fusion center, assuming that a MAP fusion rule is used. The minimization is to be carried out over the number of sensors L and transmission strategies satisfying the overall capacity constraint

$$\sum_{l=1}^L b_l \leq B, \quad (32)$$

where B is a given positive integer. Let us call the above problem Q1. This problem, in general, does not have a closed-form solution. Reference [4] finds sufficient conditions under which using B identical sensors (sensors using the same

transmission policy), each sending one bit of information, is optimal. We will apply our results to arrive at the same conditions as in [4], and also characterize the solution for special values of B .

As a first step, we will relax the constraints in problem Q1. We view each sensor z_l over the time periods $1, 2, \dots, T$, as T different sensors $z_{l,t}$, and hence remove the constraint that all $z_{l,t}$ must use the same transmission policy γ_l . Because of (32), $L \leq B$. Hence, we can imagine that we are starting with $n = TB$ sensors, some of which will be censored, and rewrite (32) as

$$\sum_{t=1}^T \sum_{l=1}^B \xi_{l,t} b_{l,t} \leq TB \quad (33)$$

where $\xi_{l,t} \in \{0, 1\}$. For each $b \in \{1, \dots, B\}$ for which Γ_b is nonempty, consider a nonempty finite subset of Γ_b , and use G to denote the union of these subsets over b . Let \mathcal{G} be the collection of all such G . With $n = TB$, we wish to minimize

$$B \inf_{G \in \mathcal{G}} \lim_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^*(G).$$

We will obtain an optimal solution to the latter problem, which we call problem Q2. If from that optimal solution we can derive a strategy that does not change the error exponent and yet meets the constraints that $\gamma_{l,t} = \gamma_l$ and $\xi_{l,t} = \xi_l$ for all t , then we will have found an optimal solution to problem Q1. In particular, sufficient conditions for problem Q2 to have all sensors using the same one-bit transmission policy are also sufficient for problem Q1 to have B identical one-bit sensors.

To put problem Q2 into our constrained censoring context, let $\rho(\gamma) = b$ for every $\gamma \in \Gamma_b$, and note that $c = 1$. Let $K_b^* = \inf_{\gamma \in \Gamma_b} \Lambda^*(\gamma)$. (If Γ_b is empty, we set $K_b^* = \infty$.)

Proposition 4: Suppose Assumptions 3 i) and ii) hold.

i) For problem Q2,

$$\inf_{G \in \mathcal{G}} \lim_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^*(G) = \min_{1 \leq b \leq B} \frac{1}{b} K_b^*.$$

In particular, using the same one-bit transmission function at all uncensored sensors is asymptotically optimal iff $K_1^* = \min_{1 \leq b \leq B} (K_b^*/b)$.

ii) Let $b^* = \arg \min_{1 \leq b \leq B} (K_b^*/b)$. For problem Q1, if B/b^* is an integer, we can restrict to using $L = B/b^*$ sensors, all of them using the same b^* -bit transmission policy, without affecting the optimal exponent.

Proof: Part i) follows from Corollary 1. For part ii), let γ_{b^*} achieve $K_{b^*}^*$ to within $\epsilon > 0$. Let each of the L sensors in problem Q1 use γ_{b^*} . This comes within ϵ of the optimal exponent for problem Q2, and therefore for problem Q1 as well. \square

Let $H = -\min_{s \in [0,1]} \log \mathbb{E}_0 \left[\left(\frac{dw_1}{dw_0} \right)^s \right]$, and note that $K_b^* \geq -H$. Hence, for any $b \in \{2, \dots, B\}$

$$\frac{1}{b} K_b^* \geq -\frac{H}{2}.$$

So if $K_1^* \leq -H/2$, we meet the sufficient conditions for problem Q2 to achieve the optimal error exponent with identical one-bit sensors. In that case, it is also optimal for problem Q1 to have B identical one-bit sensors. This recovers Proposition 2 of [4].

On the other hand, suppose that B is an even integer and that $K_2^*/2 < K_1^*$. Then, it is strictly suboptimal to use B identical one-bit sensors for problem Q1. This is the content of [4, Proposition 3]. For a general B that is not an integer multiple of b^* , the solution to Q1 involves an integer program, which can be difficult to solve for large B . However, as B increases to infinity, we can approach the optimal performance by using $\lfloor B/b^* \rfloor$ b^* -bit sensors.

