Optimal RSS Threshold Selection in Connectivity-Based Localization Schemes

Gianni Giorgetti
Electrical Engineering Dept.
Arizona State University, US
gianni.giorgetti@asu.edu

Sandeep K. S. Gupta
School of Comp. & Informatics
Arizona State University, US
sandeep.gupta@asu.edu

Gianfranco Manes
Dip. Elettronica e Telecom.
Univ. degli Studi di Firenze, IT
gianfranco.manes@unifi.it

ABSTRACT

Connectivity-based localization schemes compute the node positions using proximity information collected within the network. In many cases of practical interest, Received Signal Strength (RSS) measurements are available, and connectivity data can be obtained by comparing the RSS against a threshold. We use the Cramér-Rao bound (CRB) analysis to determine the threshold value that minimizes the localization error. The CRB is based on knowledge of the propagation model's parameters and the true node positions. Since this information is not available to a localization scheme, we approximate the optimal threshold value using a function that depends only on the number of nodes in the network. We use extensive simulations and RSS data from in-field experiments to validate the results of the proposed approach.

Categories and Subject Descriptors

C.2.2 [**Computer Communication Networks**]: Network Protocols.

General Terms

Algorithms, Performance, Theory.

Keywords

Localization, Connectivity, Signal Strength, Optimal, Threshold, Cramér-Rao Bound, Approximation.

1. INTRODUCTION

Localization is an active research area devoted to support *location awareness* in applications where the use of GPS is not cost effective (e.g. sensor networks) or technically feasible (e.g. indoor applications). Existing localization solutions can be broadly divided into *range-based* and *range-free* schemes depending on whether they compute the node positions using distances and angles estimates or proximity information such as radio connectivity. Range-based approaches are capable of accurate results, but they often rely on additional hardware (e.g. ultrasound transceivers or antenna arrays), which make them more expensive and less suitable for ad-hoc deployments. On the other hand, range-free schemes offer a coarser resolution, but are cheaper and easier to deploy.

One of the most popular range-free techniques is localization based on radio connectivity. The principle underlying this approach is simple: since each node has a limited communication range, the successful transmission of a radio packet from node A to node B implies that the two nodes are close in space. The use of connectivity information is appealing for several reasons. First, since nodes already exchange data using radio messages, connectivity information is easy to acquire or it might be already available; in fact, many contention-free MAC protocols and routing algorithms also require this information. Second, connectivity between a pair of nodes is a binary value (1 if nodes are connected, 0 otherwise); therefore this information can be efficiently communicated across the network with minimal impact on the energy budget of sensor nodes. Third, several localization schemes are available to process connectivity data on hardware with limited memory and computational resources (e.g. [2, 5]).

Another merit of connectivity-based localization schemes is that they are easy to simulate. Using the *idealized radio model*[1] widely adopted in previous research work, connectivity between nodes can be simulated regardless of the complex phenomena that regulate RF propagation. This model provides an abstraction useful in simulation studies; however, it does not define a criterion to obtain connectivity data in real world applications. In other words, system designers implementing a connectivity-based scheme will have to define their own rule to establish which nodes are to be considered neighbors. In the Centroid scheme [2], for example, nodes are regarded as neighbors if at least 80% of the message transmitted are successfully received. Unfortunately, simple rules like this will not always produce satisfactory results, especially for densely deployed networks.

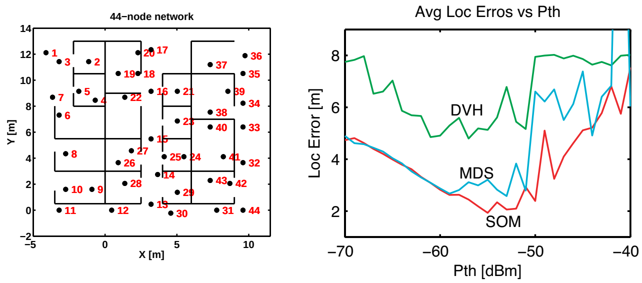
1.1 Motivating Example

Consider the case where one wants to localize the nodes in Figure 1.a. The data[2] for this network has been collected by measuring the average *Received Signal Strength* (RSS) between pairs of nodes in the cubicles of an office space [12]. Note that every node of this network is in the radio range of every other node, and no packet loss was reported (i.e. the

[1]Two nodes are connected if their distance is less than a fixed radius.

[2]`http://www.eecs.umich.edu/~hero/localize/`

MSWiM'08, October 27–31, 2008, Vancouver, BC, Canada.



(a) (b)

Figure 1: Localization errors for the 44-node network described in [12]. Nodes 3,10,35 and 44 are used as anchors in the localization process.

network is fully connected). Any localization scheme that uses proximity information will generate a large error. In fact, since all the nodes have the same neighbors, eventually, they will be associated to similar positions.

We can "artificially" reduce the connectivity by setting a threshold and considering neighbors only those pairs of nodes whose average RSS exceeds the threshold. It is not clear, however, how such a threshold should be set: a value that is too low might be ineffective in reducing the connectivity, while a value that is too high might cause the network to become disconnected and, again, result in large localization error.

What is the correct threshold value? Figure 1b provides an empirical answer to this question by reporting the average localization error of three range-free algorithms for different values of the RSS threshold. The schemes used to localize the nodes in Figure 1a are: DV-HOP [10], Multidimensional Scaling (MDS) [15] and localization using Self-Organizing Maps (SOM) [5]. The plots show that a proper threshold should be chosen between -60 dBm and -50 dBm; in fact, in this range all three algorithms produce a low error.

Although *a-posteriori* analysis of the localization error clearly shows the existence of threshold values that are more effective than others, computing the error requires knowledge of the true node coordinates. In real-world applications, an effective threshold value will have to be found using a different approach.

1.2 Outline and Contributions

The aim of our work is to compute the optimal threshold (i.e. the threshold that minimizes the localization error) when connectivity is derived from RSS measurements. Our analysis is based on the model proposed by Patwari and Hero III [13], who have derived an expression for the Cramér–Rao bound (CRB) when connectivity is obtained from the RSS data described by the log-normal shadowing model. After reviewing this model in Section 2, we introduce the Fisher information and the CRB by analyzing a localization case with a single node in one dimension (Section 3). This case provides an intuitive example of how to use the CRB analysis to determine the optimal RSS threshold for a generic connectivity-based localization scheme. In Section 3.3, the analysis is extended to localization of networks deployed in 2D and 3D spaces. Simulation examples and analysis of the localization error are used to validate the threshold choice based on the CRB analysis.

