Optimal Sensor Selection in Binary Heterogeneous Sensor Networks

Marcelino Lázaro, Member, IEEE, Matilde Sánchez-Fernández, Member, IEEE, and Antonio Artés-Rodríguez, Senior Member, IEEE

Abstract—We consider the problem of sensor selection in a heterogeneous sensor network when several types of binary sensors with different discrimination performance and costs are available. We want to analyze what is the optimal proportion of sensors of each class in a target detection problem when a total cost constraint is specified. We obtain the conditional distributions of the observations at the fusion center given the hypotheses, necessary to perform an optimal hypothesis test in this heterogeneous scenario. We characterize the performance of the tests by means of the symmetric Kullback–Leibler divergence, or $J$-divergence, and showing the linearity of the $J$-divergence with the number of sensors of each class, we found that the optimal proportion of sensors is “winner takes all” like. The sensor class with the best performance/cost ratio is selected.

Index Terms—Energy scaling, sensor networks, sensor selection.

I. INTRODUCTION

The advent of distributed sensor networks has been raising new technical challenges during the last years in multiple disciplines [1] and has triggered enormous interest and development in all the research areas related with distributed and collaborative information transmission and processing [2], [3], such as signal processing, communication theory, networking, data fusion, power control, etc. Multiple contributions have been made to characterize the fundamental performance limits in large-scale randomly deployed wireless sensor networks. In particular, the detection problem has received a lot of attention (see, for instance, [2] and [4]–[8]).

In many applications, the sensor network is intended to have a long lifetime. In this case, energy consumption becomes a crucial aspect in networks with self-powered sensors because it limits the lifetime of the network and seriously affects the maintenance cost (to replace batteries in a large network, when possible, can be a high demanding task). The lifetime of the network has been assessed in different contributions, giving birth to multiple definitions for the lifetime of a sensor network, results measuring the lifetime, and presenting methods to increase it (see, for instance, [9] and references therein).

The importance of energy consumption has raised energy control strategies at all levels of the system to preserve the limited energy budget. This global energy awareness [10] considers all components in the network. Some examples are the development of low-power hardware components (processors, converters, etc.), specific operating systems [11], the definition of efficient routing protocols [12], and the design of the network architecture; for instance, the Sensor Network with Mobile Agents (SENMA) network proposes to use a many-to-one scheme based on using mobile access points [13]. This network has gained a significant attention because of its simplicity, scalability and energy efficiency. For a more detailed description of energy consumption at the different network levels, see, for instance, [10] or [14].

Since the most energy-demanding task, according to [15], is the wireless transmission, an interesting alternative to economize energy is to employ censored transmission schemes, i.e., schemes where, according to some criteria, the transmission of the available information is avoided. Several censoring schemes have been proposed in the literature preventing transmissions when the local information is not considered informative or useful enough. Typically, transmission of the local measurement is prevented when the likelihood ratio of a given node is out of certain bounds [16]–[18]. Other approaches even avoid that some sensors, determined by the fusion center from global side information in cooperative networks, take measurements (and therefore prevent transmission of that sensors) [19]. Also, sequential tests have been proposed at the local nodes [20] to only transmit when the certainty about the decision is high enough. In this paper, we will use the term “censored” to make reference to the scenario when sensors, after taking a measurement and making a local decision, only will try to transmit if a positive local decision happens. This strategy is similar to the transmission strategies proposed in [16]–[18], as an extreme case, and also resembles the “two-level quantization” proposed in [20]. However, our setting is not limited to the use of local likelihood ratios or sequential tests, respectively, thus allowing one to employ widely used, non-parametric, learning-based local detectors, like in [21].

A distinctive feature of wireless sensor networks is the flexibility they offer in their design. A wide variety of sensor classes, all of them with different performance characteristics and costs,
can be deployed in the network. In the design phase, it has to be decided how to spend the available resources to achieve the best possible performance.

Some efforts have already been made by several authors to determine the selection of the local detection rules (the kind of sensors to be used). In 1988, Tsitsiklis [22] showed that when the number of sensors is arbitrarily large, the optimal binary decentralized detection is achieved by identical local detection rules, and this result has been recently extended [7], showing that using an identical transmitter is also optimal. In [18], authors analyze the optimal decision strategy with a resource constraint (considering measurement and transmission costs), finding that randomization over the choice of measurement and over the choice of when to transmit achieves the best performance.

An interesting question is if it is better to use a lot of cheap and relatively low-performance sensors, a few expensive and higher performance sensors, or a combination of different sensors with different costs and performance. This question has been addressed in [6] and [7], where authors indicate that usually the option of more low-cost sensors performs better if density allows one to maintain the assumption of conditional independence, and in [23], which analyzes if it is better to quantize relatively few sensors finely or as many as possible coarsely and finds that the optimal choice depends on the signal-to-noise ratio.

With respect to other important aspects, until recently [24], [25], there were no results considering the existence of transmission errors or the implementation of energy-scaling procedures, and to the best of our knowledge, there are no results about how to select the sensors when an economic cost constraint is considered, as usually happens in a real application.

In [26] and [27], we formulated the detection problem in networks with binary sensors, not necessarily based on local likelihood ratio test (LRT) decisions, by using their positions and probabilities of local detection. We introduced two alternative transmission schemes that allow the control of the energy consumption of the network. In both cases, a parameter is included to achieve energy scalability, allowing one to control the tradeoff between lifetime of sensors and performance. This parameter defines the probability that a sensor senses at a given sampling instant and can be dynamically updated to fit the current network requirements (trading off performance and network lifetime). The operation of this parameter, which is fully explained in Section III, can be easily integrated in a “sleep” and “wakeup” schedule of sensor nodes, a technique that has been demonstrated to further increase the lifetime of the network [28], [29]. In [26], to minimize energy consumption, a censoring scheme was presented based on the idea that only sensors with local positive decisions try to transmit their positions. In [27], the formulation is extended to the case where all sensors that sense in a sampling instant try to transmit, and both strategies are compared. We called these schemes, which were developed for homogeneous sensor networks, censored and uncensored transmission schemes, respectively. Furthermore, unlike in most of the settings used to analyze the detection performance in sensor networks, our schemes do not consider the common assumption that the information sent by every sensor is reliably conveyed to the fusion center. The probability of transmission errors is introduced in the definition of the schemes and therefore in the corresponding analysis (see Section III).

