Computer Networks 210 (2022) 108865

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Adaptive range-based localization algorithm based on trilateration and reference node selection for outdoor wireless sensor networks

Jari Luomala *, Ismo Hakala

University of Jyvaskyla, Kokkola University Consortium Chydenius, P.O. Box 567, FI-67701 Kokkola, Finland

A R T I C L E I N F O

Keywords:
Range-based localization
RSSI
Adaptive
Trilateration
Reference node selection
Wireless sensor network

A B S T R A C T

Locating the nodes of outdoor wireless sensor networks (WSNs) using (tri)lateration with a low-cost ranging technique, such as the received signal strength indicator (RSSI), often results in inaccurate location estimates. This can mostly be explained by the combined effect of distance estimate errors and localization geometry, both of which are subject to the reference nodes used. To develop techniques for reducing localization error, the distance estimate errors and localization geometry must be analyzed and taken into account. To address these challenges, this paper aims to seek ways to improve the quality of range-based trilateration localization for WSN nodes in varying outdoor conditions. Based on simulations, we analyze the effects of ranging error and localization geometry on localization error. For that purpose, we introduce a simple measure to evaluate the geometry of reference triangle (GRT). To improve localization accuracy and precision, we propose an adaptive range-based localization (ARBL) algorithm that is based on trilateration and reference node selection. In ARBL, the GRT values are calculated for each 3-combination of a preselected reference node set, based on which the combinations are selected. The algorithm exploits these reference node 3-combinations aiming to find the best ones at a given time using a selection criteria that is based on ranging error and localization geometry. The simulation and experimental results indicate that the proposed algorithm reduces localization error considerably. This shows that it is possible to achieve sufficient localization accuracy using range-based trilateration localization, even based on the RSSI in challenging outdoor conditions, by employing applicable techniques and information.

1. Introduction

In the last two decades, localization of nodes in wireless sensor networks (WSNs) has attracted the attention of numerous researchers. Localization is one of the core services in wireless sensor networks, and the knowledge of nodes' locations is useful or even necessary in many functions, services, and applications in WSNs [1–3]. Contrary to traditional networks, WSNs are designed for specific applications, and thus, have specific design constraints [1,4]. WSN nodes also have much tighter resource constraints (e.g., limited communication range, and limited energy, processing, memory, and storage capacity). These constraints apply to localization algorithms and protocols in WSNs as well. A technique commonly used to obtain location information is Global Navigation Satellite Systems (GNSS), such as GPS or GLONASS. However, adding a GNSS receiver to each WSN node is neither a cost-effective nor energy-efficient solution in a large scale. Further, the receiver's operation is limited in some outdoor environments, such as under dense foliage or in urban canyons. Therefore, alternative solutions are required.

In *anchor-based* localization, a few of the nodes are called reference nodes (aka anchors, beacons, landmarks, or seeds), and their locations are assumed to be known *a priori* [2,3,5–7]. Reference nodes are equipped with a GNSS receiver, or their locations are manually configured. Unknown (unlocalized) nodes use reference nodes' coordinates with distance (or angle) estimates or other information to estimate their own locations using a localization algorithm.

Typically, localization algorithms in wireless sensor networks are classified into two main categories: *range-based* and *range-free* algorithms [2,6–10]. Range-based algorithms employ estimated inter-node distances or angles in localization, whereas range-free algorithms use connectivity (e.g., hop counts) or pattern matching (fingerprinting), for example. In range-based localization, distance or angle estimates can be obtained using ranging techniques, including time of arrival (ToA), time difference of arrival (TDoA), angle of arrival (AoA), or received signal strength indicator (RSSI). Based on the distance estimates to the reference nodes, and the reference nodes' coordinates, an unknown node can estimate its location using a location computation technique,

* Corresponding author.
E-mail addresses: jari.t.luomala@jyu.fi (J. Luomala), ismo.hakala@jyu.fi (I. Hakala).

https://doi.org/10.1016/j.comnet.2022.108865
Received 8 September 2021; Received in revised form 20 January 2022; Accepted 21 February 2022
Available online 18 March 2022

such as Lateration [5,11] (trilateration or multilateration), Min–max (bounding box) [11,12], or a probabilistic approach (e.g., maximum likelihood).

Similar to the other range-based techniques, RSSI-based localization has its pros and cons. On one hand, it is a low-cost, energy-efficient technique requiring no additional hardware except a radio transceiver, which makes it suitable for sensor networks. On the other, this approach is sensitive to changes in environmental and weather conditions which often results in inaccurate range and location estimates [13–16]. Range-based localization error depends mainly on the combined effect of the ranging error and localization geometry, that is, the locations of the reference nodes relative to the unknown node [17]. Thus, the size of the localization error varies depending on which reference nodes are used, as ranging errors and localization geometry change accordingly. Furthermore, some location computation techniques are more robust to ranging errors and/or localization geometry than others [11]. Although trilateration is a widely used, low-cost technique for location computation, trilateration is sensitive to ranging errors and the locations of the reference nodes [17]. This may cause unpredictable variation in the location estimates depending on the reference node set. In addition, the same reference node combination may work for one unknown node but not necessarily for another. Therefore, selecting suitable reference nodes (i.e., those with the highest quality for localization) in each case is important for achieving acceptable localization accuracy. Furthermore, selecting only a subset of reference nodes will reduce energy consumption.

This paper aims to seek ways to improve the quality of range-based trilateration localization for WSN nodes in spatially and temporally varying outdoor conditions. Based on simulations, we analyze the effects of ranging error and localization geometry on localization error (in 2D). As a solution, we propose an adaptive range-based trilateration localization (ARBL) algorithm that is based on reference node selection. To reduce the computational load, the reference nodes are preselected based on their RSSI value and geometry. Initially, a set of n reference nodes with the best RSSI value is selected from the set of all reference nodes. Geometry of reference triangle (GRT) values are calculated for each 3-combination in this preselected reference node set. The algorithm exploits these 3-combinations and aims to find the best ones at a given time for computing the final location estimate of a node, by adapting the reference node selection to the prevailing localization geometry and ranging errors. The algorithm is scalable, and can be implemented in low-cost, resource-constrained WSN nodes.

GRT is inspired by the geometric dilution of precision (GDOP), which is commonly used in satellite positioning to describe the geometric error caused by the mutual geometry of satellites. To calculate GDOP, an estimate of the location and satellite location data is required. In contrast, the calculation of GRT is based only on the location data of the reference points in the vicinity of the unlocalized node. The distance of the reference point from the node is estimated based on the RSSI value. GRT approximates well the so-called average HDOP value, which is based on the average of the HDOP values of the different position estimates at the plane (2D).

The results from the simulations and the collected experimental data indicate that the ARBL algorithm can reduce localization error considerably, resulting in much better accuracy than the nearest-neighbors approach in a given set of reference nodes. It also produces smaller localization error than EATL [18] and RNST [19] algorithms with the applied simulation parameters. This shows that achieving usable and accurate location estimates is possible, even in challenging and varying outdoor conditions, provided that applicable techniques and information are exploited. Overall, this study provides new insights into RSSI- and range-based localization and its feasibility in wireless sensor networks.

In summary, this paper makes the following main contributions:

- We analyze the effects of ranging error and localization geometry on localization error.
- We introduce a simple measure to evaluate the geometry of reference triangle (GRT), and propose an ARBL algorithm that is based on trilateration and reference node selection.
- We evaluate the performance of the proposed algorithm using simulations and real measurement data.

The remainder of this paper is organized as follows. In Section 2, related studies are discussed briefly. In Section 3, the theoretical background related to the localization problem in scope is presented. In Section 4, the effect of localization geometry and ranging error on localization error is analyzed based on simulations. The proposed adaptive range-based localization algorithm is presented in Section 5, followed by the performance evaluation in Sections 6 and 7. Finally, after a short discussion, the conclusion is provided in Section 8.

2. Related work

To date, several survey articles about localization in WSNs have been published (e.g., [2,5–10,20,21]), classifying and presenting some of the numerous localization algorithms and techniques proposed in the past few years. In this section, we review several papers related to our work, focusing on trilateration-based localization algorithms that utilize reference node selection.

