Understanding the Limits of RF-Based Collaborative Localization

Gianni Giorgetti, *Member, IEEE*, Sandeep Kumar S. Gupta, *Senior Member, IEEE, Member, ACM*, and Gianfranco Manes, *Senior Member, IEEE*

***Abstract*—RF-based localization has gained popularity as a low-cost solution to support *position awareness* in ad hoc networks. The received signal strength (RSS) measured by pairs of nodes can be used to obtain either range estimates or connectivity information. It is not clear, however: 1) when a range-based scheme should be used in favor of a connectivity-based one, and 2) how to optimally convert the RSS into connectivity data. This paper uses analysis of the Fisher information and the Cramér–Rao bound (CRB) to answer these questions. Solutions are found by comparing the network connectivity against two values: the critical connectivity (CC) and the optimal connectivity (OC). After discussing the properties of both values, we show how their approximation can be used to improve the performance of RF-based localization systems.**

***Index Terms*—Connectivity, Cramér–Rao bound (CRB), Fisher information, localization, ranging, received signal strength (RSS), signal strength.**

I. Introduction

THE ABSOLUTE position of a wireless device can be determined by collecting measurements from *anchors* at fixed positions. For example, if a node can estimate the distance from three or more anchors, its position can be computed using *trilateration*. This approach is simple and effective, but it might fail in applications with insufficient anchor coverage such as low-power sensor networks. In these scenarios, a node might compensate for the lack of reference devices by taking measurements with other peer nodes at unknown positions. If all the nodes participate in this collaborative effort, the data can be used to localize the whole network (Fig. 1).

Several techniques have been proposed to localize wireless networks. Among the proposed solutions, collaborative localization based on radio measurements has gained significant popularity due to its low cost and flexibility. Radio-positioning techniques support virtually any wireless device, integrate well with existing networks, and operate both outdoors and indoors. Additionally, radio messages can be used to implement both *range-based* and *connectivity-based* schemes.

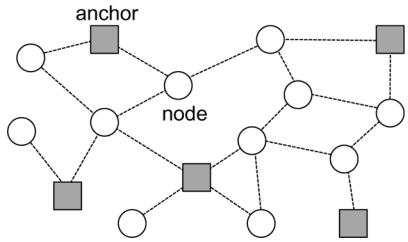

Fig. 1. Collaborative localization: The position of several wireless nodes has to be computed using peer-to-peer measurements and anchor information.

In the range-based approach, the *received signal strength* (RSS) measured by the transceiver is used to estimate the distance of the transmitting node. When a sufficient number of range estimates is available, the node positions are computed using multilateration algorithms [29], semidefinite programming [4], maximum likelihood estimators [24], or spring-mass relaxation approaches [26]. In general, any distance-based localization scheme will work with RSS range estimates.

On the other hand, connectivity-based schemes discard the RSS values, but the successful reception of radio messages indicates that two nodes are close in space. This condition is also expressed by saying that the nodes are *neighbors* or *connected*. Several connectivity-based schemes have been proposed in the literature. Examples include the simple Centroid algorithm [6] and more sophisticated approaches such as convex optimization [9], distance vector positioning (DV_HOP) [21], multidimensional scaling [30], and neural networks [10].

The number of available schemes offers many implementation alternatives, but it also makes it difficult to determine which scheme will work better for a specific application. Since direct comparison between many different solutions is impractical, successful implementation of an RF-based solution requires a deep understanding of the characteristic of range-based and connectivity-based schemes. For this purpose, we define two fundamental problems that have not been sufficiently investigated in previous literature: the *Measurement Selection* and the *Optimal Connectivity* problems.

Measurement Selection Problem (MSP): The first problem concerns the choice between range-based and connectivity-based approaches. It is well known that range-based schemes are capable of *fine-grained* positioning, while connectivity-based approaches can only produce *coarse-grained*

Manuscript received March 21, 2010; revised October 03, 2010 and February 26, 2011; accepted March 03, 2011; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor D. Starobinski. This work was supported in part by SFAz and NIH under Grant 1 R01 HD45816-01A1, Mediserve, Inc., and Intel Corporation.

G. Giorgetti was with the IMPACT Laboratory, Arizona State University, Tempe, AZ 85287 USA. He is now with Navizon, Inc., Miami Beach, FL 33139 USA (e-mail: gianni@navizon.com).

S. K. S. Gupta is with the IMPACT Laboratory, Arizona State University, Tempe, AZ 85287 USA (e-mail: sandeep.gupta@asu.edu).

G. Manes is with the Department of Electronics and Telecommunications, University of Florence, Florence 50139, Italy (e-mail: gfm@unifi.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2011.2141681

results. However, range estimates obtained from RSS values can be inaccurate due to the unpredictability of the wireless channel, and some authors have occasionally noticed that connectivity-based approaches can outperform range-based schemes in noisy environments [3], [22]. Given these premises, it is not clear which approach will work better when. *Should the radio measurements be used to estimate the internode distances or be converted into connectivity information?*

Optimal Connectivity Problem (OCP). This problem concerns how to obtain connectivity information from radio measurements. Previous research work has often assumed circular connectivity based on an *idealized* radio propagation. While this model is useful in simulation studies, in real-world applications connectivity cannot be defined as a geometric function of the (unknown) node positions. Connectivity will have to be defined as a function of the radio measurements collected between two nodes. For example, in the Centroid scheme [6], nodes are connected if at least 80% of the messages are received. Other authors have proposed schemes where the neighbors are determined by sorting the RSS values [17]. The need to find a solution to the MSP requires putting the choice of the connectivity model on a firmer basis and introduces a more general problem: Besides the heuristics proposed in the literature, *is there an optimal approach to obtain connectivity information from radio measurements?*

A. Outline and Contributions

This paper provides an analytical comparison between the performance of range-based and connectivity-based schemes. We also address the problem of how to optimally obtain connectivity information from radio measurements. While both these problems were initially investigated in our previous work [11], [12], this paper addresses the MSP and OCP using a unified approach and further supports the results with examples and analytical evidences.

The solution to MSP and OCP is based on research work that has cast localization as a parameter estimation problem. Several authors have previously used analysis of the Fisher information and Cramér–Rao bound (CRB) to define the limits that bound the minimum error for localization using noisy range estimates [7], [19], [24], [28] and connectivity measurements [25]. Our analysis provides an intuitive explanation of the previous results. Additionally, this paper aims at explaining the differences between range-based and connectivity-based localization and proposes practical solutions to the MSP and OCP that are independent of the specific scheme used.

Section II introduces a simple localization example designed to build an intuitive understanding of the Fisher information when the RF signal follows a log-normal distribution (see Table I for notation). Connectivity data are obtained by setting a threshold P_{th}: Two nodes are connected if their average RSS exceeds P_{th}, and disconnected otherwise. Using Fisher information analysis, Section III shows how to set a correct threshold value in a connectivity-based scheme. Additionally, it is shown that, unlike RSS measurements, connectivity data still carry some information when nodes are out of range.

TABLE I
NOTATION

Symbol	Description	Notes
z	Average RSS measured between two nodes	in dB or dBm
$\bar{P}_r(d)$	Average RSS expected at distance d	in dB or dBm
n_{p}	Path loss exponent	typical $n_{\text{p}} \in [2, 4]$
σ_{dB}	Shadowing standard deviation	typ. $\sigma_{\text{dB}} \in [3, 12]$ dB.
P_{th}	Connectivity threshold	in dB or dBm
d_{th}	Distance threshold	in meters
c	Connectivity between nodes	$c \in \{0, 1\}$
F_{rss}	Fisher Information for RSS measurements	1D case
F_{con}	Fisher Information for connectivity meas.	1D case
CRB_{rss}	CRB for RSS measurements.	multiparameter case
CRB_{conn}	CRB for connectivity meas.	multiparameter. case
CC	Critical Connectivity value	
OC	Optimal Connectivity value	
k	Network connectivity	average node degree

Section IV extends the same analysis to localization scenarios with multiple nodes deployed as a network. To provide a homogeneous comparison between the range-based and the connectivity-based approach, we investigate the CRB as a function of the network connectivity. CRB analysis indicates the existence of two fundamental values: the *critical connectivity* (CC) and the *optimal connectivity* (OC) values.

These values are crucial in answering the two problems addressed by this paper. Given a network with average connectivity k, CRB analysis suggests that a connectivity-based scheme should be used if $k \leq \text{CC}$, while a range-based solution will work better if $k > \text{CC}$. When a connectivity-based scheme is the only viable solution, the optimal P_{th} is the quantization value that ensures $k = \text{OC}$.

Unfortunately, computing the CRB requires knowledge of the node positions. To devise a scheme of practical utility, Section V investigates the properties of the CC and OC values in relation to a few important application parameters. Based on the results found, Section VII shows that the CC value can be approximated with sufficient accuracy by using a function of the network size, i.e., the number of network devices, and the parameters of the propagation model. The approximation of the OC value only requires knowledge of the network size. We conclude the paper discussing practical applications of the results in few typical network positioning problems.

II. 1-D Node Localization Problem

We begin our study using a simple localization scenario. This problem will help with understanding RF-based localization and the factors that affect its performance.

Consider two devices placed as in Fig. 2. The goal is to compute the position of node B, or equivalently its distance d from the origin. The position will be estimated using two solutions: one that uses RSS value collected between the two nodes (Section II-B), and another that converts them into connectivity data (Section II-C). After describing the propagation model for the RF signal, Section III will use the Fisher information to compare the two approaches.

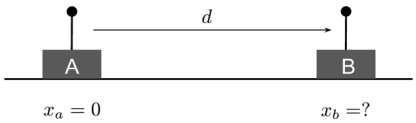

Fig. 2. 1-D localization: The distance of node B from the origin has to be estimated using RSS and connectivity information.

A. Propagation Model

To enable localization, the two nodes exchange radio messages and collect a set $\mathcal{Z}$ of RSS values: $\mathcal{Z} = \{z_1, z_2, z_3, \ldots\}$. Let z be the average of the RSS value collected

$$z = \frac{1}{|\mathcal{Z}|} \sum_{z_i \in \mathcal{Z}} z_i. \tag{1}$$

This analysis assumes z described by the *log-normal shadowing model*, a propagation model that is widely used for link budget analysis in wireless communication. Adoption of this model is supported both by theoretical analysis of the RF propagation and by measurements in indoor as well as outdoor radio channels [5], [14], [27], [32].

If the RSS is measured in dB or dBm, z is the outcome of a random variable Z with normal distribution

$$Z \sim \mathcal{N}(\bar{P}_r(d), \sigma_{\mathrm{dB}}). \tag{2}$$

The mean of Z is the *log-distance* term $\bar{P}_r(d)$ that models the attenuation of the signal as a function of the distance

$$\bar{P}_r(d) = P_0 - 10 n_p \log_{10}(d/d_0) \tag{3}$$

where P_0 is the received power measured at reference distance d_0, and n_p is the *path loss exponent* (typically $2 \le n_p \le 4$).