Notably, the CRB analysis provides a result that is independent of the scheme used. However, computing the CRB requires knowledge of the true node positions, therefore this approach does not yield a solution of practical utilization. To solve this conundrum, in Section 4 we analyze some properties of the Fisher information matrix and we show that the optimal threshold corresponds to a connectivity value that is independent from the node density and the parameters of the shadowing model. Therefore, *computing the optimal threshold can be cast as the problem of finding the optimal connectivity value for the network to localize.*

The main contribution of our work is an approximate formula to compute the optimal connectivity value as a function of the network size (see Section 4.4). Using extensive simulations, we show that this approximation, which does not use knowledge of the node positions or the parameters of the propagation model, can accurately approximate the results obtained using the CRB analysis. The proposed result has two main applications:

1. At design time, the optimal connectivity value produced by the formula can be used to plan for network deployments suitable for localization using range-free schemes.
2. At run-time, if the network is densely deployed and the connectivity is too high, the localization error can be reduced by setting a threshold on the RSS values. In Section 5.1, we demonstrate this application using data two from real-world deployments with nodes densely placed in a 2D and a 3D space.

Results in this paper indicate what network connectivity should be used to ensure a low localization error. We note that the results are independent of the localization scheme used. In addition, using a similar analysis, we have also investigated how to reduce the localization error by choosing between range-based and connectivity-based approaches when the measurements are derived from RSS data [6].

2. THE PROBLEM

RF-based localization is a popular research topic, but the problem of how to convert RSS measurements into connectivity data has not been thoroughly investigated. The solutions proposed are mostly based on heuristic approaches. For example, the already mentioned centroid scheme [2] selects the neighbors based on the packet error rate and other authors have proposed a scheme where the neighbors are determined by sorting the RSS values [7].

We aim at putting the choice of the connectivity model on a more rigorous footing and define a criterion of general applicability. Our starting point is the work by Patwari and Hero III [13], where connectivity data is obtained by comparing the average RSS values against a threshold. This model assumes that any two nodes i and j exchange messages and collect a set (possibly empty) of RSS values: $\mathcal{Z}_{ij} = \left\{ z_{ij}^{(1)}, z_{ij}^{(2)}, z_{ij}^{(3)}, \ldots \right\}$.

2.1 The log-normal shadowing model

The RSS measurements collected by each node are affected by random variations caused by multipath fading due to reflection, diffraction and scattering of the RF signal in the surrounding environment. This variability can be partly

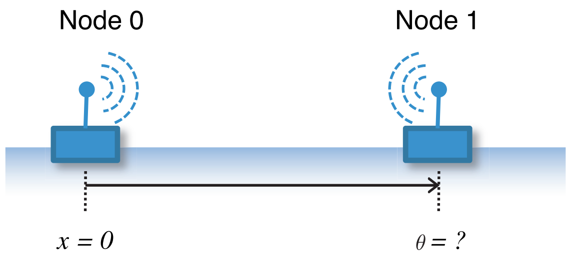


Figure 2: 1D localization: the distance of node 1 from the origin has to be estimated using connectivity information.

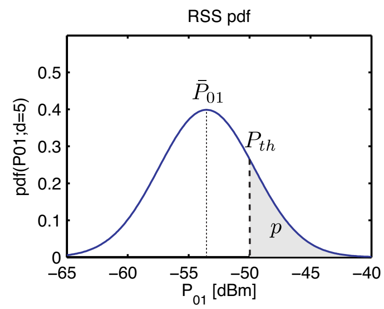


Figure 3: Probability density function of P_{01}.

reduced[3] by averaging the measured values. Let P_{ij} be the average of the RSS values measured in dB (or dBm) between nodes i and j. According to the *log-normal shadowing model* widely used for link budget analysis in wireless communication [14], P_{ij} is modeled as a random variable with normal distribution:

$$P_{ij} \sim \mathcal{N}(\bar{P}_{ij}, \sigma_{dB}) \quad (1)$$

$$\bar{P}_{ij} = P_0 + 10\, n_p \log_{10}\left(\frac{d_0}{d_{ij}}\right). \quad (2)$$

In the expression above, P_0 is the received power measured at a distance d_0, n_p is the *path loss* exponent and d_{ij} is the distance between nodes i and j.

2.2 Threshold-based connectivity

We determine that two nodes should be considered connected if P_{ij} is greater than a fixed threshold P_{th} and disconnected in the other case. If two nodes are too far to communicate, we account for the possibility of having empty sets $\mathcal{Z}_{ij}$ by computing P_{ij} as follows:

$$P_{ij} = \begin{cases} \frac{1}{|\mathcal{Z}_{ij}|} \sum_{z \in \mathcal{Z}_{ij}} z & \text{if } \mathcal{Z}_{ij} \neq \emptyset \\ -\infty & \text{if } \mathcal{Z}_{ij} = \emptyset. \end{cases} \quad (3)$$

The connectivity between two nodes is defined by a random variable $\mathbf{c_{ij}}$ which takes the following values:

$$\mathbf{c_{ij}} = \begin{cases} 0 & \text{if } P_{ij} < P_{th} \quad \text{(nodes disconnected)} \\ 1 & \text{if } P_{ij} \geq P_{th} \quad \text{(nodes connected)}. \end{cases} \quad (4)$$

Having based the connectivity model on the selection of a threshold P_{th}, the rest of the work will focus on how to reduce the localization error by properly choosing P_{th}.

3. CRAMÉR-RAO BOUND ANALYSIS

We address the problem of selecting the optimal threshold by first introducing the case where a single node has to be localized using connectivity information. This simple scenario allow us to intuitively explain the concepts that will be later used to solve the problem in the general case.

3.1 Single Node Localization

Suppose two devices placed along a line as in Figure 2. The position of node 1 (the unknown parameter θ) has to be estimated using connectivity information. The two nodes have collected RSS data by exchanging radio messages and P_{01} is the average of such values.

According to (2) and (4), the probability of the event $\mathbf{c_{01}} = 1$ (nodes connected) is the shadowed area in Figure 3 and can be computed analytically as:

$$p = \Pr\{\mathbf{c_{01}} = 1\} = 1 - G\left(\frac{P_{th} - \bar{P}_{01}}{\sigma_{dB}}\right), \quad (5)$$

where G is the CDF of a normal random variable $\mathcal{N}(0,1)$.

When the expected received power $\bar{P}_{01}$ equals P_{th}, the nodes are connected with probability $p = 0.5$. This condition occurs when the distance between the nodes equals the *threshold distance* d_{th}:

$$d_{th} = d_0 10^{\frac{P_0 - P_{th}}{10\, n_p}}. \quad (6)$$

Using the equation above, (5) can be rewritten to show the dependance of p on the true node distance θ and d_{th}:

$$p = p(\theta, d_{th}) = 1 - G\left[K_c \log\left(\frac{\theta}{d_{th}}\right)\right], \quad (7)$$

where the constant $K_c = (10\, n_p)/(\sigma_{dB} \log 10)$ depends on propagation model's parameters.