2.1. Anchor- and range-based trilateration localization algorithms

Many of the anchor- and range-based localization algorithms and techniques proposed in the past few years are based on trilateration [2,5,6,8]. Although extensively studied, most of the range-based (and range-free) approaches do not take into account the effect of localization geometry on localization error. The selection of reference nodes is an essential part of range-based localization, and particularly important for trilateration-based localization. Reference node selection greatly affects the localization quality, and therefore, cannot be ignored. Often reference nodes are selected either randomly or by using the nearest neighbors approach without considering the localization geometry. By selecting the reference nodes based on localization geometry and distance measurement error (and other metrics), localization error could be considerably reduced.

2.2. Reference node selection algorithms

Despite the prevalence of anchor-based localization, relatively few studies related to reference node selection algorithms have been conducted. This subject is covered, for instance, in papers [17–19,22–31].

In [22], a selective anchor node localization algorithm (SANLA) is proposed. In SANLA (based on DV-hop), an unknown node computes its location by trilaterations, in each of which one of the anchor nodes is fixed (the reference node), and two are the combinations of the other anchors. Next, the same anchor combinations are used with the unknown node to compute the location of the fixed reference node. These coordinates are compared with the real coordinates, and the combination which produced the smallest error can be found and reported to the unknown node. Now, the unknown node knows which of its location estimates gave the best result, and considers that one as the node's coordinate. Although the algorithm was shown to produce good results, its computation cost is increasing heavily along with the increased number of nodes or anchor ratio.

The authors in [19] proposed a reference node selection algorithm based on trilateration (RNST) for mobile nodes in indoor sensor networks. The unlocalized node computes the distances between each pair of nodes, and finds out if any of the reference triplets can form a nearly equilateral triangle. Then the location estimates are computed using all the possible equilateral triangles, and the average value is set as the final estimate. However, the performance of the algorithm was evaluated using only the equilateral triangle placement of the reference nodes.

The confidence-based iterative localization (CIL) algorithm is proposed in [17]. The algorithm is based on the quality of trilateration (QoT), a probabilistic metric that quantifies the geometric relationship of the reference nodes and ranging errors, and indicates the accuracy of the particular trilateration. In CIL, each node has a confidence value indicating the accuracy and reliability of the location estimate, and it is used as an evaluating indicator. The confidence of a node based on trilateration is defined as a product of the QoT and the confidence of the reference nodes in trilateration. The localization process is carried out from high-confidence nodes (beacons with positioning devices) to low-confidence nodes (others) iteratively using trilaterations. An unlocalized node receives location information from reference nodes (localized nodes) with different confidence values. At each stage, an unknown node selects the trilateration with the highest quality (confidence) to locate itself. At any time, the location estimate can be refined if a higher-confidence location is available. The experiment and simulation results showed that localization accuracy was significantly improved by using CIL.

In a more recent work [18], the authors propose a distributed reference node selection algorithm based on error analysis for trilateration localization (EATL). The algorithm complies with three principles for optimizing the selection of anchor nodes to be used in trilateration. Two principles are related to the distribution of the reference nodes, and state that the minimum internal angle of the reference triangle should be larger than 13° and its shortest edge as long as possible. The third principle is related to the relative position between the unknown node and the reference nodes, which also affects localization accuracy, and states that the distances between the unknown node and the reference nodes should be as similar as possible. The competitive performance of the algorithm was demonstrated with simulations. However, the distance errors applied in the simulations appear to be somewhat over-optimistic, particularly if RSSI is used for ranging, which may reduce the performance in real scenarios.

Improved trilateration localization with minimum uncertainty propagation and optimized selection of anchor nodes (ITL-MEPOSA) is proposed in [24]. The authors define the standard deviation of consecutive distance estimates between an unknown node and an anchor node as the uncertainty information. Optimized selection of anchor nodes is based on minimum uncertainty propagation; that is, three anchors that have the minimum product of the mean distance estimate and the corresponding uncertainty information (SD) are selected. These anchor nodes and the corresponding mean distance estimates are employed in trilateration. However, the effect of localization geometry is not dealt with.

To tackle the issue of biased locations of reference nodes, in [23] the authors proposed a method for selecting reference nodes based on hierarchical clustering. Based on their coordinates, reference nodes are divided into separate groups, and references belonging to different clusters are selected for trilateration.

In [25], distributed algorithms to select a subset of reference nodes to minimize the error of localization (CRLB of variance) was studied when global (global-nearest-neighbor, global-crlb) or only local (local-nearest-neighbor, local-crlb) information is available. The authors proposed an algorithm, called local-crlb, which uses distances and the CRLB on localization error for selecting the references. The simulation results showed that local-crlb achieved significantly smaller location error than the often used nearest-neighbor approach (local-nearest-neighbor). The authors extended their investigations in [26] by analyzing distributed algorithms for selecting subsets of references for localization regarding location error, energy consumption and communication. Three algorithms (RS-GC, RS-LC, RS-LC-CR) use the CRLB on localization error for selecting while two algorithms (RS-GD, RS-LD) are conventional, distance-based algorithms. For both local and global knowledge, CRLB-based algorithms achieved better accuracy than the conventional ones. The algorithms with global knowledge achieved the smallest error, but with the price of higher energy consumption and communication overhead. However, RS-LD achieved the best ratio between location error and energy consumption. While CRLB considers both ranging errors and localization geometry, it requires some prior knowledge from the network (e.g., variation of RSSI, path loss exponent), and may be computationally expensive.

A node selection algorithm based on RSS threshold, called Node-selection Least Squares (NS-LS), is proposed in [27,29]. The idea is to select the highest RSS threshold guaranteeing a mean number of anchor nodes (N_m) inside the range of all the nodes. The N_m value is chosen to optimize the trade-off in position accuracy versus energy consumption, and it was achieved when $N_m = 3$. When compared with ML estimation, NS-LS showed similar performance in localization accuracy while the reduction in energy consumption was considerable. In [28,29], the authors improved their distributed and cooperative RSS-based localization algorithm by combining the node selection with real-time path loss estimation (OLPL-NS-LS). For the node selection, two criteria depending on the path loss estimates was proposed to reduce the number of cooperating nodes: *low path loss selection* for selecting the nodes with the lowest path loss exponent and, thus, better propagation conditions; *low distance selection* for selecting the closest nodes to reduce distance error estimates, which was shown to achieve better performance. It was shown that OLPL-NS-LS outperformed ML and MDS algorithms, and that having a smaller number of cooperative nodes reduced the energy consumption without affecting the accuracy. However, computing the location estimates using the proposed iterative approach involves many steps and requires initial location estimates.

Applying game theory and utility functions, the authors in [30] proposed a reference node selection strategy based on utility function for iterative multilateration based algorithms. Utility function includes the relevant information for the node selection, and it is comprised of the benefit indicator and the cost function, and models the trade-off between them. Here, CRLB is inversely proportional to the benefit, and the cost is proportional to distance. Each reference node combination is a possible coalition, and the coalition value for each is computed based on the utility function. Then, the subset of reference nodes with the highest coalition value is used for localization. Simulation results showed that higher coalition values resulted in more accurate location estimates. As compared to random selection and closest distance selection, the utility based selection achieved smaller localization error. In [31], the authors addressed the reference node selection problem for cooperative localization by modeling the localization process as a cooperative game. They also proposed a randomized search method exploiting spatial correlation (in terms of GDOP or CRLB) among the reference nodes in coalition to reduce computation complexity. Besides ranging quality and node geometry, anchor uncertainty was included into the node selection via the squared position error bound (SPEB). In the utility function, benefit is inversely proportional to SPEB, and cost is proportional to the number of selected nodes and their distance to the coalition head. However, the SPEB is calculated using the estimated locations as input which may produce errors.

Collectively, these studies outline a critical role for selecting reference nodes in the localization process. Some of the proposed techniques consider either localization geometry [18,19,23] or ranging/localization error [22,24] in the selection, and some consider both [17,25,26,30,31]. Some of the algorithms take also the energy efficiency into account [26], or try to optimize the trade-off between localization accuracy and energy consumption [27–29]. Further, the anchor uncertainty can be utilized in the reference node selection [31]. Unfortunately, increasing the localization accuracy often reduces energy efficiency due to increased computation and/or communication cost. Some algorithms might use iterative approaches, require initial location estimates, or some prior knowledge from the network. Sometimes the algorithm is evaluated based on simulations using parameters or setups which may be unrealistic. Or the scalability of the algorithm is bad. We recognize the importance of taking the ranging error and localization geometry into account in localization. In our approach, we aim at improving the localization geometry by selecting the reference nodes for trilateration based on a simple evaluation. Further, we consider the combined effect of ranging error and localization geometry in computing the final location estimate. The proposed method was designed and evaluated based on simulations and real RSSI measurement data from an outdoor sensor network. This approach is scalable and low-cost, thus enabling its implementation in WSN nodes.

