Chapter IX
Statistical Location Detection

Saikat Ray
University of Bridgeport, USA

Wei Lai
Boston University, USA

Dong Guo
Boston University, USA

Ioannis Ch. Paschalidis
Boston University, USA

ABSTRACT

The authors present a unified stochastic localization approach that allows a wireless sensor network to determine the physical locations of its nodes with moderate resolution, especially indoors. The area covered by the wireless sensor network is partitioned into regions; the localization algorithm identifies the region where a given sensor resides. The localization is performed using an infrastructure of stationary clusterheads that receive beacon packets periodically transmitted by the given sensor. The localization algorithm exploits the statistical characteristics of the beacon signal and treats the localization problem as a multi-hypothesis testing problem. The authors provide an asymptotic performance guarantee for the system and use this metric to determine the optimal placement of the infrastructure nodes. The placement problem is NP-hard and they leverage special-purpose algorithms from the theory of discrete facility location to solve large problem instances efficiently. They also show that localization decisions can be taken in a distributed manner by appropriate collaboration of the clusterheads. The approach is validated in a Boston University testbed.

INTRODUCTION

Localization – determining the approximate physical position of a user/device on a site – can be seen as an important enabling service in *Wireless Sensor Networks (WSNs)*. The Global Positioning System (GPS) (Hofmann-Wellenhof et al. 1997) provides an effective localization technology outdoors and its popularity and the host of location-based services it has spawned is a testament to the importance of location information. The GPS technology though is unreliable in downtown urban areas and not functional indoors. Moreover, GPS receivers are expensive and power-hungry making them inappropriate for many WSN applications that emphasize very low-cost low-power sensor nodes.

A reliable indoor localization service would be extremely useful and would give rise to a plethora of innovative applications including: asset and personnel tracking in hospitals, warehouses, and other large complexes; locating faulty sensors in building automation applications; intelligent audio players in self-guided museum tours; intelligent maps for large malls and offices, smart homes (Hodes et al. 1997, Priyantha et al. 2000); as well as surveillance, military and homeland security related applications. Moreover, a location detection service is an invaluable tool for counter-action and rescue (Meissner et al. 2002) in disaster situations.

For these reasons, localization has received widespread attention in the literature and many approaches have been developed. A large class of localization systems uses special hardware (e.g., infrared sensors, ultrasound) which necessitates the deployment of a special-purpose WSN just for this purpose. Several related works are described in the following section. We are instead interested in a localization approach that can use WSN features found in virtually all existing platforms. Specifically, all WSN nodes carry a radio to communicate with each other. That radio is often rather rudimentary and the only information on the received RF signals one can obtain is signal strength. Received signal strength depends on the location of the transmitting sensor and the objective is to exploit this information to reveal the transmitter's location. At the same time, we are interested in an approach that is general enough to exploit additional RF or other information that could be obtained with more sophisticated hardware, for instance signal angle-of-arrival and signal time-of-flight. As we will see, we are able to deal with any vector of available *observations* about the transmitting sensor.

The approach we develop in this work starts with a "discretization" of the localization problem by splitting the coverage area into a set of regions. The problem is to determine the region where a sensor node we seek resides. Quantities like signal strength are highly variable indoors due to the dynamic character of the environment leading to multipath and fading in the propagation of RF signals. For example, the propagation environment inside a building is highly complex and dynamic as there are multiple reflections, doors that may be open or shut, and people (acting as RF energy absorbers and reflectors) that are constantly moving.

To accommodate this level of variability it is critical that we use a *stochastic characterization* of signal strength or other RF characteristics the localization system may rely upon. To that end, we will associate a probabilistic description of observations to every region-clusterhead pair. In some cases, e.g., for small regions, a single probability density function (pdf) of observations may suffice to characterize observations about a sensor in a certain region as received at a specific clusterhead. In other cases, especially when the region is large, a *family* of pdfs may be necessary to accurately represent observations anywhere in the region. A pdf family is intended to provide robustness with respect to the position of the sensor within a region and can be constructed from measurements taken from locations within the region. This flexible stochastic characterization of observations, ranging from a single pdf to a sufficiently rich family of pdfs combines earlier localization approaches we have developed (Ray et al. 2006, Paschalidis and Guo 2007). Once we have these probabilistic descriptors we can think of the localization problem as a hypothesis testing problem where we have to match observations to a single pdf or a pdf family. In the latter case, the problem is known as a composite hypothesis testing problem. To make decisions we will rely on likelihood ratio tests which we show to be optimal in a certain asymptotic sense. Our optimal approach increases the accuracy more than 3 times over an alternative approach; cf. Section "TESTBED AND EXPERIMENTAL RESULTS". Further, note that we are interested in locating the current position of a node; we do not assume any mobility model and as such do not attempt to predict/estimate any mobility pattern.

An advantage of the approach we advance is that we are able to characterize the performance of the localization system, quantified by the probability of error. In particular, we obtain the dominant exponent of this probability as the number of observations grows large. Having a meaningful performance metric enables us to pose the following design question: How should clusterheads be placed to minimize the probability of error? We study this optimal deployment/WSN-design problem which turns out to be NP-complete. However, we leverage results from the theory of discrete facility location and present an efficient algorithm that can solve reasonably large instances.

An important consideration in WSNs is whether the localization algorithm can run in a distributed manner by appropriate in-network processing. We demonstrate that one can organize the necessary computations so that 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 (BU) building, and our experimental results establish that we can achieve accuracy that is, roughly, on the same order of magnitude as the radius of our regions. This is, in fact, the best possible accuracy one can expect from a discretized system that turns localization into the problem of identifying the sensor's region. We report experimental results from our testbed showing great promise in using an approach of this type in a practical setting.

The rest of this chapter is organized as follows. We first discuss related work. In PROBLEM FORMULATION, we introduce our system model. In MATHEMATICAL FOUNDATION we introduce the mathematical underpinnings of hypothesis testing. In the same section we review the standard hypothesis testing problem, study the composite binary hypothesis testing problem, establish an optimality condition for the test we propose, and obtain bounds on the error exponents which allow us to optimize performance. In OPTIMUM CLUSTERHEAD PLACEMENT we consider the WSN design problem and present a fast algorithm for solving it in an efficient manner. In LOCALIZATION DECISIONS we develop the distributed decision approach and compare it to a centralized one. Results from an implementation of our approach in the testbed are reported in TESTBED AND EXPERIMENTAL RESULTS. Final remarks are in CONCLUSIONS.

RELATED WORK

Several non-GPS location detection systems have been proposed in the literature. One class of localization systems is "deterministic" and as in (Bahl et al. 2000) compares the mean signal strength from a sensor to a pre-computed signal-strength map of the coverage area. This approach though, may be unreliable indoors due to the significant variability of the RF signal landscape (due to multipath, fading, etc.). A similar system is SpotOn (Hightower et al. 2000). The *Nibble* system improves up on *RADAR* by taking the probabilistic nature of the problem into account (Castro et al. 2001). A similar approach is also found in (Yong Wu et al. 2007). Another class of systems uses trilateration or stochastic trilateration techniques as in (Patwari et al. 2003) where signal strength measurements are used to estimate the distance and location. These techniques assume a model describing how signal strength reduces with distance (path loss formula) and the modeling error can lead to inaccuracies. In the experimental results we report in this chapter, our approach is shown to significantly reduce the mean error distance compared to stochastic trilateration techniques. In (Patwari et al. 2008), a more accurate path loss model including correlated shadow loss and non-shadow loss was introduced. In (Battiti et al. 2003), the location detection problem is cast in a statistical learning framework to enhance the models. In (Lasse Klingbeil et al. 2008), a sequential Monte Carlo simulation technique was introduced to estimate the location and motion of a mobile sensor. The Monte Carlo techniques require some information about the mobility model or probability of a node's location at a given time; we assume neither. Performance trade-off and deployment issues are explored in (Prasithsangaree et al. 2002). References to many other systems can be found in the homepage of (Youssef 2008).

In addition to the related work, the works presented in this chapter not only describe and evaluate a localization system, but also characterize the performance, outline optimization approaches and propose a distributed decision making algorithm.

PROBLEM FORMULATION

In this section we introduce our system model. Consider a WSN deployed in a site for localization purposes. The reader may assume the site to be the interior of a building. We divide the site into N regions denoted by an index set $L = \{L_1,\ldots,L_N\}$. There are M distinct positions $B=\{B_1,\ldots,B_M\}$ at which we can place the fixed infrastructure nodes we call clusterheads.