In the rest of the paper, the problem of selecting the optimal threshold will focus on computing the value P_{th} or d_{th} depending on the cases. Since the two values are related by (2) and (6), computing one value or the other is indifferent.

3.2 Fisher Information

To determine the d_{th} value that minimizes the estimation error for the nodes' distance, we consider the Fisher information associated with the random variable $\mathbf{c_{01}}$. The Fisher information is a measure of the amount of information that $\mathbf{c_{01}}$ carries about the parameter θ, and it defines what is the accuracy possible in estimating θ. Let $f()$ denote the *probability mass function* (pmf) of the random variable $\mathbf{c_{01}}$:

$$f(c_{01}) = f(c_{01}; \theta, d_{th}) = \begin{cases} 1 - p(\theta, d_{th}) & \text{if } c_{01} = 0 \\ p(\theta, d_{th}) & \text{if } c_{01} = 1. \end{cases} \quad (8)$$

The Fisher information is:

$$F(\theta, d_{th}) = E\left\{\left[\frac{\partial}{\partial\theta} \log f(c_{01}; \theta, d_{th})\right]^2\right\} =$$

$$\sum_{\mathbf{c_{01}} \in \{0,1\}} \left(\frac{\frac{\partial}{\partial\theta} f(c_{01}; \theta, d_{th})}{f(c_{01}; \theta, d_{th})}\right)^2 f(c_{01}; \theta, d_{th}). \quad (9)$$

[3]By averaging the RSS values, part of the signal variability due to time-dependent sources of multi-path (people or car moving, movements of foliage due to wind, etc) can be removed.

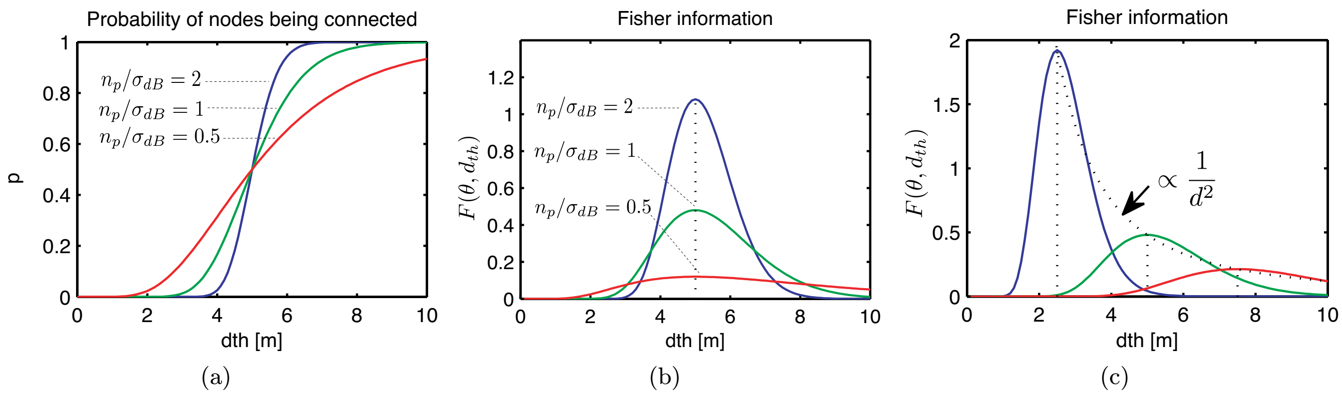


Figure 4: a) Probability $p(\theta, d_{th})$ for $\theta = 5\,\mathrm{m}$, b) Fisher Information for different ratios n_p/σ_{dB}, d) Fisher Information for nodes at distance $\theta = \{2.5, 5.0, 7.5\}$ m.

The Fisher information is a quality metric of the estimation procedure [4]. If T is an estimator for θ, then the variance of T is bounded by the inverse of F:

$$\mathrm{Var}\{T(\mathbf{c_{01}})\} \geq \frac{1}{F(\theta, d_{th})}. \quad (10)$$

The inequality above, known as Cramér–Rao bound (CRB), is the lower bound on the variance for *any* unbiased estimator[4] when using observation that are outcomes of the random variable $\mathbf{c_{01}}$. The CRB is not related to any particular estimation technique, but it only depends on the *measurement model* given by (8). Our goal is to find the value of d_{th} that maximizes the Fisher information or, equivalently, minimizes the CRB. To find the expression of F as a function of θ and d_{th}, we replace (8) in (9) and obtain:

$$F(\theta, d_{th}) = K\, I_R(\theta, d_{th}) \left(\frac{1}{\theta}\right)^2, \quad (11)$$

where $K = (2K_c^2/\pi)$ is a constant that depends on the parameters of the shadowing model and $I_R()$ is a term that depends on the ratio between θ and d_{th}:

$$I_R(\theta, d_{th}) = \frac{\exp\left[-K_c^2 \log(\theta/d_{th})^2\right]}{1 - \mathrm{erf}\left[\frac{K_c}{\sqrt{2}} \log(\theta/d_{th})\right]^2}. \quad (12)$$

Figure 4a shows the value of p as a function of the threshold distance d_{th}. The plots are computed for different values of the ratio n_p/σ_{dB}, and assuming a distance between the two nodes equal to 5 meters ($\theta = 5\,\mathrm{m}$). The information content of the measurements is proportional to the square of the ratio n_p/σ_{dB}. As shown in Figure 4a and 4b, increasing this ratio results in a sharper probability transition and larger values of the Fisher information. Intuitively, larger values of the parameter n_p imply a stronger correlation between the received power and the distance between the nodes, which is a condition that causes the estimation error to decrease. On the other hand, larger values of the parameter σ_{dB} pertains to environments where a strong shadowing noise reduces the accuracy of the distance estimates.

While the parameters of the shadowing models depend on the radio environment and are out of our control, the amount of information available can be maximized by properly choosing d_{th}. The plots in Figure 4b show that F always peaks when d_{th} equals θ, and then it rapidly decreases to zero as the difference between d_{th} and θ increases. In order to reduce the estimation error, we should set a threshold value as close as possible to the true node distance (which is unknown).

Threshold values with a large difference from θ will reduce the amount of information available and result in less accurate estimates. For example, if the nodes are five meters apart and we set a threshold distance that is too low (e.g. $d_{th} = 2\,m$), the two nodes will be disconnected with probability very close to one. The measurement carries little information about the true node distance because the nodes will almost always result as disconnected, no matter of what the actual value θ is. From a localization point of view, we can only infer that the distance between the nodes is greater than 2 m ($\theta > 2\,\mathrm{m}$).