3. Theoretical background

A sensor network can be represented by an undirected Euclidean graph $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_m, v_{m+1}, \dots, v_n\}$ is the set of n nodes, and $e = \{v_i, v_j\} \subset E$ if the nodes v_i and v_j are connected. The nodes are connected if their inter-node distance $w(e)$ is smaller than their radio range, which can be defined as distance or based on RSSI. Generally, nodes are assumed to be aware of their neighboring nodes. In V, the set of m unknown nodes $\mathcal{U} = \{v_1, v_2, \dots, v_m\}$, and the set of $n-m$ reference nodes $\mathcal{R} = \{v_{m+1}, \dots, v_n\}$. Considering two-dimensional (2D) case, the localization problem can be defined as: Given a multihop network $G = (V, E)$, a set of reference nodes $\mathcal{R}$ with the known locations (x_r, y_r) for all $r \in \mathcal{R}$, and the inter-node distances w, the aim is to find the unknown locations (x_u, y_u) for all $u \in \mathcal{U}$.

A typical range-based localization algorithm employs some ranging technique (RSSI, ToA, TDoA) to estimate the inter-node distances, and an appropriate location computation technique to compute unknown node locations. In this context, the distance estimates needed in localization can be obtained using any ranging technique. As the experimental part of this study is based on distance estimation using RSSI, in this section we describe the techniques commonly used for RSSI-based ranging and location computation. Furthermore, some of the main factors affecting RSSI- and range-based localization accuracy are briefly discussed.

3.1. RSSI-based ranging

The principle behind RSSI-based ranging techniques is that a radio signal decays (its amplitude decreases) when it travels farther from the transmitter. A popular model used to reflect the path loss of a radio signal is the *log-normal shadowing model*, which can be expressed as (see, e.g., [10,32]):

$$P_r(d) = P_r(d_0) - 10n \log\left(\frac{d}{d_0}\right) + X_\sigma, \tag{1}$$

where $P_r(d)$ (or $RSSI(d)$) is the received power [dBm] at distance d [m] from the transmitter, $P_r(d_0)$ (or $RSSI(d_0)$) is the received power [dBm] at the reference distance d_0 (usually 1 m) from the transmitter, n is the path loss exponent (PLE), and X_σ is the zero-mean Gaussian random variable with the variance of σ^2, that is, $X \sim N(0, \sigma_{\mathrm{dB}})$. The value of $P_r(d_0)$ can be estimated, for example, by using the Friis free-space equation, or it can be a predefined or empirically measured value. In addition, the PLE n can be predefined, or computed either offline or online by using, for example, fixed reference nodes with known distances.

Eq. (1) without the stochastic term X_σ is often referred to as the *log-distance path loss model*, expressing the average received power at distance d from the transmitter. The log-distance path loss model can be used to estimate the distance $\hat{d}$ [m] between two neighboring nodes as follows:

$$\hat{d} = d_0 10^{(P_r(d_0) - P_r(d))/10n}. \tag{2}$$

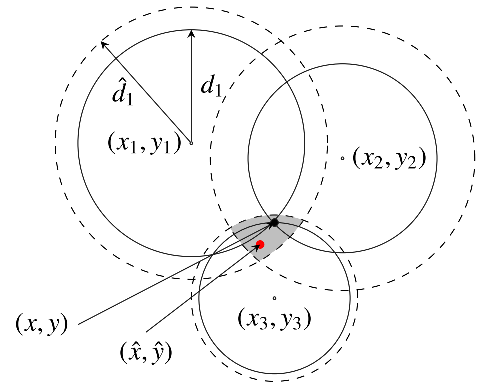


Fig. 1. An example of trilateration with accurate and inaccurate distance estimates. Using the correct distances (d_i) leads to the accurate location (x, y) while using the erroneous distances ($\hat{d}_i$) leads to the inaccurate location estimate $(\hat{x}, \hat{y})$. In the figure, $i = 1, \dots, 3$.

3.2. Lateration

Lateration is a commonly used location computation technique in WSNs, and it refers to the technique of determining an unknown node's location based on three (*trilateration*) or more (*multilateration*) reference nodes with known locations and the measured distances (e.g., based on the RSSI) to them [3,11,33]. In two-dimensional space, distances to at least three non-collinear reference nodes are required to obtain a unique solution. Next, mathematical basics for lateration are briefly described.

Assume that the locations of n reference nodes are $\mathbf{x_i} = (x_i, y_i)$, and that the distances between the unknown node $\mathbf{x} = (x, y)$ and the reference nodes are d_i. This leads to the following set of n equations:

$$(x_i - x)^2 + (y_i - y)^2 = d_i^2, \tag{3}$$

where $i = 1, \dots, n$. The equations can be linearized, for example, by subtracting the last nth equation from all the preceding ones. After rearranging the terms, we obtain the following system of linear equations:

$$\mathbf{Ax} = \mathbf{b}, \tag{4}$$

where the coefficient matrix

$$\mathbf{A} = \begin{bmatrix} 2(x_n - x_1) & 2(y_n - y_1) \\ 2(x_n - x_2) & 2(y_n - y_2) \\ \vdots & \vdots \\ 2(x_n - x_{n-1}) & 2(y_n - y_{n-1}) \end{bmatrix},$$

and the right-side vector

$$\mathbf{b} = \begin{bmatrix} d_1^2 - d_n^2 - x_1^2 - y_1^2 + x_n^2 + y_n^2 \\ d_2^2 - d_n^2 - x_2^2 - y_2^2 + x_n^2 + y_n^2 \\ \vdots \\ d_{n-1}^2 - d_n^2 - x_{n-1}^2 - y_{n-1}^2 + x_n^2 + y_n^2 \end{bmatrix}.$$

The algebraic solution for $\mathbf{Ax} = \mathbf{b}$, and thus, the location (estimate) of the unknown node is

$$\mathbf{x} = (\mathbf{A^T A})^{-1} \mathbf{A^T b}. \tag{5}$$

In practice, the distance estimates are erroneous to some extent, resulting in varying localization error. An example of trilateration with inaccurate distance estimates is illustrated in Fig. 1. To mitigate the effect of inaccurate distance estimates, more than three reference nodes and distance estimates could be used [33]. On average, the localization accuracy increases with the number of reference nodes used (to some point) provided that their observations are independent [26]. The increase of accuracy is logical since more information can be used for localization. In the case $n > 3$, $\mathbf{Ax} = \mathbf{b}$ is an overdetermined system of linear equations for which we can find a least square solution $\hat{\mathbf{x}}$ that minimizes the mean square error, that is:

$$\hat{\mathbf{x}} = \arg_{\mathbf{x}} \min \|\mathbf{Ax} - \mathbf{b}\|^2. \tag{6}$$

Although adding extra reference nodes should improve the estimate, it will also increase the computation and communication cost. The system of linear equations comes bigger and require various solving methods [33]. Further, communication to each reference node is required. In resource-limited WSNs, the trade-off between localization accuracy and efficiency often need to be done. Achieving high localization accuracy with a minimum number of reference nodes is still an open research problem [10].

3.3. Factors affecting localization accuracy

Localization accuracy is the most important metric to evaluate the performance of a localization algorithm or system. Often, the accuracy is presented as a mean absolute localization error (MAE) of the nodes in the network (or the samples for a node), that is, the difference between the estimated and ground truth locations. Localization error can be also presented as relative to radio range.