The standard deviation σ_{dB} in (2) characterizes the variability measured between pairs of nodes with the same separation distance, but placed in different locations. Obstructions in the path between the nodes and reflections of the signal due to nearby obstacles can produce significant differences in the average RSS between equidistant nodes. Typically, σ_{dB} is between 3 and 12 dB. Fig. 3 shows the probability density function (pdf) of RSS values computed for nodes with separation distance $d = 5$ m and parameters $n_p = 2.5$ and $\sigma_{\mathrm{dB}} = 3$ dBm.

B. Range-Based Localization

Ranging schemes will estimate d using the average RSS measured between the nodes. Having assumed the model (2), in absence of prior information on d, the position can be obtained with the *maximum likelihood estimator* (MLE)

$$\hat{d}_{\mathrm{ML}} = d_0 10^{(P_0 - z)/10 n_p}. \tag{4}$$

The MLE offers a simple solution to convert RSS values into range estimates and allows us to evaluate the estimation error. If the RSS measured between the two devices is $z = \bar{P}_r(d) + \delta$, where δ is a sample from the random variable $\Delta \sim \mathcal{N}(0, \sigma_{\mathrm{dB}})$, then the error is

$$e = \hat{d}_{\mathrm{ML}} - d = d\left(10^{-\frac{\delta}{10 n_p}} - 1\right). \tag{5}$$

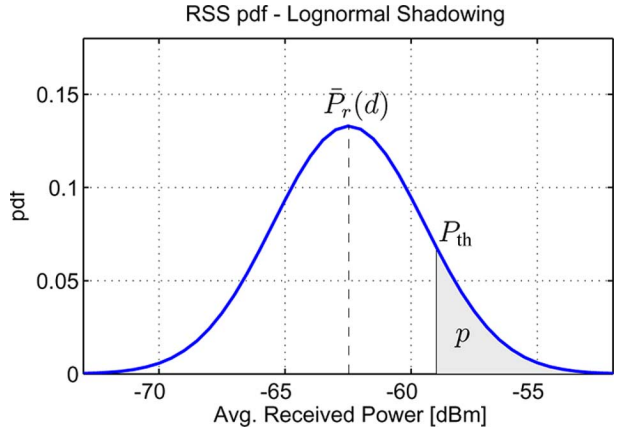

Fig. 3. PDF of the average RSS values for nodes placed at $d = 5$ m. Log-normal shadowing parameters: $d_0 = 1$ m, $P_0 = -45$ dBm, $n_p = 2.5$, $\sigma_{\mathrm{dB}} = 3$ dBm. The area p is the probability that the power exceeds the threshold P_{th}.

In the absence of shadowing effects, the MLE produces the correct estimate, but when $\delta \neq 0$, the error is proportional to the node distance. Consider, for example, the propagation parameters used in Fig. 3. The expected received power for nodes placed 5 and 10 m apart are $\bar{P}_r(5) = -62.5$ dBm and $\bar{P}_r(10) = -70$ dBm. Assuming a shadowing contribution $\delta = +3$ dBm in both cases, the absolute error will be just 1.2 m for the 5-m case, but it will double to 2.4 m for the 10-m case.

C. Connectivity-Based Localization

The second option is to estimate d using connectivity data. Connectivity-based schemes do not normally produce distance estimates, but in the scenario of Fig. 2, the node position coincides with the distance.

The first step requires converting the RSS into connectivity. A general approach is to compare z against a threshold P_{th}: The two nodes are "connected" if $z \ge P_{\mathrm{th}}$, or "disconnected" in the other case. According to this binary quantization, connectivity is described by a random variable C

$$C = \begin{cases} 0, & \text{if } Z < P_{\mathrm{th}} \quad (\text{nodes disconnected}) \\ 1, & \text{if } Z \ge P_{\mathrm{th}} \quad (\text{nodes connected}). \end{cases} \tag{6}$$

In the rest of the paper, the term *connectivity-based schemes* is used to denote those algorithms capable of localizing nodes based on binary information $C = \{0, 1\}$.

The probability the event $C = 1$ (nodes connected) is the shadowed area in Fig. 3. Its value is

$$p = \Pr\{C = 1\} = 1 - G\left(\frac{P_{\mathrm{th}} - \bar{P}_r(d)}{\sigma_{\mathrm{dB}}}\right) \tag{7}$$

where G is the cumulative distribution function (cdf) of a normal random variable $\mathcal{N}(0, 1)$. Analysis of the probability (7) has been previously used to determine the coverage area of base stations [27] and connectivity in ad hoc wireless networks [2], [15].

A connectivity-based scheme will use the intuitive assumption that two nodes are "far" if $C = 0$, and "close" if $C = 1$. This approach does not require knowing the propagation model's parameters. On the down side, computing an actual estimate for d is not straightforward. For example, unlike the

previous case, the MLE produces trivial values of scarce utility: $\hat{d}_{\rm ML} = +\infty$ for $C = 0$, and $\hat{d}_{\rm ML} = 0$ for $C = 1$.

In practice, some other solution will be used to translate the values (6) into actual distance estimates. Assuming a particular scheme, however, is not necessary. Localization approaches using RSS and connectivity data can be compared on a general basis using the Fisher information and the CRB.

III. Fisher Information and CRB: 1-D Case

The Fisher information (F) measures the uncertainty when estimating an unknown parameter on the basis of noisy measurements. In the proposed example, the unknown parameter is d, while the available measurements are modeled by the random variables Z and C.

The inverse of the Fisher information, known as the Cramér–Rao bound, sets a lower limit on the variance that can be achieved when estimating d using *any* unbiased estimator

$$\mathrm{Var}\{\hat{d}\} \geq \frac{1}{F(d)}. \tag{8}$$

By studying the Fisher information for RSS and connectivity measurements, we can compare the two approaches analytically and identify the parameters that affect their error.

A. Fisher Information for RSS Measurements

The *measurement model* that relates the RSS value to the distance is a normal distribution with pdf

$$f_Z(z;d) = \frac{1}{\sigma_{\rm dB}\sqrt{2\pi}} \exp\left(-\frac{(z-\bar{P}_r(d))^2}{2\sigma_{\rm dB}^2}\right). \tag{9}$$

The Fisher information for RSS measurements is computed by evaluating the function

$$F_{\rm rss}(d) = E\left\{\left[\frac{\partial}{\partial d}\log f_Z(z;d)\right]^2\right\}. \tag{10}$$

For the nodes in Fig. 2, substituting (9) into (10) yields

$$F_{\rm rss}(d) = {K_c}^2\frac{1}{d^2} \tag{11}$$

where the constant

$$K_c = \frac{10}{\log 10}\frac{n_p}{\sigma_{\rm dB}} \tag{12}$$

describes the effect of the the propagation model's parameters.

Fig. 4 shows $F_{\rm rss}$ as a function of d for different values of n_p and $\sigma_{\rm dB}$. The plots confirm what was already seen in Section II-B when discussing the MLE error. The amount of information available to estimate d decreases for increasing values of the distance.

The $F_{\rm rss}$ value also depends on propagation model's parameters n_p and $\sigma_{\rm dB}$. Larger n_p values imply a stronger correlation between the received power and the node distance, which is a condition that causes the estimation error to decrease. On the other hand, larger $\sigma_{\rm dB}$ values pertain to radio channels with a strong shadowing component. Shadowing can be regarded as a source of *noise* that reduces the accuracy of the estimation process.

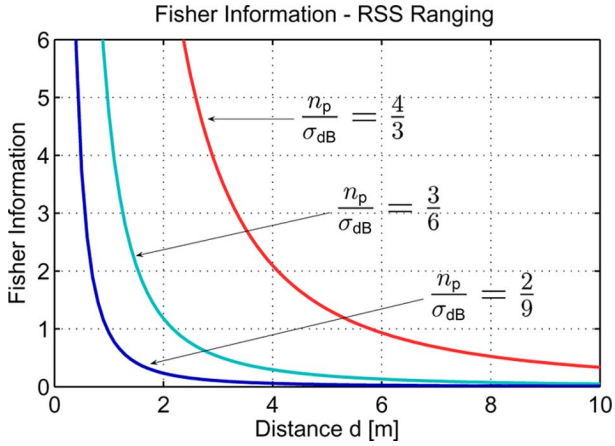

Fig. 4. Fisher information for RSS values measured at various distances.

B. Fisher Information for Connectivity Measurements

The Fisher information for connectivity data depends not only on the node distance, but also on the value of the threshold $P_{\rm th}$ used in (6). To present an expression of F similar to (11), we first need to describe how $P_{\rm th}$ relates to d.

Let $d_{\rm th}$ be the distance at which the expected received power equals the power threshold, i.e., $\bar{P}_r(d_{\rm th}) = P_{\rm th}$. Solving (3), we obtain

$$d_{\rm th} = d_0 10^{(P_0 - P_{\rm th})/(10n_p)}. \tag{13}$$

When the devices are at a distance $d = d_{\rm th}$, they will be connected with probability $p = 0.5$. Therefore, $d_{\rm th}$ can be thought as an average *threshold distance*.

Equation (7) can be rewritten to show the dependence of p on the distance d and $d_{\rm th}$

$$p = p(d, d_{\rm th}) = 1 - G\left[K_c \log\left(\frac{d}{d_{\rm th}}\right)\right]. \tag{14}$$

The probability mass function that relates connectivity measurement to the node position and the threshold distance is

$$f_C(c;d,d_{\rm th}) = \begin{cases} 1-p(d,d_{\rm th}), & \text{if } c=0 \\ p(d,d_{\rm th}), & \text{if } c=1 \\ 0, & \text{else.} \end{cases} \tag{15}$$

Replacing f_Z with f_C in (10), the expression that defines the Fisher information for connectivity measurements is

$$\begin{aligned} F_{\rm con}(d,d_{\rm th}) &= K_c^2\frac{1}{d^2}I_r(d,d_{\rm th}) \\ &= F_{\rm rss}(d)I_r(d,d_{\rm th}). \end{aligned} \tag{16}$$

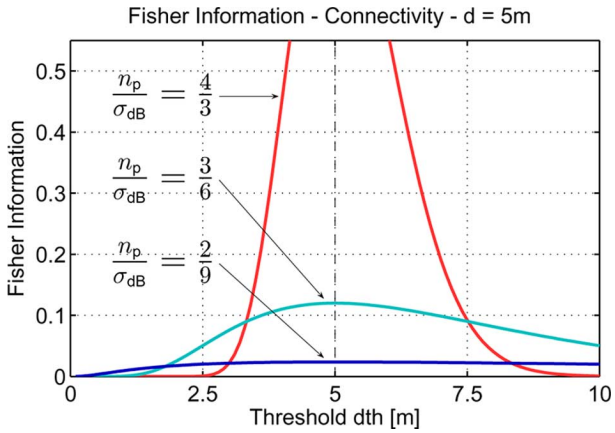

Fig. 5. Fisher information for connectivity measurements for different thresholds $d_{\rm th}$ and different ratios $n_p/\sigma_{\rm dB}$. The node distance is $d = 5$ m.