E. ρ Independent of the Transmission Function

Suppose that for every value r of the side information, all the transmission functions in Γ have the same cost, e.g., that the process of transmission under state r requires the same energy for all γ . Then, we can assume $\rho(r, \gamma) = \tilde{\rho}(r)$ for some non-negative function $\tilde{\rho}$. Suppose also that the set of transmission policies Γ is of the form $\Gamma = \prod_{r \in \mathcal{R}} \Gamma(r)$, where $\Gamma(r)$ is the set of allowed transmission policies $\gamma(\cdot, r)$, when the side information takes the value r . Let

$$\Lambda^*(s, r) = \inf_{\gamma \in \Gamma(r)} \Lambda(s, r, \gamma)$$

and

$$\Phi^*(s, \xi) = \log \mathbb{E}_0 \left[\left(\xi(R_1) e^{\Lambda^*(s, R_1)} + 1 - \xi(R_1) \right) \ell_\mu^s(R_1) \right].$$

Corollary 2: Assume that $\rho(r, \gamma) = \tilde{\rho}(r)$ for all $\gamma \in \Gamma(r)$, and that Assumption 3 holds. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{e,n}^* = \inf_{s \in [0,1]} \min_{u \in [0,1]} (u \Phi^*(s, \xi_1) + (1-u) \Phi^*(s, \xi_2)) \quad (34)$$

where the infimum is taken over all $u \in [0, 1]$ and censoring policies ξ_i that satisfy

$$u \mathbb{E}[\xi_1(R_1) \tilde{\rho}(R_1)] + (1-u) \mathbb{E}[\xi_2(R_1) \tilde{\rho}(R_1)] \leq c.$$

Furthermore, it is optimal to use the same transmission policy for all sensors.

Proof: The result is obtained from Theorem 4, by observing that the constraints do not affect the optimization with respect to γ_1 and γ_2 , and that

$$\begin{aligned} \inf_{\gamma_1, \gamma_2} (u \Phi(s, (\xi_1, \gamma_2)) + (1-u) \Phi(s, (\xi_2, \gamma_2))) \\ = u \Phi^*(s, \xi_1) + (1-u) \Phi^*(s, \xi_2). \quad \square \end{aligned}$$

In this case, we use the same transmission policy at all sensors, and at most two different censoring policies. Suppose that s^* is a minimizing value of s in (34). Then, for any $r \in \mathcal{R}$, we can use a transmission function $\gamma(\cdot, r) \in \Gamma(r)$ that minimizes $\Lambda(s^*, r, \gamma)$, if the minimum is attained.

XI. CONCLUSION

We have formulated a general framework involving censoring in a sensor network, and resource constraints. We allow the sensors to censor based on some side information, while taking into account a general cost function that depends only on the side information and the transmission policy used by the sensor. We allow the sensors to cooperate with each other and show that for

a Neyman–Pearson formulation, such cooperation is not necessary in the asymptotic regime of a large number of sensors and small Type I error. Every sensor can independently use the same (generally, randomized) local policy. An optimal policy is found by maximizing $\mathbb{E}_0[\xi_{V_1}(R_1)I(R_1, \gamma_{V_1})]$ subject to the constraint that $\mathbb{E}_0[\xi_{V_1}(R_1)\rho(R_1, \gamma_{V_1})] \leq c$. This maximization captures the *tradeoff* between the Type II error exponent and the resource constraint.

In the Bayesian context, we have shown that, in the absence of sensor cooperation, asymptotic optimality is obtained by dividing the sensors into two groups, with every sensor in each group using the same pure policy. We have also shown how to find optimal strategies in some special cases, and the relationship of our results to other works.

Most of our results can be extended in various directions. For example, we may have multiple resource constraints. With k constraints, we will generally need the local randomization variable V_1 to have a range of cardinality $k + 1$ (Neyman–Pearson case), or to divide the sensors into $k + 1$ groups, with every sensor in each group using the same policy (Bayesian case). Extensions to the case of more than two hypotheses are also possible, along the lines of [5].