Localization error $\Delta\hat{x}_i$ for node (or sample) i is defined as:

$$\Delta\hat{x}_i = \|\hat{\mathbf{x}}_\mathbf{i} - \mathbf{x}\| = \sqrt{(\hat{x}_i - x)^2 + (\hat{y}_i - y)^2}, \tag{7}$$

where $\hat{\mathbf{x}}_\mathbf{i} = (\hat{x}_i, \hat{y}_i)$ and $\mathbf{x} = (x, y)$ are the estimated and ground truth locations, respectively. The mean and the standard deviation of localization error, $\overline{\Delta\hat{x}}$ and $s_{\Delta\hat{x}}$, respectively, are defined as:

$$\overline{\Delta\hat{x}} = \frac{1}{n}\sum_{i=1}^{n} \Delta\hat{x}_i, \;\; s_{\Delta\hat{x}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\Delta\hat{x}_i - \overline{\Delta\hat{x}})^2}, \tag{8}$$

where n is the number of nodes (or location estimate samples).

In the following, we briefly introduce some of the main factors that affect the accuracy (or precision) of RSSI-based node localization. Most apply to anchor- and range-based localization in general. The magnitude of the effect on the localization error depends, for example, on the joint effect of the factors, and the localization algorithm used (see, e.g., [11,34] for a detailed analysis). We categorize the factors into two main groups: (1) *distance estimation error* and (2) *localization geometry*. In addition, *algorithmic error* has an effect on localization accuracy.

3.3.1. Distance estimation error

Ranging error. Ranging error is the difference between the estimated and the true inter-node distance.

Ranging error $\Delta\hat{d}_i$ for sample i is defined as:

$$\Delta\hat{d}_i = \hat{d}_i - d, \tag{9}$$

where $\hat{d}_i$ and d are the estimated and the true distances, respectively. The mean and the standard deviation of the ranging error, $\overline{\Delta\hat{d}}$ and $s_{\Delta\hat{d}}$, respectively, are defined as:

$$\overline{\Delta\hat{d}} = \frac{1}{n}\sum_{i=1}^{n} \left|\Delta\hat{d}_i\right|, \;\; s_{\Delta\hat{d}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\Delta\hat{d}_i - \frac{1}{n}\sum_{i=1}^{n}\Delta\hat{d}_i)^2}, \tag{10}$$

where n is the number of distance estimate samples.

Ranging has a significant role in range-based localization, and ranging accuracy impacts localization accuracy [5,10,11,35]. Ranging error is probably the most significant and characteristic factor that impairs RSSI-based localization accuracy, mainly because of the sensitivity of the RSSI to changes in environmental and weather conditions [16]. The sources of error in RSSI-based ranging can be divided into external (e.g., propagation environment, weather conditions, interference) and internal (related to node's HW) factors. Two major sources of error are path loss exponent (PLE) estimation error (see, e.g., [16,36]), and temperature change (see, e.g., [13,15,16]). Other sources of error include RF transceiver related issues, antenna characteristics, interference, and other weather conditions (e.g., humidity, rain, snow). Furthermore, the choice of which ranging technique is used affects the final localization accuracy (e.g., RSSI vs. TDoA) [5,10]. The effect of distance estimates on localization accuracy is clear, as can also be seen by looking at the equations for lateration. Reducing ranging error is the key to improved distance estimation and localization accuracy [35]. Our previous papers [15,16] focused on finding the reasons behind the RSSI-based ranging error, and proposed some error mitigation techniques.

Node degree. The degree of a node u, $\deg(u)$, is the number of its neighbors (i.e., links). The average node degree of a graph G (i.e., the network) is denoted as

$$\deg_{avg}(G) = \frac{1}{n}\sum_{u=1}^{n} \deg(u), \tag{11}$$

where n is the number of nodes [37]. The average node degree (connectivity, the average number of neighbors) affects the localization error indirectly through distance estimation error in multihop cases. The connectivity depends on the node density, $\rho = n/A$ (i.e., the number of nodes per unit area), and the radio transmission range [37]. Many localization algorithms are sensitive to the connectivity (or node density), and it has a strong effect on their localization accuracy [6,11]. Further, RSS-based ranging technique requires higher node density than time-based (ToA) to achieve good localization accuracy [32]. A bigger node degree (greater number of neighbors) results in straighter paths and increases the number of options for finding a shortest path to the reference nodes [11,34]. On average, increasing the average number of neighbors (node density) decreases localization error [11,20,32,34]. Nodes at the border areas of the network have a lower node degree, on average, than that of nodes in the middle [37].

Network topology. Network topology affects the localization error indirectly through the distance estimation error. The network may be anisotropic for an unknown node (e.g., due to holes or blocks between it and the reference nodes), which leads to overestimation of distances. Localization accuracy greatly depends on the network topology; irregular and random topology usually results in larger localization error [6,10,20]. In irregular topology, node density in some regions may differ largely from the average node density of the network [10]. Therefore, large variation in the node degree values between the nodes may indicate a bad network topology.

3.3.2. Localization geometry

The number and location of reference nodes. The number (or percentage) and location of reference nodes used in node localization define the localization geometry. Dilution of precision (DOP) is a measure that refers to the geometry of the distribution of reference nodes relative to an unknown node used to estimate its location [34,38,39]. Geometric dilution of precision, GDOP, is commonly used in GNSS applications to describe the geometric error caused by the mutual geometry of satellites. Statistically, DOP can be seen as a multiplier of the expected localization error [38]. Smaller DOP indicates better distribution of the reference nodes and therefore, better location precision. An example of two different localization geometries is illustrated in Fig. 2.

As a measure of localization geometry in the two-dimensional case, *horizontal DOP* (*HDOP*) can be used (see, e.g., [39]). First, we construct a geometry matrix $\mathbf{H}$,

$$\mathbf{H} = \begin{bmatrix} \frac{x - x_1}{R_1} & \frac{y - y_1}{R_1} \\ \frac{x - x_2}{R_2} & \frac{y - y_2}{R_2} \\ \vdots & \vdots \\ \frac{x - x_n}{R_n} & \frac{y - y_n}{R_n} \end{bmatrix}, \tag{12}$$

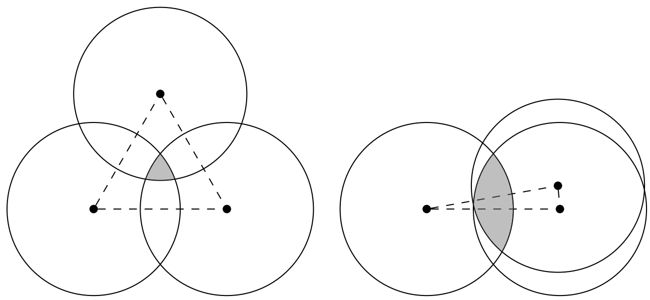


Fig. 2. An example of a good (left) and bad (right) localization geometry.

where (x, y) and (x_i, y_i), $i = 1, \ldots, n$, are the coordinates of an unknown node and the reference nodes, respectively, and R_i is the Euclidean distance between the unknown node and reference node i, defined as

$$R_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}.$$

The covariance matrix $\mathbf{G}$ can be derived as follows:

$$\mathbf{G} = (\mathbf{H}^{\mathbf{T}}\mathbf{H})^{-1}, \tag{13}$$

from which HDOP can be computed as follows:

$$HDOP = \sqrt{\sigma_x^2 + \sigma_y^2} = \sqrt{G_{1,1} + G_{2,2}}. \tag{14}$$

The location of reference nodes (their geometric relation) has a significant effect on localization accuracy [6,10,17,31]. Further, the relative location between the unknown and reference nodes affects the accuracy [10,18]. Consequently, if more than three reference nodes are available, the selection of reference nodes based on localization geometry can be used to improve localization accuracy [18,19,23,25,26,30,31]. Thus, localization error might be reduced by selecting only the most suitable references for a node. In addition, the localization accuracy depends on the number (density) of reference nodes [6,7]. Increasing the number of reference nodes tends to increase localization accuracy to some extent [10,11,18,29,34].

