Robust and Distributed Stochastic Localization in Sensor Networks: Theory and Experimental Results

IOANNIS CH. PASCHALIDIS and DONG GUO
Boston University

We present a robust localization system allowing wireless sensor networks to determine the physical location of their nodes. The coverage area is partitioned into regions and we seek to identify the region of a sensor based on observations by stationary clusterheads. Observations (e.g., signal strength) are assumed random. We pose the localization problem as a composite multi-hypothesis testing problem, develop the requisite theory, and address the problem of optimally placing clusterheads. We show that localization decisions can be distributed by appropriate in-network processing. The approach is validated in a testbed yielding promising results.

Categories and Subject Descriptors: C.2.1 [**Network Architecture and Design**]: Distributed networks, Wireless communication.

General Terms: Algorithms, Experimentation, Theory.

Additional Key Words and Phrases: Sensor networks, localization, information theory, hypothesis testing, optimal deployment, testbed.

1. INTRODUCTION

Localization is viewed as an important service in *Wireless Sensor Networks (WSNETs)* because it enables a number of innovative services, including asset and personnel tracking and locating nodes that report a critical event. The Global Positioning System (GPS) provides an effective localization technology outdoors but is expensive for many WSNET applications, unreliable in downtown urban areas, and not operational indoors.

The localization literature is large but we will restrict our attention to systems that only use RF signals from the sensors to localize. The motivation is that RF is the common denominator of all WSNET platforms since all sensors have a radio to

Ioannis Ch. Paschalidis is with the Center for Information and Systems Engineering, the Department of Electrical and Computer Engineering, and the Systems Engineering Division, Boston University, 15 St. Mary's St., Brookline, MA 02446, e-mail: `yannisp@bu.edu`, url: `http://ionia.bu.edu/`.

Dong Guo is with the Center for Information and Systems Engineering, Boston University, e-mail: `dguo@bu.edu`.

Research partially supported by the NSF under grants DMI-0330171, CNS-0435312, ECS-0426453, EFRI-0735974, and by the DOE under grant DE-FG52-06NA27490.

A preliminary version of this work has appeared in Paschalidis and Guo [2007].

communicate with each other. Moreover, most existing WSNET nodes carry very rudimentary hardware that only allows the computation of the signal strength seen by a receiver for packets transmitted by some other node. Additional RF characteristics, like time-of-flight or angle-of-arrival, are not commonly available and require more sophisticated hardware. The key idea underlying RF-based localization is as follows: when a packet is transmitted by a sensor, associated RF characteristics observed by stationary sensors – the *clusterheads* – depend on the location of the transmitting sensor. These *observations* are exploited to reveal that location. As we will see, the method we develop can localize using just RF signal strength but is general enough to accommodate additional RF characteristics should they be available.

One class of RF-based localization systems relies on a "deterministic" pattern matching approach as in Bahl and Padmanabhan [2000], Lorincz and Welsh [2006], and Kaemarungsi and Krishnamurthy [2004]. They use the signal strength (mean) values observed at a sensor for packets transmitted by a set of beacon nodes and compare these values to a pre-computed signal-strength map of the coverage area. RADAR (Bahl and Padmanabhan [2000]), for instance, one of the first localization systems developed, computes a Euclidean distance between observed signal strength values at a sensor and the corresponding values pre-recorded at a set of training locations to determine the location closest to the sensor. Such an approach may face challenges when the RF signal landscape is highly variable. This is the case in indoor environments which are very dynamic (e.g., doors opening and closing, people moving, etc.) and feature multipath and fading.

Another class of localization systems uses triangulation or stochastic triangulation techniques as in Patwari et al. [2003] where signal strength measurements are used to estimate distance and location. The approach in Madigan et al. [2005] seeks to benefit from estimating multiple locations at the same time. These techniques assume a model describing how signal strength diminishes with distance (path loss formula) and the modeling error can lead to inaccuracies. In experimental results we report in this paper our approach can reduce the mean error distance by a factor of 3.6 compared to stochastic triangulation techniques. A different triangulation-like approach that may be vulnerable to RF signal variability appeared in Yedavalli et al. [2005] and relies on a monotonicity property of signal strength as a function of distance to be satisfied most of the time.

The work closer to the approach we present is in Ray et al. [2006] which developed a stochastic localization system formulating the problem as a standard hypothesis testing problem. Specifically, signal strength measurements from a number of locations spread throughout the coverage area are used to obtain probability density functions (pdfs) of signal strength at every potential clusterhead position. To locate a sensor somewhere in the coverage area the system tries to "match" measurements for that sensor to these pdfs, hence, a hypothesis testing problem. A limitation of this approach is that when the sensor we seek is not close to a location from which we have measurements, then the observations may not match well with any of the pdfs leading to errors. One can reduce these errors by obtaining measurements from more points, but this is costly. This motivates the work in this paper.

The key idea underlying the present work is to partition the coverage area into a set of regions. The problem is to determine the region where the sensor we seek resides. To every region-clusterhead pair we associate a *family* of (signal strength) pdfs. This is intended to provide robustness with respect to the position of the sensor within a region. The pdf family can be constructed from measurements taken from locations within the region and can better represent the region than a single pdf. We still pose the localization problem as a hypothesis testing problem but now we have to match signal strength measurements to a pdf family, resulting to a *composite hypothesis testing* problem. In this new framework we consider the Generalized Likelihood Ratio Test (GLRT) decision rule and obtain a necessary and sufficient condition under which it is optimal in a Generalized Neyman-Pearson (GNP) sense, thus, generalizing earlier work in Zeitouni et al. [1992]. Another important problem we consider is that of optimally placing clusterheads – an optimal deployment/WSNET design problem – to minimize the maximum probability of error.

We further demonstrate that our system can localize in a distributed manner by appropriate in-network processing: clusterheads make observations and take local decisions which get processed as they propagate through the network of clusterheads. The final decision reaches the gateway and, as we show, there is no performance cost compared to a centralized approach. We have implemented our approach in a testbed installed at a Boston University building. [1] Our experimental results establish that we can achieve accuracy that is, roughly, on the same order of magnitude as the radius of our regions. Specifically, we have achieved a mean error distance from 8 feet down to 9 inches depending on the size of the regions we define. The price to pay for greater accuracy is the amount of measurements needed as smaller (thus, more) regions require more measurements to determine the family of pdfs corresponding to every region-clusterhead pair.

Our contributions include:

—formulating the localization problem as a composite hypothesis testing problem aiming at accommodating the stochastic nature of RF signals propagating indoors and providing robustness with respect to measurements based on which a localization decision is made;

—generalizing the GLRT optimality conditions in Zeitouni et al. [1992] to the case where both hypotheses correspond to a family of pdfs – a result which is of independent interest;

—characterizing the performance of the localization system which enables

—solving the clusterhead placement problem building on the work in Ray et al. [2006];

—devising a distributed algorithm for making the localization decision; and

—testing the proposed approach on an actual testbed.

The paper is organized as follows. In Sec. 2, we introduce our system model. In Sec. 3, we study the composite binary hypothesis testing problem, establish an optimality condition for GLRT, and obtain bounds on the error exponents which allow us to optimize the GLRT threshold. We also consider the case where the

[1] See `http://pythagoras.bu.edu/bloc/index.html`

GLRT optimality conditions are not satisfied. In Sec. 4, we consider the clusterhead placement problem. In Sec. 5, we develop the distributed decision approach and make comparisons to a centralized one. In Sec. 6, we provide results from a testbed implementation of our approach. Finally, in Sec. 7, we draw conclusions.

2. PROBLEM FORMULATION

In this section we introduce our system model. Consider a WSNET deployed in a site for localization purposes. We divide the site into N regions denoted by an index set $\mathscr{L} = \{L_1, \dots, L_N\}$. There are M distinct positions $\mathscr{B} = \{B_1, \dots, B_M\}$ at which we can place clusterheads.

Let a sensor be located in region $L \in \mathscr{L}$. A set of packets broadcasted by the sensor is received by some of the clusterheads which observe certain physical quantities associated with each packet. Often, the observed physical quantities are just the received signal strength (RSSI) and, if technology allows it, one can also observe the angle-of-arrival of the signal or other signal characteristics. Our methodology is general enough to apply to any set of physical observations.

Let $\mathbf{y}^{(i)}$ denote the vector of observations by a clusterhead at position B_i corresponding to a packet broadcasted by the sensor. These observations are assumed to be random. To simplify the analysis in the rest of the paper we will assume that the observations take values from a finite alphabet $\Sigma = \{\sigma_1, \dots, \sigma_{|\Sigma|}\}$, where $|\Sigma|$ denotes the cardinality of Σ. [2] A series of n consecutive observations are denoted by $\mathbf{y}_1^{(i)}, \dots, \mathbf{y}_n^{(i)}$ and are assumed independent and identically distributed (i.i.d.) conditioned on the region the sensor node resides. This assumption is justified when the site is dynamic enough (e.g., doors opening or closing, people moving) so that the lengths of various radio-paths between the receiver and the transmitter change on the order of a wavelength between consecutive observations. For example, if a sensor operates at the 2.4 GHz ISM band, the half-wavelength is only about 6cm, and body movements of a user who carries the sensor may alone cause observations separated in time by a few seconds to be i.i.d. Observations made by different clusterheads at about the same time need not be independent. We acknowledge that when the site and the transmitter/receiver are fairly static, observations over such short times may be correlated; a case we do not handle. The requisite theory could be developed for that case as well but to "learn" models capturing such correlation would probably require too many measurements for a practical system.