In this paper, we consider the problem of designing a network dedicated to a detection problem when different types of sensors, with different costs and characteristics, are available and when a cost limitation is imposed. Sensors will provide their binary (positive or negative) local decision to the fusion center (binary sensors). We assume that the local decision rules and working parameters (such as energy scaling parameters) have been fixed in advance. Note that the total cost defined in this setting only depends on the number of sensors, and not on the sensing and transmission policies or strategies, as in [18] or [19]. The goal is to find the network configuration, in terms of the number of sensors of each class, providing the best possible detection performance at the fusion center. The simplest case, considering only two classes of sensors, was analyzed in [30] for a censored transmission scheme. Here, we extend the analysis to the general case of \( N_c \) classes of sensors and consider two different transmission schemes, censored and uncensored, both including the modeling of transmission errors and an energy scaling parameter to control the network lifetime. To characterize any kind of binary sensor (not necessarily based on LRT), we use a function \( p_y \) to model the probability of local detection (including false alarms) for each class of sensors. This characteristic of the sensor varies as a function of the distance between the sensor and the source or target to be detected [31], and therefore of sensor position. At the fusion center, the knowledge of each \( p_y \), sensor positions along with local decisions, and the total number of sensors of each class allows one to obtain the distribution of the observations conditioned to each hypothesis. Hence, the optimal fusion rule for a detection problem is a likelihood ratio test. The figure of merit used for the selection of the best network configuration will be the symmetric Kullback–Leibler (KL) divergence [32] between the conditional distributions of the observations for each hypothesis. The optimal solution will maximize this measure subject to the cost constraint.

The rest of this paper is organized as follows. First, we formulate the problem, establish the notation, and introduce the measure that will be used to evaluate the performance of the detection test in Section II. Section III defines the two transmission schemes employed to economize energy. The conditional distributions, the definition of the observations of the hypothesis tests, and the analytical expressions for the KL divergence are obtained in Section IV. In Section V, we formulate and solve the sensor selection problem under the cost constraint. Section VI presents some simulation results. Section VII concludes this paper by summarizing and discussing its main contributions.

II. Problem Statement and Notation

We consider a binary sensor network randomly deployed with a uniform distribution over a region \( \mathcal{D} \subset \mathbb{R}^2 \) of finite area \( S \). The network is composed of binary sensors of \( N_c \) different classes, each class having different cost and characteristics: \( C^i \) denotes the cost of an individual sensor of class-\( i \), and sensors of class-\( i \), \( i \in \{1,2,\ldots,N_c\} \), are characterized by using the probability of local (at sensor level) detection function

\[
p_y^i(z|\mathbf{x},\alpha^i) = P(Y = 1|Z = z, \mathbf{X} = \mathbf{x}) \quad (1)
\]
which represents the probability of a positive local decision \((Y = 1)\) of an agent located at coordinates \(z\) for a sensor of class \(i\) located at position \(x\). Parameter \(\alpha^i\) denotes the probability of false alarm for sensors of this class. This kind of probability of local detection is a function of the distance between the sensor and the target, and its main properties, along with the joint distributions of \(X\) and \(Y\), were introduced in [31]. We assume that the position and class of every sensor are known at the fusion center. Positions and classes can be obtained at an initial calibration phase during the deployment of the network. In this case, to know the position and class of a given sensor, only a simple identification tag has to be transmitted (position and class will be implicit), and sensors do not have to incorporate a GPS chip or any other location device. We also assume that all sensors of a given class apply the same binary detection rule, not necessarily based on an LRT, and not necessarily known at the fusion center. We remark that in our setting, the only necessary information at the fusion center about a sensor of a given class is its local decision, location, and the probability of local detection for that class. In practice, this function can be obtained analytically (based on the knowledge about the physical process for measurement and local decision rule) or by means of empirical measurements. Finally, it is also assumed that the fusion center knows the number of sensors of each class that are deployed in the network. This knowledge will be used to obtain the distributions of the number of available measures at the fusion center, which will be defined in Section IV.

Let \(\ell\) denote the number of sensors of the \(i\)th class that are randomly deployed with a uniform distribution in area \(\mathcal{D}\). The exploration of \(\mathcal{D}\), which can be either automatic or beacon driven, can potentially produce the following data set:

\[
\left\{ \{x^1_J, y^1_J\}_{j=1}^{\ell_1}, \{x^2_J, y^2_J\}_{j=1}^{\ell_2}, \ldots, \{x^{N_c}_J, y^{N_c}_J\}_{j=1}^{\ell_{N_c}} \right\}
\]  

(2)

with \(x^i_J \in \mathcal{D}\) being the position of the \(j\)th sensor of class \(i\) and \(y^i_J \in \{0, 1\}\) being its binary local decision. We assume a many-to-one transmission strategy with no hops to send this information to the fusion center, although this assumption about the network architecture does not limit the scope of the results. Not all information in (2) will be available at the fusion center to perform the hypothesis test. The reason is twofold. First, transmission errors can happen. Secondly, some techniques to economize power can prevent the sensors to sense and transmit at every sampling time. This will be treated in detail in Section III.

We are interested in a detection problem that can be formulated as a binary hypothesis test. In [26] and [27], we formulated the problem of deciding if a target is present or not at a specific position \(z\) \(\in \mathcal{D}\). Here we will also treat the extended problem of detecting if a target is in a subregion \(\mathcal{D}' \subset \mathcal{D}\). It is possible to use a common framework for both scenarios by defining (abusing of the notation) an effective probability of local detection function \(p^i_z(x, \alpha^i)\). If the goal is to detect a target at \(z\), this probability is given by

\[
p^i_z(x, \alpha^i) = p^i_z(z, x, \alpha^i)
\]  

(3)

and if the goal is to detect a target in a region \(\mathcal{D}'\)

\[
p^i_z(x, \alpha^i) = \frac{1}{\mathcal{S}'} \int_{\mathcal{D}'} p^i_z(z, x, \alpha^i) dz
\]  

(4)

where \(\mathcal{S}'\) is the area of region \(\mathcal{D}'\). The problem of defining the optimal fusion rule to decide if one or several targets are present in region \(\mathcal{D}\) by fusing several local decision measures (deciding over several subregions \(\mathcal{D}'\)) is out of the scope of this paper (in fact, it is a current open research line). Therefore, the two hypotheses are defined as follows:

- Null hypothesis \(H_0\) if a target is not present at \(z\) (or in region \(\mathcal{D}'\));
- Alternative hypothesis \(H_1\) if a target is present at \(z\) (or in region \(\mathcal{D}'\)).