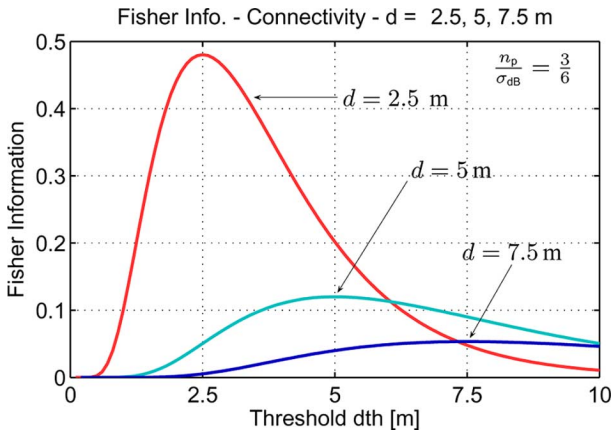

Fig. 6. $F_{\rm con}$ for nodes at distance $d = \{2.5, 5.0, 7.5\}$ m and different threshold values. The $F_{\rm con}$ value always peaks when $d_{\rm th} = d$.

The function $F_{\rm con}$ is equal to $F_{\rm rss}$ multiplied by an additional term. The term I_r depends on the ratio between d and the threshold distance $d_{\rm th}$

$$I_r(d, d_{\rm th}) = \frac{2}{\pi} \frac{\exp\left[-K_c^2 \log\left(d/d_{\rm th}\right)^2\right]}{1 - \mathrm{erf}\left[K_c \log\left(d/d_{\rm th}\right)/\sqrt{2}\right]^2}. \qquad (17)$$

Fig. 5 shows $F_{\rm con}$ as a function of $d_{\rm th}$ when $d = 5$ m for different values of $n_p/\sigma_{\rm dB}$. Fig. 6 shows $F_{\rm con}$ for increasing distances d. Connectivity measurements reach the maximum information content when $d_{\rm th}$ equals the true node distance (which is unknown). If $d_{\rm th} = d$, the probability of detecting the node as connected is 0.5, and $F_{\rm con}$ is approximately 37% lower than $F_{\rm rss}$. In fact, $I_r(d, d_{\rm th}) = 2/\pi \cong 0.63$ for $d_{\rm th} = d$.

As will be discussed in Section IV, selecting a correct threshold is a critical step in implementing a connectivity-based localization scheme.

C. Threshold Selection

An incorrect threshold will reduce the Fisher information and increase the error of a connectivity-based scheme. To understand the effect of different thresholds, consider the following problem. Assume a node that can occupy three positions A, B, and C at a distance 3, 5, and 7 m, respectively from a reference node (see Fig. 7). The goal is to identify the true node position using connectivity measurements. Three quantization thresholds are available: low, medium, and high with values $P_{\rm L}$, $P_{\rm M}$, and $P_{\rm H}$ (see Fig. 8). *Which threshold will work best?*

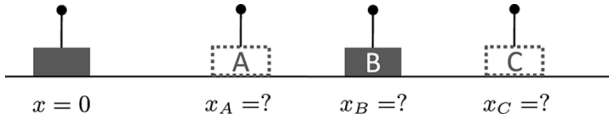

Fig. 7. Localization example with a node placed in three possible positions.

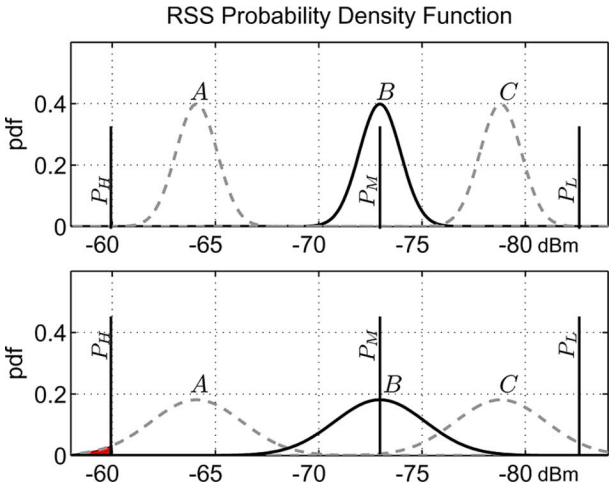

Fig. 8. Distribution of the RSS values for the example in Fig. 7. *(top)* Low $\sigma_{\rm dB}$ value. *(bottom)* Higher $\sigma_{\rm dB}$ value.

Assume B to be the true but unknown node position. The top plot in Fig. 8 shows the thresholds together with the RSS distribution for the node at different positions. Let $p_{\rm a}$, $p_{\rm b}$, and $p_{\rm c}$ be the probability of measuring the node as connected. If the low threshold is used (i.e., $P_{\rm th} = P_{\rm L}$), it will be impossible to determine the position occupied by the node because all the three cases will produce connected measurements with probability close to one ($p_{\rm a} \approx 1, p_{\rm b} \approx 1, p_{\rm c} \approx 1$). Similarly, using the high threshold $P_{\rm H}$ will result in the same measurement (nodes disconnected) for all the three cases: $p_{\rm a} \approx 0$, $p_{\rm b} \approx 0$, $p_{\rm c} \approx 0$. Given the ambiguity in the measurements, the node could be placed either at A or C without changing the connectivity. Neither $P_{\rm L}$ nor $P_{\rm H}$ allows a connectivity-based scheme to determine the correct node position.

The best threshold selection is $P_{\rm th} = P_{\rm M}$, which yields probabilities $p_{\rm a} \approx 1$, $p_{\rm b} = 0.5$, and $p_{\rm c} \approx 0$. Since this threshold maximizes the probability to obtain different measurements for nodes at different positions, the expected localization error is lower than the error in the previous cases.

D. Effect of Noise on Threshold Selection

The problem just illustrated is also useful to understand the effect of noisy RSS value on connectivity based localization.

Low values of the ratio $n_p/\sigma_{\rm dB}$ degrade the quality of range estimates using RSS and have a similar effect on connectivity measurements (see Figs. 4 and 5). However, a large RSS variability can sometimes mitigate the effect of a wrong threshold selection. Fig. 8 (bottom) shows the RSS distributions computed

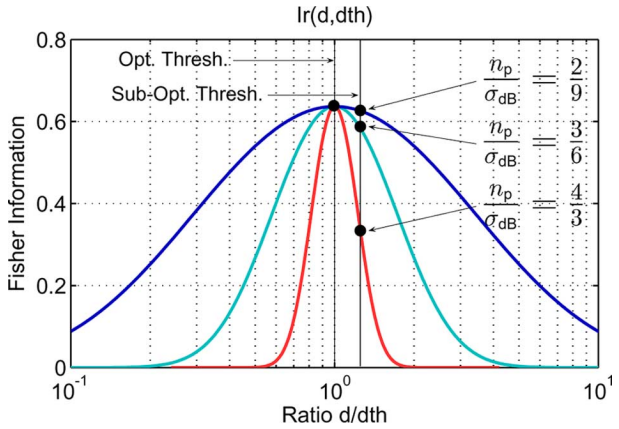

Fig. 9. Effect of decreasing values of the ratio $n_p/\sigma_{\rm dB}$ on the $I_r(d, d_{\rm th})$ term. Unlike $F_{\rm rss}$, the $I_r(d, d_{\rm th})$ term increases with the noise in the measurements.

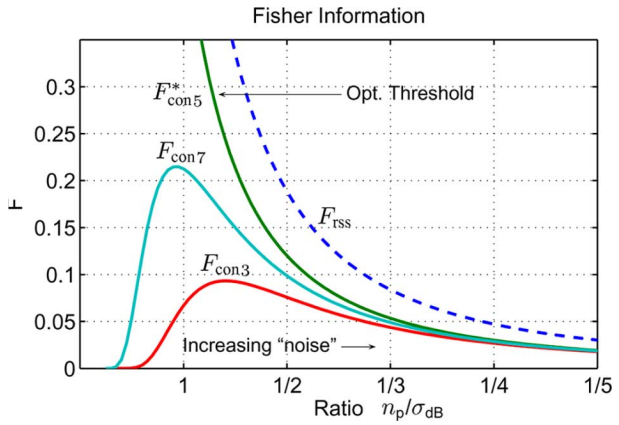

Fig. 10. Fisher information as a function of $n_p/\sigma_{\rm dB}$ for two nodes at $d = 5$ m. The differences between using an optimal threshold $F^*_{\rm con5}$ and suboptimal values $F_{\rm con3}, F_{\rm con7}$ decrease as the ratio $n_p/\sigma_{\rm dB}$ decreases. The plot also show the Fisher information $F_{\rm rss}$ for the same pair of nodes.

for the same n_p, but with a larger $\sigma_{\rm dB}$ value. The thresholds $\{P_{\rm L}, P_{\rm H}\}$, while still nonoptimal, are not as ineffective as they were in the previous case. In both cases, there is a nonzero probability to obtain a different measurement for at least one of the nodes at position A or C.

The effect of noise on connectivity-based localization can be measured by evaluating $F_{\rm con}$ for different values of $n_p/\sigma_{\rm dB}$. Fig. 9 shows the $I_r(d, d_{\rm th})$ term plotted as a function of the ratio $d/d_{\rm th}$ for different values of the parameters n_p and $\sigma_{\rm dB}$. For a fixed $d/d_{\rm th}$ value, I_r increases with increasing noise in the measurements. As a result, threshold selections that are ineffective for large values $n_p/\sigma_{\rm dB}$ will produce better results when $n_p/\sigma_{\rm dB}$ decreases.

Fig. 10 further illustrates the effect of noisy measurements on the threshold. The plots shows Fisher information values computed when nodes are 5 m apart. The values $F_{\rm con3}$, $F_{\rm con5}$, and $F_{\rm con7}$ measure the information when using a $d_{\rm th}$ equal to 3, 5, and 7 m, respectively. The optimal threshold is $d_{\rm th} = 5$ m. In fact, $F_{\rm con5}$ is always greater than $F_{\rm con3}$ and $F_{\rm con7}$. However, when the ratio $n_p/\sigma_{\rm dB}$ decreases, the differences between different choices become negligible.

E. Comparison: To Range or Not to Range?

Comparison between $F_{\rm rss}$ and $F_{\rm con}$ shows that RSS measurements always carry greater information content than connectivity ones (see Fig. 10). However, this is only true as long as the nodes are within the *radio range* of each other.