As shown above, HDOP is based on the relative locations of the unknown node and reference nodes. In [39], it was shown that HDOP values vary widely as the relative position of the node to be located to the fixed reference nodes changes. Therefore, its validity depends on how accurate the location estimate of the unknown is. To overcome this uncertainty, we defined a simple measure for localization geometry, called the geometry of reference triangle (GRT), in the case when three reference nodes are used (i.e., trilateration). Unlike HDOP, GRT depends only on the known locations of the reference nodes. Specifically, GRT defines the ratio of the edges of the reference triangle as:

$$GRT = \frac{d_{\max}}{d_{\min} + d_{\mathrm{md}}}, \tag{15}$$

where $d_{\max}$, $d_{\min}$, and d_{md} are the maximum, minimum, and median edge length of the reference triangle, respectively, and $GRT = [0.5, 1]$. In the ideal case ($GRT = 0.5$), the reference nodes form an equilateral triangle ($d_1 = d_2 = d_3, \measuredangle_{1,2,3} = 60°$), while in the worst case ($GRT = 1$), the reference nodes are collinear. In practice, the ratio of the edges, as shown in Eq. (15), determines the size of the smallest angle in the triangle. At the general level, GRT describes the effect of geometry of reference nodes on localization. As compared to HDOP, GRT is computationally much simpler. Further, when applied with inaccurate location estimates, the computed HDOP is misleading unlike GRT, which is based on the known reference node locations only. Also, the usage of HDOP requires that location estimates are already computed while GRT does not. For GRT, it is necessary only to preselect the reference nodes based on RSSI values. Therefore, GRT is more practical and energy-efficient in the real scenarios.

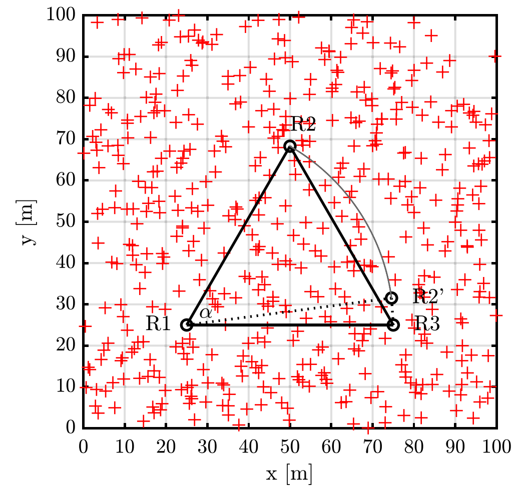


Fig. 3. Simulation setup. 500 unknown nodes (crosses) and three reference nodes (R1, R2, R3).

Error in reference nodes' locations. The effect of inaccurate reference nodes' locations on localization error is also clear, as can be seen in the lateration equations, and it results in erroneous distance and localization geometry estimates. If the reference nodes' locations used in the location computation are erroneous, whether or not the ranging error exists, the distance estimates will be incorrect. Moreover, an error in the reference nodes' locations affects the localization geometry, as the locations used in the location computation are distorted. This leads to different DOP (and GRT) and localization accuracy and precision. As stated in the literature, the accuracy (uncertainty) of the reference nodes' locations affects the localization accuracy of unknown nodes [31,34,35]. Usually reference nodes' locations are based on GNSS and may have a large variation [40]. Particularly, if regular nodes are used as references (virtual anchors) after they have obtained their own location estimates with some uncertainty, it has effect on localization accuracy [31,35]. Our previous paper [40] focused on reducing the location error of GNSS-based stationary reference nodes.

4. Localization error analysis

Although the role of ranging error and localization geometry in localization error has been widely recognized in the literature, it is necessary to know the mechanism in more detail. In this section, we analyze the effect of localization geometry (HDOP and GRT) and ranging error on localization error through simulations. The purpose is to show the need for an algorithm that takes both these aspects into account.

4.1. Simulation setup

We analyzed the effect of localization geometry and ranging error on localization error through simulation setup, as shown in Fig. 3. The setup consisted of 500 unknown nodes uniformly distributed in the area of 100×100 m $(x, y \sim U(0, 100))$, and three reference nodes (R1, R2, R3). Reference nodes R1 and R3 were fixed, and the location of R2 was changed between R2 and R2' by varying the angle α from 60° to 0.5° in step of 0.5°. Distance measurement error, Δd, was uniformly distributed, that is, $\Delta d \sim U(-\varepsilon, \varepsilon)$. where $\varepsilon = \{0.1, 0.15, 0.2, 0.25, 0.3\}d$. Thus, we assume that the distance measurement error is relative to the measured distance between the unknown and the reference node, d, and that each unknown node can directly communicate with the reference nodes (i.e., are within their radio range). For each combination of Δd and α, the location estimates of the unknown nodes were computed using trilateration, and the corresponding localization

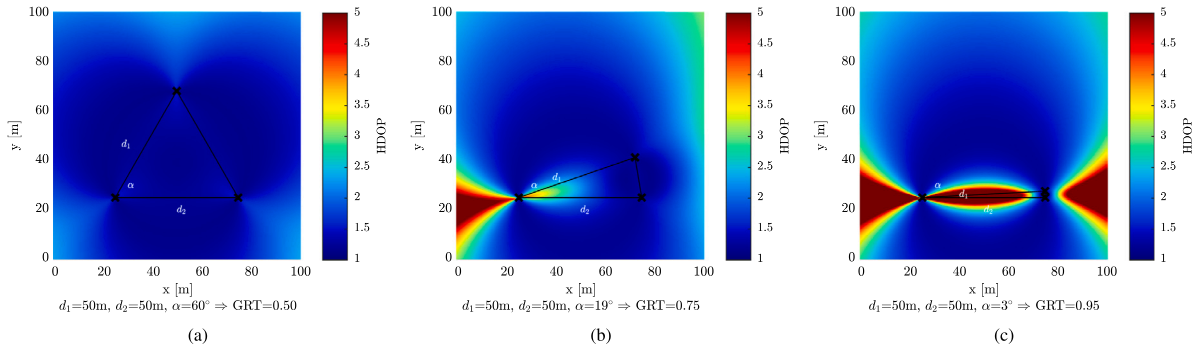


Fig. 4. An example of different localization geometries and the corresponding HDOP and GRT values. As can be seen, the HDOP value (in colors) depends also on the location of an unknown node while the GRT value does not. Smaller HDOP and GRT values indicate better localization geometry. The resolution of the grid to compute the HDOP values is 0.2×0.2 m, i.e., a total of 250000 grid points in the area of 100×100 m. The HDOP values outside the range $[1, 5]$ are clamped to the first (< 1) or last (> 5) colormap color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Parameters used in the simulations.

Parameter	Value
Network area	100×100 m
Number of unknown nodes	500
Number of reference nodes	3
Angle α	$[60.0, 0.5]°$
Ranging error ($\Delta d \sim U(-\epsilon, \epsilon)$)	$\epsilon = [0.1, 0.3]d$

errors, HDOP, and GRT were computed. To implement lateration (as described in Section 3), we used the MATLAB *mldivide* function [41] to solve the system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$. For each scenario, the mean absolute error (MAE) of the location estimates ($n = 500$) and localization error standard deviation (SD) was computed, as well as the average HDOP. The average HDOP (AvgHDOP) was computed by averaging the HDOP values of all the unknown nodes ($n = 500$). The parameters used in the simulations and their values are shown in Table 1.

4.2. Effects of ranging error and localization geometry on localization error

The geometry of the reference nodes and unknown node has a significant effect on localization. We used two metrics, HDOP and GRT, to analyze this effect. We also show the relationship between HDOP and GRT. For reference node selection, we use only GRT that is computationally lighter and thus more suitable for resource-constrained nodes. The value of HDOP depends on the locations of the unknown node and the reference nodes, while GRT depends only on the locations of the reference nodes, as shown in Fig. 4. Generally, the better geometry of the reference nodes results in better HDOP values on average, and also in smaller variation. However, there can be large variations depending on the location of the unknown node, particularly when the reference nodes' geometry is bad, as shown in Fig. 4(b) and (c), for example. As GRT defines only the ratio of the edges of the reference triangle, there can be multiple triangles with the same GRT value but different reference locations. This results in different HDOP values and localization errors, on average. Typically, the best HDOP values can be achieved by placing the reference nodes at the edges of the network in a regular shape.

The relationship between AvgHDOP and GRT values is illustrated in Fig. 5(a). As can be seen, the relationship is linear when GRT is below 0.9 approximately. This indicates that, on average, GRT approximates the AvgHDOP, and we can use GRT instead of HDOP to measure the localization geometry. Actually, the relationship between the AvgHDOP and GRT results from the fact that when HDOP is averaged over the network (nodes) it evens the effect of the unknown node location and emphasizes the effect of the reference nodes. When GRT is above 0.9 approximately, AvgHDOP starts to increase fast. This results mostly from very high HDOP values at the locations close to or in line with the reference nodes, as can be seen in Fig. 4(c). Therefore, it is important to exclude the reference node combinations with high GRT values from location computation. It should be noted that the correlation between the AvgHDOP and GRT depends on the number and distribution of the unknown nodes.

