Abstract—Localization is essential for wireless sensor networks (WSNs). It is to determine the positions of sensor nodes based on incomplete mutual distance measurements. In this paper, to measure the accuracy of localization algorithms, a ranging error model for time of arrival (TOA) estimation is given, and the Cramer-Rao Bound (CRB) for the model is derived. Then an algorithm is proposed to deal with the case where 1) ranging error accumulation exists, and 2) some anchor nodes broadcast inaccurate/wrong location information. Specifically, we first present a ranging error-tolerable topology reconstruction method without knowledge of anchor node locations. Then we propose a method to detect anchor nodes that have inaccurate/wrong position information. Simulations demonstrate the improvement of our algorithm compared to others.

Keywords – Wireless sensor networks, localization, Cramer-Rao Bound, topology reconstruction.

I. INTRODUCTION

Localization is essential for wireless sensor networks (WSNs). It is to determine the positions of sensor nodes based on incomplete mutual distance measurements. Localization algorithms can be categorized into two classes: ranging-based algorithms and ranging-free algorithms. Compared with ranging-free algorithms, ranging-based algorithms can achieve higher localization accuracy with fewer anchor nodes [1]. Traditional ranging methods are based on time of arrival/time difference of arrival (TOA/TDOA), received signal strength (RSS), and angle of arrival (AOA). Among them, TOA is preferred because of its high ranging accuracy and relatively simple hardware structure, especially with the technique of ultra-wideband (UWB) [2].

Specifically, localization is to estimate the positions of all the nodes in a network, given the positions of some nodes (referred to as anchor nodes) and partial pair-wise distance measurements between neighbor nodes [3]. Two major challenges in localization are as follows. First, the accuracy of distance measurements may be degraded by noise and non-line-of-sight (NLOS) blockage (referred to as ranging error problem). The impact of the ranging error usually depends on the geometrical relationship between sensors, the bandwidth of pulses, application scenarios, etc. [4], [5]. It is difficult to take all the factors into account. Second, the position information provided by some anchor nodes may be inaccurate/wrong. For example, in a WSN for forest fire detection, some anchor nodes may be moved by pedestrians or animals (thus referred to as inaccurately-positioned anchor nodes). But the anchor nodes still broadcast their original positions, thus leading to localization errors.

In the literature, the ranging error problem has been investigated extensively. Using distance measurements and positions of anchor nodes, other nodes’ relatively accurate positions can be calculated by Multilateration [6] or multidimensional scaling (MDS) [7] method. And to overcome the ranging errors, an optimization problem can be formulated to minimize some global error functions, such as maximum likelihood estimation [8], and semidefinite programming [3], [9]. Although these methods may guarantee estimation accuracy under certain conditions, most of them become vulnerable in the presence of inaccurately-positioned anchor nodes.

In this paper, our objective is to solve both the ranging error problem and the inaccurate anchor node position problem simultaneously. We first analyze Cramer-Rao Bound (CRB) for localization in a WSN. To measure the accuracy of a specific localization algorithm, CRB is studied in the literature [10], [11], where ranging error between any pair of nodes is usually assumed to be independent and identically-distributed. However, as shown in Section II-B of this paper, ranging error also depends on the distance between the two nodes. In this research, we present an ideal ranging error model in TOA estimation, and further derive the CRB for localization. Moreover, we propose a localization algorithm that consists of two steps: first to estimate the network topology (i.e., relative position between nodes) without the knowledge of anchor node locations, then to detect inaccurately-positioned anchor nodes and further obtain the positions of all the nodes in the network. The rest of this paper is organized as follows. The CRB for localization is derived in Section II. Section III and IV present the two steps in our localization algorithm, i.e., topology reconstruction and detection of inaccurately-positioned anchor nodes, respectively. The performance evaluation is given in Section V, followed by concluding remarks in Section VI.