The transceivers used in many wireless devices have a finite sensitivity $P_{\rm s}$ that limits their capability to receive messages. When the signal reaches the recipient with a power below $P_{\rm s}$, the communication is likely to fail. The loss of radio packets not only impedes data transfer, but also affects localization. If two nodes are unable to communicate, no RSS information will be collected, i.e., $F_{\rm rss} = 0$. A range-based approach such as the MLE will not be able to produce any position estimate. The effect of lost messages is less severe for connectivity-based localization. Since the transceiver will only receive messages with RSS $\geq P_{\rm s}$, the threshold selection problem is meaningful only for values $P_{\rm th} \geq P_{\rm s}$. According to the model (6), when nodes are unable to communicate, they can be implicity assumed to be disconnected. Therefore, even out-of-range nodes will produce valid $C = 0$ measurements with a positive information content $F_{\rm con} > 0$.

For the 1-D case studied in this section, simply observing that two nodes are out of range does not provide clues about the actual node locations. We only know that the two nodes are not close to each other. However, as shown in Sections IV–IX, taking into account the information $F_{\rm con} > 0$ from disconnected nodes will improve the results when estimating the locations of multiple nodes in a mesh network.

IV. Collaborative Network Localization

Many applications such as context-aware computing and *wireless sensor networks* (WSNs) exploit ad hoc networks of wireless devices. Typically, only a fraction of the nodes are located at known positions, while the remaining nodes must be localized using information collected within the network.

We refer to this scenario as *collaborative localization*. To compensate for the scarcity of anchors, every node attempts to take measurements with every other node in the network, even those units at unknown positions. This approach is also known as *multihop localization* because it supports localization of nodes placed several *hops* away from the anchors.

A. Cramér–Rao Bounds

Consider a 2-D network with n nodes at unknown locations and m anchors. Similarly to the previous case, localization can be cast as parameter estimation problem. The difference is that each nonanchor node i now contributes the two unknown parameters x_i and y_i. Let $\boldsymbol{\theta}$ be the vector containing all the unknown x and y node coordinates: $\boldsymbol{\theta} = [x_1, \ldots, x_n, y_1, \ldots, y_n]$.

The unknown positions are computed using measurements collected between pairs of nodes. Let $\mathbf{z} = \{z_{ij}\}$ be the set of average RSS values measured between any pair of nodes i and j. The values z_{ij} can be used to estimate the distance between nodes i and j, or quantized into connectivity data c_{ij}. If two nodes i and j are unable to communicate, they will be assumed disconnected (i.e., $c_{ij} = 0$).

The information to estimate the node positions is measured by a *Fisher information matrix* (FIM) with $2n \times 2n$ elements. For connectivity measurements, the FIM elements are

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

where $f_{\mathbf{C}}$ is the *joint probability function*

$$f_{\mathbf{C}}(\mathbf{c};\boldsymbol{\theta},d_{\text{th}}) = f(c_{11},\ldots,c_{mm};\boldsymbol{\theta},d_{\text{th}}). \tag{19}$$

Given the structure of $\boldsymbol{\theta}$, the FIM is partitioned in submatrices $\mathbf{F}_{xx}$, $\mathbf{F}_{xy}$, and $\mathbf{F}_{yy}$, each having $n \times n$ elements

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{xx} & \mathbf{F}_{xy} \\ \mathbf{F}_{xy}^t & \mathbf{F}_{yy} \end{bmatrix}. \tag{20}$$

Assuming independent measurements between different pairs of nodes, the element of each submatrix have an expression similar to the Fisher information for the 1-D case. For example, the $\mathbf{F}_{xx}$ submatrix has elements

$$[f_{xx}]_{ij} = \begin{cases} -F_{\text{con}}(d_{ij},d_{\text{th}}) \left(\frac{x_i - x_j}{{d_{ij}}^2}\right)^2, & (i \neq j) \\ \sum_{k=1}^{n+m} F_{\text{con}}(d_{ij},d_{\text{th}}) \left(\frac{x_i - x_j}{{d_{ij}}^2}\right)^2, & (i = j) \end{cases} \tag{21}$$

where $d_{ij} = \sqrt{(x_i - x_y)^2 (y_i - y_j)^2}$ is the Euclidean distance between nodes i and j. The F_{xx} elements for RSS measurements are similar, but the summation for the diagonal elements only includes nodes that are in the radio range of node i.

In the submatrices $\mathbf{F_{yy}}$ and $\mathbf{F_{xy}}$, the terms $(x_i - x_j)^2/d_{ij}^4$ are replaced by $(y_i - y_j)^2/d_{ij}^4$, and by $(x_i - x_j)(y_i - y_j)/d_{ij}^4$ in the case of $\mathbf{F_{yy}}$ and $\mathbf{F_{xy}}$, respectively. More details about how to compute the FIM are provided in the work of Patwari *et al.* [24] for RSS measurements, and Patwari and Hero III [25] for connectivity-based localization.

Anchor information contributes to the diagonal terms of each submatrix. At least three noncollinear anchors are needed for localization in 2-D. Failure to include sufficient anchor information will cause the FIM to be rank-deficient [19]. Although analysis of the CRB is still possible by using the Moore–Penrose pseudoinverse of the FIM [7], we assume that the FIM is always invertible.

The CRB for the multiparameter case states that the inverse of the FIM bounds the covariance matrix of any unbiased estimator T for $\boldsymbol{\theta}$

$$\text{Cov}\{T(\mathcal{M})\} \geq \frac{1}{\mathbf{F}} \tag{22}$$

where $\mathcal{M}$ contains the set of available measurement, RSS or connectivity, between pairs of nodes in the network.

The diagonal elements of $\mathbf{F^{-1}}$ are the lower bound for the variance on the node coordinates x_i and y_i. In the rest of our analysis, we will use CRB_{rss} and CRB_{conn} to denote the average value of the $2n$ coordinates' standard deviation

$$\text{CRB}_{\text{rss}} = \frac{1}{2n} \sum_{k=1}^{2n} \sqrt{\left[\mathbf{F}_{\text{rss}}^{-1}\right]_{kk}} \tag{23}$$

$$\text{CRB}_{\text{conn}} = \frac{1}{2n} \sum_{k=1}^{2n} \sqrt{\left[\mathbf{F}_{\text{conn}}^{-1}\right]_{kk}} \tag{24}$$

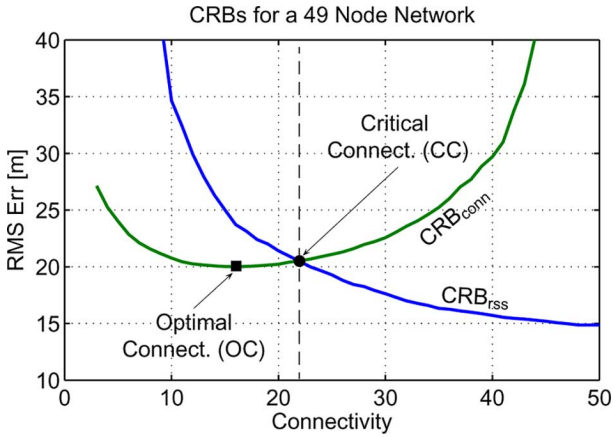

Fig. 11. CRBs for a 49-node network with four anchors on the corner of the deployment area. The critical connectivity (•) and optimal connectivity (■) values are marked on the plot.

where $\mathbf{F}_{\text{rss}}$ and $\mathbf{F}_{\text{conn}}$ are the FIMs computed with RSS and connectivity measurements.

B. CRB Analysis: RSS Measurements and Connectivity

Fig. 11 shows the CRB_{rss} and the CRB_{conn} computed for a 49-node network with four anchors on the corner of the deployment area. Different connectivity values were obtained by increasing the communication range of each node. A qualitative analysis of the two CRBs follows.

1) RSS Measurements: As shown in the previous section, each available measurements has an additive contribution to the FIM elements. A given network connectivity k, say $k = 10$, means that each node is in the radio range of 10 other nodes, hence 10 range estimates are available to compute its position. As the connectivity increases, the number of measurements increases, causing the CRB_{rss} to decrease monotonically (see Fig. 11).

Increasing the number of nodes in range becomes less beneficial for high connectivity values. Given the small amount of Fisher information obtained from distant nodes, including those range estimates improves the localization error only marginally.

2) Connectivity Measurements: Unlike the previous case, the CRB_{conn} does not decrease with the connectivity (see Fig. 11). According to the model discussed in Section II, the number of connected nodes is not related to the number of available measurements. The same connectivity value $k = 10$ means that the average RSS of 10 nodes are above the threshold P_{th}. The remaining nodes are considered disconnected, either because their signal strength is below the threshold or because they are out of range. Hence, in all cases, the number of measurements available is equal to the number of network nodes minus one.

Increasing the connectivity does not necessarily cause the CRB_{conn} to decrease. However, the connectivity depends on d_{th} and, similarly to the 1-D case, the threshold selection will affect the localization error as explained in Section V.

C. Effect of Threshold Selection

The convex shape of CRB_{conn} can be explained using analysis of the F_{con} similar to the 1-D case. We start by considering

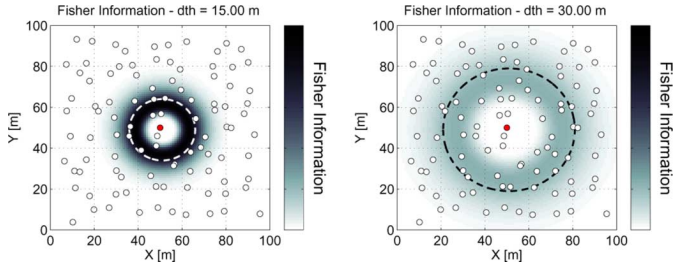

Fig. 12. Fisher information for two threshold values. The background color indicates the Fisher Information at different distances from the node in the center. Darker colors correspond to a higher information content. Nodes with distances close to the threshold contribute the most information.

the $\mathrm{CRB}_{\mathrm{conn}}$ at the boundaries of the connectivity range, and then we analyze intermediate values.

Fig. 11 shows that the $\mathrm{CRB}_{\mathrm{conn}}$ increases when the network connectivity approaches values at the extremes of the range considered. Connectivity measurements are equivalent to knowledge of the neighbor set of each node. In the extreme case of a network with connectivity equal to zero, all the neighbor sets will be empty. In a fully connected network, all the neighbor sets will contain every node. In both cases, localization will not produce meaningful results because when all the node have the same neighbor sets, no information is available to discriminate their positions.

The large error for extreme connectivity values can also be explained by analyzing the F_{con}. Only pairs of nodes with distance comparable to d_{th} contribute significant F_{con} values. When d_{th} is extremely small or extremely large, the total amount of information will be small because no pairs of nodes will have a distance similar to d_{th}. Any connectivity-based scheme will produce poor results simply because little information is available to localize the nodes.