The effect of GRT on localization error (MAE) is illustrated in Fig. 5(b). As can be seen, MAE clearly increases when GRT increases. Likewise, the increasing of distance measurement error (Δd) increases MAE. The effect of distance measurement error on localization error is bigger when localization geometry is worse (bigger GRT and AvgHDOP). This shows that localization geometry and ranging error both have a significant effect on localization error. When the localization geometry (GRT) gets worse, the variation in HDOP values increases. This results in higher variation in localization error, too, as the localization error of a node depends on its location relative to the reference nodes. To illustrate this, the localization error (MAE±SD) vs. GRT ($\epsilon = 0.1d$) is shown in Fig. 5(c). It reveals that the variation of localization error also increases when GRT increases, together with MAE. Therefore, it is very important to exclude the reference nodes with bad geometry (high GRT value). This way we can reduce the mean and standard deviation of localization error, and also reduce the effect of ranging error.

5. Adaptive range-based localization algorithm based on trilateration and reference node selection

Ranging errors and localization geometry vary throughout the network, and depend on the reference nodes applied for localization. This inevitably leads to varying localization errors for a node depending on the reference set used. Based on the findings from simulations and experiment, and further analysis, we propose an ARBL algorithm that aims to find the best combinations of reference nodes at a given time. The proposed algorithm computes an unknown node's location estimate for each combination of three reference nodes (i.e., trilateration) that fulfills the specified conditions, and selects n combinations based on a criterion to be used in the final location computation. The purpose is to reduce localization error by adaptively selecting the most appropriate reference nodes and location estimates at a particular time, while aiming to keep the algorithm as cost-effective and scalable as possible. Actually, the ARBL algorithm could be applied to the combinations of more than three reference nodes which might reduce the localization errors. In that case, however, the computation of GRT values and lateration would be computationally more expensive. In the following, a more detailed description of the algorithm is given.

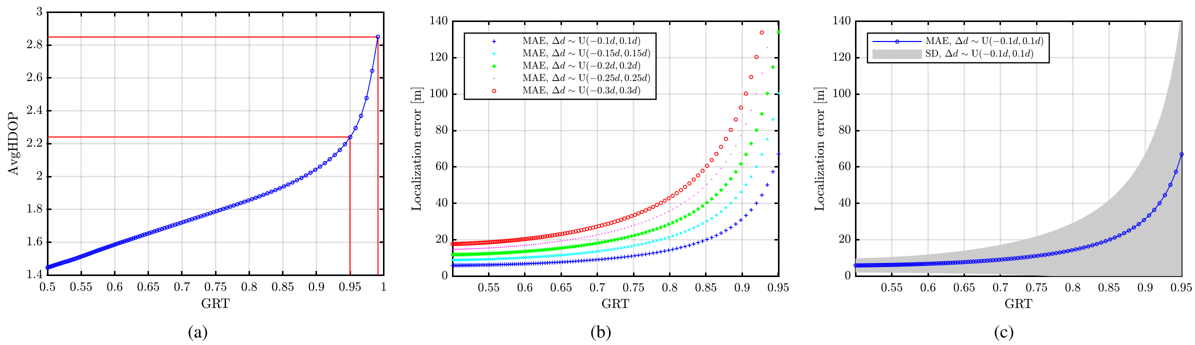


Fig. 5. The effect of localization geometry and ranging error on localization error (based on the simulation setup shown in Fig. 3). (a) AvgHDOP vs. GRT, (b) Localization error (MAE) vs. GRT, (c) Localization error (MAE±SD) vs. GRT ($n = 500$).

5.1. Algorithm description

The flow chart of the proposed algorithm (ARBL) is presented in Fig. 6. We divide the algorithm into four phases and discuss them in order.

5.1.1. Selecting reference combinations

In the beginning, an unknown node forms a reference set of the n best reference nodes based on the RSSI values; that is, $S_{REF} = \{i \mid RSSI_i \in \max_n RSSI\}$, where i is the reference node index. By ruling out the reference nodes with bad RSSIs, we are trying avoid the references that are distant or have a bad link. In general, distance estimates should be better for shorter distances [32]. Then, the node generates all possible 3-node combinations of S_{REF}; that is, $C(n, 3) = j$ combinations $(C_1, \dots, C_j)$. For example, if $n = 5$, the number of combinations $C(5, 3) = 10$. Using more than five reference nodes increases the number of combinations quickly; the total number of k-combinations of n nodes is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. If less than five reference nodes are available, the number of 3-node combinations would be $C(4, 3) = 4$ or $C(3, 3) = 1$, which would probably downgrade the performance. Therefore, for a proper functionality, at least five reference nodes are needed. For trilateration, three reference nodes at the minimum are required.

5.1.2. Evaluating reference triangles

The node evaluates the suitability of reference triangles for localization based on their geometry. For each combination C_i, the node computes the GRT value, $GRT = \frac{d_{\max}}{d_{\min} + d_{\mathrm{md}}}$, where $d_{\max}$, $d_{\min}$, and d_{md} are the maximum, minimum, and median edge length of the reference triangle C_i, respectively. To rule out possible badly formed triangles, we set the following condition for the ratio of the edges that must be fulfilled:

$$GRT_i < GRT_{th}, \tag{16}$$

where GRT_i is the GRT value for combination C_i, and GRT_{th} is the given threshold value for GRT. By changing the threshold value, GRT_{th}, the condition can be tightened or loosened. This condition rules out nearly collinear and otherwise badly formed reference node combinations which may have a negative effect on localization accuracy. The elimination of nearly collinear reference triangles is an essential part of the algorithm. Good triangles can tolerate larger distance errors and produce reasonable localization accuracy, while bad triangles may produce very large localization errors even with relatively small distance errors (see Fig. 5). It is possible that all combinations of the n closest reference nodes are unsuitable for localization. In a very bad geometry scenario ($GRT \geq GRT_{th}$), the initial location estimates will be computed for all the combinations.

In general, the unknown node should be inside the convex hull of the reference nodes. As we know only the location estimates, which are probably inaccurate, the reasoning based on them whether the node is inside or outside the convex hull is uncertain. Furthermore, convexity alone does not describe the suitability of the reference nodes. A node outside the convex hull may have a better localization geometry (HDOP) than inside it, depending on the situation (see, e.g., [39] and Fig. 4). In addition, calculating HDOP based on the location estimates can be misleading. However, as we know the shape of the reference triangles exactly, we chose to use that information to evaluate the suitability of the reference nodes. Further, GRT can be used to exclude the bad combinations before any locations are computed which is more energy-efficient.

5.1.3. Computing initial location estimates

If the conditional statement above is true for the particular reference triangle C_i, the node will compute its initial location estimate by applying the reference node coordinates and the distance estimates in trilateration. The node computes an initial location estimate, $(\hat{x}_i, \hat{y}_i)$, for each reference node combination C_i that fulfills the condition defined in Eq. (16).

5.1.4. Computing the final location estimate

To estimate the error of each location estimate $\hat{\mathbf{x}}$, we first compute the difference between the distance estimate $\hat{d}_i$ to the reference node i (e.g., based on RSSI) and the distance d'_i to the reference node i computed based on the location estimate. The difference, Δd_i, is computed as:

$$\Delta d_i = \hat{d}_i - \underbrace{\sqrt{(\hat{x} - x_i)^2 + (\hat{y} - y_i)^2}}_{d'_i}, \tag{17}$$

where $(\hat{x}, \hat{y})$ are the coordinates of the location estimate, and (x_i, y_i) is the location of reference node i. The basic principle is illustrated in Fig. 7.