With every clusterhead-region pair (B_i, L_j) we associate a family of pdfs $p_{\mathbf{Y}^{(i)}|\boldsymbol{\theta}_j}(\mathbf{y})$ where $\mathbf{Y}^{(i)}$ denotes the random variable corresponding to observations $\mathbf{y}^{(i)}$ at clusterhead B_i when the transmitting sensor is in some location within L_j. Here, $\boldsymbol{\theta}_j \in \Omega_j$ is a vector in some space Ω_j parametrizing the pdf family. As mentioned earlier, the use of a family of pdfs rather than a single pdf is intended to provide robustness with respect to the exact position of the sensor within the region L_j. As we will see later on, we will use measurements at a few locations (or even a single one) within L_j but we will associate to these measurements a family of pdfs parametrized by $\boldsymbol{\theta}_j$. For example, one could obtain an empirical pdf from the measurements and associate with L_j pdfs with the same shape as the empirical pdf

[2] This is indeed the case in practice since WSNET nodes report quantized RSSI measurements.

and a mean lying in some interval centered at the empirical mean.

Given a family of pdfs for every pair (B_i, L_j) we are interested in placing $K \leq M$ clusterheads at positions in $\mathscr{B}$ and use observations by them to determine the region in which a sensor node resides. To that end, we will (i) characterize the performance of the localization system in terms of the probability of error, (ii) develop an algorithm for placing clusterheads that provides guarantees for the probability of error, and (iii) develop approaches for determining the sensor location in a distributed manner by directing the clusterheads to do appropriate processing of their observations and only forward minimal information to the gateway.

3. BINARY COMPOSITE HYPOTHESIS TESTING

We start our analysis by considering the simpler problem of having a single clusterhead at B_k and two possible regions L_i and L_j at which the sensor may reside. Each region has an associated pdf family $p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\mathbf{y})$ and $p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y})$, respectively. [3] The clusterhead makes n i.i.d. observations $\mathbf{y}^{(k),n} = (\mathbf{y}_1^{(k)}, \ldots, \mathbf{y}_n^{(k)})$ from which we need to determine the region L_i vs. L_j. We will be using the notation $p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\mathbf{y}^{(k),n}) = \prod_{l=1}^{n} p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\mathbf{y}_l^{(k)})$.

The problem at hand is a binary composite hypothesis testing problem for which the so called *Generalized Likelihood Ratio Test (GLRT)* is commonly used. The GLRT compares the normalized generalized log-likelihood ratio

$$X_{ijk}(\mathbf{y}^{(k),n}) = \frac{1}{n} \log \frac{\sup_{\boldsymbol{\theta}_i \in \Omega_i} p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\mathbf{y}^{(k),n})}{\sup_{\boldsymbol{\theta}_j \in \Omega_j} p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y}^{(k),n})}$$

to a threshold λ and declares L_i whenever

$$\mathbf{y}^{(k),n} \in \mathscr{S}_{ijk,n}^{GLRT} \triangleq \{\mathbf{y}^n \mid X_{ijk}(\mathbf{y}^n) \geq \lambda\},$$

and L_j otherwise. There are two types of error (referred to as type I and type II, respectively) associated with a decision with probabilities

$$\alpha_{ijk,n}^{GLRT}(\boldsymbol{\theta}_j) = \mathbf{P}_{\boldsymbol{\theta}_j}[\mathbf{y}^{(k),n} \in \mathscr{S}_{ijk,n}^{GLRT}], \qquad \beta_{ijk,n}^{GLRT}(\boldsymbol{\theta}_i) = \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{y}^{(k),n} \notin \mathscr{S}_{ijk,n}^{GLRT}],$$

where $\mathbf{P}_{\boldsymbol{\theta}_j}[\cdot]$ (resp. $\mathbf{P}_{\boldsymbol{\theta}_i}[\cdot]$) is a probability evaluated assuming that $\mathbf{y}^{(k),n}$ is drawn from $p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\cdot)$ (resp. $p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\cdot)$). We use a similar notation and write $\alpha_{ijk,n}^{\mathscr{S}}(\boldsymbol{\theta}_j)$ and $\beta_{ijk,n}^{\mathscr{S}}(\boldsymbol{\theta}_i)$ for the error probabilities of any other test that declares L_i whenever $\mathbf{y}^{(k),n}$ is in some set $\mathscr{S}_{ijk,n}$. In the sequel, we will often consider the asymptotic rate according to which these probabilities approach zero as $n \to \infty$. We will use the term *exponent* to refer to the quantity $\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}[\cdot]$ for some probability $\mathbf{P}[\cdot]$; if the exponent is d then the probabilities approaches zero as e^{-nd}.

Zeitouni et al. [1992] have established conditions for the optimality of the GLRT in a Neyman-Pearson sense for general Markov sources. The analysis in Zeitouni et al. [1992] is carried out for the case where one hypothesis corresponds to a single pdf and the other to a pdf family. We provide a generalization (in an i.i.d. setting)

[3]We note that the pdf families are associated with a region-clusterhead pair. Thus, $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_i$ depend on k as well but we elect to suppress this dependence in the notation for simplicity. We will be usually referring to the triplet i, j, k, hence, it will be evident to which $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_i$ we refer.

to the situation of interest where both hypotheses correspond to a family of pdfs. More specifically, we will establish a necessary and sufficient condition for the GLRT to satisfy the *generalized Neyman-Pearson* optimality criterion given below.

Definition 1
Generalized Neyman-Pearson (GNP) Criterion*: We will say that the decision rule* $\{\mathscr{S}_{ijk,n}\}$ *is optimal if it satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log \alpha_{ijk,n}^{\mathscr{S}}(\boldsymbol{\theta}_j) < -\lambda, \qquad \forall \boldsymbol{\theta}_j \in \Omega_j, \tag{1}$$

and maximizes $-\limsup_{n\to\infty} \frac{1}{n} \log \beta_{ijk,n}^{\mathscr{S}}(\boldsymbol{\theta}_i)$ *uniformly for all* $\boldsymbol{\theta}_i \in \Omega_i$*.*

For any sequence of observations $\mathbf{y}^n = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$, the empirical measure (or type) is given by $\mathbf{L}_{\mathbf{y}^n} = (L_{\mathbf{y}^n}(\sigma_1), \ldots, L_{\mathbf{y}^n}(\sigma_{|\Sigma|}))$, where

$$L_{\mathbf{y}^n}(\sigma_i) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{\mathbf{y}_j = \sigma_i\}, \qquad i = 1, \ldots, |\Sigma|,$$

and $\mathbf{1}\{\cdot\}$ denotes the indicator function. We will denote the set of all possible types of sequences of length n by $\mathscr{L}_n = \{\boldsymbol{\nu} \mid \boldsymbol{\nu} = \mathbf{L}_{\mathbf{y}^n}$ for some $\mathbf{y}^n\}$ and the type class of a probability law $\boldsymbol{\nu}$ by $T_n(\boldsymbol{\nu}) = \{\mathbf{y}^n \in \Sigma^n \mid \mathbf{L}_{\mathbf{y}^n} = \boldsymbol{\nu}\}$, where Σ^n denotes the cartesian product of Σ with itself n times. Let

$$H(\boldsymbol{\nu}) = -\sum_{i=1}^{|\Sigma|} \boldsymbol{\nu}(\sigma_i) \log \boldsymbol{\nu}(\sigma_i),$$

be the entropy of the probability vector $\boldsymbol{\nu}$ and

$$D(\boldsymbol{\nu} \| \boldsymbol{\mu}) = \sum_{i=1}^{|\Sigma|} \boldsymbol{\nu}(\sigma_i) \log \frac{\boldsymbol{\nu}(\sigma_i)}{\boldsymbol{\mu}(\sigma_i)},$$

the relative entropy of $\boldsymbol{\nu}$ with respect to another probability vector $\boldsymbol{\mu}$.

Lemma 3.5.3 in Dembo and Zeitouni [1998] states that it suffices to consider functions of the empirical measure when trying to construct an optimal test (i.e., the empirical measure is a sufficient statistic). Considering hereafter tests that depend only on $\mathbf{L}_{\mathbf{y}^n}$, the so called generalized Hoeffding [1965] test is optimal according to the GNP criterion and accepts L_i when $\mathbf{y}^{(k),n}$ is in the set

$$\mathscr{S}_{ijk,n}^{*} = \{\mathbf{y}^n | \inf_{\boldsymbol{\theta}_j} D(\mathbf{L}_{\mathbf{y}^n} \| \mathbf{P}_{\boldsymbol{\theta}_j}) \geq \lambda\},$$

where $\mathbf{P}_{\boldsymbol{\theta}_j}$ denotes the probability law induced by $p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\cdot)$. The following lemma generalizes Hoeffding's result and a similar result in Zeitouni et al. [1992]; the proof is in Appendix A.