Denoting by \(\Theta\) a vector of random variables modelling the observations, and by \(\theta\) an instance of this variable, the detection is based on the likelihood ratio, typically expressed in logarithmic scale [log-likelihood ratio (LLR)] as follows:

\[
\Gamma = \ln \frac{f_\Theta[H_1|H_1]}{f_\Theta[H_0|H_0]} \geq \tau.
\]  

(5)

Threshold \(\tau\) can be selected according to the classical Bayes or Neyman–Pearson (NP) criteria [33], [34]. In many real applications, the prior probability of each hypothesis is rarely known, and probability of false alarm (PFA) is a serious concern. In this case, the NP test is the natural choice: threshold \(\tau\) is selected to fix a given PFA, and the test provides the higher probability of detection for this PFA. In the following, we will assume that the fusion center performs an NP test.

With the premises described above, the goal is to obtain the optimal proportion of sensors of different classes when the total cost dedicated to sensors is constrained by a maximum cost \(C\)

\[
\sum_{i=1}^{N_c} \ell_i \cdot C^i \leq C,
\]  

(6)

Analytically evaluating the performance of a hypothesis test (probability of false alarm or probability of misdetection) is frequently intractable. For that reason, either bounds or measures that are somehow related to those performance parameters are typically used [34]. The Ali–Silvey class of distance measures [35], particularly, the Bhattacharyya distance and the \(J\)-divergence, have been used as tractable design criteria because of their link with performance [36]–[38]. In our approach, we propose to use the \(J\)-divergence, also known as symmetric KL divergence [32], between the conditional probability density functions \(f_\Theta[H_1|H_1]\) and \(f_\Theta[H_0|H_1]\) as a performance-related measure. In the following, we will sometimes use the notation \(J(H_0|H_1)\) for short. The \(J\)-divergence has been used as a figure of merit, for instance, in problems such as the design of generalized quantizers for binary decision systems [37], or more recently in the design of power allocation in distributed detection with wireless sensor networks [39], where, similarly to our approach, this measure is used to define the power-allocation problem as a constrained optimization problem. This divergence is the symmetrized extension of the conventional KL divergence

\[
J(H_0|H_1) = D(H_0||H_0) + D(H_0||H_1)
\]  

(7)

where \(D(H_i||H_j)\) denotes the (nonsymmetric) KL divergence between the conditional probability density functions under hypotheses \(H_i\) and \(H_j\). From the definition of the KL divergence
The expected value under hypothesis $H_j$, it can be seen that the $J$-divergence is

$$J(H_0|H_1) = E_{H_1} \{ \Gamma \} - E_{H_0} \{ \Gamma \} \tag{8}$$

i.e., the difference between the mean value of the LLR in (5) under both hypotheses. In order to further validate this figure of merit in our particular setting, simulation experiments will be provided in Section VI. In all these experiments, the proposed figure of merit provides the same results (sensor selection) as the misdetection probability for NP tests.

III. DEFINITIONS FOR THE TRANSMISSION STRATEGIES TO ECONOMIZE ENERGY

As we explained before, power consumption is a key aspect in networks with sensors powered by batteries. The replacement of batteries is critical in the maintenance of such a network, and to maximize the mean lifetime of the power source of a given sensor becomes a serious concern in order to reduce the maintenance cost. In [26] and [27], we introduced two different transmission schemes aimed at increasing the lifetime of every sensor and thus the lifetime of the whole network. We call them censored and uncensored transmission schemes, respectively, and these were formulated for homogeneous sensor networks, built with identical sensors. Now, we will extend them to heterogeneous sensor networks using sensors from $N_c$ different classes. Fig. 1 represents the block diagram that depicts the way observations are generated for each of the proposed transmission schemes.

In both schemes we introduce a parameter to control the tradeoff between lifetime and performance, thus providing an energy-scaling mechanism. We call this parameter the probability of sensing $p_s$. At every sampling instant, not all sensors in the network will sense but every sensor will randomly decide to sense or not with a probability $p_s$ of sensing. Decreasing this parameter allows one to enlarge the lifetime of the batteries of each sensor [by a factor $(1-p_s)$, approximately], at the price of a lower detection performance. This parameter can easily be integrated in a sleep and wakeup schedule. Taking this into account, we have $\ell^*_s \leq \ell^i$ sensors of class-$i$ that sense at a given sampling instant. The censored scheme implements another mechanism for energy scaling: only the $\ell^*_s \leq \ell^i$ sensors with a positive local decision ($y^i_j = 1$) will try to access the channel.

Also, in both schemes, we consider another parameter $p_e$ to model the probability of transmission error (meaning that the information does not reach the fusion center or is discarded if errors are detected). This parameter models the joint effect of the medium access, transmission, and error detection (with erasure). Thus the number of sensors with available data at the fusion center is $\ell^*_e \leq \ell^i$.

We assume that all sensors of a given class have the same $p_e$. This is reasonable if a mobile access point is used (as proposed in SENMA [13], for instance), but for fixed access points the probability of transmission error will be dependent on sensor locations. The provided results can be easily extended for the case of individual values of $p_e$ for each sensor. This will only change (36) (the second binomial will be replaced by the product with individual values of $p_e$), (40), (43)–(46), and (50), where (1–$p_e$) will be replaced by the averaged term considering the values of $p_e$ for each sensor.

To simplify notation, in the following, we will consider a common value of $p_s$ and $p_e$ for all classes of sensors. The extension to different values for each class is straightforward.