A similar situation occurs if we select a threshold that is too large compared to the actual node distance (e.g. $d_{th} = 8\,\mathrm{m}$). The optimal choice is $d_{th} = 5\,\mathrm{m}$, which corresponds to the case where nodes are connected with probability $p = 0.5$. If we increase the distance between the two nodes, the optimal threshold is still achieved by setting $d_{th} = \theta$, but the information obtained from connectivity measurements decreases with the square of the distance between the two nodes (see Figure 4c). In other words, distance estimates for nearby nodes will be more accurate than distance estimates for nodes that are far from each other.

3.3 Collaborative Localization

We now study localization in networks with several nodes placed in 2D or 3D spaces. We refer to this scenario as *collaborative localization*. Even if a node is not in the radio range of any anchors, the proximity of other nodes (all placed at unknown locations) provide information to locate the node itself. This approach is also known as *multi-hop localization* because it supports localization of nodes placed several *hops* away from the anchors.

3.3.1 CRB Analysis

Consider a network with n nodes at unknown locations and m anchors. Similarly to the previous case, nodes collect RSS measurements and obtain connectivity values c_{ij} by comparing the average received power P_{ij} against a thresh-

[4]If $\hat{\theta}$ is an estimate of the unknown parameter θ obtained using T ($\hat{\theta} = T(C)$), then the estimator T is unbiased if $E\{\hat{\theta}\} = \theta$.

old P_{th}. Let $\mathcal{C}$ be the set of all the random variables associated with the measurements:

$$\mathcal{C} = \{\mathbf{c_{ij}} : \mathbf{c_{ij}} \in \{0, 1\},\ 1 \leq i, j \leq n + m\} . \tag{13}$$

The connectivity measurements are used to compute the n unknown node positions. The unknown coordinates can be arranged in a vector $\boldsymbol{\theta}$ with the following structure:

$$\boldsymbol{\theta} = \begin{cases} [\boldsymbol{\theta}_x, \boldsymbol{\theta}_y] & \text{if } d = 2 \\ [\boldsymbol{\theta}_x, \boldsymbol{\theta}_y, \boldsymbol{\theta}_z] & \text{if } d = 3, \end{cases} \tag{14}$$

where d is the dimensionality of the deployment space and the vectors $\boldsymbol{\theta}_x, \boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_z$ contains the unknown coordinates: $\boldsymbol{\theta}_x = [x_1, \ldots, x_n], \boldsymbol{\theta}_y = [y_1, \ldots, y_n]$ and $\boldsymbol{\theta}_z = [z_1, \ldots, z_n]$.

Again, we will use analysis of the Fisher information and the CRB to determine a threshold that minimize the estimation error for $\boldsymbol{\theta}$. In the case of collaborative localization, the measurement model is the *joint probability function*

$$f(\mathcal{C}; \boldsymbol{\theta}, d_{th}) = f(c_{11}, \ldots, c_{mm}; \boldsymbol{\theta}, d_{th}), \tag{15}$$

which relates the connectivity measurements to the node positions defined by $\boldsymbol{\theta}$ and the threshold distance d_{th}. If we assume that the RSS measurements between each pair of nodes are independent from all other pairs, then the joint probability (15) can be written as:

$$f(\mathcal{C}; \boldsymbol{\theta}, d_{th}) = \prod_{i,j=1}^{n+m} f(c_{ij}; \mathbf{v_i}, \mathbf{v_j}, d_{th}), \tag{16}$$

where $\mathbf{v_i}$ and $\mathbf{v_i}$ are the vectors with the coordinates of nodes i and j; $\mathbf{v_i} = [x_i, y_i]^t$ for $d = 2$, and $\mathbf{v_i} = [x_i, y_i, z_i]^t$ for $d = 3$. Each pmf in (16) is similar to (8). In particular, two nodes i and j are connected with probability

$$p_{ij} = \Pr\{\mathbf{c_{ij}} = 1\} = 1 - G\left[K_c \log\left(\frac{d_{ij}}{d_{th}}\right)\right], \tag{17}$$

where $d_{ij} = \sqrt{(\mathbf{v_i} - \mathbf{v_j})^t(\mathbf{v_i} - \mathbf{v_j})}$ is the Euclidean distance between the nodes. All the other symbols have the same meaning discussed in Section 3.1.

In the multi-parameter case, the information is measured by the *Fisher Information Matrix* (FIM) with elements

$$[F(\boldsymbol{\theta})]_{ij} = E\left\{\frac{\partial}{\partial\theta_i} \log f(\mathcal{C}; \boldsymbol{\theta}, d_{th})\ \frac{\partial}{\partial\theta_j} \log f(\mathcal{C}; \boldsymbol{\theta}, d_{th})\right\}. \tag{18}$$

The FIM has $(2n \times 2n)$ elements for nodes placed in 2D spaces, and $(3n \times 3n)$ elements when localization computes 3D coordinates. Given the structure of the parameter vector defined in (14), the FIM is partitioned in sub-matrices $\mathbf{F}_{xx}, \mathbf{F}_{xy}, \cdots, \mathbf{F}_{zz}$ with $n \times n$ elements each:

$$\mathbf{F} = \begin{cases} \begin{bmatrix} \mathbf{F}_{xx} & \mathbf{F}_{xy} \\ \mathbf{F}_{xy}^t & \mathbf{F}_{yy} \end{bmatrix} & \text{if } d = 2 \\ \begin{bmatrix} \mathbf{F}_{xx} & \mathbf{F}_{xy} & \mathbf{F}_{xz} \\ \mathbf{F}_{xy}^t & \mathbf{F}_{yy} & \mathbf{F}_{yz} \\ \mathbf{F}_{xy}^t & \mathbf{F}_{yz}^t & \mathbf{F}_{zz} \end{bmatrix} & \text{if } d = 3. \end{cases} \tag{19}$$

More details on how to compute the FIM for the 2D case are given by Patwari and Hero III [13]. For our analysis, it suffices to note that each sub matrix has elements similar to (11). For example, the elements of the sub-matrix $\mathbf{F_{xx}}$ are:

$$[f_{xx}]_{ij} = \begin{cases} -K \cdot I_R(d_{ij}, d_{th}) \frac{(x_i - x_j)^2}{d_{ij}^4} & (i \neq j) \\ K \cdot \sum_{k=1}^{n+m} I_R(d_{ik}, d_{th}) \frac{(x_i - x_k)^2}{d_{ik}^4} & (i = j) \end{cases} \tag{20}$$

The sub-matrices $\mathbf{F_{yy}}$ and $\mathbf{F_{xy}}$ have a similar structure, but the terms $(x_i - x_j)^2$ are replaced by $(y_i - y_j)^2$ in the case of $\mathbf{F_{yy}}$, and by $(x_i - x_j)(y_i - y_j)$ in the case of $\mathbf{F_{xy}}$. Similarly, the terms in the sub-matrices $\mathbf{F_{xz}}, \mathbf{F_{yz}}$ and $\mathbf{F_{zz}}$ are: $(x_i - x_j)(z_i - z_j)$, $(y_i - y_j)(z_i - z_i)$ and $(z_i - z_j)^2$ respectively.