II. CRB FOR LOCALIZATION

A. Network Model

Assume that all the nodes in a WSN are fixed on a plane. There are $M$ sensors (with IDs from 1 to $M$) without position information and $N - M$ sensors (with IDs from $M + 1$ to $N$) with known position information, referred to as non-anchor nodes and anchor nodes, respectively. Anchor nodes form a set $\mathcal{P}$, while non-anchor nodes form a set $\mathcal{Q}$. Node $i$ can communicate with node $j$ and thus can measure the distance to node $j$ if and only if the distance between nodes $i$ and $j$, i.e., $d_{i,j}$, is less than the visible radius $R$. All the nodes that
can communicate with node $i$ form a set denoted as $\mathcal{F}(i)$. Define a distance measurement set $\mathcal{D}$ in which
\begin{equation}
\mathcal{D} = \{d_{i,j}|i \in \mathcal{P} \cup \mathcal{Q}, j \in \mathcal{F}(i)\}
\end{equation}
where $d_{i,j}$ is the measured distance between nodes $i$ and $j$. Our objective is to use set $\mathcal{D}$ and position information of anchor nodes, $\mathcal{W}_p$, where
\begin{equation}
\mathcal{W}_p = \{(x_i, y_i)^T|i \in \mathcal{P}\}
\end{equation}
to reconstruct the set
\begin{equation}
\mathcal{W}_q = \{(x_i, y_i)^T|i \in \mathcal{Q}\}
\end{equation}
where $(x_i, y_i)^T$ is coordinate vector of node $i$, and $\mathcal{W}_p$ and $\mathcal{W}_q$ denote the set of coordinate vectors of anchor nodes and non-anchor nodes, respectively. Throughout the paper, if $z$ means a coordinate of a node or the distance between two nodes, we use $\hat{z}$ to denote the measured or estimated value of $z$ in the localization.

B. Ranging Error Model

As localization is based on ranging results among nodes, a well-built ranging error model can facilitate localization in the process of refinement. Ranging accuracy depends on a number of factors, such as application scenarios, TOA estimation algorithms, and the bandwidth of pulses. Generally it is difficult to construct a general ranging error model to take into account all the factors. So we use the lower bound of TOA estimation error to approximate the ranging error. With additive Gaussian white noise, the CRB of time delay estimation using TOA ranging is [12]
\begin{equation}
\sigma_\varepsilon^2 \geq \frac{1}{8\pi^2 \gamma \beta_f^2}
\end{equation}
where $\sigma_\varepsilon^2$ is the variance of time delay estimation error, $\beta_f$ is the bandwidth of ranging signal, and $\gamma$ is signal-to-noise ratio (SNR). For omnidirectional antennas with only line-of-sight, we have
\begin{equation}
P_r = \frac{P}{4\pi d_{i,j}^4}
\end{equation}
where $P_r$ is received power, $P$ is transmit power, and $d_{i,j}$ is the distance between the transmitter (node $i$) and receiver (node $j$). When in a multipath channel, we have
\begin{equation}
P_r = \frac{P}{4\pi d_{i,j}^{4n}}
\end{equation}
where $n$ is multipath factor that is decided by specific application scenarios.

The SNR of the received signal is
\begin{equation}
\gamma = \frac{P_r}{N_0 \beta_f}
\end{equation}
where $N_0$ is the noise power spectral density. From (4), (6), and (7) we can get
\begin{equation}
\sigma_\varepsilon^2 \geq \frac{N_0 \sigma_\delta^2}{2\pi \beta_f P}.
\end{equation}

Take the equality in (8), and multiply both sides with $c^2$ where $c$ denotes the speed of light, then we get:
\begin{equation}
E\{(\hat{d}_{i,j} - d_{i,j})^2\} = c^2 \cdot \sigma_\varepsilon^2 = \frac{c^2 N_0 \sigma_\delta^2}{2\pi \beta_f P}
\end{equation}
where $E\{\cdot\}$ means expectation. Similar to [11], it is assumed that the ranging error follows a zero-mean normal distribution. Let $\sigma_0 = \sqrt{\frac{c^2 N_0 \sigma_\delta^2}{2\pi \beta_f P}}$, then the ranging error model can be described as:
\begin{equation}
\hat{d}_{i,j} = d_{i,j} + \sigma_0 \varepsilon_0
\end{equation}
where $\sigma_0$ is defined as noise factor, and $\varepsilon_0$ is a standard normal random variable denoted as $\varepsilon_0 \sim N(0,1)$.