A. Uncensored Transmission Scheme

In this transmission scheme, all sensors that sense try to transmit their information to the fusion center. To model the errors in the medium access and transmission procedures, we use a probability of transmission error $p_e$. Finally, only the information of $\ell^*_s \leq \ell^i$ sensors of class-$i$, $i \in \{1, \ldots, N_c\}$, is available at the fusion center. Instead of all the potential information given by (2), the information available at the fusion center is

$$\left\{ \{x^1_{j}, y^1_{j} \}_{j=1}^{\ell^*_s}, \{x^2_{j}, y^2_{j} \}_{j=1}^{\ell^*_s}, \ldots, \{x^{N_c}_{j}, y^{N_c}_{j} \}_{j=1}^{\ell^*_s} \right\} \tag{9}.$$

The readings of sensors of a given class can be divided in $\ell^*_a$ positive readings ($y^i_j = 1$) and $\ell^*_n$ negative readings ($y^i_j = 0$), where $\ell^*_a = \ell^*_e + \ell^*_n$. Without loss of generality, we will organize the received readings with the positive ones before the negative ones, i.e.,

$$\left\{ \{x^1_{j}, y^1_{j} \}_{j=1}^{\ell^*_a}, \{x^2_{j}, y^2_{j} \}_{j=1}^{\ell^*_a}, \ldots, \{x^{N_c}_{j}, y^{N_c}_{j} \}_{j=\ell^*_a+1}^{\ell^*_a} \right\} \tag{10}.$$

B. Censored Transmission Scheme

The most demanding task in terms of energy consumption in a sensor is the wireless transmission [15]. For the sake of saving energy, in this transmission scheme only sensors with a positive local decision ($y^i_j = 1$) try to transmit their data to the fusion center. As in the uncensored transmission scheme, we consider that each sensor that tries to transmit has a probability of transmission error $p_e$. At the fusion center, only the positions of the $\ell^*_a \leq \ell^i$ sensors of class-$i$ that achieved a successful transmission are available. In this scheme, $\ell^*_a = \ell^*_e$, because now $\ell^*_n = 0$ by definition. Therefore, the information available at the fusion center are those positions

$$\left\{ \{x^1_{j} \}_{j=1}^{\ell^*_a}, \{x^2_{j} \}_{j=1}^{\ell^*_a}, \ldots, \{x^{N_c}_{j} \}_{j=1}^{\ell^*_a} \right\} \tag{11}.$$
and the knowledge of the positive local decision $y_j = 1$ for all those sensors.

IV. DISTRIBUTIONS AND DIVERGENCES FOR CENSORED AND UNCENSORED SCHEMES

In this section, we will present the conditional distributions for the observations under each hypothesis $f_{\Theta|H_u}(\Theta|H_u)$, $u \in \{0, 1\}$, and the performance measure $J(H_0||H_1)$.

Some definitions are helpful to the further developments. Let $L = [L_1, \ldots, L_{N_c}]$ denote the vector of random variables modelling the number of sensors of each class deployed in the network, $L_u = [L_{a_1}, \ldots, L_{a_{N_c}}]$ be the vector of random variables modelling the number of successful transmissions for each class, and $L_d = [L_{a_1}^{d_1}, L_{a_2}^{d_2}, \ldots, L_{a_{N_c}}^{d_{N_c}}]$ denote the vector of random variables modelling the number of sensors with a successful transmission and positive or negative readings (only useful for the uncensored scheme). Vectors $\ell_a = [\ell_a^1, \ldots, \ell_a^{N_c}]$, $\ell_d = [\ell_{a_1}^{d_1}, \ell_{a_2}^{d_2}, \ldots, \ell_{a_{N_c}}^{d_{N_c}}]$ are used to denote their corresponding realizations for a given observation.

A. Observations and Distributions—Uncensored Scheme

For the uncensored scheme, the hypothesis test is based on the following vector of observations:

$$\Theta = [\Theta_1, \ldots, \Theta_{a_1}, \ldots, \Theta_1^{N_c}, \ldots, \Theta_a^{N_c}, \ell_a, \ell_d]$$

and we denote by $\Theta$ the random variable used to model these observations. The vector of observations consists of, ordered by class, the positions of the sensors of each class that achieved a successful transmission up to the fusion center (ordered inside each class by allocating first the positions of sensors with positive readings and then those with negative readings), along with the number of readings, positive and negative, of each class. We want to remark that the decisions $y_j$ do not have to be included in the observation vector because the arrangement of sensor positions and vector $\ell_d$ make them implicit.

Assuming conditional independence of the sensors under each hypothesis, and taking into account the arrangement of sensor positions, the conditional distribution of the observations under hypothesis $H_u$ is

$$f_{\Theta|H_u}(\Theta|H_u) = \prod_{i=1}^{N_c} \prod_{j=1}^{\ell_{a_i}} f_{X_i|H_u}(x_i|H_u, 1)$$

$$\cdot \prod_{k=\ell_{a_1}+1}^{\ell_a} f_{X_k|H_u}(x_k|H_u, 0) \cdot f_{L_d|L_H}(\ell_d|\ell, H_u).$$

(13)

Appendix A derives the expressions for $f_{X_i|H_u}(x_i|H_u, y_j)$ and $f_{L_d|L_H}(\ell_d|\ell, H_u)$.

B. Observations and Distributions—Censored Scheme

The hypothesis test for the censored scheme is based on the following vector of observations:

$$\Theta = [\Theta_1, \ldots, \Theta_{a_1}, \ldots, \Theta_1^{N_c}, \ldots, \Theta_a^{N_c}, \ell_a]$$

which includes, in this order, the positions of the $\ell_{a_i}$ sensors of class-$i$, $i \in \{1, \ldots, N_c\}$, that achieved a successful transmission, ordered by class, and the number of sensors of each class that are available at the fusion center $\ell_{a_i}$, $i \in \{1, \ldots, N_c\}$.

Assuming conditional independence of the sensors under each hypothesis, and taking into account that the positions correspond to sensors with a positive local decision, the conditional distribution of the observations under hypothesis $H_u$ is

$$f_{\Theta|H_u}(\Theta|H_u) = \prod_{i=1}^{N_c} \prod_{j=1}^{\ell_{a_i}} f_{X_i|H_u}(x_i|H_u, 1) \cdot f_{L_d|L_H}(\ell_d|\ell, H_u).$$

(15)

Appendix A contains the expression for $f_{L_d|L_H}(\ell_d|\ell, H_u)$.