Lemma 3.1 *The generalized Hoeffding test satisfies the GNP criterion.*

Next, we will determine the exponent of $\beta_{ijk,n}^{\mathscr{S}^*}(\boldsymbol{\theta}_i)$. Define the set $\mathscr{A}_{ijk} = \{\mathbf{Q} | \inf_{\boldsymbol{\theta}_j} D(\mathbf{Q} \| \mathbf{P}_{\boldsymbol{\theta}_j}) < \lambda\}$. We have

$$\beta_{ijk,n}^{\mathscr{S}^*}(\boldsymbol{\theta}_i) = \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{y}^{(k),n} \notin \mathscr{S}_{ijk,n}^{*}] = \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{L}_{\mathbf{y}^{(k),n}} \in \mathscr{A}_{ijk} \cap \mathscr{L}_n].$$

Due to Sanov's theorem (Dembo and Zeitouni [1998, Chap. 2])

$$\inf_{\mathbf{Q}\in\mathscr{A}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \le -\limsup_{n\to\infty}\frac{1}{n}\log\beta^{\mathscr{S}^*}_{ijk,n}(\boldsymbol{\theta}_i)$$
$$\le -\liminf_{n\to\infty}\frac{1}{n}\log\beta^{\mathscr{S}^*}_{ijk,n}(\boldsymbol{\theta}_i) \le \inf_{\mathbf{Q}\in\mathscr{A}^o_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}), \quad (2)$$

where $\mathscr{A}^o_{ijk}$ denotes the interior of $\mathscr{A}_{ijk}$. Since $\mathscr{A}_{ijk}$ is an open set the upper and lower bounds match and $\inf_{\mathbf{Q}\in\mathscr{A}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i})$ is the exponent of $\beta^{\mathscr{S}^*}_{ijk,n}(\boldsymbol{\theta}_i)$.

The following theorem establishes a necessary and sufficient condition for the optimality of GLRT under the GNP criterion. The proof is in Appendix B.

Theorem 3.2 *The GLRT with a threshold λ is asymptotically optimal under the GNP criterion, if and only if*

$$\inf_{\mathbf{Q}\in\mathscr{C}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \ge \inf_{\mathbf{Q}\in\mathscr{A}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}), \quad (3)$$

for all $\boldsymbol{\theta}_i$, where

$$\mathscr{C}_{ijk} = \{\mathbf{Q}|\inf_{\boldsymbol{\theta}_j} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j}) - \inf_{\boldsymbol{\theta}_i} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) < \lambda \le \inf_{\boldsymbol{\theta}_j} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j})\}.$$

Furthermore, assuming that (3) is in effect

$$\limsup_{n\to\infty}\frac{1}{n}\log\alpha^{GLRT}_{ijk,n}(\boldsymbol{\theta}_j) \le -\lambda, \qquad \forall\boldsymbol{\theta}_j\in\Omega_j, \quad (4)$$

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta^{GLRT}_{ijk,n}(\boldsymbol{\theta}_i) \le -\inf_{\mathbf{Q}\in\mathscr{A}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}),\ \forall\boldsymbol{\theta}_i\in\Omega_i. \quad (5)$$

3.1 Determining the optimal threshold

Assuming that the condition of Thm. 3.2 is in effect, it can be seen from (4) and (5) that the exponent of the type I error probability is increasing with λ but the exponent of the type II error probability is nonincreasing with λ. We have no preference on the type of error we make, thus, we wish to balance the two exponents and determine the value of λ at which they become equal. In this subsection we detail how this can be done and obtain a λ^*_{ijk} that bounds the worst case (over Ω_j and Ω_i) exponents of the type I and type II error probabilities. To simplify the exposition we will be assuming that Ω_j and Ω_i are discrete sets; this is also the case in the experimental setup we describe later on.

Let us consider the exponent of the type II GLRT error probability (cf. (5)):

$$\begin{aligned} Z_{ijk}(\lambda,\boldsymbol{\theta}_i) = \ & \min_{\mathbf{Q}}\ D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \\ & \text{s.t. } \min_{\boldsymbol{\theta}_j} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j}) \le \lambda. \end{aligned} \quad (6)$$

The worst case exponent over $\boldsymbol{\theta}_i\in\Omega_i$ is given by

$$Z_{ijk}(\lambda) = \min_{\boldsymbol{\theta}_i} Z_{ijk}(\lambda,\boldsymbol{\theta}_i).$$

Note that $Z_{ijk}(\lambda)$ is nonincreasing in λ, $Z_{ijk}(0) = \min_{\boldsymbol{\theta}_i}\min_{\boldsymbol{\theta}_j} D(\mathbf{P}_{\boldsymbol{\theta}_j}\|\mathbf{P}_{\boldsymbol{\theta}_i})$, and $\lim_{\lambda\to\infty} Z_{ijk}(\lambda) = 0$. Assuming that $Z_{ijk}(0) > 0$, there exists a $\lambda^*_{ijk} > 0$ such that

$Z_{ijk}(\lambda^*_{ijk}) = \lambda^*_{ijk}$. Furthermore, both error probability exponents in (4) and (5) are no smaller than λ^*_{ijk}.

Now consider the clusterhead at B_k observing $\mathbf{y}^{(k),n}$ and seeking to distinguish between L_i and L_j. Assume that the GLRT using $X_{ijk}(\mathbf{y}^{(k),n})$ satisfies condition (3) and, also, the GLRT using $X_{jik}(\mathbf{y}^{(k),n})$ satisfies the symmetric condition. The clusterhead has the option of using the GLRT by comparing $X_{ijk}(\mathbf{y}^{(k),n})$ to the threshold λ^*_{ijk}, or comparing $X_{jik}(\mathbf{y}^{(k),n})$ to a threshold λ^*_{jik} that can be obtained in exactly the same way as λ^*_{ijk}. Let

$$d_{ijk} = \max\{\lambda^*_{ijk}, \lambda^*_{jik}\}, \tag{7}$$

and set $(\bar{i}, \bar{j}) = (i, j)$ if λ^*_{ijk} is the maximizer above; otherwise set $(\bar{i}, \bar{j}) = (j, i)$. Define the maximum probability of error as

$$P^{(e)}_{ijk,n} \stackrel{\triangle}{=} \max\{\max_{\boldsymbol{\theta}_{\bar{j}}} \alpha^{GLRT}_{\bar{i}\bar{j}k,n}(\boldsymbol{\theta}_{\bar{j}}), \max_{\boldsymbol{\theta}_{\bar{i}}} \beta^{GLRT}_{\bar{i}\bar{j}k,n}(\boldsymbol{\theta}_{\bar{i}})\}.$$

The discussion above leads to the following proposition.

Proposition 3.3 *Assume that the GLRT using $X_{ijk}(\mathbf{y}^{(k),n})$ satisfies condition (3) and, also, the GLRT using $X_{jik}(\mathbf{y}^{(k),n})$ satisfies the symmetric condition. Then, when the clusterhead at B_k compares $X_{\bar{i}\bar{j}k}(\mathbf{y}^{(k),n})$ to d_{ijk} the maximum probability of error satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log P^{(e)}_{ijk,n} \leq -d_{ijk}.$$

One of the challenges computing d_{ijk} is that the problem in (6) is nonconvex. Specifically, the relative entropy in the constraint is convex in $\mathbf{Q}$ but minimization over $\boldsymbol{\theta}_j$ yields a piecewise convex function. This may not be an issue when there are relatively few possible values of $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_i$ but for large sets Ω_j and Ω_i computing d_{ijk} becomes expensive. To address this issue, we will next develop a lower bound (through duality) to $Z_{ijk}(\lambda, \boldsymbol{\theta}_i)$.

Let $\tilde{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$ the optimal value of the dual to (6); by weak duality it follows $Z_{ijk}(\lambda, \boldsymbol{\theta}_i) \geq \tilde{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$. We have

$$\tilde{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i) = \max_{\mu\geq 0} \left[\min_{\boldsymbol{\theta}_j} \min_{\mathbf{Q}} [D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) + \mu D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j})] - \mu\lambda \right]. \tag{8}$$

Note that the optimization over $\mathbf{Q}$ is convex and the optimization over μ is concave, thus, this problem can be solved efficiently. (In fact, the optimization over $\mathbf{Q}$ can be solved analytically.) It can be seen that $\tilde{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$ is convex and nonincreasing in λ for all $\boldsymbol{\theta}_i$. Furthermore, the exponent of the type II GLRT error probability is no smaller than $\tilde{Z}_{ijk}(\lambda) = \min_{\boldsymbol{\theta}_i} \tilde{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$. Note that $\tilde{Z}_{ijk}(\lambda)$ is also nonincreasing in λ, $\tilde{Z}_{ijk}(0) = \min_{\boldsymbol{\theta}_i} \min_{\boldsymbol{\theta}_j} D(\mathbf{P}_{\boldsymbol{\theta}_j}\|\mathbf{P}_{\boldsymbol{\theta}_i})$, and $\lim_{\lambda\to\infty} \tilde{Z}_{ijk}(\lambda) = 0$. Assuming that $\tilde{Z}_{ijk}(0) > 0$, there exists a $\tilde{\lambda}^*_{ijk} > 0$ such that $\tilde{Z}_{ijk}(\tilde{\lambda}^*_{ijk}) = \tilde{\lambda}^*_{ijk}$. Furthermore, both error exponents in (4) and (5) are no smaller than $\tilde{\lambda}^*_{ijk}$.

Following the same line of development as before we set

$$\tilde{d}_{ijk} = \max\{\tilde{\lambda}^*_{ijk}, \tilde{\lambda}^*_{jik}\}, \tag{9}$$

and define $\bar{i}$, $\bar{j}$, and $P^{(e)}_{ijk,n}$ in the same way as earlier. It can be seen that $\tilde{d}_{ijk} \le d_{ijk}$. We arrive at the following proposition which provides a weaker but more easily computable probabilistic guarantee on the probability of error.