Let a sensor be located in region $l \in L$. A series of packets broadcast by the sensor are received by the clusterheads (not necessarily all of them) which observe certain physical quantities associated with each packet. In most existing WSN platforms the observed physical quantities are just the received signal strength indicator (RSSI), which is related to the voltage observed at the receiver's antenna circuit.

Let $y^{(i)}$ 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 we will assume that the observations take values from a finite alphabet $\Sigma = \{\sigma_1, \ldots, \sigma_{|\Sigma|}\}$, where $|\Sigma|$ denotes the cardinality of Σ. In practice, this is indeed the case since WSN nodes report quantized RSSI measurements. A series of n consecutive observations are denoted by $y_1^{(i)}, \ldots, y_n^{(i)}$ and are assumed independent and identically distributed (i.i.d.) conditioned on the region the sensor node resides. This assumption is well justified for fairly dynamic sites where the various radio-paths between the receiver and the transmitter change rapidly; for typical indoor sensor networks observations separated by a few seconds can be i.i.d. Observations made by different clusterheads at about the same time need not be independent.

With every clusterhead-region pair (B_i, L_j) we associate a family of pdfs $p_{Y^{(i)}|\theta_j}(y)$ where $Y^{(i)}$ denotes the random variable corresponding to observations $y^{(i)}$ at clusterhead B_i when the transmitting sensor is in some location within L_j. Here, $\theta_j \in \Omega_j$ is a vector in some space Ω_j parameterizing the pdf family. We allow the possibility that for some (typically small) regions the pdf family degenerates into a single pdf. We will be writing $p_{Y^{(i)}|L_j}(y)$ in that case. Such a pdf can be obtained by using measurements at a single position within the region and constructing an empirical distribution. For larger regions, a single pdf may not be an appropriate representation for all positions within the region. As we will see, 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 θ_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 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 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.

MATHEMATICAL FOUNDATION

In this section we take the clusterhead locations as given and formulate the hypothesis testing problem for determining the location of sensors. First we describe the simpler case where there is only one pdf associated with each location. Then we expand on the case of composite hypothesis testing where there are more than one pdf per region.

Binary Hypothesis Testing

Suppose we place clusterheads in K out of the M available positions in B. Without loss of generality let these positions be $B_1,\ldots,B_K$. Suppose also that a sensor is in some location $l \in L$ and transmitting packets. As before, let $y^{(i)}$ be the vector of observations at each clusterhead $i = 1, \ldots, K$; we write $y = (y^{(1)}, \ldots, y^{(K)})$ for the vector of observations at all K clusterheads. These observations are random; let Y denote the random variable corresponding to y and $p_{Y|L_j}(y)$ the pdf of Y conditional on the sensor being in location $L_j \in L$. Observations $y^{(i)}$ and $y^{(j)}$ made at the same instant need not be independent. If they are, however, it follows that $p_{Y|L_j}(y) = p_{Y^{(1)}|L_j}(y^{(1)}) \cdots p_{Y^{(K)}|L_j}(y^{(K)})$. Suppose that the clusterheads make n consecutive observations $y_1, \ldots, y_n$, which are assumed i.i.d. Based on these observations we want to determine the location l of the sensor.

The problem at hand is a standard N-ary hypothesis testing problem. It is known that the maximum *a posteriori* probability (MAP) rule is *optimal* in the sense of minimizing the probability of error. More specifically, we declare $l = L_j$ if

$$j = \arg\max_{i=1,\ldots,N} [\pi_i p_{Y|L_i}(y_1) \cdots p_{Y|L_i}(y_n)] \tag{1}$$

(ties are broken arbitrarily), where π_i denotes the prior probability that the sensor is in location L_i.

Next we turn our attention to binary hypothesis testing for which tight asymptotic results on the probability of error are available. These results will be useful in establishing performance guarantees for our proposed clusterhead placement later on.

Suppose that the sensor's position is either L_i or L_j. A clusterhead located at B_k makes n i.i.d. observations $y_1^{(k)}, \ldots, y_n^{(k)}$. Let

$$X_{ijk}(y) = \log[p_{Y^{(k)}|L_i}(y) / p_{Y^{(k)}|L_j}(y)]$$

be the log-likelihood ratio. Define

$$\Lambda_{ijk}(\lambda) = \log \mathrm{E}_{L_j}[e^{\lambda X_{ijk}(y)}] \tag{2}$$

The expectation is taken with respect to the density $p_{Y^{(k)}|L_j}(y)$. It follows that

$$\Lambda_{ijk}(\lambda) = \log \int_{-\infty}^{\infty} p^{\lambda}_{Y^{(k)}|L_i}(y) p^{1-\lambda}_{Y^{(k)}|L_j}(y)\mathrm{d}y \tag{3}$$

The function $\Lambda_{ijk}(\lambda)$ is the log-moment generating function of the random variable X_{ijk}, hence convex (see Dembo and Zeitouni 1998, Lemma 2.2.5 for a proof). Let d_{ijk} be the Fenchel-Legendre transform (or convex dual) of $\Lambda_{ijk}(\lambda)$ evaluated at zero, i.e.,

$$d_{ijk} = \sup_{\lambda \in [0,1]} [-\Lambda_{ijk}(\lambda)]. \tag{4}$$

d_{ijk} is the so called Chernoff information or distance (it is nonnegative, symmetric, but it does not satisfy the triangle inequality) between the densities $p_{Y^{(k)}|L_i}(y)$ and $p_{Y^{(k)}|L_j}(y)$ (Dembo and Zeitouni 1998 § 3.4, Chernoff 1952).

Consider next the probability of error in this binary hypothesis testing problem when we only use the observations made by clusterhead B_k. Suppose we make decisions optimally and let S^n denote the optimal decision rule (i.e., a mapping of $y_1^{(k)}, \ldots, y_n^{(k)}$ onto either "accept L_i" or "accept L_j"). We have two types of errors with probabilities

$$\alpha_n = \mathrm{P}_{L_j}[S^n \text{rejects } L_j] \;, \quad \beta_n = \mathrm{P}_{L_i}[S^n \text{ rejects } L_i] \tag{5}$$

The first probability is evaluated under $p_{Y^{(k)}|L_j}(y)$ and the second under $p_{Y^{(k)}|L_i}(y)$. The probability of error, $P_n^{(e)}$, of the rule S^n is simply $P_n^{(e)} = \pi_i \alpha_n + \pi_j \beta_n$. Large deviations asymptotics for the probability of error under the optimal rule S^n have been established by (Chernoff 1952, Dembo and Zeitouni 1998 Corollary 3.4.6) and are summarized in the following theorem.

Theorem IV.1 (Chernoff's bound) *If* $0 < \pi_i < 1$*, then*

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n = \lim_{n\to\infty} \frac{1}{n} \log \beta_n = \lim_{n\to\infty} \frac{1}{n} \log P_n^{(e)} = -d_{ijk}$$

In other words, all these probabilities approach zero exponentially fast as n grows and the exponential decay rate equals the Chernoff distance d_{ijk}. Intuitively, these probabilities behave as $f(n)e^{-nd_{ijk}}$ for sufficiently large n, where $f(n)$ is a slowly growing function in the sense that $\lim_{n\to\infty}(\log f(n))/n = 0$. In the sequel, we will often consider the asymptotic rate according to which probabilities approach zero as $n \to \infty$. We will use the term *exponent* to refer to the quantity $\lim_{n\to\infty}(1/n)\log P[\cdot]$ for some probability P[·]; if the exponent is d then the probability approaches zero as e^{-nd}. When the Maximum Likelihood (ML) rule is optimal (i.e., prior probabilities of the hypotheses are equal), we have

Theorem IV.2 *Suppose* $\pi_i = 1/N \quad \forall i$*. Then* $P_n^{(e)} \le e^{-nd_{ijk}}$ *for all n.*

The proof is given in Appendix A. Note the interesting fact that Chernoff distances, and thus the exponents of the probability of errors do not depend on the priors π_i. This observation is important since prior probabilities are not within the control of the system designer, but conditional pdfs are.

The Chernoff distances between the joint densities of the data observed by all the clusterheads can be defined similarly by replacing $p_{Y^{(k)}|L_i}(y)$ and $p_{Y^{(k)}|L_j}(y)$ by $p_{Y|L_i}(y)$ and $p_{Y|L_j}(y)$, respectively. The clusterhead placement problem consists of choosing the placement so as to maximize these Chernoff distances. However, the optimum clusterhead placement that maximizes distances between the joint densities turns out to be a nonlinear problem with integral constraints. It quickly becomes intractable with increasing problem size and the optimum clusterhead placement for realistic sites cannot be computed using such a formulation. Optimization of the clusterhead placement in our formulation, on the other hand, reduces to a linear optimization problem (although still with integral constraints), for which large problem instances can be solved within reasonable time. For the resultant placement, we derive bounds on the probability of error of the decision rule that uses joint distributions.