For intermediate connectivity values, the choice of d_{th} determines which measurements are emphasized in the estimation process. Fig. 12 shows the Fisher information available to locate a node in the center of the network. The nodes are plotted against a background that shows the amount of information at a given distance. Comparing the plots for $d_{\mathrm{th}} = 15$ m and $d_{\mathrm{th}} = 30$ m shows that *the choice of d_{th} determines a tradeoff between obtaining high-quality measurements from a few nearby nodes or obtaining less valuable data for a larger number of nodes that are farther away*. Increasing d_{th} increases the number of nodes whose distance is similar to d_{th}. However, since $F_{\mathrm{con}} \propto 1/d^2$, these nodes contribute individually less information as the distance increases.

V. Critical and Optimal Connectivity Values

Fig. 11 indicates the existence of two important connectivity values. The CC is the point where the two CRB lines cross. The OC is the connectivity value that minimizes the $\mathrm{CRB}_{\mathrm{conn}}$.

For a system designer, knowing the CC and OC values is important to solve both the MSP and the OCP.

1) *Critical Connectivity:* Assume a network with average node connectivity k. Comparing k against CC determines which localization approach should be used. A connectivity-based scheme is preferable for $k < \mathrm{CC}$, while a range-based scheme will be potentially more effective for $k \geq \mathrm{CC}$. According to the CRB analysis, this choice minimizes the localization error.
2) *Optimal Connectivity:* If the same network has to be localized using a connectivity-based scheme, the error can be reduced by adjusting the threshold P_{th} until a connectivity $k = \mathrm{OC}$ is reached. Setting such a threshold ensures a condition of maximum information content and avoids the large error of connectivity-based schemes in highly connected networks.

For a given network topology, the CC and OC values are found by computing the $\mathrm{CRB}_{\mathrm{rss}}$ and the $\mathrm{CRB}_{\mathrm{conn}}$ as a function of the network connectivity. Unfortunately, this approach cannot be used in real applications because evaluating the CRB requires knowledge of the propagation model's parameters and, most importantly, the unknown node positions.

Sections VI–IX will investigate some of the properties of the CRB. The goal is to gain a better understanding of the CC and OC values and find solutions to approximate these values for use in practical applications.

VI. Properties of the CC and OC Values

In previous work, the CRB for range-based localization was shown to be invariant under global translation, rotation, or reflection of the network [7]. The extra terms $I_R(\cdot,\cdot)$ in the $\mathrm{CRB}_{\mathrm{conn}}$ only depends on the ratio d/d_{th}, therefore the same properties hold for connectivity-based localization.

This section investigates the effect of other application parameters on the CC and OC values. The parameters considered are: 1) the number of network nodes; 2) the ratio n_p/σ_{dB}; 3) the scaling factor for the node coordinates; and 4) the number of anchor nodes.

For the CC value, we are interested in conditions that alter the difference between the $\mathrm{CRB}_{\mathrm{rss}}$ and the $\mathrm{CRB}_{\mathrm{conn}}$. When the relative position of the two CRBs changes, the CC value will also change. For the OC value, the analysis will focus on conditions that alter the position of the minimum of the $\mathrm{CRB}_{\mathrm{conn}}$. For a given network, the optimal connectivity is achieved by setting a threshold P^*_{th}, which in turn corresponds to distance threshold d^*_{th}. Studying the properties of the OC value is equivalent to investigating changes in the d^*_{th} value for different application parameters.

A. Number of Network Nodes

Both the CC and the OC values increase with the number of nodes in the network (see Fig. 13).

1) Critical Connectivity: Assume a network with average connectivity k. Further assume that the number of nodes is increased without changing k. If the deployment area is fixed, this is possible only if the communication range of each node and the RSS threshold are reduced. In this new network, a lower $\mathrm{CRB}_{\mathrm{rss}}$ is expected because now the same number of measurements are available from nodes that are closer.[1]

The $\mathrm{CRB}_{\mathrm{conn}}$ decreases even more noticeably than the $\mathrm{CRB}_{\mathrm{rss}}$. In fact, for a connectivity-based scheme, the error reduces not only because nodes are closer to each other, but also because there are more measurements available from the

[1]More in general, a proof by Patwari *et al.* [24] gives sufficient conditions for a decreasing CRB when new nodes are added to the network.

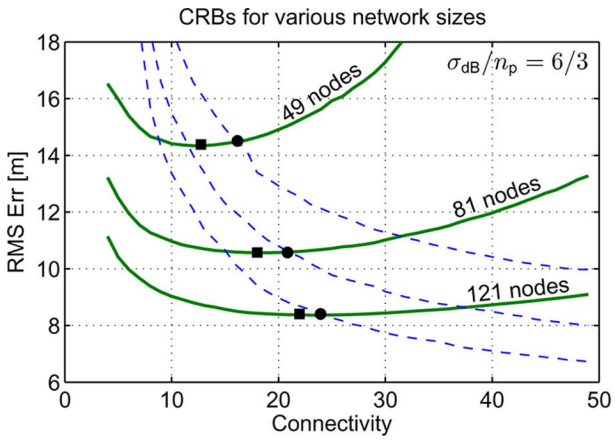

Fig. 13. Critical connectivity (•) and optimal connectivity (■) for different network sizes.

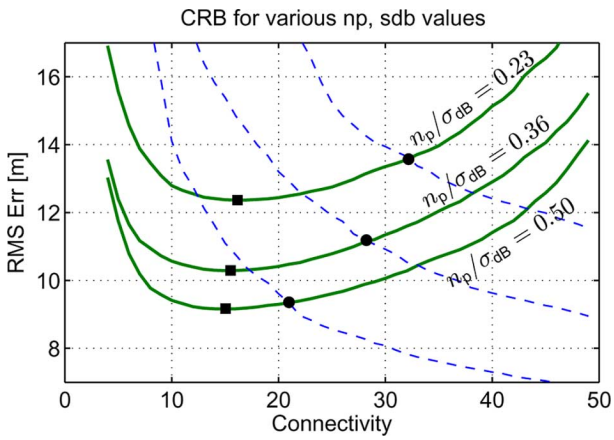

Fig. 14. Critical connectivity (•) and optimal connectivity (■) values for a 64-node network with different values of the ratio n_p/σ_{dB}.

disconnected nodes. Fig. 13 shows the two CRBs computed for networks with 49, 81, and 121 nodes deployed in a square region with each side equal to 100 m. While both CRBs decrease, the reduction is more evident for the $\mathrm{CRB}_{\mathrm{conn}}$. As a result, the CC value increases with the network size.

2) Optimal Connectivity: Fig. 13 also shows that OC increases with the network size. Consider again the example in Fig. 12, and assume that the threshold $d_{\mathrm{th}} = 30$ m marked on the bottom plot is optimal, i.e., $d^*_{\mathrm{th}} = 30$ m. This threshold maximizes the available Fisher information from neighboring nodes. If the network size is increased by adding nodes at distances $d \geq d_{\mathrm{th}}$, i.e., without affecting the current connectivity level, the current threshold will no longer be optimal. Since this new topology has more nodes at $d \geq d_{\mathrm{th}}$, the available information can be increased by using a larger threshold value, thus increasing the connectivity. An informal proof is presented in the Appendix.

B. Propagation Model Parameters n_p/σ_{dB}

1) Critical Connectivity: The CC value also increases when the ratio n_p/σ_{dB} decreases. As discussed in Section III, this term describes the quality of the RSS measurements. As the noise increases, larger localization errors are expected. In fact, both the $\mathrm{CRB}_{\mathrm{rss}}$ and the $\mathrm{CRB}_{\mathrm{conn}}$ increase as shown in Fig. 14. However, for connectivity measurements, parts of the losses are compensated by the term $I_r(\cdot,\cdot)$, which was shown to increase with larger noise levels (see Section III-D). Since noise has a less severe impact on connectivity-based schemes, the CC increases for lower values of n_p/σ_{dB} [see Fig. 14(a)].

2) Optimal Connectivity: Unlike the CC, the OC is not sensitive to the ratio n_p/σ_{dB}. The Appendix provides an informal proof based on an analysis similar to the one used in the previous section. The argument used is that changes in the ratio n_p/σ_{dB} will not alter the optimality of the d_{th} value. Fig. 14 empirically supports the same argument by showing that the OC values are not sensitive to the ratio n_p/σ_{dB}.

C. Coordinate Scaling

Both the CC and OC values do not change when the network coordinates are scaled by a factor S. This property follows from the equations that describe the Fisher information for RSS and

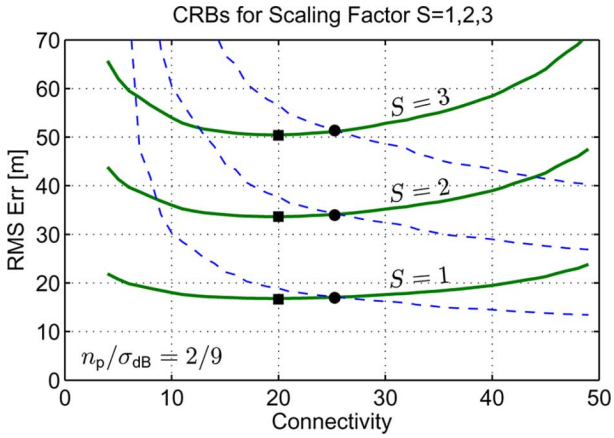

Fig. 15. Critical connectivity (•) and optimal connectivity (■) values for a 64-node network with coordinates scaled by a factor $S = \{1, 2, 3\}$.

connectivity measurements. Consider the terms F_{rss} and F_{con} discussed in Section III, and assume that all the node distances are multiplied by a factor S. Also assume that d_{th} is scaled by the same factor, so the network connectivity remains constant. Under these conditions, both the F_{rss} and the F_{con} terms will be scaled by a factor S^{-2}

$$\begin{aligned} F_{\mathrm{rss}}(Sd) &= S^{-2} F_{\mathrm{rss}}(d) \\ F_{\mathrm{con}}(Sd, Sd_{\mathrm{th}}) &= S^{-2} F_{\mathrm{con}}(d, d_{\mathrm{th}}). \end{aligned} \tag{25}$$

When considering the multiparameter case, scaling the network coordinates has the same effect on the FIM elements for RSS and connectivity measurements. Since both matrices will be multiplied by the same constant factor S^{-2}, the relative positions of the two CRB and the minimum of the $\mathrm{CRB}_{\mathrm{conn}}$ will not change (see Fig. 15).

D. Number of Anchor Nodes

Our simulations also show that increasing the number of anchor nodes causes both the CRBs to decrease, but without significantly affecting the position of CC and OC. As discussed in Section IV, anchor information contributes to the diagonal elements of the FIM. While a larger number of anchors will lower the error, our simulations did not show a significant correlation between the number of anchors and the CC and OC values.