APPENDIX

Proof of Lemma 8: Let

$$a = \sup \left\{ \int f(v)\mu(dv) : \mu \in \mathcal{M}, \int g(v)\mu(dv) \leq c \right\}$$

and

$$b = \sup \{ u f(v_1) + (1 - u) f(v_2) : u \in [0, 1], v_1, v_2 \in \mathcal{V}_0, u g(v_1) + (1 - u) g(v_2) \leq c \}.$$

Clearly, $a \geq b$. It remains to show that $a \leq b$. Assume that $a < \infty$, and fix an $\epsilon > 0$. Choose a $\mu \in \mathcal{M}$ such that $\int g(v)\mu(dv) \leq c$, and

$$a \leq \int f(v)\mu(dv) + \epsilon. \quad (35)$$

Let C be the convex hull of the set $\{(f(v), g(v)) : v \in \mathcal{V}_0\}$. It can be shown that the point $(\int f(v)\mu(dv), \int g(v)\mu(dv)) \in C$ (see, for example, [15, p. 25]). Therefore, there exists a finite set of points $\{v_1, \dots, v_k\} \subset \mathcal{V}_0$, and nonnegative scalars $\alpha_1, \dots, \alpha_k$ summing to 1, such that

$$\int f(v)\mu(dv) = \sum_{i=1}^k \alpha_i f(v_i)$$

and

$$\int g(v)\mu(dv) = \sum_{i=1}^k \alpha_i g(v_i).$$

Consider the linear program in which we maximize $\sum_{i=1}^k \alpha_i f(v_i)$ over the nonnegative scalars $\alpha_1, \dots, \alpha_k$ so that $\sum_{i=1}^k \alpha_i g(v_i) \leq c$ and $\sum_{i=1}^k \alpha_i = 1$. From a well-known result in linear programming [11], there exists an optimal solution to this linear program with at most two of the α_i being nonzero. Hence, $\int f(v)\mu(dv) \leq b$. From (35), we have

$a \leq b + \epsilon$, and since ϵ is arbitrary, we obtain $a \leq b$. The case where $a = \infty$ has a similar proof. The proof is now complete.

Proof of Lemma 9: For any fixed pure policy π and a given $s \in I = (\tau_1, \tau_2)$, it is well known that the first and second derivatives of $\Phi(s, \pi)$, w.r.t. s , are finite. What needs to be proved is that these derivatives are *uniformly* bounded for all policies π .

For a given $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0}$, let

$$\begin{aligned} \varphi(s) &= \mathbb{E}_0 \left[(\xi_V(R_1) e^{\Lambda(s, R_1, \gamma_V)} + 1 - \xi_V(R_1)) \ell_\mu^s(R_1) \right] \\ &= \mathbb{E}_0 \left[\xi_V(R_1) (\ell_{10}^V(Y_1 | R_1) \ell_\mu(R_1))^s \right. \\ &\quad \left. + (1 - \xi_V(R_1)) \ell_\mu^s(R_1) \right] \end{aligned} \quad (36)$$

so that $\Phi(s, \pi) = \log \varphi(s)$. Then, for each $s \in I$,

$$\frac{d^2}{ds^2} \Phi(s, \pi) = \frac{1}{\varphi(s)} \frac{d^2}{ds^2} \varphi(s) - \left(\frac{d}{ds} \Phi(s, \pi) \right)^2. \quad (37)$$

To prove the lemma, it suffices to show that for all policies π and all $s \in I$, $\varphi(s)$ is uniformly bounded away from 0, and $\frac{d}{ds} \varphi(s)$ and $\frac{d^2}{ds^2} \varphi(s)$ are uniformly bounded. We do this in several steps below. To keep the notation simple, we will abbreviate $\ell_{10}(X_1 | R_1)$ to ℓ_{10} , $\ell_{10}^V(Y_1 | R_1)$ to ℓ_{10}^V , and $\ell_\mu(R_1)$ to ℓ_μ .

a) For every pure transmission policy $\gamma \in \Gamma$, and every $s \in [0, 1]$ and $r \in \mathcal{R}$, we have $\Lambda(s, r, \gamma) \leq 0$, because $\Lambda(0, r, \gamma) = \Lambda(1, r, \gamma) = 0$, and $\Lambda(s, r, \gamma)$ is a convex function of s for each r (see [7, Lemma 2.2.5]). Therefore, using Jensen's inequality

$$\begin{aligned} \varphi(s) &\geq \mathbb{E}_0 \left[e^{\Lambda(s, R_1, \gamma_V)} \ell_\mu^s \right] = \mathbb{E}_0 \left[(\ell_{10}^V \ell_\mu)^s \right] \\ &\geq \mathbb{E}_0 \left[(\ell_{10} \ell_\mu)^s \right] = \psi(s). \end{aligned} \quad (38)$$