Binary Composite Hypothesis Testing

We now consider the case where regions *i* and *j* have an associated pdf family $p_{Y^{(k)}|\theta_i}(y)$ and $p_{Y^{(k)}|\theta_j}(y)$, respectively, corresponding to observations at clusterhead B_k (θ_j and θ_i depend on *k* as well but we elect to suppress this dependence in the notation for simplicity). The clusterhead makes *n* i.i.d. observations $y^{(k),n} = (y_1^{(k)}, \ldots, y_n^{(k)})$ from which we need to determine the region L_i vs. L_j. We will be using the notation $p_{Y^{(k)}|\theta_i}(y^{(k),n}) = \prod_{l=1}^{n} p_{Y^{(k)}|\theta_i}(y_l^{(k)})$.

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

$$X_{ijk}(y^{(k),n}) = \frac{1}{n}\log\frac{\sup_{\theta_i\in\Omega_i} p_{Y^{(k)}|\theta_i}(y^{(k),n})}{\sup_{\theta_j\in\Omega_j} p_{Y^{(k)}|\theta_j}(y^{(k),n})}$$

to a threshold λ and declares L_i whenever

$$y^{(k),n} \in S_{ijk,n}^{GLRT} = \{y^{(k),n} \mid X_{ijk}(y^{(k),n}) \ge \lambda\}$$

and L_j otherwise. Note that in the case one of the pdf families, say the one corresponding to the (B_k, L_i) clusterhead-region pair, is a singleton $p_{Y^{(k)}|L_i}(y)$ the supremum in the numerator is moot. When both pdf families are singletons GLRT becomes the standard LRT and a threshold $\lambda = 0$ should be used. There are two types of error (referred to as type I and type II, respectively) associated with a decision with probabilities

$$\alpha_{ijk,n}^{GLRT} = \mathrm{P}_{\theta_j}[y^{(k),n} \in S_{ijk,n}^{GLRT}], \qquad \beta_{ijk,n}^{GLRT} = \mathrm{P}_{\theta_i}[y^{(k),n} \notin S_{ijk,n}^{GLRT}]$$

where $P_{\theta_j}[\cdot]$ (resp. $P_{\theta_i}[\cdot]$) is a probability evaluated assuming that $y^{(k),n}$ is drawn from $p_{Y^{(k)}|\theta_j}(y)$ (resp. $p_{Y^{(k)}|\theta_i}(y)$). We use a similar notation and write $\alpha^S_{ijk,n}$ and $\beta^S_{ijk,n}$ for the error probabilities of any other test that declares L_i whenever $y^{(k),n}$ is in some set $S_{ijk,n}$.

Since we have two probabilities of error we can not minimize both at the same time. A natural objective is to minimize one (type II) subject to a constraint on the other (type I). This is known as the *generalized Neyman-Pearson* optimality criterion and is given below.

Definition 1

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

$$\limsup_{n\to\infty} \frac{1}{n} \log \alpha^S_{ijk,n}(\theta_j) < -\lambda, \quad \forall\, \theta_j \in \Omega_j \tag{6}$$

and maximizes $-\limsup_{n\to\infty} \frac{1}{n} \log \beta^S_{ijk,n}(\theta_i)$ *uniformly for all* $\theta_i \in \Omega_i$.

Zeitouni et al. 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 to the situation of interest where both hypotheses correspond to a family of pdfs. We will establish a necessary and sufficient condition for the GLRT to satisfy the GNP criterion.

Let us introduce some additional notation, which is common in information theory. For any sequence of observations $y^n=(y_1,\ldots,y_n)$, the empirical measure (or type) is given by $L_{y^n} = (L_{y^n}(\sigma_1),\ldots,L_{y^n}(\sigma_{|\Sigma|}))$, where

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

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

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

be the entropy of the probability vector ν and

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

the divergence or relative entropy of ν with respect to another probability vector μ.

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). Let P_{θ_j} denote the probability law induced by $p_{Y^{(k)}|\theta_j}(\cdot)$. Considering hereafter tests that depend only on L_{y^n}, the so called generalized Hoeffding test (Hoeffding 1965) that accepts L_i when $y^{(k),n}$ is in the set

$$S^*_{ijk,n} = \{y^n \mid \inf_{\theta_j} D(L_{y^n} \,\|\, P_{\theta_j}) \geq \lambda\},$$

and accepts L_j otherwise, is optimal according to the GNP criterion. The following lemma generalizes Hoeffding's result and a similar result in (Zeitouni et al. 1992); the proof is in Appendix B.

Lemma IV.3 *The generalized Hoeffding test satisfies the GNP criterion.*

Next, we will determine the exponent of $\beta_{ijk,n}^{S^*}(\theta_i)$. Define the set

$A_{ijk} = \{Q \mid \inf_{\theta_j} D(Q \| P_{\theta_j}) < \lambda\}.$

We have

$\beta_{ijk,n}^{S^*}(\theta_i) = P_{\theta_i}[y^{(k),n} \notin S_{ijk,n}^*] = P_{\theta_i}[L_{y^{(k),n}} \in A_{ijk} \cap L_n].$

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

$$\inf_{Q \in A_{ijk}} D(Q \| P_{\theta_i}) \le -\limsup_{n\to\infty} \frac{1}{n} \log \beta_{ijk,n}^{S^*}(\theta_i) \le -\liminf_{n\to\infty} \frac{1}{n} \log \beta_{ijk,n}^{S^*}(\theta_i) \le \inf_{Q \in A_{ijk}^o} D(Q \| P_{\theta_i}) \tag{7}$$

where A_{ijk}^o denotes the interior of A_{ijk}. Since A_{ijk} is an open set the upper and lower bounds match and $\inf_{Q\in A_{ijk}} D(Q \| P_{\theta_i})$ is the exponent of $\beta_{ijk,n}^{S^*}(\theta_i)$.

The following theorem establishes a necessary and sufficient condition for the optimality of GLRT under the GNP criterion. The proof is omitted; we refer the interested reader to (Paschalidis and Guo 2007).

Theorem IV.4 *The GLRT with a threshold* λ *is asymptotically optimal under the GNP criterion, if and only if*

$$\inf_{Q \in C_{ijk}} D(Q \| P_{\theta_i}) \ge \inf_{Q \in A_{ijk}} D(Q \| P_{\theta_i}) \tag{8}$$

for all θ_i, where

$C_{ijk} = \{Q \mid \inf_{\theta_j} D(Q \| P_{\theta_j}) - \inf_{\theta_i} D(Q \| P_{\theta_i})\} < \lambda \le \inf_{\theta_j} D(Q \| P_{\theta_j})\}.$

Furthermore, assuming that (8) is in effect

$$-\liminf_{n\to\infty} \frac{1}{n} \log \alpha_{ijk,n}^{GLRT}(\theta_j) \le -\lambda, \quad \forall\, \theta_j \in \Omega_j \tag{9}$$

$$-\liminf_{n\to\infty} \frac{1}{n} \log \beta_{ijk,n}^{GLRT}(\theta_i) \le -\inf_{Q \in A_{ijk}} D(Q \| P_{\theta_i}), \quad \forall \theta_i \in \Omega_i. \tag{10}$$

Although, Thm. IV.4 is an interesting theoretical result, in practice it is not trivial to verify whether condition (8) is satisfied or not. To that end, the following theorem derives bounds on the type I and type II error probability exponents in the absence of condition (8).

Theorem IV.5 *The GLRT with a threshold* λ *satisfies*

$$-\liminf_{n\to\infty} \frac{1}{n} \log \alpha_{ijk,n}^{GLRT}(\theta_j) \leq -\lambda, \quad \forall\, \theta_j \in \Omega_j \tag{11}$$

$$-\liminf_{n\to\infty} \frac{1}{n} \log \beta_{ijk,n}^{GLRT}(\theta_i) \leq -\inf_{Q \in D_{ijk}} D(Q \,\|\, P_{\theta_i}), \quad \forall \theta_i \in \Omega_i. \tag{12}$$

where

$$D_{ijk} = \{Q \mid \inf_{\theta_j} D(Q \,\|\, P_{\theta_j}) - \inf_{\theta_i} D(Q \,\|\, P_{\theta_i})\} < \lambda\}.$$

Proof: For $y^{(k),n} \in S_{ijk,n}^{GLRT}$

$$\begin{aligned}
\lambda &\leq \frac{1}{n} \log \sup\nolimits_{\theta_i \in \Omega_i} p_{Y^{(k)}|\theta_i}(y^{(k),n}) - \frac{1}{n} \log \sup\nolimits_{\theta_j \in \Omega_j} p_{Y^{(k)}|\theta_j}(y^{(k),n}) \\
&= \sup_{\theta_i}[-H(L_{y^{(k),n}}) - D(L_{y^{(k),n}} \,\|\, P_{\theta_i})] - \sup_{\theta_j}[-H(L_{y^{(k),n}}) - D(L_{y^{(k),n}} \,\|\, P_{\theta_j})] \\
&= -\inf_{\theta_i} D(L_{y^{(k),n}} \,\|\, P_{\theta_i}) + \inf_{\theta_j} D(L_{y^{(k),n}} \,\|\, P_{\theta_j})
\end{aligned} \tag{13}$$

This implies that $y^{(k),n} \in S_{ijk,n}^{GLRT}$ is equivalent to $y^{(k),n} \notin D_{ijk}$.