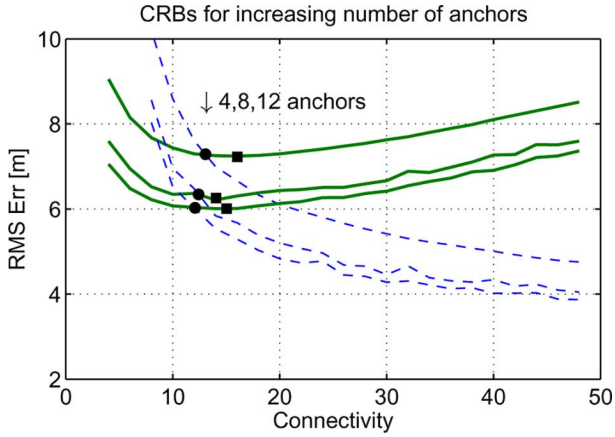

Fig. 16. Critical connectivity (•) and optimal connectivity (■) values for a 100-node network with coordinates 4, 8, and 12 anchors.

TABLE II
EFFECT OF PARAMETERS ON CC AND OC VALUES

Parameter	Effect on CC	Effect on OC
N. of Nodes	YES	YES
Ratio n_p/σ_{dB}	YES	Negligible
Scaling Factor S	NO	NO
No. of Anchors	Negligible	Negligible

Fig. 16 shows the CRBs for a 100-node topology with 4, 8, and 12 anchors.

VII. SIMULATIONS AND APPROXIMATIONS

The effect of the application parameters considered in Section VI is summarized in Table II. This section uses extensive simulations to model the dependence of CC and OC values on the parameters that affect their values.

We generated more than 500 topologies with a number of nodes between 20 and 400 nodes randomly deployed in a square region measuring 50×50 m^2. Four nodes on the corners of the network were used as anchors. For each topology, the parameters n_p and σ_{dB} were sampled from the intervals $[2, 4]$ and $[3, 9]$ dBm respectively, resulting in values of the ratio n_p/σ_{dB} between and 0.22 and 1.33 dBm^{-1}.

A. Critical Connectivity

Fig. 17 shows the simulation results for the CC values. The values, which are plotted against the number of nodes and the ratio n_p/σ_{dB}, appear to lie on a smooth surface. We interpolate the CC values using a function that is found empirically

$$\tilde{\mathrm{CC}}(n, r) = a_0 + a_1 n + a_2 r + a_3 nr + + a_4 \log n + a_5 \exp(-r) \quad (26)$$

where n is the number of nodes and r is the value of $(n_p/\sigma_{\mathrm{dB}})^{-1}$. The values of the coefficient a_i, obtained by least squares fitting, are the following: $a_0 = -37.1022$, $a_1 = -0.0732$, $a_2 = 8.8506$, $a_3 = 0.0377$, $a_4 = 6.0667$, $a_5 = 41.8567$. The mean squared error between $\tilde{\mathrm{CC}}(n, r)$ and the data points is equal to 6.18, while the average error is equal to 1.88. We find this error sufficiently small for practical application of (26) in approximating the CC value.

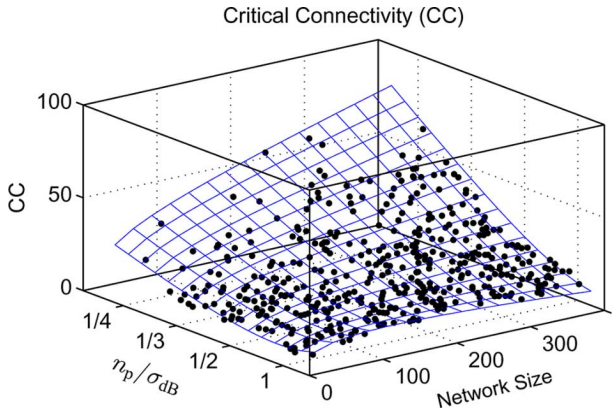

Fig. 17. Simulation results and approximation of the CC values.

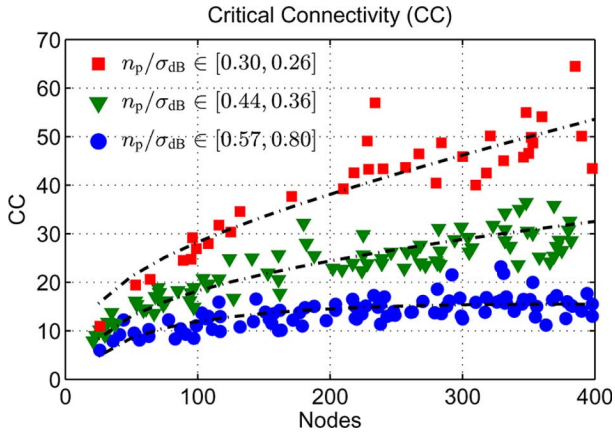

Fig. 18. Approximation of the CC values in selected n_p/σ_{dB} ranges.

Fig. 18 shows the CC values for different intervals of the ratio n_p/σ_{dB}. The dotted lines are computed using (26) for n_p/σ_{dB} equal to the central value of ranges considered. For higher n_p/σ_{dB} values, the CC value stabilizes around 15. As the ratio n_p/σ_{dB} decreases, however, there is a stronger correlation between the network size and CC values. Therefore, range-based schemes are beneficial only in highly connected networks. These results confirm the observations of other authors, who have occasionally noted that connectivity-based schemes outperform range-based ones in conditions of low connectivity [8] or when the ranges are estimated using noisy measurements [3], [22].

B. Optimal Connectivity

The simulation results for the OC values are shown in Fig. 19. The values only depend on the network size and can be interpolated with a function of $\sqrt{n}$

$$\tilde{\mathrm{OC}}(N) = -3.8290 + 2.3922\sqrt{n} \quad (27)$$

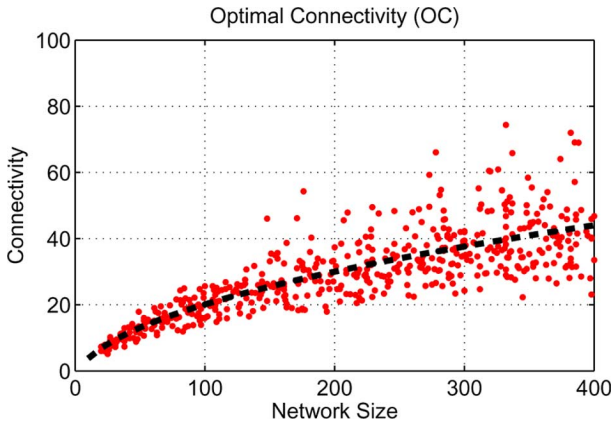

Fig. 19. Simulation results and approximation of the OC values.

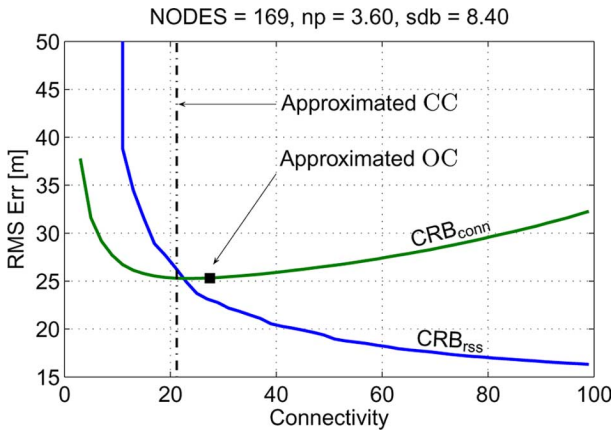

Fig. 20. Approximation of the CC and OC values for a 196-node network deployed in a square region with side 450 m and $n_p/\sigma_{\mathrm{dB}} = 3.6/8.4$. The approximated CC value is close to the intersection of the two CRBs, and the approximated OC value is close to the absolute minimum of $\mathrm{CRB}_{\mathrm{conn}}$.

where n is the network size. The equation coefficients were determined by least squares fitting; the average error between $\tilde{\mathrm{OC}}(n)$ and the data points is equal to 5.22.

C. Example

Assume one has to localize a sensor network with 196 devices deployed in a square area with each side 450 m. We generated such a topology by computing the node positions using a *noisy grid model* and sampling the RSS values from a log-normal distribution with $n_p = 3.6$, $\sigma_{\mathrm{dB}} = 8.4$ dB. These values are based on measurements collected in a tall grassy field with sparse bushes [32].

If the parameters of the propagation model are unknown at run-time, the system designer will use a connectivity-based scheme. The optimal connectivity value approximated by (27) is $\tilde{\mathrm{OC}}(196) = 27.26$. If the actual network connectivity is larger than 27.26, it can be reduced by applying a threshold P_{th} to the RSS measurements. If the connectivity is significantly lower than this value, the localization result might be improved by adopting transceiver with increased range.

If the parameters n_p and σ_{dB} were estimated beforehand, the other option is to use a range-based approach. The CC value approximated using (26) is $\tilde{\mathrm{CC}}(196, 8.4/3.6) = 21.22$. Therefore, the use of an RSS ranging solution will be effective only for network connectivity above 21.22.

Fig. 20 validates the results of the choice based on approximated equations (26) and (27) by showing the two CRBs computed for the test topology considered. Recall that computing the CRB is possible only because this is a simulated scenario and the node positions are known. The plots show that both the approximated values are close to the CC and OC determined by CRB analysis.

VIII. Related Work

This paper uses CRB analysis to solve the MSP and the OCP defined in the Introduction. The CRB for collaborative localization has been previously investigated by a number of authors, and some other works have reported empirical comparison between range-based and connectivity-based schemes.

A. Study of the Localization Bounds

Analysis of the CRB has been previously used to characterize the error bound of localization algorithms, especially for ranging approaches (angle or distances) affected by Gaussian noise. Moses *et al.* [19] have derived the CRB for localization based on signals emitted by a set of sources and nodes capable of measuring the time of arrival (ToA) or the angle of arrival (AoA). A study of the CRB for distance-based localization under various topology conditions has been proposed by Savvides *et al.* [1]. The results for various levels of measurement noise and different nodes and beacon density are qualitatively similar to the $\mathrm{CRB}_{\mathrm{rss}}$ plots shown in Section VI.

Wang *et al.* have defined a Bayesian bound (BB) that is the covariance of a posterior distribution computed from the sensor observation [34]. This bound is equivalent to the CRB for measurements with Gaussian error, but it is computationally less demanding. Analysis of the CRB has been proposed by Patwari *et al.* for the case of collaborative localization using distance estimates obtained by ToA and RSS [24] and for localization using angle estimates [23]. Localization using connectivity information or quantized RSS levels has been studied by Patwari and Hero III [25].

B. Heuristic Solutions for the MSP and the OCP

Our work has defined two fundamental problems that are relevant to practical implementation of RF-based positioning schemes. In the past, some authors have occasionally discussed the effect of connectivity on the localization error. For example, it has been noticed that range-based schemes perform well only with connectivity level of 10 or more [8], [18], [20]. Other authors have noted that connectivity-based localization can outperform range-based localization in sparse networks [8], [10], [33] or when the measurements are noisy [3], [22]. Such observations concord with our results presented here.