C. Error Bound for Localization

With the ranging error model, the CRB for localization can be derived. Let
\begin{equation}
\theta_n = (x_1, y_1, ..., x_M, y_M)^T
\end{equation}
be the coordinate vector of non-anchor nodes, and
\begin{equation}
\hat{\theta}_n = (\hat{x}_1, \hat{y}_1, ..., \hat{x}_M, \hat{y}_M)^T
\end{equation}
be the estimated coordinate vector of non-anchor nodes. Then the error covariance matrix is
\begin{equation}
\Sigma = E\{(\hat{\theta}_n - \theta_n)(\hat{\theta}_n - \theta_n)^T\}.
\end{equation}
The error covariance matrix is bounded by the CRB, i.e.,
\begin{equation}
\Sigma \preceq \text{CRB}.
\end{equation}

On the other hand, let $J$ denote the Fisher Information Matrix (FIM) of the WSN, whose elements can be proved as follows.
\begin{equation}
J_{2i-1,2i-1} = \sum_{j \in F(i)} \left( \frac{2n^2 (x_i - x_j)^2}{d_{i,j}^4} + \frac{1}{\sigma_0^2} \frac{(x_i - x_j)^2}{d_{i,j}^{2n+2}} \right)
\end{equation}
\begin{equation}
J_{2i,2i} = \sum_{j \in F(i)} \left( \frac{2n^2 (y_i - y_j)^2}{d_{i,j}^4} + \frac{1}{\sigma_0^2} \frac{(y_i - y_j)^2}{d_{i,j}^{2n+2}} \right)
\end{equation}
\begin{equation}
J_{2i-1,2i-1} = J_{2i-1,2i-1} = \sum_{j \in F(i)} \left( \frac{2n^2 (x_i - x_j)(y_i - y_j)}{d_{i,j}^4} + \frac{1}{\sigma_0^2} \frac{(x_i - x_j)(y_i - y_j)}{d_{i,j}^{2n+2}} \right);
\end{equation}
for $j \neq i$, if $j \in F(i)$
\begin{equation}
J_{2i-1,2j-1} = J_{2j-1,2i-1} = -\left( \frac{2n^2 (x_i - x_j)^2}{d_{i,j}^4} + \frac{1}{\sigma_0^2} \frac{(x_i - x_j)^2}{d_{i,j}^{2n+2}} \right)
\end{equation}
\begin{equation}
J_{2i,2j} = J_{2j,2i} = -\left( \frac{2n^2 (y_i - y_j)^2}{d_{i,j}^4} + \frac{1}{\sigma_0^2} \frac{(y_i - y_j)^2}{d_{i,j}^{2n+2}} \right)
\end{equation}
\[ J_{2i-1,2j} = J_{2j,2i-1} = J_{2i,2j-1} = J_{2j-1,2i} = -\left(\frac{2n^2(x_i - x_j)(y_i - y_j)}{d_{i,j}^4} + \frac{1}{\sigma_0^2}, \frac{(x_i - x_j)(y_i - y_j)}{d_{i,j}^{2n+2}}\right); \]
and other entries in \( J \) are 0’s.

When there are i) more than 3 anchor nodes, and ii) a sufficient number of measured distances between nodes in the WSN, the inverse of \( J \) exists, and the CRB can be represented as

\[ \text{CRB} = J^{-1}. \] (21)

In an actual application, the locations provided by some anchor nodes may not be accurate. Still taking these anchor nodes as reference points will lead to inaccurate or even catastrophic results. Thus we propose to first obtain the whole network’s topology (i.e., a set of mutual distances of all the nodes) without use of anchor node location information, then detect and remove inaccurately-positioned anchor nodes, and finally get the positions of all the nodes. So for the ranging error bound, we view the \( N - M \) anchor nodes as non-anchor nodes, and thus have a \( 2N \times 2N \) FIM \( J \). The FIM \( J \) has a rank of \( 2N-3 \) [10]. So the inverse of FIM does not exist. This means lower bound of \( E((\tilde{x}_i - x_i)^2) \) or \( E((\tilde{y}_i - y_i)^2) \) for node \( i \) cannot be obtained. However, a lower bound of the summation of \( E((\tilde{x}_i - x_i)^2 + (\tilde{y}_i - y_i)^2) \), \( i = 1,2, ..., N \) can be calculated as follows. A lower bound of \( \sum_{i=1}^{N} E((\tilde{x}_i - x_i)^2 + (\tilde{y}_i - y_i)^2) \) is equal to the trace (i.e., sum of diagonal elements) of \( J^t \), where \( J^t \) is a Moore-Penrose pseudo-inverse of \( J \). Let \( \lambda_1, \lambda_2, ..., \lambda_{2N-3} \) be the non-zero eigenvalues of \( J \), then the trace of \( J^t \) can be expressed as [10]