Next note that the right hand side of (13) is upper bounded by $\inf_{\theta_j} D(L_{y^{(k),n}} \,\|\, P_{\theta_j})$ which implies that $y^{(k),n} \in S_{ijk,n}^{*}$ as well. It follows that $\alpha_{ijk,n}^{GLRT}(\theta_j) \leq \alpha_{ijk,n}^{*}(\theta_j)$ which establishes that the GLRT satisfies (11) due to Lemma IV.3.

To compute the type II exponent note that

$$\beta_{ijk,n}^{GLRT}(\theta_i) = P_{\theta_i}[y^{(k),n} \notin S_{ijk,n}^{GLRT}] = P_{\theta_i}[y^{(k),n} \in D_{ijk}]$$

An immediate application of Sanov's theorem (Dembo and Zeitouni 1998, Chap. 2) yields (12). ∎

Determining the Optimal Threshold

It can be seen from (11) and (12) 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 would like 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. (12)):

$$Z_{ijk}(\lambda, \theta_i) = \min\nolimits_Q D(Q \,\|\, P_{\theta_i})$$

$$\text{s.t. } \min\nolimits_{\theta_j} D(Q \,\|\, P_{\theta_j}) - \min\nolimits_{\theta_i} D(Q \,\|\, P_{\theta_i}) \leq \lambda,$$

which is equivalent to

$$\begin{aligned} &Z_{ijk}(\lambda,\theta_i) = \min_Q D(Q \| P_{\theta_i}) \\ \text{s.t. } &\min_{\theta_j} D(Q \| P_{\theta_j}) - D(Q \| P_{\theta_i}) \le \lambda \quad \forall \theta_i \end{aligned} \tag{14}$$

The worst case exponent over $\theta_i \in \Omega_i$ is given by

$$Z_{ijk}(\lambda) = \min_{\theta_i} Z_{ijk}(\lambda, \theta_i).$$

Note that $Z_{ijk}(\lambda)$ is nonincreasing in λ, 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 (11) and (12) are no smaller than λ^*_{ijk}.

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

$$\hat{d}_{ijk} = \max\{\lambda^*_{ijk},\ \lambda^*_{jik}\} \tag{15}$$

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

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

The discussion above leads to the following proposition.

Proposition IV.6 *Suppose that the GLRT at clusterhead B_k compares $X_{\bar{i}\,\bar{j}k}(y^{(k),n})$ to $\hat{d}_{ijk}$. The maximum probability of error satisfies*

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

One of the challenges computing $\hat{d}_{ijk}$ is that the problem in (14) is nonconvex. This may not be an issue when there are relatively few possible values of θ_j and θ_i but for large sets Ω_j and Ω_i computing $\hat{d}_{ijk}$ becomes expensive. To address this issue, we will next develop a lower bound to $Z_{ijk}(\lambda, \theta_i)$ using nonlinear duality.

Let $\bar{Z}_{ijk}(\lambda,\theta_i)$ be the optimal value of the dual of (14); by weak duality it follows that $\bar{Z}_{ijk}(\lambda,\theta_i) \ge Z_{ijk}(\lambda,\theta_i)$ We have

$$\bar{Z}_{ijk}(\lambda,\theta_i) = \max_{\mu_{\theta_i}\ge 0}[\min_{\theta_j} \min_Q [D(Q \| P_{\theta_i}) + \sum_{\theta_i} \mu_{\theta_i} D(Q \| P_{\theta j}) - \sum_{\theta_i} \mu_{\theta_i} D(Q \| P_{\theta i})] - \sum_{\theta_i} \mu_{\theta_i} \lambda]$$

By simple algebra, we have

$$\bar{Z}_{ijk}(\lambda,\theta_i) = \max_{\mu_{\theta_i}\ge 0}[\min_{\theta_j} \min_Q [\sum_{r=1}^{|\Sigma|} Q(\sigma_r) \log(Q(\sigma_r) A(\sigma_r))] - \sum_{\theta_i} \mu_{\theta_i} \lambda] \tag{16}$$

where $A(\sigma_r) = \dfrac{1}{P_{Y^{(k)}|\theta_i}(\sigma_r)} \cdot \prod_{\theta_i} \left(\dfrac{P_{Y^{(k)}|\theta_i}(\sigma_r)}{P_{Y^{(k)}|\theta_j}(\sigma_r)} \right)^{\mu_{\theta_i}}$.

Note that the optimization over Q is convex and the optimization over μ_{θ_i} is concave, thus, this problem can be solved efficiently. In fact, the optimization over Q can be solved analytically yielding

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

$\overline{Z}_{ijk}(\lambda, \theta_i)$ is convex and nonincreasing in λ for all θ_i. Furthermore, the exponent of the type II GLRT error probability is no smaller than $\overline{Z}_{ijk}(\lambda) = \min_{\theta_i} \overline{Z}_{ijk}(\lambda, \theta_i)$. Note that $\overline{Z}_{ijk}(\lambda)$ is also nonincreasing in λ, and $\lim_{\lambda \to \infty} \overline{Z}_{ijk}(\lambda) = 0$. Assuming that $\overline{Z}_{ijk}(\lambda) > 0$, there exists a $\overline{\lambda}^*_{ijk} > 0$ such that $\overline{Z}_{ijk}(\overline{\lambda}^*_{ijk}) = \overline{\lambda}^*_{ijk}$. Furthermore, both error probability exponents in (11) and (12) are no smaller than $\overline{\lambda}^*_{ijk}$

Following the same line of development as before, set

$$\overline{d}_{ijk} = \max\{\overline{\lambda}^*_{ijk}, \overline{\lambda}^*_{jik}\} \tag{17}$$

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

Proposition IV.7 *Suppose that the GLRT at clusterhead B_k compares $X_{\overline{i}\,\overline{j}k}(y^{(k),n})$ to $\overline{d}_{ijk}$. The maximum probability of error satisfies*

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

OPTIMUM CLUSTERHEAD PLACEMENT

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

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 $y^{(k),n} = (y^{(k)}_1, \ldots y^{(k)}_n)$. Let d_{ijk} be the GLRT threshold obtained in the previous section for each region pair (i, j), $i < j$, and clusterhead k. Specifically, d_{ijk} is the Chernoff distance if we compare regions each with a single pdf associated to observations at B_k; in that case as we explained earlier we apply the LRT decision rule. If however, one of the hypothesis is composite, i.e., there is a pdf family associated with at least one of the two regions, then we apply the GLRT and the error exponent is obtained from either (15), or (17), depending on which optimization problem we elect to solve.

Figure 1. Clusterhead placement MILP formulation

$$\max \quad \epsilon \tag{18}$$

$$\text{s.t.} \quad \sum_{k=1}^{M} x_k = K \tag{19}$$

$$\sum_{k=1}^{M} y_{ijk} = 1, \; i, j = 1, \ldots, N, i < j, \tag{20}$$

$$y_{ijk} \le x_k, \; \forall i, j, i < j, k = 1, \ldots, M, \tag{21}$$

$$\epsilon \le \sum_{k=1}^{M} d_{ijk} y_{ijk}, \; \forall i, j, i < j, \tag{22}$$

$$y_{ijk} \ge 0, \; \forall i, j, i < j, \forall k, \tag{23}$$

$$x_k \in \{0, 1\}, \; \forall k. \tag{24}$$

We make $N-1$ binary decisions with the LRT or 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 the previous section 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.

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.

In this formulation, the decision variables are x_k, y_{ijk}, and ε where $k = 1,\ldots,M,\ i, j = 1,\ldots,N,\ i < j$. x_k is the indicator function of a clusterhead been placed at position B_k. Equation (19) represents the constraint that K clusterheads are to be placed. Constraint (22) enforces that for every region pair there exist a clusterhead k with d_{ijk} larger than ε . Let x_k^*, y_{ijk}^*, and ε^* $(k = 1, \ldots, M,\ i, j = 1, \ldots, N,\ i < j)$ be an optimal solution of this MILP. Although this problem is NP-hard (Ray et al. 2006), it can be solved efficiently for sites with more than 100 locations by using a special purpose algorithm proposed in the sequel.