Proposition 3.4 *Assume that the GLRT using $X_{ijk}(\mathbf{y}^{(k),n})$ satisfies condition (3) and, also, the GLRT using $X_{jik}(\mathbf{y}^{(k),n})$ satisfies the symmetric condition. Then, when the clusterhead at B_k compares $X_{\bar{i}\bar{j}k}(\mathbf{y}^{(k),n})$ to $\tilde{d}_{ijk}$ the maximum probability of error satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log P^{(e)}_{ijk,n} \le -\tilde{d}_{ijk}.$$

Next, we tackle the case when the GLRT optimality condition (3) is not satisfied.

3.2 When the GLRT is not optimal

Define the set $\mathscr{D}_{ijk} = \{\mathbf{Q} | \inf_{\boldsymbol{\theta}_j} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j}) - \inf_{\boldsymbol{\theta}_i} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) < \lambda\}$ and note that by an argument similar to (29) $\mathbf{y}^{(k),n} \in \mathscr{S}^{GLRT}_{ijk,n}$ is equivalent to $\mathbf{L}_{\mathbf{y}^{(k),n}} \notin \mathscr{D}_{ijk}$. Hence, by Sanov's theorem

$$\limsup_{n\to\infty} \frac{1}{n} \log \alpha^{GLRT}_{ijk,n}(\boldsymbol{\theta}_j) \le - \inf_{\mathbf{Q}\in\overline{\mathscr{D}_{ijk}}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j}), \forall \boldsymbol{\theta}_j \in \Omega_j, \tag{10}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta^{GLRT}_{ijk,n}(\boldsymbol{\theta}_i) \le - \inf_{\mathbf{Q}\in\mathscr{D}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}),\ \forall \boldsymbol{\theta}_i \in \Omega_i. \tag{11}$$

Using the same argument as in the proof of Thm. 3.2 we can show that (4) still holds. The exponent of the type II GLRT error probability (cf. (11)) is

$$\begin{aligned} \hat{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i) &= \min\nolimits_{\mathbf{Q}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \\ \text{s.t. } \min\nolimits_{\boldsymbol{\theta}_j} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j}) &- \min\nolimits_{\boldsymbol{\theta}_i} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \le \lambda, \end{aligned}$$

which is equivalent to

$$\begin{aligned} \hat{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i) &= \min\nolimits_{\mathbf{Q}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \\ \text{s.t. } \min\nolimits_{\boldsymbol{\theta}_j} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_j}) &- D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \le \lambda, \quad \forall \boldsymbol{\theta}_i. \end{aligned} \tag{12}$$

The worst case exponent over $\boldsymbol{\theta}_i \in \Omega_i$ is given by

$$\hat{Z}_{ijk}(\lambda) = \min_{\boldsymbol{\theta}_i} \hat{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i).$$

$\hat{Z}_{ijk}(\lambda)$ is nonincreasing in λ, and $\lim_{\lambda\to\infty} \hat{Z}_{ijk}(\lambda) = 0$. Assuming that $\hat{Z}_{ijk}(0) > 0$, there exists a $\hat{\lambda}^*_{ijk} > 0$ such that $\hat{Z}_{ijk}(\hat{\lambda}^*_{ijk}) = \hat{\lambda}^*_{ijk}$. Furthermore, both error probability exponents in (10) and (11) are no smaller than $\hat{\lambda}^*_{ijk}$.

Following the same argument as before we set

$$\hat{d}_{ijk} = \max\{\hat{\lambda}^*_{ijk}, \hat{\lambda}^*_{jik}\}, \tag{13}$$

and define $\bar{i}$, $\bar{j}$, and $P^{(e)}_{ijk,n}$ in the same way as earlier. The discussion above leads to the following proposition.

Proposition 3.5 *Suppose that the clusterhead at B_k uses the GLRT and compares $X_{\bar{i}\bar{j}k}(\mathbf{y}^{(k),n})$ to $\hat{d}_{ijk}$. Then, the maximum probability of error satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log P^{(e)}_{ijk,n} \leq -\hat{d}_{ijk}.$$

Problem in (12) is nonconvex; we will again use dual relaxation to obtain a quantity that is easier to compute. Let $\bar{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$ the optimal value of the dual of (12); by weak duality it follows $\hat{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i) \geq \bar{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$. After some algebra

$$\bar{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i) \quad = \quad \max_{\mu_{\boldsymbol{\theta}_i}\geq 0}[\min_{\boldsymbol{\theta}_j} \min_{\mathbf{Q}}[\sum_{r=1}^{|\Sigma|} Q(\sigma_r)\log(Q(\sigma_r)A(\sigma_r)) \; - \; \textstyle\sum_{\boldsymbol{\theta}_i} \mu_{\boldsymbol{\theta}_i}\lambda], \quad (14)$$

where $A(\sigma_r) = \frac{1}{\mathbf{P}_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\sigma_r)} \cdot \prod_{\boldsymbol{\theta}_i} \left(\frac{\mathbf{P}_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\sigma_r)}{\mathbf{P}_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\sigma_r)}\right)^{\mu_{\boldsymbol{\theta}_i}}$. Note that the optimization over $\mathbf{Q}$ is convex and the optimization over $\mu_{\boldsymbol{\theta}_i}$ is concave, thus, this problem can be solved efficiently. In fact, the optimization over $\mathbf{Q}$ can be solved analytically yielding

$$Q(\sigma_l) = \tfrac{1}{A(\sigma_l)} / \left(\textstyle\sum_{r=1}^{|\Sigma|} \tfrac{1}{A(\sigma_r)} \right), \; l = 1, \ldots, |\Sigma|.$$

$\bar{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$ is convex and nonincreasing in λ for all $\boldsymbol{\theta}_i$. Furthermore, the exponent of the type II GLRT error probability is no smaller than $\bar{Z}_{ijk}(\lambda) = \min_{\boldsymbol{\theta}_i} \bar{Z}_{ijk}(\lambda, \boldsymbol{\theta}_i)$. Note that $\bar{Z}_{ijk}(\lambda)$ is also nonincreasing in λ, and $\lim_{\lambda\to\infty} \bar{Z}_{ijk}(\lambda) = 0$. Assuming that $\bar{Z}_{ijk}(0) > 0$, there exists a $\bar{\lambda}^*_{ijk} > 0$ such that $\bar{Z}_{ijk}(\bar{\lambda}^*_{ijk}) = \bar{\lambda}^*_{ijk}$. Furthermore, both error exponents in (10) and (11) are no smaller than $\bar{\lambda}^*_{ijk}$.

Following the same approach as before, set

$$\bar{d}_{ijk} = \max\{\bar{\lambda}^*_{ijk}, \bar{\lambda}^*_{jik}\}, \quad (15)$$

and define $\bar{i}$, $\bar{j}$, and $P^{(e)}_{ijk,n}$ in the same way as earlier. It can be seen that $\hat{d}_{ijk} \geq \bar{d}_{ijk}$. We arrive at the following proposition which provides a weaker but more easily computable probabilistic guarantee on the probability of error.

Proposition 3.6 *Suppose that the clusterhead at B_k uses the GLRT and compares $X_{\bar{i}\bar{j}k}(\mathbf{y}^{(k),n})$ to $\bar{d}_{ijk}$. Then, the maximum probability of error satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log P^{(e)}_{ijk,n} \leq -\bar{d}_{ijk}.$$

4. LOCALIZATION AND CLUSTERHEAD PLACEMENT

In this section, we focus on how to place the $K \leq M$ clusterheads at positions in $\mathscr{B}$ to facilitate localization. We start by considering the multiple composite hypothesis testing problem of identifying the region $L \in \mathscr{L}$ in which the sensor we seek resides.

4.1 Multiple composite hypothesis testing

We assume, without loss of generality, that we have placed clusterheads in positions $B_1, \ldots, B_K$, each one making n i.i.d. observations $\mathbf{y}^{(k),n} = (\mathbf{y}^{(k)}_1, \ldots, \mathbf{y}^{(k)}_n)$. Let d_{ijk} be the GLRT threshold obtained in Sec. 3 for each region pair (i, j), $i < j$, and clusterhead k. (d_{ijk} is obtained from either (7), (9), (13), or (15), depending on which optimization problem we elect to solve.)

We make $N-1$ binary decisions with the GLRT rule to arrive at a final decision. Specifically, we first compare L_1 with L_2 to accept one hypothesis, then compare the accepted hypothesis with L_3, and so on and so forth. For each one of these L_i vs. L_j decisions we use a single clusterhead B_k as detailed in Sec. 3 and the exponent of the corresponding maximum probability of error is bounded by d_{ijk}. All in all we make $N-1$ binary hypothesis decisions.