\[ Tr(J^t) = \sum_{i=1}^{2N-3} \frac{1}{\lambda_i}. \] (22)
And further, the average CRB of ranging errors can be obtained, given by \( Tr(J^t)/N \).

### III. Topology Reconstruction

To obtain a network topology, MDS [7] method that only takes distance matrix as the input is a good candidate because of its capability of parallel computing and ranging error tolerance. However, it requires that any pair of nodes should be able to reach each other, which is not practical. On the other hand, in [13], a method is given for topology reconstruction from incomplete distance information. Let \( D \) be a distance matrix in which \( D_{i,j} = d_{i,j}^2 \). It is proved in [13] that the rank of \( D \) is at most 4. As discussed in [13], suppose the first 4 rows of \( D \) are linearly independent and are known, which also means that each of the first 4 nodes can reach all other nodes. Also assume that in each of other rows of \( D \), at least 4 entries are known, i.e., each node can range the distance with at least 4 other nodes. Then the network topology can be reconstructed, i.e., the value of each entry in \( D \) can be obtained. One problem in the topology reconstruction is how to find 4 nodes, each with measured distance to any other node. Furthermore, in the topology reconstruction method, it is assumed that \( p_{i,j} \) (the probability that node \( i \) can measure the distance to node \( j \)) is known for any \( i \) and \( j \). The assumption makes the method impractical in a real network. Another drawback of the topology reconstruction method comes from the fact that the accuracy of localization with the method depends largely on the “best guess” of the distance between nodes \( i \) and \( j \) when it cannot be measured. In the following, we propose a new topology reconstruction method to solve these problems.

As aforementioned, it may be difficult to find 4 nodes in a WSN, each with measured distance to any other node. Generally, this condition can be satisfied in a local area, namely a sub-network. In other words, in the group of nodes in a sub-network, it is possible to find 4 nodes that can reach any other node in the group, and then to obtain the distance matrix of all nodes in the sub-network by the preceding procedure. Thus the sub-network becomes all-connected, which means the distance between any pair of nodes is known. For a node in an adjacent area of the sub-network, if it can reach at least 4 nodes in the all-connected sub-network, its distance to any node in the sub-network can be obtained, and thus it can be included into the all-connected sub-network. This means the all-connected sub-network is gradually enlarged, and eventually it includes all the nodes in the WSN.

As ranging errors are inevitable, the performance of topology reconstruction can be degraded due to accumulated ranging errors. To deal with the error accumulation issue, we have the following requirements.

- The initially selected sub-network should be an all-connected sub-network rather than one with only 4 nodes that can reach any other node;
- Once adjacent nodes are included into the sub-network and new topology is generated (e.g., by MDS method), distance refinement is necessary to address error accumulation in the search for adjacent nodes. We choose BFGS method [14] (a quasi-Newton method) to reduce the cumulative errors.

Define a connection matrix as \( W \) in which \( w_{i,j} \) is equal to 1 if the distance between node \( i \) and \( j \) is measurable, and equal to 0 otherwise. Our method is stated as follows.

**Step 1: Setup of initial all-connected sub-network** – Set \( k = 0 \), find an all-connected sub-network \( A_k \), in which \( |A_k| \geq 4 \), i.e., the number of elements in set \( A_k \) is no less than 4, and generate distance matrix \( D_0 \) for the sub-network.

**Step 2: Inclusion of new adjacent nodes** – According to connection matrix \( W \), search the set \( B_k = \{i \mid i \notin A_k, |F(i) \cap A_k| \geq 4\} \), i.e., each node in \( B_k \) has at least 4 measured distances with nodes in \( A_k \). All the nodes in \( B_k \) are to be added into the all-connected network, shown as follows.