An approach to formalize the conversion of RSS values into connectivity data has been previously proposed by Li *et al.* [17]. Their work implements a *Partial range information* (PRI)

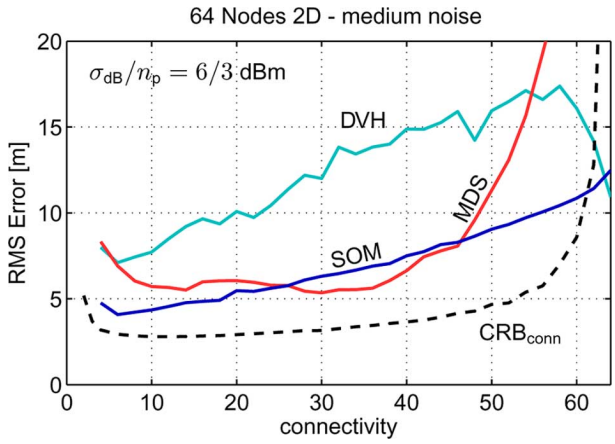

Fig. 21. RMS error achieved by three different connectivity-based localization schemes on a 64-node network with four anchor nodes. The schemes considered are SOM [10], MDS [30], and DV-HOP [21].

scheme that derives "sub-hop" information used to improve localization. The motivation for our approach is similar to that of the PRI. In fact, we also try to improve the localization accuracy by choosing a proper quantization threshold for the RSS values. Another scheme combining connectivity and RSS measurements is the Weighted Centroid [31].

IX. Conclusion

This paper has defined and proposed solutions for the MSP and the OCP, two important problems that have not been formally investigated in the previous literature. Using our results, a system engineer can determine if a localization scenario is better addressed using a range-based or a connectivity-based scheme. If the choice favors a connectivity-based scheme, the localization error can be reduced by setting an opportune quantization threshold on the RSS values. Since the results are expressed in terms of "optimal connectivity," this approach will also aid those applications with devices unable to accurately report the RSS (e.g., Bluetooth or RFID). In these applications, the units will have to be deployed ensuring an average connectivity approximately equal to OC.

Analysis of the $\mathrm{CRB}_{\mathrm{conn}}$ has shown that enough information is available to localize even networks with low connectivity, e.g., $k \leq 10$. In this case, results close to the bound can be achieved by designing connectivity-based schemes capable of taking into account the information associated to all the disconnected node pairs ($c_{ij} = 0$). For example, Fig. 21 shows a case where the SOM algorithm [10] achieves results close to the CRB for low connectivity values. On the other hand, schemes that only take into account connected nodes ($c_{ij} = 1$) will perform poorly for sparsely connected networks. How to efficiently use information from both connected and disconnected nodes is an interesting research direction.

Additional improvements are possible by combining range and connectivity information. For example, range-based schemes can impose constraints on the minimum separation distance between disconnected nodes (e.g., [16]). Similarly, connectivity-based schemes can use RSS information to "sort" one-hop neighbors [17]. In both cases, using additional information will cause the localization error to decrease, and the values computed using (26) and (27) will not necessarily correspond to the true CC and OC values. Analysis of these cases will be the focus of our future research work.

Finally, our conclusions are based on analysis of the CRB, which refer to unbiased schemes that efficiently use all the available information. In practice, solutions will often produce biased results and will not be efficient estimators. Fig. 21 shows how the RMS error for three connectivity-based schemes can be substantially different from the CRB. Given these differences, not every connectivity-based scheme will perform better than every range-based scheme for connectivity values below CC. Similarly, the best results might be achieved for a connectivity different from OC. However, study of the CRB allows one to identify conditions in which the measurements provide maximum information content.

The results in this paper can benefit both practitioners and theoreticians in the localization field. Specifically, readers interested in practical deployments will be able to use some of our findings to improve the performance of their systems or, at least, to avoid poor results due to choices that yield measurements with little information. On the other hand, readers interested in theoretical analysis should be able to use this work as a starting point for future research that attempts to analytically compare broad classes of localization schemes.

Appendix
Some Properties of the OC Value

This section provides an informal proof to explain the dependence of the OC value on the number of network nodes and the ratio n_p/σ_{dB}. Before discussing such properties, we remark that the CRB depends on the node distances and the geometrical configuration of the network nodes. However, when nodes have an approximately uniform distribution, the properties of the CRB can be understood using a qualitative analysis similar to the one used in Section III for the 1-D case.

To understand the effect of different numbers of nodes on the OC value, consider the information available to estimate the position of a generic node in the network. For simplicity, we will consider a node in the center of the network, similar to the case in Fig. 12. Assuming independent RSS measurements, the available information is given by summation of the nodes' contribution at different distances

$$F_{\mathrm{tot}}(d_{\mathrm{th}}) = \sum_i F_{\mathrm{con}}^{(i)}(d_i, d_{\mathrm{th}}) = \sum_i \frac{{K_c}^2}{d_i^2} I_r(d_i, d_{\mathrm{th}}) \quad (28)$$

where d_i is the distance of the ith neighbors. The optimal threshold d_{th}^* is the value that maximizes the information F_{tot}. Since the $F_{\mathrm{con}}^{(i)}(d_i, d_{\mathrm{th}})$ terms are nonnegative, and their value reduces to zero when $d \ll d_{\mathrm{th}}$ or $d \gg d_{\mathrm{th}}$, the optimal threshold is found by equating the derivative to zero

$$\frac{\partial F_{\mathrm{tot}}}{\partial d_{\mathrm{th}}}(d_{\mathrm{th}}^*) = \sum_i \frac{\partial {F_{\mathrm{con}}}^{(i)}}{\partial d_{\mathrm{th}}}(d_i, d_{\mathrm{th}}^*) = 0. \quad (29)$$

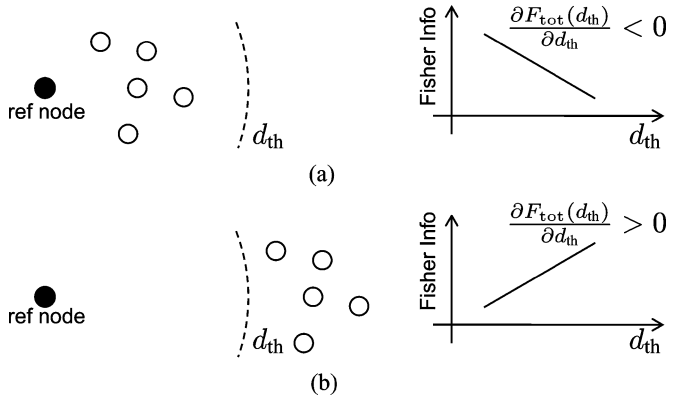

Fig. 22. Derivative of the Fisher information in two selected cases. (a) All the neighbors are at a distance less than the threshold $d_{\rm th}$. (b) All the neighbors are at a distance greater than the threshold $d_{\rm th}$.

Differentiation of the terms I_r's in (28) yields a complicated expression, but the results can be simplified by considering an approximated form for I_r

$$\tilde{I}_r(d, d_{\rm th}) \approx \frac{2}{\pi} \exp\left(-\frac{\log\left(d/d_{\rm th}\right)^2}{K_a \left(\frac{n_p}{\sigma_{\rm dB}}\right)^2} \right) \tag{30}$$

where $K_a \approx 0.13$ is a constant that was numerically determined using least square fitting. The intuition for using the above approximation is that the I_r term closely resembles a Gaussian kernel when plotted on a logarithmic scale (see Fig. 9). When the terms $\tilde{I}_r$ are used in place of I_r, the terms in (29) have a more tractable expression

$$\frac{\partial F_{\rm con}^{(i)}}{\partial d_{\rm th}}(d_i, d_{\rm th}) = \frac{2{K_c}^2}{0.13(n_p/\sigma_{\rm dB})^2} \frac{1}{d_{\rm th} d_i^2} \tilde{I}_r(d_i, d_{\rm th}) \log\left(\frac{d_i}{d_{\rm th}}\right). \tag{31}$$

The sign of each derivative only depends on d_i and $d_{\rm th}$

$$\operatorname{sign}\left(\frac{\partial F_{\rm con}^{(i)}}{\partial d_{\rm th}}\right) = \operatorname{sign}\log\left(\frac{d_i}{d_{\rm th}}\right) = \begin{cases} -1 & \text{if } d_i < d_{\rm th} \\ 0 & \text{if } d_i = d_{\rm th} \\ +1 & \text{if } d_i > d_{\rm th}. \end{cases} \tag{32}$$

This result concords with the intuitive notion of the optimal threshold built so far. If all the neighbors are at distances less than $d_{\rm th}$, the derivative will be negative [see Fig. 22(a)]. The $F_{\rm tot}$ value can be increased by reducing $d_{\rm th}$, i.e., moving it closer to the neighbors. If all the neighbors are at distances greater than the threshold, the derivative of $F_{\rm tot}$ will be positive. To obtain more information, $d_{\rm th}$ needs to be increased.

Consider now a network in which the threshold is optimal, i.e., $\partial F_{\rm tot}/\partial d_{\rm th} = 0$. If nodes are added at a distance greater than $d_{\rm th}$ (i.e., without increasing the current connectivity level), the contribution of the new units will cause the derivative to become positive, thus violating the condition of optimality. To bring the derivative to zero, some nodes at a distance less than $d_{\rm th}$ also need to be introduced, causing the optimal connectivity to increase. In conclusion, as shown in Fig. 13, the OC value will increase for increasing values of the network size.

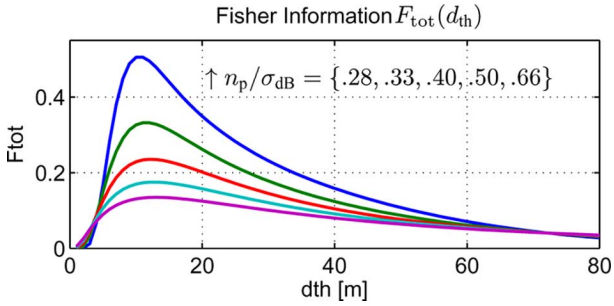

Fig. 23. Value of the function $F_{\rm tot}(d_{\rm th})$ computed for a node in the center of a 64-node topology and increasing values of the ratio $n_p/\sigma_{\rm dB}$.