Note that anchor information contributes to the diagonal terms of each submatrix. At least three anchors are needed for localization in 2D, while four non-collinear anchor nodes are necessary for localization in 3D. Failure to include sufficient anchor information will cause the FIM to be rank deficient [9]. In this case, analysis of the CRB is possible by using the Moore-Penrose pseudoinverse of the FIM [3]. In our analysis we assume that the FIM is always invertible.

The inverse of the FIM bounds the covariance matrix of any unbiased estimator for $\boldsymbol{\theta}$ that uses observation from the set of random variables $\mathcal{C}$:

$$\text{Cov}\{T(\mathcal{C})\} \geq \frac{1}{\mathbf{F}}, \tag{21}$$

The diagonal elements of $\mathbf{F}^{-1}$ are the lower bound for the variance on the node coordinates x_i, y_i and z_i: $\sigma_{xi}^2 = [\mathbf{F}^{-1}]_{i,i}$, $\sigma_{yi}^2 = [\mathbf{F}^{-1}]_{i+N,i+N}$, and $\sigma_{zi}^2 = [\mathbf{F}^{-1}]_{i+2N,i+2N}$. The variance on the position of each sensor location is obtained by summation of the variance of the single coordinates:

$$\sigma_i^2 = \begin{cases} \sigma_{ix}^2 + \sigma_{iy}^2 & \text{if } d = 2 \\ \sigma_{ix}^2 + \sigma_{iy}^2 + \sigma_{iz}^2 & \text{if } d = 3. \end{cases} \tag{22}$$

If the same topology has to be localized in different environments (different realization of the random variables P_{ij}'s), then the terms (22) are a lower bound for the RMS error on the position of each node. Assuming that $\hat{\mathbf{v}}_\mathbf{i}^{(1)}, \cdots, \hat{\mathbf{v}}_\mathbf{i}^{(k)}$ are K estimates for the position of node i, then:

$$\text{RMS}(i) = \sqrt{\frac{1}{K} \sum_{i=1}^{K} (\hat{\mathbf{v}}_\mathbf{i}^{(i)} - \mathbf{v_i})^t (\hat{\mathbf{v}}_\mathbf{i}^{(i)} - \mathbf{v_i})} \geq \sigma_i. \tag{23}$$

3.3.2 Measurement Model

In the analysis above we assumed that *every* node makes connectivity measurements with *every other* node in the network. In practice, it may happen that two nodes are too far from each other to exchange messages and collect RSS information. In absence of external interferences, this situation occurs when the RF signal reaches the recipient with a power that is below the transceiver's sensitivity P_s.

The sensitivity P_s can be regarded as an implicit threshold set by the hardware. Since the probability of receiving messages with RSS lower than P_s is low, the threshold selection problem is meaningful only for values $P_{th} > P_s$. When nodes are unable to exchange radio messages, their RSS is lower than P_s and, consequently, lower than P_{th}. It follows that even pairs of nodes that are out of their radio range produce valid connectivity measurements. According to the discussion in this section and the model in Section 2.2, these nodes are always associated with the event "node disconnected".

4. OPTIMAL THRESHOLD

The lower limit on the variance of the node positions can be found by computing the inverse of the FIM. Since the Fisher information depends on the choice of the threshold distance d_{th}, the values (22) will also depend on d_{th}.

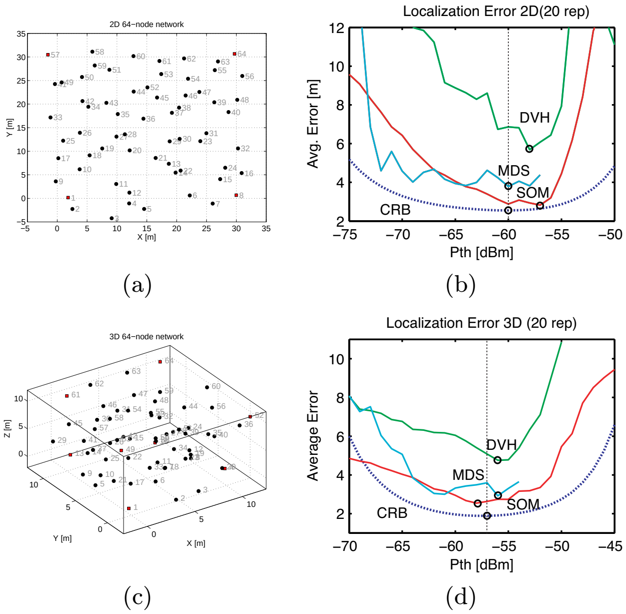


Figure 5: CRB and average localization error for localization in 2D and 3D spaces using three range-free schemes.

To determine an optimal value for the threshold, we consider the average of the standard deviation of the node positions and we take the value that minimizes it:

$$d^*_{th} = \arg\min_{d_{th}} \mathrm{CRB}(d_{th}) \tag{24}$$

$$\mathrm{CRB}(d_{th}) = \frac{1}{n}\sum_{i=1}^{n} \sigma_i, \tag{25}$$

where the values σ_i depends on d_{th} as shown in the previous section.

Before discussing some properties of the FIM, we use two localization examples to validate the choice of the threshold computed using (24). To get a sense of how localization occurs in practice, we have computed the average error of three schemes that are based on diverse approaches to range-free localization. The schemes considered are DV-HOP [10], MDS [15] and the SOM-based localization algorithm described in [5]. Figures 5a and 5c show two sample topologies with nodes placed in a 2D and a 3D space respectively. The node positions were generated by arranging the nodes in a regular configuration (e.g. on the vertex of a regular grid for the 2D case), and then perturbing the initial positions with zero mean Gaussian noise. The CRB for each network is plotted in Figures 5b and 5d together with the error achieved by the three localization algorithms. The values were computed by averaging the localization error[5] over 20 repetitions obtained for different realization of the values P_{ij}'s. As shown in Figure 5b and 5d, the three schemes achieve different localization errors, however, in all of the cases, the minimum error is reached when the power threshold P_{th} is close to the one that produces the minimum value of the CRB.

[5]For a meaningful analysis, the CRB should be compared against the the RMS errors defined in (23), but in Figure 5 we show the average localization error because it provides a more intuitive metric.