For each \( i \in B_k \), define \( T_i \) as the set of nodes in \( A_k \) that have measured distances with \( i \), denoted as \( T_i = \{j \mid j \in F(i) \cap A_k\} \). Obviously \( |T_i| \geq 4 \). Let \( R_i = A_k \setminus T_i \). So we need to know the distances of \( i \) to the nodes in \( R_i \). Without loss of generality, we assume \( |A_k| = m, |T_i| = l \geq 4 \), the nodes in \( T_i \) are with IDs from 1 to \( l \), and other nodes in \( A_k \) are with IDs from \( l+1 \) to \( m \). So the distance matrix of the set \( A_k \cup \{i\} \)
is given by

\[ \mathbb{D}_k = \begin{pmatrix}
  d_{1,1}^2 & d_{1,2}^2 & \cdots & d_{1,m}^2 & d_{1,1} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  d_{l,1}^2 & d_{l,2}^2 & \cdots & d_{l,m}^2 & d_{l,1} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  d_{m,1}^2 & d_{m,2}^2 & \cdots & d_{m,m}^2 & d_{m,1} \\
  d_{2,1}^2 & d_{2,2}^2 & \cdots & d_{2,m}^2 & d_{2,1} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  d_{t,1}^2 & d_{t,2}^2 & \cdots & d_{t,m}^2 & d_{t,1} \\
\end{pmatrix}, \]

All the diagonal entries are 0’s. The unknown entries in \( \mathbb{D}_k \) are \( d_{l,t+1}^2, d_{l,t+2}^2, \ldots, d_{l,m}^2 \). Because the rank of \( \mathbb{D} \) is at most 4 and \( l \geq 4 \), the last row in \( \mathbb{D}_k \) can be represented by a linear combination of the first \( l \) rows, i.e.,

\[ \begin{pmatrix}
  d_{1,1}^2 \\
  d_{1,2}^2 \\
  \vdots \\
  d_{l,1}^2 \\
  d_{l,m}^2 \\
  d_{2,1}^2 \\
  \vdots \\
  d_{t,1}^2 \\
\end{pmatrix} = \begin{pmatrix}
  c_1 \\
  c_2 \\
  \vdots \\
  c_l \\
  c_l \\
  \vdots \\
  c_l \\
\end{pmatrix} + \begin{pmatrix}
  \delta_{1,1} \\
  \delta_{1,2} \\
  \vdots \\
  \delta_{l,1} \\
  \delta_{l,m} \\
  \delta_{2,1} \\
  \vdots \\
  \delta_{t,1} \\
\end{pmatrix}, \quad (23) \]

From (23) we can obtain the values of \( c_1, c_2, \ldots, c_l \), and further get the values of \( d_{1,t+1}^2, d_{l,t+2}^2, \ldots, d_{l,m}^2 \).

After all nodes in \( B_k \) are added into \( A_k \), \( k \) is increased by 1.

Step 3: Topology refinement – Use MDS method on the present \( \mathbb{D}_k \) (for nodes in the present \( A_k \)) to get the estimated coordinate vectors of nodes in \( A_k \) (e.g., \( (\hat{x}_i, \hat{y}_i) \) for node \( i \)). Then use the BFGS method to adjust the coordinate vectors of the nodes such that the following objective function

\[ \sum_{i < j \in F(i), i \in A_k, j \in A_k} \left( \hat{d}_{i,j} - \sqrt{(\hat{x}_i - \hat{x}_j)^2 + (\hat{y}_i - \hat{y}_j)^2} \right)^2 \]

is minimized. And update entries of \( \mathbb{D}_k \) accordingly.

Step 4: Proceed to Step 2, until all the nodes in the WSN have been included in \( A_k \).

IV. DETECTION OF INACCURATELY-POSITIONED ANCHOR NODES

Once the network topology is reconstructed (as discussed in the previous section), the positions of all non-anchor nodes can be obtained based on the reconstructed network topology and known anchor node positions. However, since some anchor nodes may be inaccurately positioned, it is necessary to detect these anchor nodes and prevent them from serving as reference nodes. Note that the inaccurately-positioned anchor nodes are still able to measure the distances to their neighboring nodes and broadcast them correctly. So the reconstructed network topology is assumed to be accurate because the position information of anchor nodes is not used in the reconstruction procedure.