C. KL Divergences—Uncensored Scheme

As explained in Section II, the performance measure is the symmetric Kullback–Leibler divergence between the conditional distributions under the null and alternative hypotheses. From the distributions given in Section IV-A and Appendix A, the (nonsymmetric) divergences included in $J(H_0||H_1)$ [see (7)] for the uncensored scheme are

$$D(H_0||H_1) = \sum_{\ell_d} f_{L_d|L_H}(\ell_d|\ell, H_0)$$

$$\cdot \left\{ \ln \frac{f_{L_d|L_H}(\ell_d|\ell, H_0)}{f_{L_d|L_H}(\ell_d|\ell, H_1)} + \ln S \cdot \sum_{i=1}^{N_c} (\ell_{a_i} + \ell_{d_i})$$

$$- \frac{1}{S} \sum_{i=1}^{N_c} \ell_{a_i} \int_D f_{X|H_1}(x|H_1, 1) dx$$

$$+ \frac{1}{S} \sum_{i=1}^{N_c} \ell_{d_i} \int_D f_{X|H_1}(x|H_1, 0) dx \right\}$$

(16)

and

$$D(H_1||H_0) = \sum_{\ell_d} f_{L_d|L_H}(\ell_d|\ell, H_1)$$

$$\cdot \left\{ \ln \frac{f_{L_d|L_H}(\ell_d|\ell, H_1)}{f_{L_d|L_H}(\ell_d|\ell, H_0)} + \ln S \cdot \sum_{i=1}^{N_c} (\ell_{a_i} + \ell_{d_i})$$

$$+ \sum_{i=1}^{N_c} \ell_{d_i} \int_D f_{X|H_1}(x|H_1, 1) \ln f_{X|H_1}(x|H_1, 1) dx$$

$$+ \sum_{i=1}^{N_c} \ell_{a_i} \int_D f_{X|H_1}(x|H_1, 0) \ln f_{X|H_1}(x|H_1, 0) dx \right\}$$

(17)
where we have denoted

\[
\sum_{\ell} \equiv \sum_{\ell_1=0}^{\ell_1} \sum_{\ell_2=0}^{\ell_2} \sum_{\ell_3=0}^{\ell_3} \cdots \sum_{\ell_{N_c}=0}^{\ell_{N_c}} \prod_{n=0}^{N_c} \ell_{n}^{\ell_{n}} = 0 .
\]

(18)

D. KL Divergences—Censored Scheme

Using the distributions given in Section IV-B and Appendix A, the divergences for the censored scheme are

\[
\begin{align*}
D(H_0||H_1) & = \sum_{\ell} f_{\ell|H_0}(\ell|H_0) \left\{ \ln \frac{f_{\ell|H_1}(\ell|H_1)}{f_{\ell|H_0}(\ell|H_0)} - \ln S \cdot \sum_{i=1}^{N_c} \ell_{i} \right. \\
& \left. - \frac{1}{S} \sum_{i=1}^{N_c} \ell_{i} \int_{D} \ln f_{X|H,Y}(x|H_1,1) dx \right\}
\end{align*}
\]

and

\[
\begin{align*}
D(H_1||H_0) & = \sum_{\ell} f_{\ell|H_1}(\ell|H_1) \left\{ \ln \frac{f_{\ell|H_1}(\ell|H_1)}{f_{\ell|H_0}(\ell|H_0)} + \ln S \cdot \sum_{i=1}^{N_c} \ell_{i} \\
& + \sum_{i=1}^{N_c} \ell_{i} \int_{D} f_{X|H,Y}(x|H_1,1) \\
& \times \ln f_{X|H,Y}(x|H_1,1) dx \right\}
\end{align*}
\]

(19)

(20)

We have denoted

\[
\sum_{\ell} \equiv \sum_{\ell_1=0}^{\ell_1} \sum_{\ell_2=0}^{\ell_2} \sum_{\ell_3=0}^{\ell_3} \cdots \sum_{\ell_{N_c}=0}^{\ell_{N_c}} \prod_{n=0}^{N_c} \ell_{n}^{\ell_{n}}
\]

(21)

V. OPTIMAL SENSOR SELECTION

To optimize the sensor selection strategy, it is necessary first to characterize the symmetric KL divergence \( J(H_0||H_1) \) in terms of the parameters of interest for the uncensored and censored schemes. For both cases, it is shown in Appendix B that the divergence is a linear function of \( \ell_i \) (i.e., \( \ell_i \in \{1, \ldots, N_c\} \), that maximize \( J(H_0||H_1) \) subject to the cost constraint

\[
\sum_{i=1}^{N_c} \ell_i \cdot C^i \leq C.
\]

(22)

Given that the divergence \( J(H_0||H_1) \) is a linear function of \( \ell_i \), the optimal sensor selection problem becomes an integer linear programming (ILP) problem in its standard form [40] with only one linear constraint, which can be written as

\[
\text{Maximize } A^T \ell
\]

\[
\text{subject to } B^T \ell \leq C
\]

\[
\ell \geq 0 \text{ and integer,}
\]

(23)
Vector \( \mathbf{A} = [A^1, A^2, \ldots, A^{N_c}] \) contains the coefficients defining the linear relationship of the objective function with \( \ell^i \) [see (23) and (24)].

Vector \( \mathbf{B} \) defines the linear constraint and contains the individual costs of every class \( \mathbf{B} = [C^1, C^2, \ldots, C^{N_c}] \).

In contrast to linear programming (LP), which can be solved efficiently in the worst case, integer programming problems are in many practical situations (those with bounded variables) NP-hard. However, using the fact that only a linear constraint is applied, the problem can be solved through the linear relaxation problem (LR), which is the linear programming problem obtained from the ILP problem just dropping the integrality restrictions. It is well known that the solution of the LR problem provides a bound for the function to maximize in the ILP problem. Moreover, if the LR problem is optimized for integer variables, then that solution is feasible and optimal for the ILP problem. This is the case in our problem when just a simple assumption is done: all the individual costs of the sensor classes \( C^k \) accomplish that \( C^k / C^{i} \) is integer. In a real and large network, with a large value of \( C \) with respect to the individual costs of sensors, this assumption does not have practical implications.