As a criterion for trying to find the best location estimates, we use the average of the absolute distance differences, $\overline{\Delta d}$ (the smaller the better). Specifically, $\overline{\Delta d}$ is computed for each approved reference node combination of an unknown node as follows:

$$\overline{\Delta d} = \frac{1}{n} \sum_{i=1}^{n} \left| \hat{d}_i - \sqrt{(\hat{x} - x_i)^2 + (\hat{y} - y_i)^2} \right|, \tag{18}$$

where $n = 3$ is the number of reference nodes. The rationale behind the criterion is the hypothesis that the smaller the $\overline{\Delta d}$, probably the smaller the localization error, on average. This is actually close to the idea of the least squares estimation in localization, in which the sum of the squared residuals, that is, the differences between the location estimate and distance estimates, is minimized. A similar approach was applied in [11], for instance, where the average of the distance differences was used to check the validity of a location estimate. To justify this hypothesis, the relationship between GRT, average $\overline{\Delta d}$, and localization error is illustrated in Fig. 8 (based on the setup in Section 4). As can be seen in Fig. 8(a), $\overline{\Delta d}$, averaged over the nodes, increases with GRT, that is, when the localization geometry gets worse. Accordingly, localization

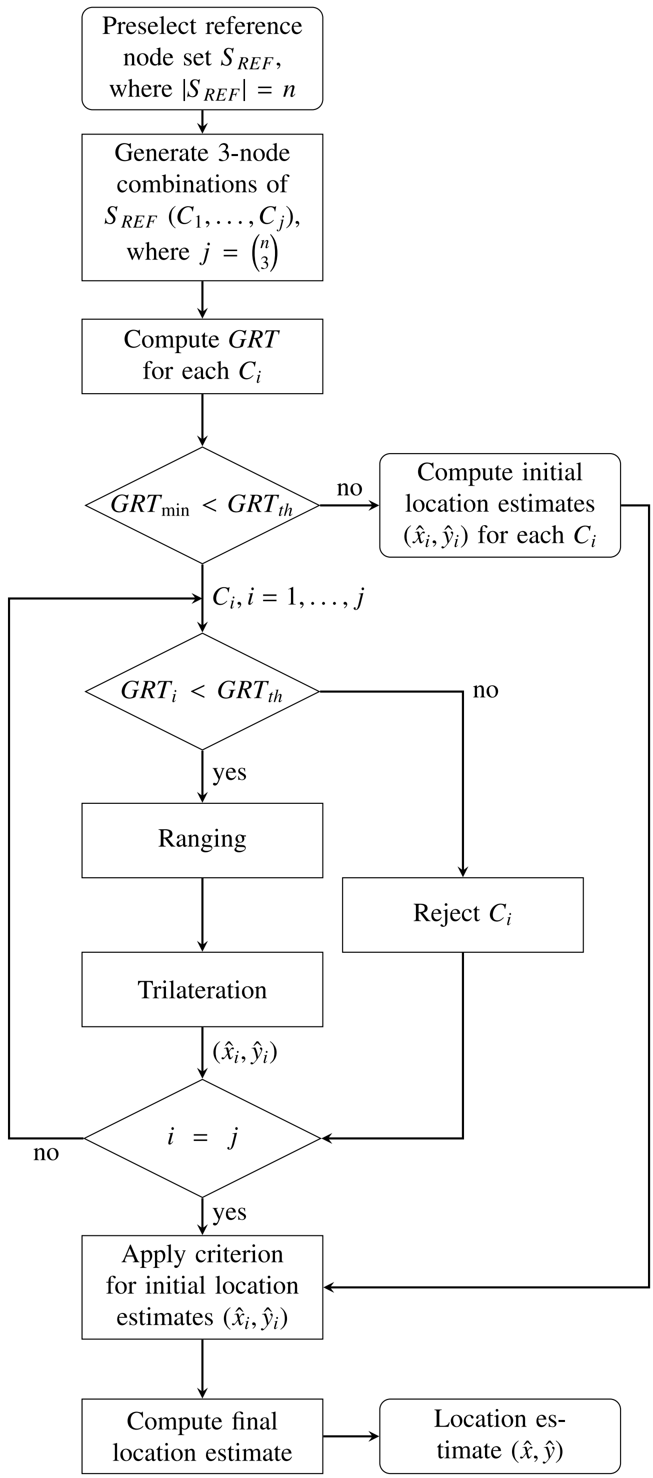


Fig. 6. Flow chart of the ARBL algorithm.

error (MAE) increases with average $\overline{\Delta d}$, on average. Bigger ranging error combined with bad geometry results in larger average $\overline{\Delta d}$ and, consequently, larger localization error (MAE). Therefore, $\overline{\Delta d}$ can be used as a justified approximation of localization error.

The final location estimate, $(\hat{x}, \hat{y})$, is then computed by averaging the n initial location estimates from the combinations that have the smallest

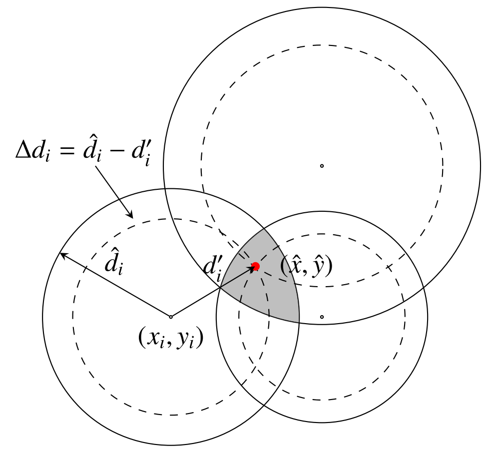


Fig. 7. Principle of computing the difference (Δd_i) between the distance estimate ($\hat{d}_i$) and the distance based on the location estimate (d'_i).

$\overline{\Delta d}$ as follows:

$$(\hat{x}, \hat{y}) = \left(\frac{1}{n} \sum_i \hat{x}_i, \frac{1}{n} \sum_i \hat{y}_i \right)_{|i \in \min_n \overline{\Delta d}}, \tag{19}$$

where i is the index of the reference set, and $(\hat{x}_i, \hat{y}_i)$ is the corresponding location estimate. To improve robustness, we use averaging instead of simply taking the location estimate with the smallest $\overline{\Delta d}$.

5.2. Implementation issues

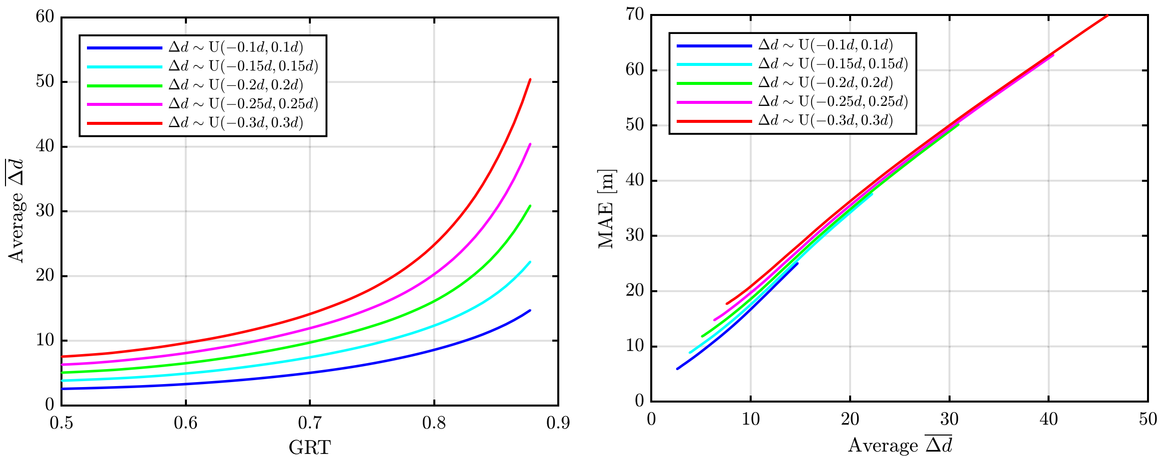
Although implementation of the proposed algorithm (ARBL) is not in the scope of this paper, we briefly discuss some of its aspects next. Overall, the algorithm is scalable and low cost, and it can be implemented in typical resource-constrained WSN nodes.

5.2.1. Selecting reference combinations

A node can initiate the localization process either by sending a request to the sink or reference nodes (active) or without request by receiving periodic beacon messages (passive). Either way, the sink or reference nodes broadcast beacon messages, which are received and forwarded by other nodes. This can be repeated consecutively for different radio channels (if used). Based on the link RSSI values, the node forms a reference set of the n best reference nodes, and generates its three-node combinations, that is, a total of $\binom{n}{3}$ combinations. Restricting the number of reference nodes to n (e.g., $n = 5$) makes the algorithm scalable. Moreover, using only three reference nodes (i.e., trilateration) makes the algorithm computationally lightweight and energy efficient. This is noteworthy, as the computational complexity of lateration increases rapidly with the number of reference nodes, thus making it excessively heavy for low-cost WSN nodes.