The next proposition establishes a useful property for the optimal solution and value of the MILP in Figure 1. In preparation for that result consider an arbitrary placement of K clusterheads. More specifically, let Y be any subset of the set of potential clusterhead positions B with cardinality K. Let $x(Y) = (x_1(Y), \ldots, x_M(Y))$ where $x_k(Y)$ is the indicator function of B_k being in Y. Define:

$$\varepsilon(Y) = \min_{\substack{i,j=1,\ldots,N \\ i<j}} \max_{k|x_k(Y)=1} d_{ijk} \tag{25}$$

We can interpret $\max_{k|x_k(Y)=1} d_{ijk}$ as the best decay rate for the probability of error in distinguishing between locations L_i and L_j from some clusterhead in Y. Then $\varepsilon(Y)$ is simply the worst pairwise decay rate.

Proposition V.1 *For any clusterhead placement Y we have*

$$\varepsilon^* \geq \varepsilon(Y) \tag{26}$$

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

$$\varepsilon^* = \min_{\substack{i,j=1,\ldots,N \\ i<j}} \max_{k|x_k^*=1} d_{ijk} \tag{27}$$

Proof: Consider the placement Y and let

$$y_{ijk} = \begin{cases} 1, \text{ if } k = \arg\max\limits_{k|x_k(Y)=1} d_{ijk}, \\ 0, \text{ otherwise,} \end{cases} ,\forall i\ \ j, i<j, \forall k.$$

If more than one y_{ijk} are 1 for a given pair (i, j), we arbitrarily set all but one of them to 0 to satisfy Eq. (20). Then

$$\min_{\substack{i,j \\ i<j}} \sum_{k=1}^{M} d_{ijk} y_{ijk}^* = \min_{\substack{i,j \\ i<j}} \max_{k|x_k(Y)=1} d_{ijk} = \varepsilon(Y).$$

Observe that $x(Y)$, y_{ijk}*'s* (as defined above), and $\varepsilon(Y)$ form a feasible solution of the MILP in Figure 1. Clearly, the value of this feasible solution can be no more than the optimal ε^*, which establishes (26).

Next note that (22) is the only constraint on ε . So, we have

$$\varepsilon^* = \min_{\substack{i,j \\ i<j}} \sum_{k=1}^{M} d_{ijk} y_{ijk}^* = \min_{\substack{i,j \\ i<j}} \sum_{k|x_k^*=1}^{M} d_{ijk} y_{ijk}^* . \tag{28}$$

The second equality is due to (21). The final observation is that the right hand side of the above is maximized when

$$y_{ijk}^* = \begin{cases} 1, \text{ if } k = \arg\max\limits_{k|x_k^*=1} d_{ijk}, \\ 0, \text{ otherwise,} \end{cases} \quad \forall i\ \ j, i<j, \forall k.$$

(Again, at most one y_{ijk}^* is set to 1 for a given (i, j) pair.) Thus, an optimal solution satisfies the above. This, along with (28) establishes (27). $\blacksquare$

As before, $\max_{k|x_k^*=1} d_{ijk}$ is the best decay rate for the probability of error in distinguishing between locations L_i and L_j from some clusterhead in the set Y^*. Then ε^* is simply the worst such decay rate over all pairs of locations. Moreover, according to Proposition V.1, this worst decay rate is no worse than the corresponding quantity $\varepsilon(Y)$ achieved by any other clusterhead placement Y.

Efficient Computation of the Proposed MILP

In this section, we propose an algorithm that solves the MILP presented in Figure 1 faster than a general purpose MILP-solver such as CPLEX (CPLEX 8.0 2002). Our approach is to construct an alternate formulation of the proposed MILP first, and then solve it using an iterative algorithm. The computational advantage of this approach lies in the fact that we solve a *feasibility* problem in each iteration that contains only $O(M)$ variables and $O(N^2)$ constraints instead of $O(N^2M)$ variables and $O(N^2M)$ constraints that appear in the formulation in Figure 1, and thus can be solved much faster.

Alternate Formulation

Let us sort the d_{ijk}'s, in nonincreasing order, and let b_{ijk} denote the index of d_{ijk}. We let equal distances have the same index. Note that b_{ijk} is a positive integer upper bounded by $MN(N-1)/2$. Now consider the MILP problem shown in Figure 2. This problem is actually the MILP formulation of the *vertex K -center problem* (Daskin 1995). The following proposition establishes that the formulations of Figure 1 and Figure 2 are indeed equivalent.

Proposition V.2 *Suppose* (s^*, t^*, π^*) *is an optimal solution to the problem in Figure 2. Then* $(x^* = s^*, y^* = t^*, \varepsilon^* = d_{i^*j^*k^*})$ *is an optimal solution to the MILP problem in Figure 1, where* (i^*, j^*, k^*) *is such that* $b_{i^*j^*k^*} = \pi^*$.

Proof: A proof analogous to the one of Prop. V.1 establishes

$$\pi^* = \max_{\substack{i,j \\ i<j}} \min_{k|s_k^*=1} b_{ijk} \qquad (29)$$

Let (i^*, j^*, k^*) be such that $\pi^* = b_{i^*j^*k^*}$. Then, $(x^*, y^*, \varepsilon^*) = (s^*, t^*, d_{i^*j^*k^*})$ is an optimal solution of the MILP problem in Figure 1. To that end, observe that x^*, y^* satisfy constraints (19)–(21), (23) and (24). Moreover, since b_{ijk} was defined as the index of d_{ijk}, Eqs. (29) and (27) imply the optimality of $(x^*, y^*, \varepsilon^*)$; namely, the min-max of the d_{ijk}'s is equivalent to the max-min of their rank. ∎

Figure 2. Equivalent formulation of the MILP of Figure 1

$$
\begin{aligned}
\min \quad & \pi && (30)\\
\text{s. t.} \quad & \sum_{k=1}^{M} s_k = K, && (31)\\
& \sum_{k=1}^{M} t_{ijk} = 1, \quad \forall i,j = 1,\ldots,N, i<j, && (32)\\
& t_{ijk} \le s_k, \quad \forall i<j,\ k = 1,\ldots,M, && (33)\\
& \pi \ge \sum_{k=1}^{M} b_{ijk} t_{ijk}, \quad \forall i,j, i<j, && (34)\\
& t_{ijk} \ge 0, \quad \forall i,j,k, i<j, && (35)\\
& s_k \in \{0,1\}, \quad k = 1,\ldots,M && (36)
\end{aligned}
$$

Figure 3. The feasibility problem. c_{ijk}'s are defined by Eq. (41)

$$\max \quad 0 \tag{37}$$

$$\text{s.t.} \quad \sum_{k=1}^{M} c_{ijk} w_k \geq 1, \ \forall i,j = 1,\ldots,N, i < j, \tag{38}$$

$$\sum_{k=1}^{M} w_k = K, \tag{39}$$

$$w_k \in \{0,1\}, \quad \forall k = 1,\ldots,M, \tag{40}$$

We remark that it is also true that there is a corresponding optimal solution to the problem of Figure 2 for every optimal solution to the problem of Figure 1.

Iterative Algorithm

Proposition V.2 allows us to solve the problem of Figure 2 instead of the problem of Figure 1. So we will concentrate on the former. Our approach is to solve this problem by an iterative *feasibility* algorithm along the lines proposed in (Daskin 1995). In particular, we use a slightly modified version of a *two-phase* algorithm proposed in (IIhan et al. 2006, Ozsoy et al. 2005).

The core idea of the iterative algorithm is to solve the feasibility problem shown in Figure 3. The problem of Figure 3 depends on a parameter θ by the following equation:

$$c_{ijk} = \begin{cases} 1, & \text{if } b_{ijk} \leq \theta, \\ 0, & \text{otherwise.} \end{cases} \tag{41}$$

Intuitively, θ represents the index of some d_{ijk} distance in the nonincreasingly sorted list that we initially created, and the feasibility problem checks whether all pairs of locations can be distinguished (by at least one clusterhead) with an error exponent greater than or equal to the d_{ijk} distance pointed by θ. If not, θ is increased, which means that it now points to a smaller d_{ijk} distance, and the process is repeated. At termination, θ, which corresponds to the largest feasible d_{ijk} distance, provides the optimal value of the problem in Figure 2. The formal iterative algorithm is shown in Figure 4. It is clear that this algorithm terminates in a finite number of steps. In particular, if $\theta = MN(N-1)/2$, we see that $c_{ijk} = 1$ for all i, j, k, and the feasibility conditions of problem of Figure 3 are trivially satisfied. Next we show that at termination, we obtain the optimal solution to the problem of Figure 2.