4.2 Clusterhead placement

Our objective is to minimize the worst case probability of error. To that end, for every pair of regions L_i and L_j we need to find a clusterhead that can discriminate between them with a probability of error exponent larger than some ϵ and then maximize ϵ. This is accomplished by the mixed integer linear programming problem (MILP) formulation of Figure 1.

$$\max \quad \epsilon \tag{16}$$
$$\text{s.t.} \quad \sum_{k=1}^{M} x_k = K \tag{17}$$
$$\sum_{k=1}^{M} y_{ijk} = 1,\ i,j = 1,\dots,N, i<j, \tag{18}$$
$$y_{ijk} \le x_k,\ \forall i,j, i<j, k=1,\dots,M, \tag{19}$$
$$\epsilon \le \sum_{k=1}^{M} d_{ijk} y_{ijk},\ \forall i,j, i<j, \tag{20}$$
$$y_{ijk} \ge 0,\ \forall i,j, i<j, \forall k, \tag{21}$$
$$x_k \in \{0,1\},\ \forall k. \tag{22}$$

Fig. 1. Clusterhead placement MILP formulation.

In this formulation, the decision variables are x_k, y_{ijk}, and ϵ where $k = 1,\dots,M$, $i,j = 1,\dots,N, i<j$. x_k is the indicator function of a clusterhead been placed at position B_k. Equation (17) represents the constraint that K clusterheads are to be placed. Constraint (20) enforces that for every region pair there exist a clusterhead k with d_{ijk} larger than ϵ. Let x_k^*, y_{ijk}^*, and ϵ^* ($k=1,\dots,M$, $i,j=1,\dots,N$, $i<j$) be an optimal solution of this MILP. Although this problem is NP-hard, it can be solved efficiently for sites with more than 100 regions and potential clusterhead positions by using a special purpose algorithm from Ray et al. [2006].

Consider an arbitrary placement of K clusterheads. More specifically, let $\mathscr{Y}$ be any subset of the set of potential clusterhead positions $\mathscr{B}$ with cardinality K. Let $\mathbf{x}(\mathscr{Y}) = (x_1(\mathscr{Y}),\dots,x_M(\mathscr{Y}))$ where $x_k(\mathscr{Y})$ is the indicator function of B_k being in $\mathscr{Y}$. Define:

$$\epsilon(\mathscr{Y}) = \min_{\substack{i,j=1,\dots,N\\ i<j}} \max_{k:x_k(\mathscr{Y})=1} d_{ijk}. \tag{23}$$

We can interpret $\max_{k:x_k(\mathscr{Y})=1} d_{ijk}$ as the best exponent for the probability of error in distinguishing between regions L_i and L_j from some clusterhead in $\mathscr{Y}$. Then $\epsilon(\mathscr{Y})$ is simply the worst pairwise exponent. The following result is from Ray et al. [2006].

Proposition 4.1 *For any clusterhead placement $\mathscr{Y}$ we have*

$$\epsilon^* \geq \epsilon(\mathscr{Y}). \tag{24}$$

Moreover, the selected placement achieves equality; i.e.,

$$\epsilon^* = \min_{\substack{i,j=1,\ldots,N \\ i<j}} \max_{k:x_k^*=1} d_{ijk}, \tag{25}$$

and the optimal solution satisfies

$$y_{ijk}^* = \begin{cases} 1, & \text{if } k = \arg\max_{k:x_k^*=1} d_{ijk}, \\ 0, & \text{otherwise}, \end{cases} \quad \forall i,j,i<j,\forall k, \tag{26}$$

where at most one y_{ijk}^* is set to 1 for a given (i,j) pair.

To summarize, the positions where clusterheads are to be placed are in the set $\mathscr{Y}^* \triangleq \{B_k : x_k^* = 1\}$. For every region pair L_i and L_j, $\max_{k:x_k^*=1} d_{ijk}$ is the best exponent for the probability of error in distinguishing between these regions and the clusterhead that will be responsible for that decision is the one corresponding to $y_{ijk}^* = 1$; we will be denoting by k_{ij}^* the index of this clusterhead.

4.3 Performance guarantee

We will use the decision rule outlined in Sec. 4.1 and for every region pair (i,j) we will rely on the clusterhead at $B_{k_{ij}^*}$ to make the corresponding decision. The following theorem establishes a performance guarantee.

Proposition 4.2 *Let $\mathbf{x}^*, \mathbf{y}^*$ be an optimal solution of the MILP in Fig. 1 with corresponding optimal value ϵ^*. Place clusterheads according $\mathscr{Y}^* \triangleq \{B_k | x_k^* = 1\}$ and for every (i,j) select one clusterhead with index k_{ij}^* so that $y_{ijk_{ij}^*}^* = 1$. Then, the worst case probability of error for the decision rule described in Sec. 4.1, $P_n^{(e),opt}$, satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n^{(e),opt} \leq -\epsilon^*. \tag{27}$$

Proof. Recall the results of Props. 3.3, 3.4, 3.5, and 3.6 for the case where d_{ijk} is defined either by (7), (9), (13), or by (15), respectively. Define $(\bar{i},\bar{j})$ as in Sec. 3. The clusterhead with index k_{ij}^* will use the GLRT which compares $X_{\bar{i}\bar{j}k_{ij}^*}(\mathbf{y}^{(k_{ij}^*),n})$ to $d_{ijk_{ij}^*}$, thus, achieving a maximum probability of error with exponent no smaller than $d_{ijk_{ij}^*}$. Now, for every i and $j \neq i$ define $E_n(i,j)$ as the event that the GLRT employed by the clusterhead at $B_{k_{ij}^*}$ will decide L_j under $\mathbf{P}_{\boldsymbol{\theta}_i}$. For all $\delta_n > 0$ and large enough n we have

$$\mathbf{P}_{\boldsymbol{\theta}_i}[\text{error}] \leq \mathbf{P}_{\boldsymbol{\theta}_i}[\cup_{j\neq i} E_n(i,j)] \leq \sum_{j\neq i} e^{-n(d_{ijk_{ij}^*}+\delta_n)} \leq (N-1)e^{-n(\epsilon^*+\delta_n)}.$$

The 2nd inequality above is due to Props. 3.3, 3.4, 3.5, or 3.6 and the last inequality above is due to (25). Since the bound above holds for all i we obtain (27). □

5. DISTRIBUTED LOCALIZATION

In this section we consider the implementation of the decision rule described in Sec. 4. We assume that the WSNET has a single gateway. We seek to devise a distributed localization algorithm in order to minimize the information that needs to be exchanged between clusterheads and the gateway. The primary motivation is that in WSNETs communication is, in general, more expensive than processing. For the remainder of this section we will assume that the clusterheads and the gateway form a connected network. Otherwise, one can simply add a sufficient number of relays.

5.1 Centralized approach

To set the stage and establish a benchmark to which we will compare other approaches we first describe a naive, centralized, approach. According to this approach, every clusterhead observes $\mathbf{y}^{(k),n} = (\mathbf{y}_1^{(k)}, \ldots, \mathbf{y}_n^{(k)})$ and transmits this information to the gateway. The clusterheads do not need to store anything and perform no processing; they are simple sensors that transmit their measurements. Letting S_1 the message size (in bits) needed to encode the measurement $\mathbf{y}_l^{(k)}$, for some l, the total amount of information that needs to be transported is $O(S_1 nK)$ bits. Each one of these bits has to be sent over multiple hops to reach the gateway; in the worst case over K hops. Thus, the worst case communication cost is $O(S_1 nK^2)$ bits. In our setting, we are interested in deploying as few clusterheads as possible, thus, the resulting clusterhead network is sparse and likely to have a linear topology. This implies that the worst case communication cost of $O(K)$ per transmission may be typical. If the clusterhead network is more dense, a binary tree may be a good assumption for its topology which implies a communication cost of $O(\log K)$ per transmission. In that best case scenario the total communication cost becomes $O(S_1 nK \log K)$. Once this information is received, the gateway can apply the decision rule discussed in Sec. 4 to identify the region at which the sensor in question resides.

5.2 Distributed approach

Next we describe a distributed implementation for the decision rule.

We start with an arbitrary pair of regions, say L_1 vs. L_2. The clusterhead at $B_{k^*_{12}}$ based on the observations $\mathbf{y}^{(k^*_{12}),n}$ uses the GLRT to make the decision; let L_{l_1} the hypothesis accepted. The clusterhead at $B_{k^*_{12}}$ sends the information that l_1 is accepted to the clusterhead at $B_{k^*_{l_1,3}}$ which follows up with the decision L_{l_1} vs. L_3, and so on and so forth. Let now L_{l_i} denote the hypothesis accepted at stage i of the algorithm, for $i = 1, \ldots, N-1$, where we set $l_0 = 1$. At the ith stage, the clusterhead at $B_{k^*_{l_{i-1},(i+1)}}$ makes the decision $L_{l_{i-1}}$ vs. L_{i+1} and sends the result to the clusterhead at $B_{k^*_{l_i,(i+2)}}$, where the clusterhead at $B_{k^*_{l_{N-1},(N+1)}}$ is the gateway. All in all this procedure takes $N-1$ stages and $L_{l_{N-1}}$ is the final accepted hypothesis.

Each clusterhead is responsible for a set of region pairs and needs to store the corresponding pdfs and thresholds d_{ijk} as well as the necessary information to decide where to forward its decision. At every stage $i = 1, \ldots, N-1$ it takes $O(n)$

work to perform the GLRT, yielding an overall $O(nN)$ processing effort distributed to the K clusterheads. In terms of communication cost, $N-1$ messages get exchanged each consisting of $O(\log N)$ bits needed to encode the decision. Each of these messages can, in the worst case, be sent over $O(K)$ hops if two distant clusterheads need to communicate, yielding an overall worst case communication cost of $O(KN \log N)$. However, one can sequence the regions in such a way that geographically close regions are close in the sequence. As a result, it will often be the case that clusterheads responsible for region pairs close in the sequence will be geographically close.