4.1 Properties of the FIM and the CRB

Analysis of the localization error in the previous example supports our choice to compute the optimal threshold value using the CRB. We now investigate how the optimal threshold changes as the original network topology is transformed or the parameters of the propagation model change. Our intention is to demonstrate that an approximation for the optimal threshold can be computed using a function that depends only on the network size rather than relying on the analytical approach shown so far. Besides the computational burden incurred in computing the inverse of a potentially large matrix, the CRB analysis does not have practical value because it is based on exact knowledge of the propagation model's parameters, and, above all, the unknown node positions.

In previous work, the CRB has been presented for the case of localization using estimates of the inter-node distances. In that context, it was shown that the CRB is invariant under global translation, rotation or reflection of the network [3]. If we exclude the terms $I_R(\cdot,\cdot)$, the FIMs for distance and connectivity measurements have the same structure, therefore the same properties hold for connectivity-based localization. In the next section we analyze the effect of global scaling of the network coordinates.

4.2 Coordinate Scaling

Consider an arbitrary network and let d^*_{th} be the optimal threshold computed using (24). If we multiply the network coordinates by a factor S, each element of the FIM will be altered. We limit our attention to the terms of the sub-matrix $\mathbf{F_{xx}}$; the term of the other sub-matrices will change similarly. After a coordinate scaling by a factor S, the new terms will be:

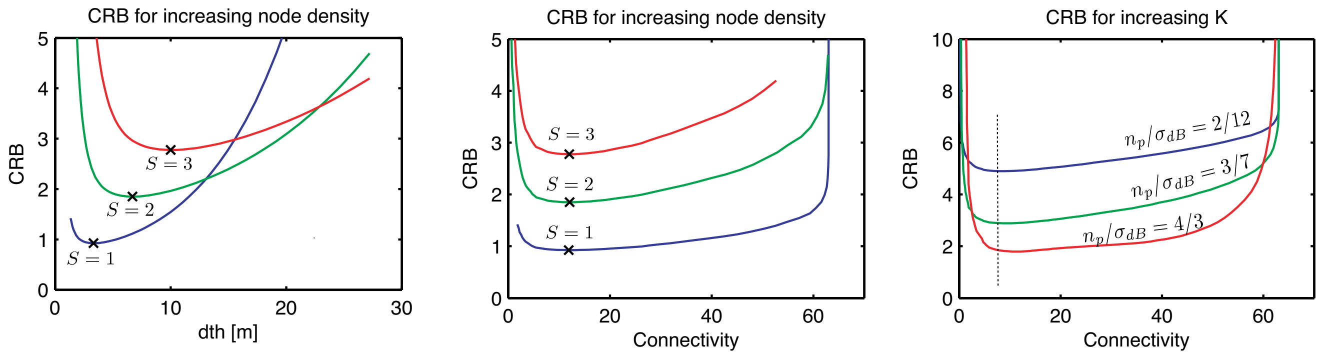
$$[f_{xx}]^{(S)}_{i\neq j} = -\left(\frac{1}{S^2}\right) K\, I_R(Sd_{ij}, d_{th}) \frac{(x_i - x_j)^2}{d^4_{ij}}. \tag{26}$$

We note that each element is multiplied by a factor $1/S^2$, which derives from the ratio $(x_i - x_j)^2/d^4_{ij}$, and also that the value of the term $I_R(\cdot,\cdot)$ is altered as a result of scaling the node distances d_{ij}. However, since $I_R(\cdot,\cdot)$ only depends on the ratio between d_{ij} and d_{th} (see (12)), if we select a new threshold such that $d^{*(S)}_{th} = S\, d^*_{th}$, the value of I_R will remain unchanged. As a result, the FIM will be multiplied by a constant factor S^{-2}, which does not affect the position of the minimum in (24). We conclude that the value $d^{*(S)}_{th} = S\, d^*_{th}$ will be the optimal threshold for the new topology.

We show the invariance of the optimal threshold to scaling in Figure 6a by computing the CRB for a 64-node topology in 2D. We note that since the threshold is scaled together with the network coordinates, the optimal threshold is achieved in correspondence of a constant value of the network connectivity (see Figure 6b). We denote this optimal connectivity value using c^*.

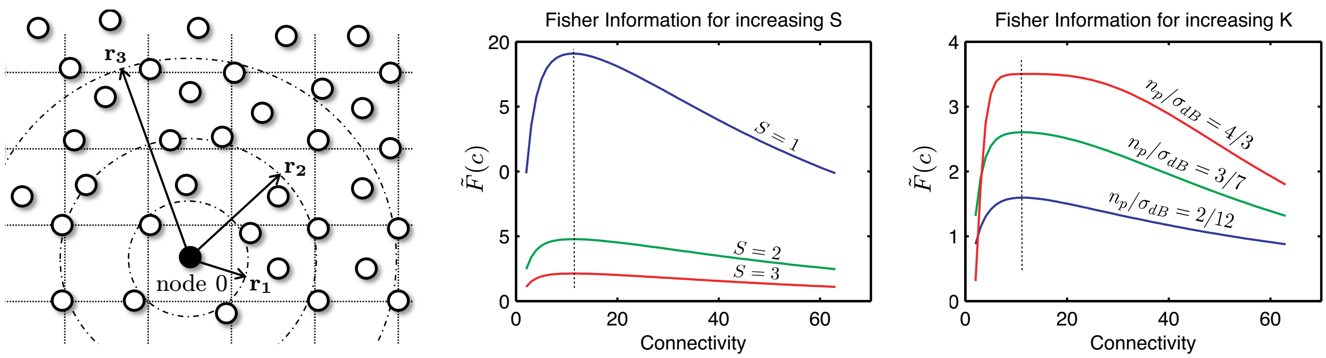
4.3 Invariance to changes in the parameters of the propagation model

In Section 3.1 we have shown that the value of the ratio n_p/σ_{db} increases or decreases the Fisher information, but it does not change the position of the optimal threshold (see Figure 4b). Using simulations, we have verified that a similar result holds for the case of collaborative localization. Changing the value of n_p/σ_{db} does not significantly alter



(a) (b) (c)

Figure 6: Properties of the CRB.



(a) (b) (c)

Figure 7: Approximation of the Fisher information using (29).

the value of d_{th}^* and, consequently, the value c^*. Although we believe that this property can be proved analytically, the differentiation of the terms in the FIM leads to a complex expression whose analysis is beyond the scope of this work. In Figure 6c, we plot the representative example of a 64-node network and the CRB computed for different ratios n_p/σ_{dB} (we recall that n_p is typically between 2 and 4, and σ_{db} between 3 and 12 dBm). Similarly to what was seen for the case of coordinate scaling, the optimal threshold produces a connectivity value that does not change with the ratio n_p/σ_{dB}.

4.4 Approximation of the Optimal Connectivity Value

In the previous section we have seen that given a network topology, the optimal threshold d_{th}^* yields a connectivity value c^* independent from the node density and the ratio n_p/σ_{dB}. Here we exploit this evidence to compute an approximation for the optimal value c^*.