Proposition V.3 *Let θ^* be the value of θ when the algorithm of Figure 4 terminates and w^* the optimal solution to problem of Figure 3 at the last iteration. Then $s^* = w^*$ induces an optimal solution to the problem of Figure 2 with optimal objective function value $\pi^* = \theta^*$.*

Proof: First note that at the last iteration, for any i, j ($i < j$), there exists at least one k such that $b_{ijk} = \theta^*$ with $w^*_k = 1$; otherwise the problem is infeasible. Next we construct a feasible solution (s^*, t^*, π^*) to the problem of Figure 2 as follows. Given a pair (i, j), we select one k such that $b_{ijk} = \theta^*$ with $w^*_k = 1$. Then we set $t^*_{ijk} = 1$ and $t^*_{ijl} = 0$ for any $l \neq k$. We repeat this process for all pairs (i, j). Finally, we set s^* =

Figure 4. Iterative feasibility algorithm

1) Set $\theta = 1$.
2) Determine c_{ijk}, $\forall i, j, k$ and solve the problem shown in Fig. 3.
    a) If the problem of Fig. 3 is infeasible, set $\theta := \theta + 1$ and go to step 2.
    b) Else, if the problem is feasible, stop.

w^* and $\pi^* = \theta^*$. Then, the triplet (s^*, t^*, π^*) satisfies all the constraints of the problem of Figure 2 and is therefore a feasible solution.

Next we prove the optimality of (s^*, t^*, π^*) by contradiction. Suppose that there exists a feasible solution $(\tilde{s}, \tilde{t}, \tilde{\pi})$ to the problem of Figure 2 such that $\tilde{\pi} < \pi^*$. Then according to the algorithm in Figure 4, there is a step where $\theta = \tilde{\pi}$. This implies that the corresponding problem of Figure 3 was infeasible; otherwise the algorithm would not have increased the value of θ beyond $\tilde{\pi}$. However, since $(\tilde{s}, \tilde{t}, \tilde{\pi})$ is feasible for the problem of Figure 2 we have

$$\tilde{\pi} \geq \sum_{k|\tilde{s}_k=1} b_{ijk}\tilde{t}_{ijk} \quad \forall i \ \ j, i < j,$$

which implies that for all (i, j) with $i < j$ there exists at least one k with $\tilde{s}_k = 1$ such that $b_{ijk} \leq \tilde{\pi}$. Hence, $w = \tilde{s}$ is feasible for the problem of Figure 3 when $\theta = \tilde{\pi}$. We arrived at a contradiction. $\blacksquare$

To expedite the convergence in the actual implementation, we use the two-phase algorithm shown in Figure 5. In the first part (*LP phase*), we construct the linear programming (LP) relaxation of the problem of Figure 3 by replacing the binary constraint (40) by

$$w_k \in [0,1], \quad \forall k = 1, \ldots, M \tag{42}$$

Then we solve the relaxed problem and compute the smallest integer $\theta = \theta_0$ such that the LP relaxation is feasible. In the second part (*IP phase*), we compute θ^* by executing the iterative algorithm starting from $\theta = \theta_0$, but this time we solve the integer programming problem instead of the LP relaxation. The important difference between our algorithm and the algorithm proposed in (IIhan et al. 2006, Ozsoy et al. 2005) is that we employ binary search both in the LP-phase and IP-phase whereas the authors of (IIhan et al. 2006, Ozsoy et al. 2005) use linear search in the LP-phase. Our use of binary search in the IP-phase further decreases the computation time for large problem instances almost by an order of magnitude.

Performance Guarantee

We will use the decision rule outlined in the beginning of this section and for every region pair (i, j) we will rely on the best positioned clusterhead $B_{k^*_{ij}}$ to make the corresponding decision. The following theorem establishes a performance guarantee.

Proposition V.4 *Let x^*, y^* be an optimal solution of the MILP in Figure 1 with corresponding optimal value ε^*. Place clusterheads according $Y^* = \{B_k \mid x^*_k = 1\}$ and for every (i, j) select one clusterhead with*

Figure 5. Two-phase iterative feasibility algorithm

1) (LP phase) Set $l := 1$, $u := \frac{MN(N-1)}{2}$.
2) Let $\theta := \lceil \frac{l+u}{2} \rceil$, determine c_{ijk} $\forall i, j, k$ and solve the relaxed version of the problem of Fig. 3.
   a) If the relaxed version of the problem of Fig. 3 is feasible, set $u := \theta$.
   b) Else, set $l := \theta$.
3) If $u - l \leq 1$, go to step 4. Else go to step 2.
4) (IP phase) If the relaxed version of the problem of Fig. 3 is feasible, set $\theta := l$, else set $\theta := u$. Then set $l := \theta, u := \frac{MN(N-1)}{2}$.
5) Determine c_{ijk}, $\forall i, j, k$ and solve the problem of Fig. 3.
   a) If the problem of Fig. 3 is infeasible, set $l := \theta, \theta := \lceil \frac{l+u}{2} \rceil$ and go to step 5.
   b) Else, if the problem is feasible
      i) If $\theta \neq u$, set $u := \theta, \theta := \lceil \frac{l+u}{2} \rceil$ and go to step 5.
      ii) Otherwise, stop.

index k_{ij}^* so that $y_{ijk_{ij}^*}^* = 1$. Then, the worst case probability of error for the decision rule described in the beginning of this section, $P_n^{(e),opt}$, satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n^{(e),opt} \leq -\varepsilon^* . \tag{43}$$

Proof: Recall the results of Theorem. IV.1 and Propositions IV.6 and IV.7 for the case where d_{ijk} is defined either by (4), (15), or by (17), respectively. Define $(\bar{i}, \bar{j})$ as in the previous section. The clusterhead with index k_{ij}^* will use the GLRT which compares $X_{\bar{i}\,\bar{j}k_{ij}^*}(y^{(k_{ij}^*),n})$ 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 P_{θ_i}. For all $\delta_n > 0$ and large enough n we have

$$P_{\theta_i}[\text{error}] \leq P_{\theta_i}[\bigcup_{j\neq i} E_n(i, j)] \leq \sum_{j\neq i} e^{-n(d_{ijk_{ij}^*}+\delta_n)} \leq (N-1)e^{-n(\varepsilon^*+\delta_n)} .$$

The 2nd inequality above is due to Thm. IV.1 or Props. IV.6, or IV.7 and the last inequality above is due to (27). Since the bound above holds for all I we obtain (43). $\blacksquare$

LOCALIZATION DECISIONS

In this section we consider the implementation of the decision rule described in the previous section. We assume that the WSN 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 WSNs 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.

Centralized Approach

We first describe a naive, centralized, approach. Every clusterhead observes $y^{(k),n} = (y_1^{(k)}, \dots, y_n^{(k)})$ 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 $y_l^{(k)}$, 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. Once this information is received, the gateway can apply the decision rule discussed in the previous section to identify the region at which the sensor in question resides.

Distributed Approach

In this subsection 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 $B_{k^*_{1,2}}$ based on the observations $y^{(k^*_{12}),n}$ uses the GLRT to make the decision; let L_{l_1} be the hypothesis accepted. The clusterhead at $B_{k^*_{1,2}}$ 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 $i = 1, \dots, N-1$, where we set $l_0 = 1$. At the i-th 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 $i = 1, \dots, N-1$ 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 resulting 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 1 compares the centralized and distributed approaches. In the distributed case we report both the best and worst case in terms of the communication cost based on the discussion in this subsection. Some observations are in order. The total processing cost is the same 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. It follows that $O(N \log N)$ is much preferable to $O(S_1 nK^2)$.

Note that both the centralized and the distributed approach guarantee the performance of the system obtained in Prop. V.4, i.e., the savings from the distributed approach come with no performance loss.