Define a matrix \( \mathbb{V} \) in which an entry \( u_{i,j} \) means the estimated distance (in the topology reconstruction) of anchor node \( i \) to anchor node \( j \), i.e., \( d_{i,j} \). Define a matrix \( \mathbb{V} \) in which an entry \( v_{i,j} \) means the declared distance of anchor node \( i \) to anchor node \( j \), i.e., the distance calculated based on position information broadcast from the anchor nodes. Also define a difference matrix \( \mathbb{Q} \) in which

\[ q_{i,j} = \begin{cases} 
  1, & \text{if } |u_{i,j} - v_{i,j}| < \tau \\
  0, & \text{otherwise} 
\end{cases} \quad (25) \]

where \( \tau \) is a pre-specified threshold. It can be seen that the more the position error of an anchor node \( i \), the larger the likelihood of \( q_{i,j} \) (with another anchor node \( j \)) being 0. Note that for an inaccurately-positioned anchor node \( i \), the \( q_{i,j} \)'s may not be all 0’s. Therefore, in order to detect inaccurately-positioned anchor nodes, we propose to find the set of accurately-positioned anchor nodes instead, as described in the following. The principle in the detection of accurately-positioned anchor nodes is the fact that \( q_{i,j} \) should be 1 if anchor nodes \( i \) and \( j \) are both accurately positioned. For the simplicity of presentation, if \( q_{i,j} = 1 \), we say anchor nodes \( i \) and \( j \) are connected. Thus our objective is to find the maximum set of all-connected anchor nodes. This can be achieved by first getting a small set of all-connected anchor nodes and then gradually adding new anchor nodes into the set until no more new anchor nodes can be added.

Once the set of accurately-positioned anchor nodes is obtained, the locations of non-anchor nodes and inaccurately-positioned anchor nodes can be calculated based on the reconstructed network topology and the locations of the accurately-positioned anchor nodes, using a similar method in [15].

V. PERFORMANCE EVALUATION

We evaluate the performance of our proposed algorithm through simulations. A number \( N = 100 \) nodes are randomly placed in a unit square. The multipath factor \( n \) is equal to 1.

A. Evaluation of Topology Reconstruction

We use root mean square distance (RMSD) [3] to measure estimation error

\[ RMSD = \frac{1}{\sqrt{N}} \left( \sum_{i=1}^{N} \left( (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right) \right)^{\frac{1}{2}}. \quad (26) \]

Fig. 1 shows the estimation error when the noise factor \( \sigma_0 \) increases from 0.05 to 0.4. In the simulation, visible radius is set to be 0.5. It can be seen that the localization error increases with the noise factor. The average CRB as derived in Section II (i.e., \( Tr(\mathbb{J})/N \)) is also shown in Fig. 1 for comparison. The result demonstrates that the estimation error in our method can keep 2-3 times of corresponding CRB.

B. Evaluation of Inaccurately-Positioned Anchor Node Detection

We show the comparison of our inaccurately-positioned anchor node detection with Multilateration [6]. In the simulation, noise factor is fixed at 0.2, visible radius is 0.5, the number of anchor nodes is 10 (with IDs from 1 to 10), and...
node detection can improve the accuracy of localization.

For a non-anchor node, the estimated location is marked by a triangle, and is connected with the true location by a dashed line. For a non-anchor node, the estimated location is marked by a circle, and is connected with the true location by a solid line. It can be seen that our inaccurately-positioned anchor node detection can improve the accuracy of localization.

VI. CONCLUSIONS

In this paper, to measure localization accuracy in WSNs, we first present a ranging error model for TOA estimation, and derive CRB for localization accordingly. Ranging error resistance and reference location error tolerance are two challenging tasks in localization. We address these issues by a topology reconstruction method without use of anchor node location information and a method to detect inaccurately-positioned anchor nodes. Simulation results show that our proposed topology reconstruction method can approach CRB, and inaccurately-positioned anchor node detection method can improve localization accuracy significantly.

REFERENCES


Fig. 1. Estimation error vs. noise factor.

Fig. 2. A comparison between Multilateration algorithm and our topology reconstruction algorithm with inaccurately-positioned anchor node detection.

(a) Localization result of Multilateration algorithm.

(b) Localization result of our topology reconstruction algorithm with inaccurately-positioned anchor node detection.