We first recall that the Fisher information reaches its maximum for nodes with distance approximately equal to d_{th} (see Section 3.2). The information also decreases quadratically with the distance; therefore, nodes that are far contribute less information than nodes that are closer.

Now assume we want to estimate the position of node 0 using connectivity information from surrounding nodes (see Figure 7a). If we use a small threshold value (e.g. $d_{th} = r_1$), only the few nodes with distance similar to r_1 will contribute a significative amount of information. If we use a larger value (e.g $d_{th} = r_3$), the number of nodes with distance similar to the threshold increases, but these nodes contributes individually less information because they are far from node 0. *In setting the optimal threshold, we need to consider the trade-off between receiving high-quality information from a small number of nearby nodes, or receiving less valuable information from a larger number of nodes that are far.*

Our goal is to estimate of the amount of information available when using an arbitrary d_{th} value. To this purpose, we partition the nodes in Figure 7a using a sequence $r_1, r_2, r_3, \cdots$ of increasing radius values. Assuming node positions sampled from a two-dimensional Poisson point process, the number of nodes c_i within a radius r_i from node 0 is:

$$c_i = \lambda \pi r_i^2, \tag{27}$$

where λ is the density of the process.

There are $(c_2 - c_1)$ nodes at a distance between r_1 and r_2, $(c_3 - c_2)$ nodes between r_2 and r_3 and so on. As previously discussed, these nodes contribute different amounts of information depending on their distance from node 0 and on the ratio between such distance and d_{th}. Since each measurement contribute additively to the elements of the FIM, we approximate the total amount of information as follows:

$$\tilde{F}(c) = K \sum_i I_R(r_i, d_{th}) \frac{1}{r_i^2}(c_i - c_{i-1}), \tag{28}$$

where c is the value of network connectivity achieved when d_{th} is used, and the term I_R is the one discussed in Section 3.2. Note that, under the hypothesis of Poisson distribution, $c = \lambda \pi d_{th}^2$.

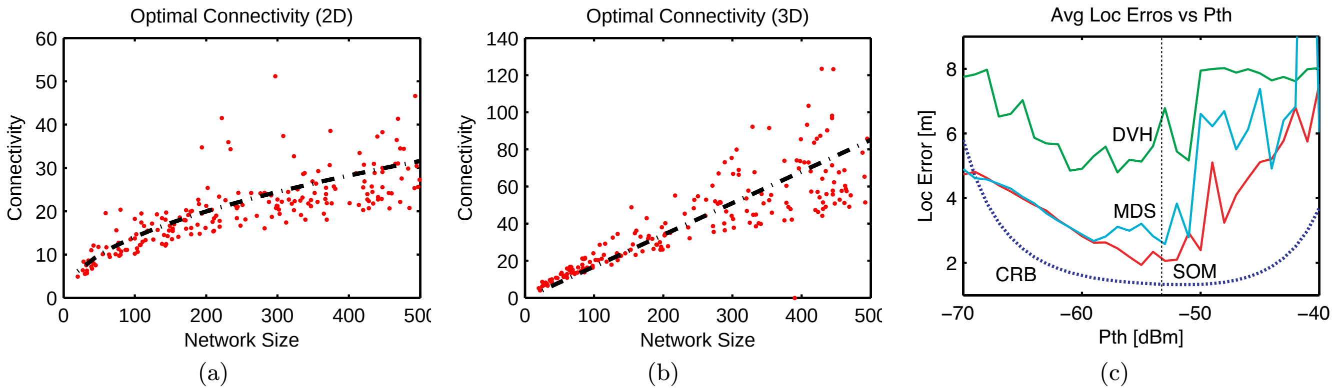


Figure 8: a) Optimal Connectivity for 2D networks, b) Optimal Connectivity for 3D networks, c) Localization error for the network in [12].

If we reduce the distance between the r_i values to an infinitesimal value, the summation in (28) becomes the following integral:

$$\tilde{F}(c) = K \int_{2}^{n+m} I_R(r_i, d_{th}) \frac{1}{r_i^2} dc_i. \tag{29}$$

In the equation above, the value c_i is a function of r_i as described in (27). The limit of integration are chosen so that that there are a minimum of two nodes within the smallest radius (the reference node and another one) and a maximum of $n + m$ nodes. Using (29) to measure the available information, the approximation of the optimal connectivity is the value that maximizes the integral:

$$\tilde{c}^* = \arg\max_c \tilde{F}(c). \tag{30}$$

Using simulation, we have empirically verified that the approximate values share the same properties of the optimal connectivity value discussed in previous sections. The value computed using (30) is independent of the node density and the ratio n_p/σ_{dB}. These properties are illustrated in Figures 7b and 7c, where the $\tilde{F}(c)$ is computed for a 64-node network (2D) with different parameters of node density and ratio n_p/σ_{dB}. *Equation* (30) *allows us to compute the optimal connectivity using a function that depends only on the network size.*

Results for the 3D case are obtained following a similar approach. The only difference is that nodes are partitioned using spheres of radius r_i. The number of nodes contained in each of such spheres is:

$$c_i = \frac{4}{3}\lambda\pi r_i^3. \tag{31}$$

5. SIMULATIONS

Using a larger simulation set, we have evaluated the quality of the approximation given by (30) for networks with nodes in 2D and 3D spaces. For each case, we simulated 200 random networks with a number of nodes between 20 and 500. The deployment areas were fixed: nodes were placed inside a square region 50 m × 50 m for 2D networks, and in cube with side measuring 50 m for 3D networks. Four and eight nodes in the corner of the network were used as anchors for localization in 2D and 3D deployments respectively. For each network the parameters of the propagation model were uniformly sampled in the following intervals: $n_p \in [2, 4]$ and $\sigma_{db} \in [3, 12]$ dBm.

Figures 8a and 8b show the simulation results. The clouds of points correspond to the optimal connectivity values c^* computed using the CRB. The dashed lines represent the approximate values $\tilde{c}^*$ found using (30). These values were computed in MATLAB by numerically integrating (`quadl`) the integral in (29) and then finding the minimum of its inverse (`fminsearch`).

The plots show that (30) produces a good approximation of the results obtained using CRB analysis. For each of the simulated networks, we compared the difference between the absolute minimum of the bound and the value computed using the approximate threshold:

$$\text{Bound Error } \% = 100\, \frac{\text{CRB}(\tilde{d_{th}}) - \text{CRB}(d_{th}^*)}{\text{CRB}(d_{th}^*)}, \tag{32}$$

where $\tilde{d_{th}}$ is the threshold that realizes the connectivity value $\tilde{c}^*$ computed using (30). On average, the bound error was 0.53% for 2D localization and 0.34% for 3D localization. The standard deviation of such error was 1.15% for 2D localization and 0.34% for the 3D case.