Table 1. Comparing the centralized and distributed approaches

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

TESTBED AND EXPERIMENTAL RESULTS

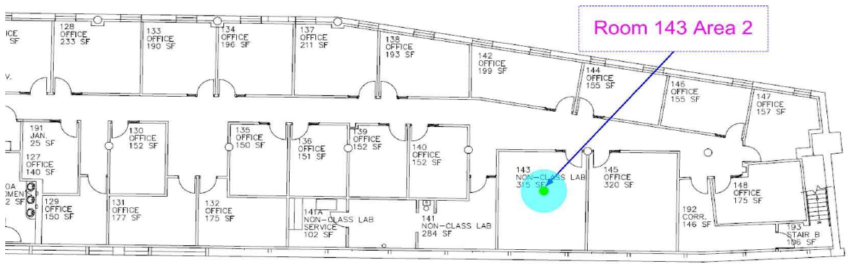
Next, we provide experimental results from a localization testbed we have installed at Boston University (BU) see Figure 6. We have appropriately named our system the *Boston University Statistical Localization System (BLoc)* (http://pythagoras.bu.edu/bloc/index.html). The testbed has a web interface through which one can poll a specific WSN node (identified by an ID). The system responds with a building floor map like the one shown in Figure 7 highlighting the room number and the region where the node was found.

The testbed uses MICAz motes manufactured by Crossbow Inc. We covered 16 rooms and corridors and defined 60 regions. Within each region we placed a mote on some furniture or on the wall. These 60 positions make up the set 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 Figure 8. Phase 1 can be carried out automatically by scheduling the motes so that when one is broadcasting the others are listening. For Phase 2 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. The pdfs

Figure 6. Floor plan for the bestbed

Figure 7. A sample response from BLoc



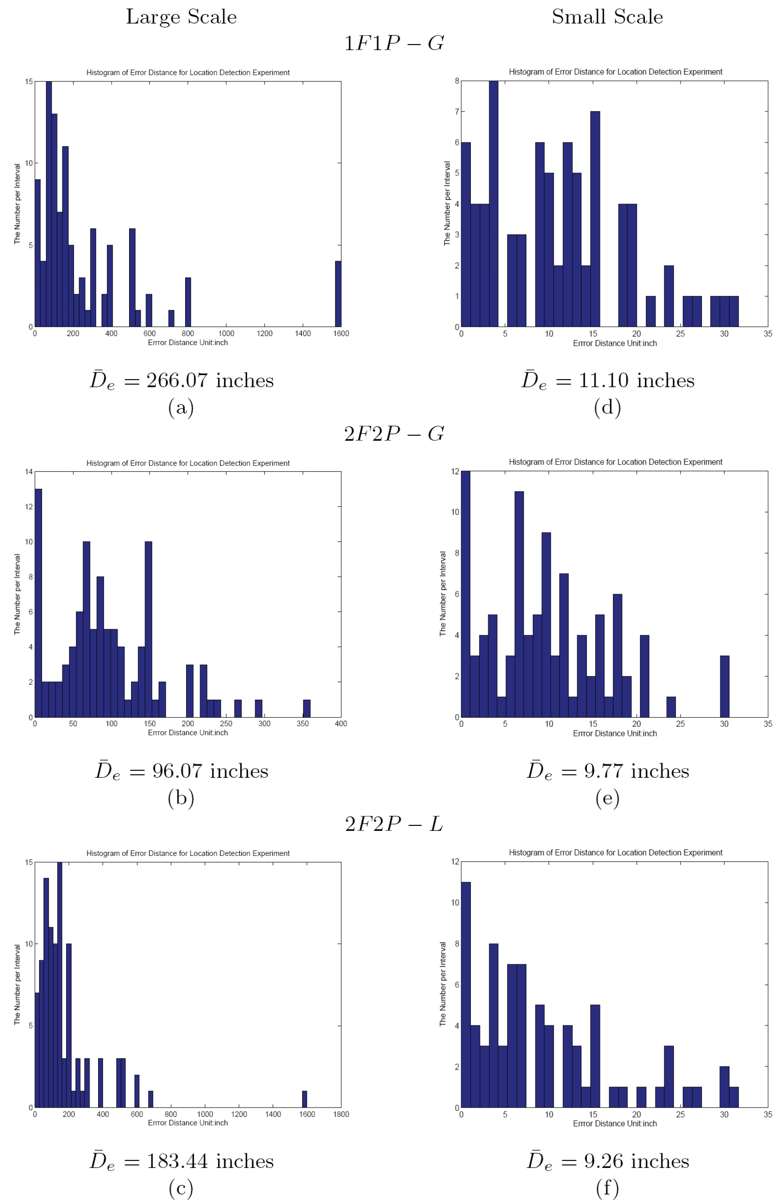
were estimated by histograms, and for each pdf we define an interval $[m_{jk} - \hat{m}_{jk}, m_{jk} + m_{jk}]$ and select points $\theta_{j,1},\ldots, \theta_{j,R}$ in this interval. We construct the family $\{P_{Y^{(k)}|\theta_j}(y), \theta_j \in \Omega_j\}$ so that the l-th member has the same shape as $P_{Y^{(k)}|B_j}(y)$ 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 $[m_{jk} - \hat{m}_{jk}, m_{jk} + m_{jk}]$ is maximized and there is no overlap. In the optimal placement obtained in Phase 4 we used 12 clusterheads to achieve a small enough probability of error and have some built-in redundancy in the clusterhead network.

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 want to locate transmits packets at a single frequency and a single power level and the system uses the GLRT (we write 1F1P − G to indicate Ver. 1 in Figure 9) to determine the region where the mote resides based on RSSI observations at the clusterheads. In Version 2 (denoted by 2F2P −G) RSSI observations are made for packets transmitted in two different frequencies and two different power levels and the GLRT is again used. Version 3 (denoted by 2F2P − L) is 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 Figure 9(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 is almost never the one at which RSSI measurements have been made in Phase 1.

Figure 8. Phases of the experimental validation

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} the corresponding mean.
2) For each (B_k, B_j) construct a pdf or a pdf family $\{p_{\mathbf{Y}^{(k)}|\theta_j}(\mathbf{y}), \theta_j \in \Omega_j\}$ to characterize transmissions from positions within L_j.
3) Compute the exponent d_{ijk} as described in Sec. IV.
4) Determine the clusterhead placement by the algorithm in Sec. V-B.
5) Determine the location of any mote in the coverage area by the decision rule of Sec. V-A.

Figure 9. Results for various versions of the system



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 $\overline{D}_e$= 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 in cost savings compared to proximity-based systems that need a higher density of observers.

For comparison purposes, we also used the same testbed and the exactly same tests with the stochastic trilateration 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 $Y(k) \mid B_j$, is a random variable with a Gaussian distribution. The mean of RSSI satisfies the path loss formula $\overline{Y}^{(k)} \mid B_j = Y_0 - 10 n_p \log_{10}(\zeta_{kj} / \zeta_0)$, 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 exactly 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 exactly same procedure as before. The results of this "small scale" localization experiment are in Figure 9(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 region is small enough. 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.

CONCLUSION

We have presented a unified robust and distributed approach for locating the area (region) where sensors of a WSN reside. We posed the problem of localization as a multiple hypothesis testing problem and proposed a combined LRT- or GLRT-based decision rule depending on the appropriate probabilistic characterization of a region.

We developed asymptotic results on the type I and type II error exponents which are critical in posing and solving the problem of optimally placing a given number of clusterheads to minimize the probability of error. We devised a mixed integer linear programming formulation to determine the optimal clusterhead placement, and a fast algorithm for solving it. We evaluated the scalability of the proposed MILP as well as the quality of the resultant placement. Our implementation of the proposed fast algorithm shows that the proposed MILP is capable of solving realistic problems within reasonable time-frames. Furthermore, we proposed a distributed approach to implement our localization algorithm 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 a combined LRT- and GLRT-based system provides significant robustness (and improved performance) compared to a simpler LRT-based system.

Furthermore, our approach leads to significantly improved accuracy compared to a stochastic trilateration technique like the one in (Patwari et al. 2003). We showed that we can achieve an accuracy on the same order of magnitude as the mean radius, which is the best possible accuracy one can achieve with a discrete system. As a result, smaller regions (and more clusterheads) lead to better accuracy but at the expense of more initial measurements (training) and a 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.

ACKNOWLEDGMENT

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

REFERENCES

Bahl, P., & Padmanabhan, V. (2000, March). RADAR: An in-building RF-based user location and tracking system. *In Proceedings of IEEE INFOCOM, 2,* 775-784. Tel-Aviv, Israel.

Battiti, R., Brunato, M., & Villani, A. (2003, January). *Statistical learning theory for location fingerprinting in wireless LANs*, University of Trento, Department of Information and Communication Technology, Trento, Italy, DIT 02-086.

Castro, P., Chiu, P., Kremenek, T., & Muntz, R. (2001, September). A Probabilistic Room Location Service for Wireless Networked Environments. *Proceedings of the 3rd international conference on Ubiquitous Computing* (pp.18-34), Atlanta, Georgia, USA

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist., *23*, 493-507.

Daskin, M. (1995). *Network and Discrete Location*. New York: Wiley.