We note that this "locality" property is plausible since the signal landscape is primarily influenced by the structure of the site. Hence, it is reasonable to expect the best clusterheads for nearby regions to be geographically close. To see that, consider a large deployment with a radius much larger than the range of the sensor nodes. The clusterheads responsible for nearby regions should be able to listen to sensors within these regions, which implies that they are geographically close compared to the overall size of the deployment.

This results in messages between clusterheads traveling a few hops. It follows that the overall communication cost will often be $O(N \log N)$.

Based on the preceding analysis, Table I compares the centralized and distributed approaches. A couple of remarks are in order. The total processing cost is the same

Table I. Comparing the centralized and distributed approaches. Typically $K = O(N)$.

	Communication cost (bits)	Processing cost
Centralized	worst: $O(S_1 n K^2)$ best: $O(S_1 n K \log K)$	$O(nN)$ at the gateway
Distributed	worst: $O(KN \log N)$ best: $O(N \log N)$	$O(nN)$ at the Kclusterheads

for both approaches but in the distributed case the work is distributed among the K clusterheads. To compare the communication costs note that typically $K = O(N)$ to ensure reasonable performance (e.g., one clusterhead for a fixed number of regions). Moreover, S_1 is the message size for the raw measurements at a clusterhead corresponding to a packet sent from the transmitting sensor, while n can be large enough (e.g., 20-30) so that the probability of error becomes small enough. Furthermore, based on the discussion earlier, we expect the worst case to be typical in the centralized approach while the best case should be typical in the distributed approach. It follows that the distributed approach leads to communication cost (and energy) savings.

Note that both the centralized and the distributed approach guarantee the performance of the system obtained in Prop. 4.2, i.e., the savings from the distributed approach come with no performance loss.

6. EXPERIMENTAL RESULTS

Next, we provide experimental results from a localization testbed we have installed at Boston University (BU).

Fig. 2. Floor plan for the testbed.

The localization testbed was implemented in a BU building and covered the areas shown on the floorplan of Fig. 2. The coverage area included typical faculty and student offices, several large rooms used as labs, two conference rooms, and an equipment bay with heavy machining equipment located on a basement level (lower middle shaded section in Fig. 2) whereas the rest of the covered area is located on the 1st floor.

The testbed uses MPR2400 (MICAz) motes from Crossbow Technology Inc. The MPR2400 (2400–2483.5 MHz band) uses the Chipcon CC2420, IEEE 802.15.4 compliant, ZigBee-ready radio frequency transceiver integrated with an Atmega128L micro-controller. Its radio can be tuned within the IEEE 802.15.4 channels, numbered from 11 (2.405 GHz) to 26 (2.480 GHz), each separated by 5 MHz. The RF transmission power is programmable from 0 dBm (1 mW) to -25 dBm. We covered 16 rooms and corridors and defined 60 regions. Within each region we placed a mote; the centers of these positions are identified by either a green circle or a red square on the floor plan. These 60 positions make up the set $\mathscr{B}$ of possible clusterhead positions. Hence, in our testbed $N = M = 60$ and B_j can be thought as the center of L_j. All 60 motes are connected to a base MICAz through a mesh network. The base mote is docked on a programming board which is connected to a laptop acting as a server.

The experimental validation of our localization approach can be divided into the five phases outlined in Fig. 3. Phase 1 can be carried out automatically by scheduling the motes so that when one is broadcasting the others are listening. We construct our pdf databases by measuring 200 packets for each pair of motes sent over two frequency channels and with two different power levels. These packets

(1) For each pair of positions (B_k, B_j) estimate the pdf $p_{\mathbf{Y}^{(k)}|B_j}(\mathbf{y})$ of RSSI at B_k when the mote at B_j is transmitting. Let m_{jk} denote the corresponding mean.
(2) For each (B_k, B_j) construct a pdf family $\{p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y}), \boldsymbol{\theta}_j \in \Omega_j\}$ to characterize transmissions from positions within L_j.
(3) Compute the exponent d_{ijk} as described in Sec. 3.1.
(4) Determine the clusterhead placement by the algorithm in Sec. 4.2.
(5) Determine the location of any mote in the coverage area by the decision rule of Sec. 4.1.

Fig. 3. Phases of the experimental validation.

are sent over a long enough time interval to capture the environment in different "states" and thus account for the variability in RSSI measurements. For Phase 2 we define an interval $[m_{jk} - \hat{m}_{jk}, m_{jk} + \hat{m}_{jk}]$ and select points $\theta_{j,1}, \ldots, \theta_{j,R}$ in this interval. We construct the family $\{p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y}), \boldsymbol{\theta}_j \in \Omega_j\}$ so that the lth member has the same shape as $p_{\mathbf{Y}^{(k)}|B_j}(\mathbf{y})$ but a mean equal to $\theta_{j,l}$, for $l = 1, \ldots, R$. $\hat{m}_{jk}$ is selected appropriately so that the union over j, k of the intervals $[m_{jk} - \hat{m}_{jk}, m_{jk} + \hat{m}_{jk}]$ is maximized and there is no overlap. The value of R determines how rich are the pdf families; in our experiments $\theta_{j,1}, \ldots, \theta_{j,R}$ were selected to include all integers in the interval $[m_{jk} - \hat{m}_{jk}, m_{jk} + \hat{m}_{jk}]$. It can be seen that to construct the pdf families we only used measurements from a single point (the center) within a region. Therefore, the measurement campaign is not necessarily more expensive than the one required by the approach in Ray et al. [2006] which uses a single (rather than a family) pdf per region. For Phase 3 we were not able to verify the GLRT optimality condition (cf. Thm. 3.2), so we obtained d_{ijk} by computing the type II exponent as in (15). The optimal placement obtained in Phase 4 is shown in Fig. 2 where we used 12 clusterheads placed at the positions of the red squares on the graph. The number of clusterheads was selected to achieve a small enough probability of error (cf. Prop. 4.2). The training phase (Steps 1 and 2 of Fig. 3) takes about a day. Step 3 depends on the hardware used to solve the corresponding optimization problems. It took about 2 days for our testbed. Step 4 takes just about half an hour. Note that these steps are performed once, assuming that the environment does not change structurally in a very dramatic manner. The detection phase (Step 5) takes on the order of 40 seconds. Finally, in terms of storage requirements, the distributed algorithm needs to store about 2 Kbytes in each of the clusterheads (pdf families, where to forward decisions, etc.).

We obtained results for three versions of the localization system. We made 100 localization tests in positions spread within the covered area. Each test used 20 packets (RSSI measurements) broadcasted by the mote to be located (5 over each channel and power level pair for the $2F2P$ cases described below). In Version 1 the mote we wish to locate transmits packets over a single frequency (2.410 GHz) and a single power level (0 dBm) and the system uses the GLRT (we write $1F1P-G$ to indicate Ver. 1 in Fig. 4) to determine the region of the mote. In Version 2 (denoted by $2F2P-G$) RSSI observations are made for packets transmitted over two different frequencies (2.410 GHz and 2.460 GHz) and two different power levels (0 dBm and -10 dBm) and the GLRT is again used. Version 3 (denoted by $2F2P - L$) is

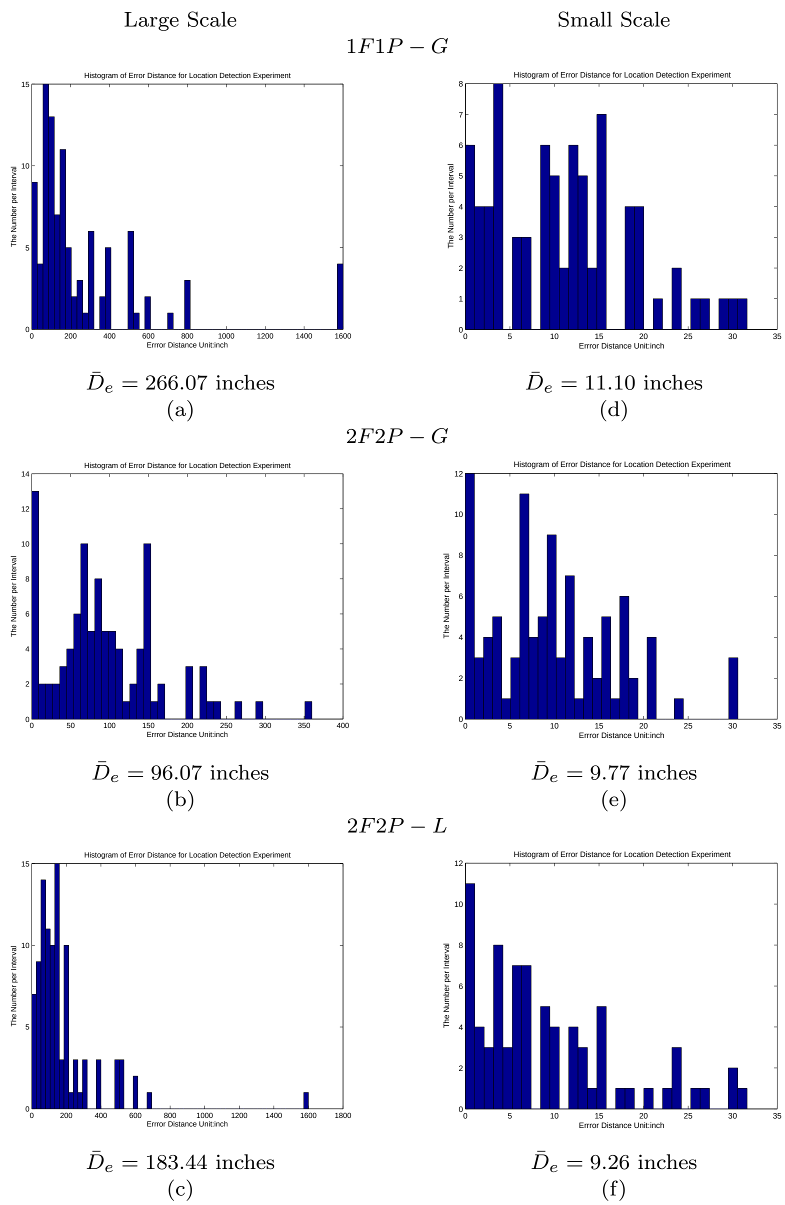

Fig. 4. Results for for various versions of the system.