Using the same technique as in the proof of Proposition 3 of [5], we can show that $\psi(s) \geq \epsilon \in (0, 1)$, for all $s \in I$, hence, $\varphi(s) \geq \epsilon$. This implies that $|\Phi(s, \pi)| \leq |\log \epsilon|$, for all $s \in I$.

b) Let $\Psi(s) = \log \psi(s)$. Then, from (38), for all $s \in [0, 1]$

$$\Phi(s, \pi) = \log \varphi(s) \geq \Psi(s).$$

It is easily shown that $\Phi(0, \pi) = \Phi(1, \pi) = \Psi(0) = \Psi(1) = 0$, and that both functions are convex in $s \in [0, 1]$. Then, for $s \in I$

$$\begin{aligned} \left| \frac{d}{ds} \Phi(s, \pi) \right| &\leq \sup_{s \in I} \left| \frac{d}{ds} \Psi(s) \right| \\ &= \sup_{s \in I} \frac{1}{\psi(s)} |\mathbb{E}_0 [(\ell_{10} \ell_\mu)^s \log(\ell_{10} \ell_\mu)]| \\ &\leq \frac{1}{\epsilon} \max \{ |\mathbb{E}_j [(\ell_{10} \ell_\mu)^{\tau_k} \log(\ell_{10} \ell_\mu)]| : k=1, 2 \} < \infty. \end{aligned} \quad (39)$$

(We used here the convexity of Ψ , which implies that the magnitude of its derivative is maximized at one of the end

points τ_k .) The finiteness of the RHS of (39) follows from Assumption 3 v).

- c) From b) above, $|\frac{d}{ds}\varphi(s)| < \infty$ for all $s \in I$. So, for the same reason as in [7, Lemma 2.2.5 (c)], we can differentiate $\varphi(s)$ twice under the expectation operator. Hence, from (36), we have

$$\begin{aligned} \frac{d^2}{ds^2}\varphi(s) &= \mathbb{E}_0 [\xi_V(R_1)(\ell_{10}^V \ell_\mu)^s \log^2(\ell_{10}^V \ell_\mu)] \\ &\quad + \mathbb{E}_0 [(1 - \xi_V(R_1)) \ell_\mu^s \log^2 \ell_\mu] \\ &\leq \mathbb{E}_0 [(\ell_{10}^V \ell_\mu)^s \log^2(\ell_{10}^V \ell_\mu)] + \mathbb{E}_0 [\ell_\mu^s \log^2 \ell_\mu] \\ &= \mathbb{E}_0 \left[\frac{d^2}{ds^2} \mathbb{E}_0 [(\ell_{10}^V \ell_\mu)^s | V] \right] + \frac{d^2}{ds^2} \mathbb{E}_0 [\ell_\mu^s] \leq 2b \end{aligned}$$

from Assumption 3 vi).

The steps above show that $\frac{d^2}{ds^2}\Phi(s, \pi)$ is uniformly bounded and completes the proof of the lemma.

ACKNOWLEDGMENT

We wish to thank the Associate Editor and the anonymous reviewers for their suggestions and careful reading of the manuscript. We also wish to thank Prof. Devavrat Shah for his helpful comments.

REFERENCES

- [1] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.

- [2] S. Appadwedula, V. V. Veeravalli, and D. Jones, "Energy-efficient detection in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 693–702, Apr. 2005.
- [3] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Select. Areas Commun., Special Issue on Wireless Sensor Networks*, vol. 22, no. 6, pp. 1007–1015, Jun. 2004.
- [4] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.
- [5] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, vol. 1, pp. 167–182, 1988.
- [6] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.
- [7] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer-Verlag, 1998.
- [8] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.
- [9] R. R. Bahadur, S. L. Zabell, and J. C. Gupta, "Large deviations, tests, and estimates," in *Asymptotic Theory of Statistical Tests and Estimation: In Honor of Wassily Hoeffding*, I. M. Chakravarti, Ed. New York: Academic, 1980.
- [10] R. Durrett, *Probability: Theory and Examples*, 2nd ed. New York, NY: Duxbury, 1995.
- [11] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, 1st ed. Belmont, MA: Athena Scientific, 1997.
- [12] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. New York: Cambridge Univ. Press, 2005.
- [13] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels, I," *Inf. Contr.*, vol. 10, pp. 65–103, 1967.
- [14] D. Stroock, *Probability Theory An Analytic View*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
- [15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.