In an LP problem, geometrically the linear constraints define a convex polyhedron, which is called the feasible region. Since the objective function is linear, and hence a convex function, all local optima are automatically global optima. The linearity of the objective function also implies that an optimal solution can only occur at a boundary point of the feasible region, unless the objective function is constant, when any point is a global maximum. In this case, having only one linear constraint, the feasible region for the LR problem is the hyperplane bounded by joining all edges in the \( N_c \)-dimensional axis given by \( \{\ell^i = C/C^k \} \). Therefore, taking into account the linearity of the objective function, its maximum can only be at one of those edges (see [40, Part VII, Ch. 29]). In this case, the solution takes the form

\[
\begin{align*}
\left\{ \ell^i = \frac{C}{C^k}, \ell^j = 0, j \neq i \right\}
\end{align*}
\]

for some \( i \). If \( \ell^i \) is integer, this is also the optimal solution of the original ILP problem.

Therefore, under these premises, the solution can be easily found evaluating the objective function \( J(H_0 || H_1) \) for \( N_c \) (all possible values of \( i \)) homogeneous networks

\[
J(H_0 || H_1) = A^i \cdot \ell^i = A^i \cdot \frac{C}{C^k}, \; i \in \{1, \ldots, N_c\}
\]

and choosing the class giving the largest divergence.

Without the integer constraint in \( C/C^k \), which, as said before, does not have practical implications in networks with a large number of sensors, the solution of the problem would require applying conventional ILP techniques, such as “branch and bound” [41]. In this case, the solution will be basically the same one but sensors from another class can be selected to complete the cost.

VI. SIMULATION EXPERIMENTS

We have performed some experiments to validate the suitability of the proposed figure of merit for the selection procedure. We have simulated a scenario with two classes of sensors having the following cost parameters:

\[
C = 1000, \quad C^1 = 20, \quad C^2 = 10.
\]

An exponential squared-norm model has been used for the probability of local detection function \( p^i_H( \mathbf{z}, \mathbf{x}, \alpha^i ) \) [30], [42]. This is a typical model that is commonly used in the literature because many physical procedures employed for sensing are characterized by an exponential decay with distance (typical in isotropic propagation). This function is given by

\[
p^i_H( \mathbf{z}, \mathbf{x}, \alpha^i ) = \alpha^i + (1 - \alpha^i - \beta^i ) e^{-\gamma ||z-x||^2}
\]

where \( \alpha^i \) is the PFA, \( \beta^i \) is the probability of misdetection of an agent placed at the same position of the sensor, and \( \gamma \) parameterizes the decreasing exponential rate of \( p^i_H \).

A circular region \( D \) of normalized radius \( R = 1 \), and parameters \( p_s = 0.5 \) and \( p_e = 0.01 \), have been employed. Two different scenarios have been considered.

- **Scenario A**
  - Class 1: \( \alpha^1 = 0.03, \beta^1 = 0.03, \gamma^1 = 2.5 \)
  - Class 2: \( \alpha^2 = 0.1, \beta^2 = 0.1, \gamma^2 = 2.5 \)

- **Scenario B**
  - Class 1: \( \alpha^1 = 0.04, \beta^1 = 0.04, \gamma^1 = 2.5 \)
  - Class 2: \( \alpha^2 = 0.1, \beta^2 = 0.1, \gamma^2 = 3.5 \)

In both scenarios, sensors of Class 1 have better general properties than sensors of Class 2, and therefore the cost per unit is higher. Fig. 2 plots the divergence \( J(H_0 || H_1) \) as a function of the number of sensors of Class 1 \( \ell^1 \). Note that the number of sensors of Class 2 \( \ell^2 \) is given by

\[
\ell^2 = \frac{C - \ell^1 \cdot C^1}{C^2}.
\]

In the censored scheme, it can be seen that for Scenario A, the best choice is to select 100 sensors of Class 2 and zero of Class 1, i.e., to select the option “cheaper sensors”; while for Scenario B, the best choice is the option “more expensive sensors.” The reason is that the performance/cost ratio for each sensor is different in each scenario. We have included the divergences for the uncensored scheme in order to compare it with its censored counterpart. Obviously, the uncensored scheme provides a higher value for \( J(H_0 || H_1) \).

We have also simulated the performance of the Neyman–Pearson tests (5) in both scenarios using a single observation, \( \theta \) in (12) or (14), to decide. These tests use only one observation (of the many sensors). We have selected a fixed probability of false alarm 0.01 for the global test performed at the fusion center, and the corresponding threshold \( \tau \) has been heuristically obtained by simulation to provide this PFA in the test for each configuration (combination of different number of sensors of each class). Fig. 3 compares the
probability of misdetection obtained in each case through $10^6$ independent Monte Carlo experiments with sensors randomly and uniformly distributed in region $D$. Experimental results confirm the conclusions obtained by using $J(H_0||H_1)$ to perform the sensor selection: the “winner takes all” strategy provides the best performance. Figs. 2 and 3 show that increasing (decreasing) the KL divergence corresponds to decrease (increase) the probability of misdetection in the simulated configurations. With the censored scheme, the lower probability of misdetection is achieved by using only sensors of Class 2 in Scenario A and by using sensors of Class 1 in Scenario B. Again, the uncensored scheme clearly outperforms the censored one, at the price of requiring more energy expenditure (it requires more transmissions).

Finally, with the same costs of the previous experiments for the two classes of sensors, again $\gamma^1 = 2.5$ and $\gamma^2 = 3.5$, and considering $\beta^1 = \alpha^1$ and $\beta^2 = \alpha^2$, we have simulated the performance of the two schemes for different values of $\alpha^1$ and $\alpha^2$. Tables I and II show the probability of misdetection and divergence for each pair of values of $\alpha^1$ and $\alpha^2$. The three values in each cell present the results for the option of choosing all sensors of Class 1, to divide the total cost equally between both classes, and having only sensors of Class 2, respectively. It can be seen that the performance is monotonic in all cases and that the maximum value of $J(H_0||H_1)$ fits in all cases the minimum probability of misdetection in both the censored and uncensored schemes.

VII. CONCLUSION

In this paper, we have analyzed the problem of sensor selection in a cost-constrained network with the goal of obtaining the highest possible performance with the specified cost. Two different transmission schemes, aimed at energy saving, have been considered. These schemes include a parameter to model the probability of errors in the transmission procedure and another parameter allowing control of the network lifetime by defining the probability of sensing of a sensor.