identical to Version 2 but the LRT rather than the GLRT is used where every region is represented by just the pdf observed in Phase 1 (rather than a pdf family). For each Version 1–3 results are reported in Fig. 4(a)–(c), respectively. In each of these figures we plot the histogram of the error distance (in inches) based on 100 trials. If the system identifies region L_j as the one where the transmitting mote is located then the error distance is defined as the distance between the transmitting mote and

B_j. For each system we also report the corresponding mean error distance ($\bar{D}_e$). We stress that for each trial the location of the transmitting mote is randomly selected and almost never one at which RSSI measurements have been made in Phase 1.

The results show that the $2F2P-G$ system, which exploits frequency and power diversity, outperforms the $1F1P-G$ system. Clearly, RSSI measurements at multiple power and frequency levels contain more information about the transmitter location. Also, the $2F2P-G$ system outperforms the $2F2P-L$ system which uses the standard LRT decision rule. This demonstrates that, as envisioned, the GLRT provides robustness leading to better performance. The issue with the LRT is that a single pdf can not adequately represent a relatively large region. We also note that the total coverage area was 5258 feet^2, that is, about 87 feet^2 per region. With a mean error distance of $\bar{D}_e = 8$ feet the mean area of "confusion" was $8^2 = 64$ feet^2. From these results it is evident that we were able to achieve accuracy on the same order of magnitude as the mean area of a region. That is, the system was identifying the correct or a neighboring region most of the time. Put differently, we can say that the achieved mean error distance is about the same as the radius of a region, defined as radius $= \sqrt{\text{area}}$ (for our experiments $\sqrt{87} = 9.3$ feet which is in fact larger than the mean error distance of 8 feet). We used a clusterhead density of 1 clusterhead per $5258/12 = 438$ feet^2. Note that our system is *not* localizing based on "proximity" to a clusterhead; one clusterhead corresponds to about 5 regions thus resulting into cost savings compared to proximity-based systems that need a higher density of observers.

An interesting question is whether the pdf families constructed during the training phase remain valid after a long period of time or need very frequent updating (which is costly). To answer this question, we performed another (smaller) set of 56 localization tests after about one year from the time we derived our pdf families. This second set of tests yielded a mean error distance of 87.32 inches for the 2F2P-G system, quite similar to the earlier tests. During this year there have been modest changes in the building with labs and conference rooms been reorganized and several faculty moving to new offices.

For comparison purposes, we also used the same testbed and the exact same tests with the stochastic triangulation method of Patwari et al. [2003]. Patwari et al. [2003] assumes that the RSSI (in db) at B_k when the mote at B_j is transmitting, say $Y^{(k)}|B_j$, is a random variable with a Gaussian distribution. The mean of RSSI satisfies the path loss formula $\bar{Y}^{(k)}|B_j = Y_0 - 10n_p \log_{10}(\zeta_{kj}/\zeta_0)$, where ζ_{ij} is the distance between B_k and B_j and ζ_0 is a normalizing constant. From prior measurements we obtained $n_p = 3.65$ and $Y_0 = -48.62$ dBm for $\zeta_0 = 3$ feet. The location estimation is obtained by maximum likelihood estimation. Applying this method and using our clusterheads in the exact same position as before resulted in a mean error distance of 341.72 inches (29 feet) which is much larger (a factor of 3.6!) than the 8 feet obtained by our method.

These results raised the question whether smaller regions can lead to better accuracy. To that end, we placed 12 motes on a table (two rows of 6 motes each). Two neighboring motes in one row (or in one column) were 6 inches apart. We defined a 36 inches^2 region around each mote and followed the exact same procedure as before. The results of this "small scale" localization experiment are in Fig. 4(d)–(f). As before frequency and power diversity improve performance. Here, however, the GLRT does not make a difference compared to LRT and this is because every trial point in the coverage area is very close to a point we have measurements from. With the LRT we can achieve a mean error distance of 9.26 inches, that is, we can again achieve an accuracy on the same order of magnitude as the mean area of a region.

7. CONCLUSIONS

In this paper, we presented a robust and distributed approach for locating the area (region) where sensors of a WSNET reside. We posed the problem of localization as a multiple composite hypothesis testing problem and proposed a GLRT-based decision rule. We established a necessary and sufficient condition for the GLRT to be optimal in a generalized Neyman-Pearson sense but also considered the case where such optimality conditions are not met. Developing asymptotic results on the type I and type II error exponents, we described how an optimal GLRT threshold can be obtained. We then turned to the problem of optimally placing a given number of clusterheads to minimize the probability of error. We devised a placement algorithm that provides a probabilistic guarantee on the probability of error. Furthermore, we proposed a distributed approach to implement the GLRT-based decision rule and demonstrated that this can lead to savings in the communication cost compared to a centralized approach.

We validated our approach using testbed implementations involving MICAz motes manufactured by Crossbow. Our experimental results demonstrate that the GLRT-based system provides significant robustness (and improved performance) compared to an LRT-based system such as the one in Ray et al. [2006]. Furthermore, our approach leads to significantly improved accuracy compared to a stochastic triangulation technique like the one in Patwari et al. [2003] – by a factor of 3.6 in our tests. We showed that we can achieve an accuracy on the same order of magnitude as the mean area of a region. This represents the best possible accuracy for a system which identifies the region of the mote rather than estimating the exact location. Smaller regions (and more clusterheads) lead to better accuracy but at the expense of more initial measurements (training) and higher equipment cost. This provides a rule of thumb for practical systems: define as small regions as possible given a tolerable amount of initial measurements and cost.

Acknowledgments

We would like to thank Binbin Li for implementing the stochastic triangulation approach which was compared to ours.

REFERENCES

Bahl, P. and Padmanabhan, V. 2000. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of the IEEE INFOCOM Conference*. IEEE, Tel-Aviv, Israel.

Dembo, A. and Zeitouni, O. 1998. *Large Deviations Techniques and Applications*, 2nd ed. Springer-Verlag, NY.

Hoeffding, W. 1965. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist. 36*, 369–401.

Kaemarungsi, K. and Krishnamurthy, P. 2004. Modeling of indoor positioning systems based on location fingerprinting. In *Proceedings of the IEEE INFOCOM Conference*.

Lorincz, K. and Welsh, M. 2006. Motetrack: A robust, decentralized approach to rf-based location tracking. *Springer Personal and Ubiquitous Computing, Special Issue on Location and Context-Awareness*, 1617–4909.

Madigan, D., Elnahrawy, E., Martin, R. P., Ju, W.-H., Krishnan, P., and Krishnakumar, A. S. 2005. Bayesian indoor positioning systems. In *Proceedings of the IEEE Infocom Conference*. Miami, Florida.

Paschalidis, I. C. and Guo, D. 2007. Robust and distributed localization in sensor networks. In *Proceedings of the 46th IEEE Conference on Decision and Control*. New Orleans, Louisiana, 933–938.

Patwari, N., Hero, A. O., Perkins, M., Correal, N. S., and O'Dea, R. J. 2003. Relative location estimation in wireless sensor networks. *IEEE Transanctions on Signal Processing 51,* 8, 2137–2148.

Ray, S., Lai, W., and Paschalidis, I. C. 2006. Statistical location detection with sensor networks. *Joint special issue IEEE/ACM Trans. Networking and IEEE Trans. Information Theory 52,* 6, 2670–2683.

Yedavalli, K., Krishnamachari, B., Ravula, S., and Srinivasan, B. 2005. Ecolocation: A sequence based technique for rf-only localization in wireless sensor networks. In *The Fourth International Conference on Information Processing in Sensor Networks*. Los Angeles, CA.

Zeitouni, O., Ziv, J., and Merhav, N. 1992. When is the generalized likelihood ratio test optimal? *IEEE Trans. Inform. Theory 38,* 5, 1597–1602.