5.1 Case Studies

Using (30) we are finally able to give an answer to the question regarding the optimal threshold for the 44 node network of Figure 1a. The optimal connectivity value for a 44 node network is 9.27; for the network considered, this connectivity is achieved when $P_{th} = -53.3$ dBm. In Figure 8c, the optimal threshold (the vertical dashed line) is plotted together with the error of the DV-HOP, MDS and SOM algorithm. For this network we also report the CRB computed using the estimated values for the propagation model's parameters ($n_p = 1.7, \sigma_{dB} = 3.91$ dBm). The connectivity value given by (30) is close to the minimum of the CRB and close to the absolute minima of the MDS and SOM errors.

In the second case study, we have used the RSS data from a 38 node network deployed in a 3D space [8]. The data is freely available on the ENALAB web site[6]. The optimal connectivity value found using (30) is 11.093, which for this network is achieved by setting a threshold $P_{th} = -34.3$ dBm. The error of the three localization algorithms for this net-

[6] `http://www.eng.yale.edu/enalab/XYZ/data_set_1.htm`

work is reported in Figure 9. Again, we note that the estimated threshold result in an error that is close to the absolute minimum error for the three localization schemes.

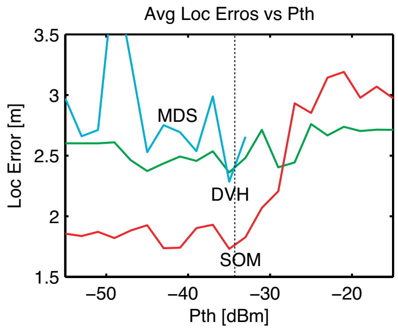


Figure 9: Localization error for the network in [8].

6. RELATED WORK

Over the past few years, analysis of the CRB have been used by a number of authors to characterize the error bound of localization algorithms, especially for the case of range measurements (angle or distances) affected by Gaussian noise. Moses et al. [9] derive the CRB for the case where localization is based on signals emitted by a set of sources, and nodes can measure the Time of Arrival (ToA) or the Angle of Arrival (AoA). A study of the CRB under various conditions of node and beacon density is proposed by Savvides et al. [1]. Wang et al. define a Bayesian Bound (BB) that is the covariance of a posterior distribution computed from the sensor observation [16]. This bound is equivalent to the CRB for measurements with Gaussian error, but it is computationally less demanding. Analysis of the CRB is proposed by Patwari et al. for the case of collaborative localization using distance estimates obtained by ToA and RSS [12], and for localization using angle estimates [11]. The case of localization using connectivity information or quantized RSS levels is studied by Patwari and Hero III [13]. The idea to obtain connectivity data from RSS value is also used by Li et al. to implement a *Partial Range Information* (PRI) scheme that derives sub-hop information useful in improving the localization accuracy [7]. This idea is somewhat similar to the one proposed in this work, since we also try to improve the localization accuracy by choosing a threshold for the RSS values.

Finally, in a work complementary to the one in this paper, we have used the CRB analysis to determine the conditions under which a range-free scheme can potentially outperform a range-based scheme that uses RSS values to obtain distance estimates [6].

7. CONCLUSIONS

Using CRB analysis we have shown how to define an optimal threshold value when connectivity information is obtained from RSS measurements. As a result of our analysis, we have found that the optimal threshold is related to the connectivity (i.e. average number of neighbors per node) and that this value can be accurately approximated using a function that depends only on the number of network nodes

The results presented are independent from the localization algorithm used and provide a simple and practical rule to determine the optimal connectivity level for range-fre localization. After having presented results for the 2D and 3D cases, we are currently investigating the 1D case.

8. ACKNOWLEDGEMENTS

This work was supported in part by grants from SFAz and NIH (1 R01 HD45816-01A1), and gifts from Mediserve Inc. and Intel Corp. The authors would like to thank the anonymous reviewers for their helpful comments.

9. REFERENCES

[1] W. Andreas Savvides, R. Sachin Adlakha, and M. Srivastava. On the Error Characteristics of Multihop Node Localization in Ad-Hoc Sensor Networks. *IPSN*, 2003.

[2] N. Bulusu, J. Heidemann, and D. Estrin. GPS-less low-cost outdoor localization for very small devices. *IEEE Personal Communications*, 7(5):28–34, 2000.

[3] C. Chang and A. Sahai. Cramér-Rao-type bounds for localization. *EURASIP Journal on Applied Signal Processing*, 2006(1):166–166, 2006.

[4] B. Frieden. *Science from Fisher Information: A Unification*. Cambridge University Press, 2004.

[5] G. Giorgetti, S.K.S. Gupta, and G. Manes. Wireless localization using self-organizing maps. *IPSN*, 2007.

[6] G. Giorgetti, S.K.S. Gupta, and G. Manes. Localization Using Signal Strength: To Range or Not To Range? *ACM MELT*, 2008.

[7] X. Li, H. Shi, and Y. Shang. A partial-range-aware localization algorithm for ad-hoc wireless sensor networks. *IEEE Conf. on Local Computer Networks*, 2004.

[8] D. Lymberopoulos, Q. Lindsey, and A. Savvides. An Empirical Analysis of Radio Signal Strength Variability in IEEE 802.15. 4 Networks using Monopole Antennas. *ENALAB Tech. Report 050501*.

[9] R. Moses, D. Krishnamurthy, and R. Patterson. A Self-Localization Method for Wireless Sensor Networks. *EURASIP Journal on Applied Signal Processing*, 2003(4):348–358, 2003.

[10] D. Niculescu and B. Nath. DV Based Positioning in Ad Hoc Networks. *Telecommunication Systems*, 22(1):267–280, 2003.

[11] N. Patwari, J. Ash, S. Kyperountas, A. Hero III, R. Moses, and N. Correal. Locating the nodes: cooperative localization in wireless sensor networks. *Signal Processing Magazine, IEEE*, 22(4):54–69, 2005.

[12] N. Patwari, A. Hero, M. Perkins, N. Correal, and R. O'Dea. Relative location estimation in wireless sensor networks. *IEEE Transactions on Signal Processing*, 51(8):2137–2148, 2003.

[13] N. Patwari and A. Hero III. Using proximity and quantized RSS for sensor localization in wireless networks. *WSNA*, 2003.

[14] T. Rappaport. *Wireless Communications: Principles and Practice*. IEEE Press Piscataway, NJ, USA, 1996.

[15] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz. Localization from mere connectivity. *MobiHoc*, 2003.

[16] H. Wang, L. Yip, K. Yao, and D. Estrin. Lower bounds of localization uncertainty in sensor networks. *ICASSP*, 2004.