The obtained results show that a “winner takes all” like strategy is the best option and provide the measure to decide the winner class. Thus, to design a network with sensors uniformly distributed, when several classes of binary sensors are available, the optimal solution is to design a homogeneous network using only one class of sensors. This class is the one providing the best performance/cost ratio, which is measured by using the symmetric KL divergence, or $J$-divergence. We also provide the mechanism to select the best sensor class considering performance (given by the probability of local detection function for that class and parameters $p_h$ and $p_e$) and cost. Simulation results validate the use of the $J$-divergence as a figure of merit to perform the selection. All these results are in consonance with those obtained in different scenarios for detection problems, like [22].

Here, we have used the same probability of sensing $p_h$ and probability of error $p_e$ in the transmission procedure for all classes of sensors. The main reason has been to avoid an
overloaded notation, but all results can be straightforwardly extended to include a different value of these parameters for each class, if necessary.

Finally, the framework used in this paper can be used in several related applications. Finding the optimal (or minimal) subset of sensors necessary to obtain a given performance, the tuning of the \( p_a \) parameter, key to the energy consumption, to achieve a given performance under energy constraints, or the optimal positioning of new sensors in a network, are only some examples.

APPENDIX A

DISTRIBUTIONS OF POSITIONS AND NUMBER OF SENSORS

The conditional distribution for the position of the sensors \( f_X^j|Y_j(x_j^j, y_j^j) \) is obtained by using the definition of the probability of local detection \( p_d(x, \alpha^c) \). Under the alternative hypothesis, positions of sensors with positive readings are distributed proportionally to \( p_d(x, \alpha^c) \)

\[
f_X^j|Y_j(x_j^j|H_1, 1) = \frac{p_d(x_j^j, \alpha^c)}{\int_D p_d(x, \alpha^c) \, dx}
\]

(31)

and positions of sensors with negative readings are distributed proportionally to one minus \( p_d(x, \alpha^c) \)

\[
f_X^j|Y_j(x_j^j|H_1, 0) = \frac{1 - p_d(x_j^j, \alpha^c)}{\int_D (1 - p_d(x, \alpha^c)) \, dx}.
\]

(32)

Under the null hypothesis, the positions of sensors are uniformly distributed for both kinds of readings

\[
f_X^j|Y_j(x_j^j|H_0, 1) = f_X^j|Y_j(x_j^j|H_0, 0) = \frac{1}{S}.
\]

(33)

Given the number of sensors of each class that are deployed, the number of sensors of each class for which measures are available at the fusion center are independent. In this case, the joint conditional distributions of the number of available measurements are probabilistically modeled as follows:

\[
\prod_{i=1}^{N_c} \prod_{\ell_a=0}^{\ell_i^c} \prod_{\ell_d=0}^{\ell_d^c} L_{\ell_d} \prod_{\ell_d=0}^{\ell_d^c} L_{\ell_d} \prod_{\ell_d=0}^{\ell_d^c} L_{\ell_d} = \prod_{i=1}^{N_c} f_{X_i|Y_i}(x_i^j|y_i, \ell_i, H_i)
\]

(34)

and

\[
\prod_{i=1}^{N_c} f_{X_i|Y_i}(x_i^j|y_i, \ell_i, H_i) = \prod_{i=1}^{N_c} L_{\ell_i} H_{\ell_d} L_{\ell_i} H_{\ell_d}
\]

(35)

for the uncensored and censored schemes, respectively. The distribution for the number of sensors of each class for the uncensored scheme is (see [27])

\[
f_{X_i|Y_i}(x_i^j|y_i, \ell_i, H_i) = \sum_{\ell_a=0}^{\ell_i^c} \left( \frac{\ell_a}{\ell_i} \right) p_a^{\ell_a} (1 - p_a) \ell_i - \ell_a
\]

\[
\cdot \left( \frac{\ell_i}{\ell_a} \right) (1 - p_a) \ell_a + p_c \ell_i - \ell_a \cdot \left( \frac{\ell_d}{\ell_a} \right) \ell_d (1 - p_d) \ell_d^c
\]

(36)

The first binomial models the number of sensors that sense, given \( \ell_i \) and the probability of sensing, \( p_a \); the second binomial represents the number of sensors achieving a successful transmission, given the number of sensors that sense (and therefore that try to transmit), and the probability of transmission error \( p_c \); the third binomial models the number of sensors with a positive local decision that achieve a successful transmission, given the total number of successful transmissions and the probability of detection of a sensor in region \( D \) under the underlying hypothesis \( H_i \).

\[
p_{D|\ell_i} = \frac{1}{S} \int_D p_d(x^j, \alpha^c) \, dx
\]

(37)

and

\[
p_{D|\ell_i} = \alpha^c
\]

(38)

For the censored scheme, the distribution of the number of successful transmissions is (see [27] for details)

\[
f_{L_{\ell_d} | \ell_i, H_i}(\ell_d^c | \ell_i, H_i) = \sum_{\ell_a=0}^{\ell_i} \left( \frac{\ell_a}{\ell_i} \right) (1 - p_a)^{\ell_i - \ell_a} \left( \frac{\ell_i}{\ell_a} \right) p_a^{\ell_a} (1 - p_d)^{\ell_d^c - \ell_a}
\]

(39)

where \( p_{\ell_i, u} \), \( u \in \{0, 1\} \), denotes the probability of a sensor of class \( i \) having a successful transmission under hypothesis \( H_i \). Taking into account that in this scheme only sensors with a positive local decision try to transmit, the probability of a successful transmission is given by

\[
p_{D|\ell_i} = (1 - p_a) \cdot p_{D|\ell_i}
\]

(40)

Taking this into account, the first binomial models the number of successful transmissions when \( \ell_i \) sensors have sensed, and the second binomial models the number of sensor that sense when \( \ell_i \) sensors are deployed and each one senses with probability \( p_a \).