A. PROOF OF LEMMA 3.1

Proof. For all $\boldsymbol{\theta}_j \in \Omega_j$ we have

$$\begin{aligned}
\alpha^{\mathscr{S}^*}_{ijk,n}(\boldsymbol{\theta}_j) &= \mathbf{P}_{\boldsymbol{\theta}_j}[\mathbf{y}^{(k),n} \in \mathscr{S}^*_{ijk,n}] \\
&= \sum_{\{\mathbf{L}_{\mathbf{y}^{(k)},n} | T_n(\mathbf{L}_{\mathbf{y}^{(k)},n}) \subseteq \mathscr{S}^*_{ijk,n}\}} |T_n(\mathbf{L}_{\mathbf{y}^{(k)},n})| \; p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y}^{(k),n}) \\
&\leq \sum_{\{\mathbf{L}_{\mathbf{y}^{(k)},n} | T_n(\mathbf{L}_{\mathbf{y}^{(k)},n}) \subseteq \mathscr{S}^*_{ijk,n}\}} e^{nH(\mathbf{L}_{\mathbf{y}^{(k)},n})} e^{-n[H(\mathbf{L}_{\mathbf{y}^{(k)},n}) + D(\mathbf{L}_{\mathbf{y}^{(k)},n} \| \mathbf{P}_{\boldsymbol{\theta}_j})]} \\
&= \sum_{\{\mathbf{L}_{\mathbf{y}^{(k)},n} | T_n(\mathbf{L}_{\mathbf{y}^{(k)},n}) \subseteq \mathscr{S}^*_{ijk,n}\}} e^{-nD(\mathbf{L}_{\mathbf{y}^{(k)},n} \| \mathbf{P}_{\boldsymbol{\theta}_j})} \\
&\leq (n+1)^{|\Sigma|} e^{-n\lambda},
\end{aligned}$$

which establishes (1). For the first inequality above note that the size of the type class of $\mathbf{L}_{\mathbf{y}^{(k)},n}$ is upper bounded by $e^{nH(\mathbf{L}_{\mathbf{y}^{(k)},n})}$ and that the probability of a sequence can be written in terms of the entropy and the relative entropy of its type (see Dembo and Zeitouni [1998, Chap. 2]). In the last inequality above we used the definition of $\mathscr{S}^*_{ijk,n}$ and the fact that the set of all possible types, $\mathscr{L}_n$, has cardinality upper bounded by $(n+1)^{|\Sigma|}$ (Dembo and Zeitouni [1998, Chap. 2]).

Let now $\mathscr{S}_{ijk,n}$ be some other decision rule satisfying constraint (1), hence, for all $\epsilon > 0$ and all large enough n

$$\alpha^{\mathscr{S}}_{ijk,n}(\boldsymbol{\theta}_j) \leq e^{-n(\lambda+\epsilon)}. \tag{28}$$

Meanwhile for all $\epsilon > 0$, all large enough n, and any $\mathbf{y}^{(k),n} \in \mathscr{S}_{ijk,n}$

$$\begin{aligned}\alpha^{\mathscr{S}}_{ijk,n}(\boldsymbol{\theta}_j) &= \sum_{\{\mathbf{L}_{\mathbf{y}^{(k),n}} | T_n(\mathbf{L}_{\mathbf{y}^{(k),n}}) \subseteq \mathscr{S}_{ijk,n}\}} |T_n(\mathbf{L}_{\mathbf{y}^{(k),n}})| \; p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y}^{(k),n}) \\ &\geq \sum_{\{\mathbf{L}_{\mathbf{y}^{(k),n}} | T_n(\mathbf{L}_{\mathbf{y}^{(k),n}}) \subseteq \mathscr{S}_{ijk,n}\}} (n+1)^{-|\Sigma|} e^{-nD(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j})} \\ &\geq e^{-n[D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j}) + \epsilon]},\end{aligned}$$

where the first inequality above uses Dembo and Zeitouni [1998, Lemma 2.1.8]. Comparing the above with (28) it follows that if $\mathbf{y}^{(k),n} \in \mathscr{S}_{ijk,n}$ then for all $\boldsymbol{\theta}_j$ $D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j}) \geq \lambda$, hence, $\mathbf{y}^{(k),n} \in \mathscr{S}^*_{ijk,n}$ and $\mathscr{S}_{ijk,n} \subseteq \mathscr{S}^*_{ijk,n}$. Consequently, for all $\boldsymbol{\theta}_i$ $\beta^{\mathscr{S}}_{ijk,n}(\boldsymbol{\theta}_i) \geq \beta^{\mathscr{S}^*}_{ijk,n}(\boldsymbol{\theta}_i)$ which establishes that the generalized Hoeffding test maximizes the exponent of the type II error probability. We conclude that it satisfies the GNP criterion. □

B. PROOF OF THEOREM 3.2

PROOF. We first show that $\mathscr{S}^{GLRT}_{ijk,n} \subseteq \mathscr{S}^*_{ijk,n}$. For $\mathbf{y}^{(k),n} \in \mathscr{S}^{GLRT}_{ijk,n}$

$$\begin{aligned}\lambda \leq& \frac{1}{n} \log \sup_{\boldsymbol{\theta}_i} p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\mathbf{y}^{(k),n}) - \frac{1}{n} \log \sup_{\boldsymbol{\theta}_j} p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y}^{(k),n}) \\ =& \sup_{\boldsymbol{\theta}_i} [-H(\mathbf{L}_{\mathbf{y}^{(k),n}}) - D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_i})] \\ & - \sup_{\boldsymbol{\theta}_j} [-H(\mathbf{L}_{\mathbf{y}^{(k),n}}) - D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j})] \\ =& - \inf_{\boldsymbol{\theta}_i} D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_i}) + \inf_{\boldsymbol{\theta}_j} D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j}) \qquad (29) \\ \leq& \inf_{\boldsymbol{\theta}_j} D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j}),\end{aligned}$$

which implies that $\mathbf{y}^{(k),n} \in \mathscr{S}^*_{ijk,n}$. It follows that $\alpha^{GLRT}_{ijk,n}(\boldsymbol{\theta}_j) \leq \alpha^{\mathscr{S}^*}_{ijk,n}(\boldsymbol{\theta}_j)$ which establishes that the GLRT satisfies (1) and (4) due to Lemma 3.1.

For the type II error probability we have

$$\begin{aligned}\beta^{GLRT}_{ijk,n}(\boldsymbol{\theta}_i) =& \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{y}^{(k),n} \notin \mathscr{S}^{GLRT}_{ijk,n}] \\ =& \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{y}^{(k),n} \notin \mathscr{S}^*_{ijk,n}] + \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{y}^{(k),n} \in \mathscr{S}^*_{ijk,n} \cap \overline{\mathscr{S}^{GLRT}_{ijk,n}}], \qquad (30)\end{aligned}$$

where $\overline{\mathscr{S}^{GLRT}_{ijk,n}}$ denotes the complement of $\mathscr{S}^{GLRT}_{ijk,n}$.

Now, if $\mathbf{y}^{(k),n} \in \overline{\mathscr{S}^{GLRT}_{ijk,n}}$ then due to (29)

$$\begin{aligned}\lambda >& \frac{1}{n} \log \sup_{\boldsymbol{\theta}_i} p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_i}(\mathbf{y}^{(k),n}) - \frac{1}{n} \log \sup_{\boldsymbol{\theta}_j} p_{\mathbf{Y}^{(k)}|\boldsymbol{\theta}_j}(\mathbf{y}^{(k),n}) \\ =& \inf_{\boldsymbol{\theta}_j} D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j}) - \inf_{\boldsymbol{\theta}_i} D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_i}),\end{aligned}$$

and if $\mathbf{y}^{(k),n} \in \mathscr{S}^*_{ijk,n}$ then $\lambda \leq \inf_{\boldsymbol{\theta}_j} D(\mathbf{L}_{\mathbf{y}^{(k),n}} \| \mathbf{P}_{\boldsymbol{\theta}_j})$, which implies that if $\mathbf{y}^{(k),n} \in$

$\mathscr{S}^*_{ijk,n} \cap \overline{\mathscr{S}^{GLRT}_{ijk,n}}$ then $\mathbf{L}_{\mathbf{y}^{(k),n}} \in \mathscr{C}_{ijk}$. Sanov's theorem yields

$$-\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{y}^{(k),n} \in \mathscr{S}^*_{ijk,n} \cap \overline{\mathscr{S}^{GLRT}_{ijk,n}}] \geq \inf_{\mathbf{Q}\in\mathscr{C}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}) \quad (31)$$

for all $\boldsymbol{\theta}_i$. Combining (2) and (31)

$$\begin{aligned}
-\limsup_{n\to\infty} \frac{1}{n} \log \beta^{GLRT}_{ijk,n}(\boldsymbol{\theta}_i) &= \min\Bigg[-\limsup_{n\to\infty} \frac{1}{n} \log \beta^{\mathscr{S}^*}_{ijk,n}(\boldsymbol{\theta}_i), \\
&\qquad -\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}_{\boldsymbol{\theta}_i}[\mathbf{y}^{(k),n} \in \mathscr{S}^*_{ijk,n} \cap \overline{\mathscr{S}^{GLRT}_{ijk,n}}\Bigg] \\
&\geq \min\Bigg[\inf_{\mathbf{Q}\in\mathscr{A}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}), \inf_{\mathbf{Q}\in\mathscr{C}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i})\Bigg] \\
&= \inf_{\mathbf{Q}\in\mathscr{A}_{ijk}} D(\mathbf{Q}\|\mathbf{P}_{\boldsymbol{\theta}_i}),
\end{aligned}$$

where the last equality holds under condition (3). Thus, the type II error probability of GLRT has the same exponent as the generalized Hoeffding test if and only if condition (3) holds. Since the generalized Hoeffding test satisfies the GNP optimality condition (Lemma 3.1) so does the GLRT under condition (3). This also establishes (5). □