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

Hightower, J., Want, R., & Borriello, G. (2000, February). *SpotON: An indoor 3d location sensing technology based on RF signal strength*. University of Washington, Department of Computer Science and Engineering, Seattle, WA, UW CSE 00-02-02.

Hodes, T. D., Katx, R. H., Schreiber, E. S., & Rowe, L. (1997, September). Composable ad hoc mobile services for universal interaction. *Proceedings of the 3rd annual ACM/IEEE international conference on Mobile computing and networking* (pp. 1-12) Budapest, Hungary.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist., 36*, 369.401.

Hofmann-Wellenhof, B., Lichtenegger, H., & Collins, J. (1997). *Global Positioning System: Theory and Practice*. 4th ed. Springer-Verlag.

Ilhan, T., & Pinar, M. (2001). *An efficient exact algorithm for the vertex p-center problem*. http://www.optimization-online.org/ DB-HTML/2001/09/376.html

ILOG CPLEX 8.0, ILOG, Inc., Mountain View, California, July 2002, http://www.ilog.com.

Klingbeil, L., & Wark, T. (2008). A wireless sensor netwok for real-time indoor localisation and motion monitoring. *Proceedings of 2008 International Conference on Information Processing in Sensor Networks* (pp. 39-50).

Meissner, A., Luckenbach, T., Risse, T., Kirste, T., & Kirchner, H. (2002). Design challenges for an integrated disaster management communication and information system. *1st IEEE Workshop on Disaster Recovery Networks*. New York, USA

Ozsoy, F. A., & Pinar, M. C. (2004, November). An exact algorithm for the capacitated vertex p-center problem. *Computers and Operations Research, 33*(5), 1420-1436.

Paschalids, I. C., & Guo, D. (2007, December). Robust and distributed localization in sensor networks. *Proceedings of 46th IEEE Conference on Decision and Control*, (pp. 933-938), New Orleans, Louisiana

Patwari, N., Hero, A. O., Perkins, M., Correal, N. S., & O'Dea,R. J. (2003). Relative location estimation in wireless sensor networks. *IEEE Transactions on signal processing, 51*(8), 2137-2148.

Patwari, N., & Agrawal, P. (2008). Effects of correlated shadowing: connectivity, localization, and RF Tomography. *In 2008 International Conference on IPSN* (pp. 82-93).

Prasithsangaree, P., Krishnamurthy, P., & Chrysanthis, P. K. (September 2002). On indoor position location with wireless LANs. *Proceedings of 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications* (pp. 720-724).

Priyantha, N. B., Chakraborty, A., & Balakrishnan, H. (2000). The cricket location-support system. *In Mobile Computing and Networking* (pp. 32-43).

Ray, S., Lai, W., & Paschalidis, I. C. (2006). Statistical location detection with sensor networks. *IEEE Transactions on Information Theory, Joint special issue with IEEE/ACM Transactions on Networking on Networking and Information Theory, 52*(6), 2670-2683.

Wu, Y., Hu, J. B., & Chen, Z. (2007). Radio map filter for sensor network indoor localization systems. *5th IEEE International Conference on INdustrial Informatics* (pp. 63-68).

Youssef, M. (2008). *Collection about Location Determination Papers available online* http://www.cs.umd.edu/~moustafa/location_papers.htm

Zeitouni, O., Ziv, J., & Merhav, N. (1992 September). When is the generalized likelihood ratio test optimal. *IEEE Transactions on Information Theory, 38*(2), 1597-1602.

APPENDIX A: UPPER BOUND FOR BINARY CASE

Theorem A.1 $P_n^{(e)} \le e^{-nd_{ijk}}$ *for all n.*

Proof: From our assumption that $y_1^{(k)},\ldots,y_n^{(k)}$ are i.i.d, we have

$$X_{ijk}(y_1^{(k)},\ldots,y_n^{(k)}) = \sum_{i=1}^{n} X_{ijk}(y_i^{(k)}) \tag{44}$$

Since ML rule rejects L_j if the likelihood ratio is greater than 0, for any n,

$$\begin{aligned}
\beta_n &= P_{L_j}[X_{ijk}(y_1^{(k)},\ldots,y_n^{(k)}) > 0] \\
&= P_{L_j}[\lambda \cdot X_{ijk}(y_1^{(k)},\ldots,y_n^{(k)}) > \lambda \cdot 0] \\
&= P_{L_j}[e^{\lambda \sum_{i=1}^{n} X_{ijk}(y_i^{(k)})} > e^{\lambda \cdot 0}] \\
&\le E_{L_j}[e^{\lambda \sum_{i=1}^{n} X_{ijk}(y_i^{(k)})}] \\
&= \left[E_{L_j}[e^{\lambda X_{ijk}(y_i^{(k)})}] \right]^n \\
&= e^{n \log\left(E_{L_j}[e^{\lambda X_{ijk}(y_i^{(k)})}] \right)}
\end{aligned}$$

Optimizing with respect to λ gives $\beta_n \le e^{-n \cdot d_{ijk}}$. Noting that d_{ijk} is symmetric with respect to the conditional distributions, $\alpha_n \le e^{-n \cdot d_{ijk}}$, and thus $P_n^{(e)} \le e^{-n \cdot d_{ijk}}$. ■

APPENDIX B: PROOF OF LEMMA IV.3

Proof: For all $\theta_j \in \Omega_j$ we have

$$\begin{aligned}
\alpha_{ijk,n}^{S^*}(\theta_j) &= P_{\theta_j}[y^{(k),n} \in S_{ijk,n}^*] \\
&= \sum_{\{L_{y^{(k)},n} | T_n(L_{y^{(k)},n} \subseteq S_{ijk,n}^*)\}} |T_n(L_{y^{(k)},n})| \, P_{Y^{(k)}|\theta_j}(y^{(k),n}) \\
&\le \sum_{\{L_{y^{(k)},n} | T_n(L_{y^{(k)},n} \subseteq S_{ijk,n}^*)\}} e^{nH(L_{y^{(k)},n})} e^{-n[H(L_{y^{(k)},n}) + D(L_{y^{(k)},n} \| P_{\theta_j})]} \\
&= \sum_{\{L_{y^{(k)},n} | T_n(L_{y^{(k)},n} \subseteq S_{ijk,n}^*)\}} e^{-nD(L_{y^{(k)},n} \| P_{\theta_j})} \\
&\le (n+1)^{|\Sigma|} e^{-n\lambda},
\end{aligned}$$

which establishes (6). For the first inequality above note that the size of the type class of $L_{y^{(k)},n}$ is upper bounded by $e^{nH(L_{y^{(k)},n})}$ 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 $S_{ijk,n}^*$ and the fact that the set of all possible types, L_n, has cardinality upper bounded by $(n+1)^{|\Sigma|}$ (Dembo and Zeitouni 1998, Chap. 2).

Let now $S_{ijk,n}$ be some other decision rule satisfying constraint (6), hence, for all $\varepsilon > 0$ and all large enough n

$$\alpha^{S}_{ijk,n}(\theta_j) \le e^{-n(\lambda+\varepsilon)} \tag{45}$$

Meanwhile for all $\varepsilon > 0$, all large enough n, and any $y^{(k),n} \in S_{ijk,n}$

$$\begin{aligned}
\alpha^{S}_{ijk,n}(\theta_j) &= \sum_{\{L_{y^{(k)},n} | T_n(L_{y^{(k)},n} \subseteq S_{ijk,n})\}} |T_n(L_{y^{(k)},n})| \, P_{Y^{(k)}|\theta_j}(y^{(k),n}) \\
&\ge \sum_{\{L_{y^{(k)},n} | T_n(L_{y^{(k)},n} \subseteq S_{ijk,n})\}} (n+1)^{-|\Sigma|} e^{-nD(L_{y^{(k)},n} \| P_{\theta_j})} \\
&\ge e^{-n[D(L_{y^{(k)},n} \| P_{\theta_j})+\varepsilon]},
\end{aligned}$$

where the first inequality above uses (Dembo and Zeitouni 1998, Lemma 2.1.8). Comparing the above with (45) it follows that if $y^{(k),n} \in S_{ijk,n}$ then for all θ_j $D(L_{y^{(k)},n} \| P_{\theta_j}) \ge \lambda$, hence, $y^{(k),n} \in S^{*}_{ijk,n}$ and $S_{ijk,n} \subset S^{*}_{ijk,n}$. Consequently, for all θ_i $\beta^{S}_{ijk,n}(\theta_i) \ge \beta^{S^*}_{ijk,n}(\theta_i)$ which establishes that the generalized Hoeffding test maximizes the exponent of the type II error probability. We conclude that it satisfies the GNP criterion. ∎