APPENDIX B

DIVERGENCE LINEARITY WITH THE NUMBER OF SENSORS OF EACH CLASS

Uncensored Scheme: We can clearly identify four terms in (16). The first one is the discrete KL-divergence between \( f_{L_{\ell_d} | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}}(\ell_d | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}) \) and \( f_{L_{\ell_d} | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}}(\ell_d | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}) \). Taking into account the independence of the conditional distributions of the number of sensors

\[
\sum_{\ell_d} f_{L_{\ell_d} | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}}(\ell_d | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}) \ln \frac{f_{L_{\ell_d} | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}}(\ell_d | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d})}{f_{L_{\ell_d} | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d}}(\ell_d | L_{\ell_d} H_{\ell_d} L_{\ell_d} H_{\ell_d})}
\]

(41)
where $D^i_{KL}(H_0||H_1)$ is the discrete KL-divergence for the joint distributions of the number of available positive and negative readings of sensors of class-i in the uncensored scheme

$$
D^i_{KL}(H_0||H_1) = \sum_{\ell^i_{ap}} \sum_{\ell^i_{an}} f_{L^i_{ap},L^i_{an} | L} \ell^i_{ap}, \ell^i_{an} | \ell^i, H_0) \cdot \ln \frac{f_{L^i_{ap},L^i_{an} | L} \ell^i_{ap}, \ell^i_{an} | \ell^i, H_0)}{f_{L^i_{ap},L^i_{an} | L} \ell^i_{ap}, \ell^i_{an} | \ell^i, H_1).}
$$

Taking the value for $\ell^i = 0$, for $\ell^i = 1$ and then extending it by induction, it can be shown that

$$
D^i_{KL}(H_0||H_1) = \ell^i \cdot p_s(1 - p_c)
$$

which is also linear with $\ell^i$.

The second term in (16) is $\ln S$ times the addition of the expected value of $\ell^i_{ap} = \ell^i_{ap} + \ell^i_{an}$ for all $i$ given the number of sensors of each class under the null hypothesis. It is straightforward to see that

$$
\sum_{\ell^i_d} f_{L^i_d | L} \ell^i_d, H_0) \cdot \sum_{i=1}^{N_c} \ell^i_{ap} = \sum_{i=1}^{N_c} \ell^i \cdot p_s(1 - p_c)
$$

which is also linear with $\ell^i$.

The last two terms in (16) are also linear with $\ell^i$ because the involved terms are proportional to the expected value of $\ell^i_{ap}$ and $\ell^i_{an}$ given the number of sensors of each class under the null hypothesis, which are

$$
\sum_{\ell^i_d} f_{L^i_d | L} \ell^i_d, H_0) \cdot \sum_{i=1}^{N_c} \ell^i_{ap} = \sum_{i=1}^{N_c} \ell^i \cdot p_s \alpha^i(1 - p_c)
$$

and

$$
\sum_{\ell^i_d} f_{L^i_d | L} \ell^i_d, H_0) \cdot \sum_{i=1}^{N_c} \ell^i_{an} = \sum_{i=1}^{N_c} \ell^i \cdot p_s (1 - \alpha^i)(1 - p_c)
$$

Note that $\int f_{x|H_1,Y}(x|H_1,Y) \, dx \, Y \in \{0, 1\}$, are constants once the probability of local detection function of a given class is defined.

To see that (17) is linear with the number of sensors, taking into account its similarity with (16) is now straightforward.

A. Censored Scheme

There are three terms in (19). The first one is the discrete KL-divergence between $f_{L^i_d | L} \ell^i_d | \ell^i, H_0)$ and $f_{L^i_d | L} \ell^i_d | \ell^i, H_1)$ taking into account the independence of the conditional distributions of the number of sensors

$$
\sum_{\ell^i_d} f_{L^i_d | L} \ell^i_d, H_0) \ln \frac{f_{L^i_d | L} \ell^i_d | \ell^i, H_0)}{f_{L^i_d | L} \ell^i_d | \ell^i, H_1) = \sum_{i=1}^{N_c} D^i_{KL}(H_0||H_1)
$$

where $D^i_{KL}(H_0||H_1)$ is the discrete KL-divergence for the distribution of the number of available readings of sensors of class-i in the censored scheme

$$
D^i_{KL}(H_0||H_1) = \sum_{\ell^i_d} f_{L^i_d | L} \ell^i_d, H_0) \ln \frac{f_{L^i_d | L} \ell^i_d | \ell^i, H_0)}{f_{L^i_d | L} \ell^i_d | \ell^i, H_1) = \sum_{i=1}^{N_c} D^i_{KL}(H_0||H_1)
$$

Taking the value for $\ell^i = 0$, for $\ell^i = 1$ and then extending it by induction, it can be shown that

$$
D^i_{KL}(H_0||H_1) = \ell^i \cdot p_s(1 - p_c)
$$

which is also linear with $\ell^i$.

The last two terms in (19) are also linear with $\ell^i$ because the involved terms are proportional to the expected value of $\ell^i_{ap}$ and $\ell^i_{an}$ given the number of sensors of each class under the null hypothesis, which are

$$
\sum_{\ell^i_d} f_{L^i_d | L} \ell^i_d, H_0) \cdot \sum_{i=1}^{N_c} \ell^i_{ap} = \sum_{i=1}^{N_c} \ell^i \cdot p_s \alpha^i(1 - p_c)
$$

and

$$
\sum_{\ell^i_d} f_{L^i_d | L} \ell^i_d, H_0) \cdot \sum_{i=1}^{N_c} \ell^i_{an} = \sum_{i=1}^{N_c} \ell^i \cdot p_s (1 - \alpha^i)(1 - p_c)
$$

which is also linear with $\ell^i$, and the last term in (19) is equivalent to the third one in (16), linear with $\ell^i$.

To see that (20) is linear with the number of sensors, taking into account its similarity with (19), is now straightforward.

ACKNOWLEDGMENT

The authors wish to thank the reviewers of the manuscript for this paper for their valuable comments that have notably helped to improve its quality.

REFERENCES


[7] J. F. Chamberland and V. V. Veeravalli, “Asymptotic results for de-


[23] S. Marano, V. Matta, and P. Willett, “Quantizer precision for dis-


[26] A. Artés-Rodríguez, M. Lázaro, and M. Sánchez-Fernández, “Decen-

[27] M. Lázaro, A. Artés-Rodríguez, and M. Sánchez-Fernández, “Decen-


erogeneous sensor networks,” in Proc. ICASSP, Honolulu, HI, Apr. 2007.