Analysis of the terms (31) also serves to understand why the OC value is not sensibly affected by changes in $n_p/\sigma_{\rm dB}$. Consider $p > 1$ and nodes placed at distance $d_1 = d_{\rm th}/p$ and $d_2 = d_{\rm th}p$. By replacing these values in (31), we obtain

$$\frac{\partial F_{\rm con}^{(i)}}{\partial d_{\rm th}}\left(\frac{d_{\rm th}}{p}, d_{\rm th}\right) = -p^4 \frac{\partial F_{\rm con}^{(i)}}{\partial d_{\rm th}}(p d_{\rm th}, d_{\rm th}). \tag{33}$$

For example, if $p = 2$, the presence of a node at distance $d_1 = d_{\rm th}/2$ can be balanced by placing $2^4 = 16$ nodes at distance $d_2 = 2d_{\rm th}$. The contribution $\partial {F_{\rm con}}^{(i)}/\partial d_{\rm th}$ of a node at distance $d_1 = d_{\rm th}/3$ can be balanced by placing $3^4 = 81$ nodes at distance $d_2 = 3d_{\rm th}$, and so on. Note that if nodes were placed according to the rule above, the OC would be exactly the same independently from $n_p/\sigma_{\rm dB}$.

In a typical ad hoc network, it is unlikely that the node distances will follow the distribution described. Depending on the value of the ratio $d/d_{\rm th}$, altering $n_p/\sigma_{\rm dB}$ will cause some of the terms $\partial {F_{\rm con}}^{(i)}/\partial d_{\rm th}$ to grow more than others, possibly causing $\partial F_{\rm tot}/\partial d_{\rm th}$ to become different than zero. However, given the symmetry[2] of the terms $\tilde{I}_r$ around the value $d = d_{\rm th}$, variations of $n_p/\sigma_{\rm dB}$ in the typical range measured in wireless applications seems not to significantly alter the position of the OC value. Fig. 23 provides support to this evidence by reporting the $F_{\rm tot}(d_{\rm th})$ values for a node in the center of a 64-node random deployment. Different values of the ratio $n_p/\sigma_{\rm dB}$ do not significantly alter the position of the $F_{\rm tot}$ maxima.

REFERENCES

[1] A. Savvides, W. Garber, S. Adlakha, R. Moses, and M. Srivastava, "On the error characteristics of multihop node localization in ad-hoc sensor networks," in *Proc. ACM/IEEE IPSN*, Palo Alto, CA, 2003, pp. 317–332.

[2] C. Bettstetter and C. Hartmann, "Connectivity of wireless multihop networks in a shadow fading environment," *Wireless Netw.*, vol. 11, no. 5, pp. 571–579, 2005.

[3] J. Beutel, *Location Management in Wireless Sensor Networks. Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*. Boca Raton, FL: CRC Press, 2005.

[4] P. Biswas, T. Lian, T. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *Trans. Sensor Netw.*, vol. 2, no. 2, pp. 188–220, 2006.

[5] J. Boccuzzi, *Signal Processing for Wireless Communications*. New York: McGraw-Hill, 2008.

[6] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low-cost outdoor localization for very small devices," *IEEE Pers. Commun.*, vol. 7, no. 5, pp. 28–34, Oct. 2000.

[7] C. Chang and A. Sahai, "Cramér–Rao-type bounds for localization," *J. Appl. Signal Process.*, vol. 2006, pp. 166–166, 2006.

[2]Symmetry is intended in the sense that $\tilde{I}R(d_{\rm th}/p, d_{\rm th}) = \tilde{I}_r(d_{\rm th}p, d_{\rm th})$.

[8] K. Chintalapudi, A. Dhariwal, R. Govindan, and G. Sukhatme, "Ad-hoc localization using ranging and sectoring," in *Proc. IEEE INFOCOM*, 2004, vol. 4, pp. 2662–2672.
[9] L. Doherty, K. pister, and L. E. Ghaoui, "Convex position estimation in wireless sensor networks," in *Proc. IEEE INFOCOM*, 2001, vol. 3, pp. 1655–1663.
[10] G. Giorgetti, S. K. S. Gupta, and G. Manes, "Wireless localization using self-organizing maps," in *Proc. ACM/IEEE IPSN*, 2007, pp. 293–302.
[11] G. Giorgetti, S. K. S. Gupta, and G. Manes, "Localization using signal strength: To range or not to range?," in *Proc. 1st ACM MELT*, 2008, pp. 91–96.
[12] G. Giorgetti, S. K. S. Gupta, and G. Manes, "Optimal RSS threshold selection in connectivity-based localization schemes," in *Proc. 11th MSWiM*, New York, NY, 2008, pp. 220–228.
[13] G. Giorgetti, "RF-based localization in GPS-denied applications," Ph.D. dissertation, Dept. Elect. Eng., Arizona State Univ., Tempe, AZ, 2009.
[14] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[15] R. Hekmat and P. Van Mieghem, "Connectivity in wireless ad-hoc networks with a log-normal radio model," *Mobile Netw. Appl.*, vol. 11, no. 3, pp. 351–360, 2006.
[16] Y. Kwon, K. Mechitov, S. Sundresh, W. Kim, and G. Agha, "Resilient localization for sensor networks in outdoor environments," in *Proc. 25th IEEE ICDCS*, 2005, pp. 643–652.
[17] X. Li, H. Shi, and Y. Shang, "A partial-range-aware localization algorithm for ad-hoc wireless sensor networks," in *Proc. 29th IEEE Int. Conf. Local Comput. Netw.*, 2004, pp. 77–83.
[18] D. Moore, J. Leonard, D. Rus, and S. Teller, "Robust distributed network localization with noisy range measurements," in *Proc. 2nd SenSys*, 2004, pp. 50–61.
[19] R. Moses, D. Krishnamurthy, and R. Patterson, "A self-localization method for wireless sensor networks," *J. Appl. Signal Process.*, vol. 2003, no. 4, pp. 348–358, 2003.
[20] R. Nagpal, H. Shrobe, and J. Bachrach, "Organizing a global coordinate system from local information on an ad hoc sensor network," in *Proc. ACM/IEEE IPSN*, 2003, pp. 333–348.
[21] D. Niculescu and B. Nath, "Ad hoc positioning system (APS) using AOA," in *Proc. 22nd IEEE INFOCOM*, 2003, vol. 3, pp. 1734–1743.
[22] D. Niculescu and B. Nath, "Error characteristics of ad hoc positioning systems (APS)," in *Proc. 5th ACM MobiHoc*, 2004, pp. 20–30.
[23] N. Patwari, J. Ash, S. Kyperountas, A. Hero, III, R. Moses, and N. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.
[24] N. Patwari, A. Hero, M. Perkins, N. Correal, and R. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.
[25] N. Patwari and A. Hero, III, "Using proximity and quantized RSS for sensor localization in wireless networks," in *Proc. 2nd ACM Int. Conf. Wireless Sensor Netw. Appl.*, 2003, pp. 20–29.
[26] N. Priyantha, H. Balakrishnan, E. Demaine, and S. Teller, "Anchor-free distributed localization in sensor networks," in *Proc. 1st SenSys*, 2003, pp. 340–341.
[27] T. Rappaport, *Wireless Communications: Principles and Practice*. Piscataway, NJ: IEEE Press, 1996.
[28] A. Savvides, W. Garber, S. Adlakha, R. Moses, and M. Srivastava, "On the error characteristics of multihop node localization in ad-hoc sensor networks," in *Proc. ACM/IEEE IPSN*, 2003, pp. 317–332.
[29] A. Savvides, C. Han, and M. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proc. 7th MobiCom*, 2001, pp. 166–179.
[30] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from mere connectivity," in *Proc. 4th ACM MobiHoc*, 2003, pp. 201–212.
[31] X. Shen, Z. Wang, P. Jiang, R. Lin, and Y. Sun, "Connectivity and RSSI based localization scheme for wireless sensor networks," *Lecture Notes in Computer Science*, vol. 3645, no. 578, pp. 578–587, 2005.
[32] K. Sohrabi, B. Manriquez, and G. Pottie, "Near ground wideband channel measurement in 800–1000 MHz," in *Proc. 49th IEEE Veh. Technol. Conf.*, 1999, vol. 1, pp. 571–574.
[33] R. Stoleru and J. Stankovic, "Probability grid: A location estimation scheme for wireless sensor networks," in *Proc. 1st IEEE SECON*, 2004, pp. 430–438.
[34] H. Wang, L. Yip, K. Yao, and D. Estrin, "Lower bounds of localization uncertainty in sensor networks," in *Proc. IEEE ICASSP*, 2004, vol. 3, pp. 917–920.

Gianni Giorgetti (M'10) received the Laurea degree in computer science from the University of Florence, Florence, Italy, in 2003, and the Ph.D. degree in electrical engineering under a joint program between the University of Florence and Arizona State University, Tempe, in 2007 and 2009, respectively.

He has held positions with Motorola Laboratories, Qualcomm, Inc., and other companies. He is currently working on indoor positioning systems with Navizon, Inc., Miami Beach, FL. His research interests include low-power sensor networks, remote monitoring, mobile computing, RF-based collaborative positioning, and angle of arrival (AOA) estimation using directional antennas.

Sandeep Kumar S. Gupta (S'93–M'95–SM'00) received the B.Tech. degree in computer science and engineering (CSE) from the Institute of Technology, Banaras Hindu University, Varanasi, India, the M.Tech. degree in CSE from the Indian Institute of Technology, Kanpur, India, and the M.S. and Ph.D. degrees in computer and information science from the Ohio State University, Columbus.

He is a Professor with the School of Computing and Informatics, Arizona State University, Tempe. He heads the Intelligent Mobile and Pervasive Applications and Computing Technologies (IMPACT) Laboratory, Arizona State University. He has coauthored the book *Fundamentals of Mobile and Pervasive Computing* (McGraw-Hill, 2004). His current research focuses on dependable, criticality aware, adaptive distributed systems with emphasis on wireless sensor networks, thermal and power-aware computing and communication, and pervasive healthcare. For information about his recent research projects and publications, please visit http://impact.asu.edu.

Dr. Gupta is a member of the Association for Computing Machinery (ACM). He is currently on the Editorial Board of the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE COMMUNICATION LETTERS, *Sustainable Computing*, and *Wireless Networks*. His research awards include a Best 2009 SCIDSE Senior Researcher and a Best Paper Award for "Security for Pervasive Health Monitoring Application."

Gianfranco Manes (M'01–SM'02) is currently a Full Professor with the Electronics and Telecommunication Department, University of Florence, Florence, Italy. He is active in the field of microwave engineering and wireless technology, including wireless sensor networks. He is the founder and Head of the MIDRA Consortium, a joint venture between the University of Florence and Motorola SpA, and the Head of the Research Centre for the ICT for Food Quality and Safety, taking scientific responsibility for leading both national and international research projects. He holds key positions in major European Union (EU) projects of the six FWs, namely, the Network of Excellence TARGET and the Integrated Project GoodFood. He has authored or coauthored over 150 paper published in books, society journals, and referenced international conferences.

Prof. Manes has been a Technical Program Committee (TPC) member and session chairman from 2002 to 2005 for the IEEE Microwave Theory and Techniques Society (IEEE MTT-S) International Microwave Symposium (IMS). He is currently an IEEE MTT-S transactions reviewer, a European Microwave Week reviewer, and a TPC member of the IEEE Radio and Wireless Symposium.