5.2.2. Evaluating reference triangles

The node needs to know the edge lengths of a reference triangle to evaluate its suitability for trilateration. The lengths can be computed based on the reference node coordinates received in the beacon messages. By changing the value of the threshold, GRT_{th}, it is possible to tighten or loosen the condition. Too small GRT_{th} would exclude too many triangles while too big GRT_{th} would include badly formed triangles. Further, this choice has an effect on localization accuracy, as was shown in Section 4.



(a) (b)

Fig. 8. An example of the relationship between GRT, average $\overline{\Delta d}$, and MAE based on the setup in Section 4. (a) The average of the absolute distance differences, $\overline{\Delta d}$, averaged over the nodes ($n = 500$) vs. GRT, (b) the mean localization error (MAE) vs. average $\overline{\Delta d}$ with different ranging errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2.3. Computing initial location estimates

The distance estimates needed in trilateration can be computed in many ways. Regardless, distance estimates can be easily computed in typical resource-constrained WSN nodes. In addition, if available, distance estimates can be computed using other ranging techniques than RSSI, such as ToA or TDoA, or together with RSSI as a hybrid approach.

In trilateration, the node uses the estimated link distances and the reference nodes' coordinates to compute the location estimate. There are alternative techniques for implementing the trilateration in WSN nodes, which may differ in accuracy and computational complexity.

5.2.4. Computing the final location estimate

For evaluating each initial location estimate, the node needs to know the distance estimates to particular reference nodes, and the coordinates of the location estimate and reference nodes. These are already known; thus, no extra communication is required. The final location estimate is then computed by averaging the coordinates of the n best initial location estimates based on the applied criterion.

5.3. Localization efficiency and scalability

The computation complexity of the ARBL algorithm is low. It uses only low-complex techniques that can be easily computed in resource-constrained sensor nodes. The computation of GRT for each reference node combination is very simple, requiring only basic arithmetic operations. Trilateration is easy to compute as it uses only three reference nodes and the related distance estimates in location computation. Further, trilaterations are computed only for the best combinations based on the GRT value. Computing the evaluation criterion for initial location estimates and the final location estimate is also simple. Using only a subset of nearby reference nodes and trilateration makes ARBL energy efficient in terms of computation and communication.

The scalability of ARBL is good. It is a distributed algorithm using only local information. Increasing the network size (number of nodes) or density does not affect the computation complexity, as the size of the reference node set is fixed.

6. Performance evaluation by simulation

The performance of the proposed algorithm was evaluated through MATLAB simulations. In the following, we first describe the data and methods used for evaluating and comparing the localization accuracy. Then, we present and analyze the results.

Table 2
Simulation parameters of the trilateration-based ARBL algorithm.

Parameter	Default value	Value/Range
Network area	—	100×100 m
Number of setups	—	200
Number of unknown nodes (N_{unk})	—	100
Total number of reference nodes (N_{ref})	15	$[6, 25]$
Number of nodes in reference set (N_{ref_set})	6	$[5, 10]$
GRT threshold value (GRT_{th})	0.95	$[0.65, 1]$
Number of estimates in final estimation (N_{final_est})	6	$\leq [1, 20]$
Ranging error ($\Delta d \sim U(-\varepsilon, \varepsilon)$)	$\varepsilon = 0.2d$	$\varepsilon = [0.1, 0.3]d$

6.1. Simulation data and methods

To evaluate the applicability of the ARBL algorithm in different scenarios, we analyzed its localization accuracy in several randomly generated networks. Further, we studied the effect of different parameters on its localization accuracy. The localization errors, averaged over the nodes, were computed for 200 different setups. Each setup consisted of 100 unlocalized nodes ($N_{unk} = 100$) and $6 - 25$ reference nodes ($N_{ref} = 6 - 25$), uniformly distributed in the area of 100×100 m ($x, y \sim U(0, 100)$). Distance measurement error, Δd, was uniformly distributed, that is, $\Delta d \sim U(-\varepsilon, \varepsilon)$, where $\varepsilon = xd$. Thus, we assume that the distance measurement error is relative to the distance between the unlocalized and the reference node. For simplicity, we also assume that each unlocalized node can directly communicate with the reference nodes (i.e., one-hop neighbors). The ARBL algorithm always uses reference nodes' 3-combinations in trilateration localization. The evaluation was performed with MATLAB (R2020a) [41]. The parameters used in the simulations and their values/ranges are shown in Table 2.

6.2. Simulation results

To evaluate the performance, we first compared ARBL to different localization approaches in terms of localization error (MAE). These were EATL [18], RNST [19], and lateration with a nearest-neighbors approach. For EATL and RNST, we applied few different parameters' values that were selected to be feasible in our setups. However, we did not use iterations for EATL as in [18]. In the nearest-neighbors approach (NN(k)), for each unlocalized node, we always selected k closest reference nodes for lateration. Here, we applied the NN(k) approach with the values of $k = \{3, 6, 9, 12\}$. In this scenario, the total number of reference nodes (N_{ref}) was 15, the number of nodes in a reference set

Table 3
Localization error [m] for ARBL, lateration with the nearest-neighbors approach (NN(k), $k = \{3, 6, 9, 12\}$), EATL [18], and RNST [19]. MAE and 90th Percentiles of localization error for the nodes, averaged over the setups.

Algorithm	Average[a]		EATL parameters[b]			RNST parameter[c]	Not located
	MAE	90th Percentile	R	r	$threshold_a$	equilateral triangle criteria	
ARBL	5.57	11.14					
Lateration with NN(3)	38.34	52.42					
Lateration with NN(6)	7.70	15.10					
Lateration with NN(9)	8.91	16.95					
Lateration with NN(12)	10.25	19.20					
EATL	7.07	12.63	50 m	$0.3R$	$0.8R$		50.1%
EATL	9.19	17.48	50 m	$0.3R$	$0.5R$		23.0%
EATL	18.12	34.90	100 m	$0.3R$	$0.4R$		1.3%
EATL	18.80	37.95	100 m	$0.15R$	$0.25R$		0%
RNST	7.59	14.28				$< 25°$	
RNST	6.73	12.63				$< 50°$	
RNST	6.60	12.28				$< 75°$	
RNST	6.83	12.74				$< 100°$	

[a]Averaged over 200 setups, 100 nodes in each.
[b]R = radio range, r = inner circle of the ring, $threshold_a$ = shortest edge length.
[c]Equilateral triangle criteria = $\sum_{i=1}^{3} |60° - \alpha_i|$.

(N_{ref_set}) was 6 (i.e., the closest reference nodes), GRT threshold value (GRT_{th}) was 0.95, and the number of initial location estimates chosen for the final location estimation based on the criterion (N_{final_est}) was 6. The distance measurement error $\Delta d \sim U(-0.2d, 0.2d)$. The localization error (MAE) of ARBL, EATL, RNST, and NN(k) with $k = 6$ (which gave the smallest average error) in each setup is shown in Fig. 9. As can be seen, ARBL outperforms the other approaches over the range of setups, with only few exceptions (it is the worst one in 2 out of 200 (1%), and the best one in 182 out of 200 (91%) setups). The high peaks are mostly due to few nodes that had very bad localization geometry. For example, the median error for the setup that has the highest MAE is about 5 m. The localization errors (MAE and the 90th percentiles of the errors) averaged over the setups for ARBL, EATL, RNST, and the lateration with nearest-neighbors approach ($k = \{3, 6, 9, 12\}$) are shown in Table 3. As can be seen, ARBL produces smaller localization error than by using the closest reference nodes, regardless the number of references. The localization error of ARBL is also smaller than that of EATL and RNST, although RNST gets quite close. For EATL, it should be noted that the high number of nodes is not located as we did not use any iterations. However, the results should give an estimate of its accuracy. Moreover, the performance of the algorithms depends on the parameter values used which makes their comparison quite challenging. Regardless, this shows that taking the effect of localization geometry and ranging error into account when selecting the references results in better outcome, in general. Furthermore, it is computationally cheaper to use only minimum number of references (= 3) in lateration.

Next, we investigated how different parameters affect the performance of ARBL when $\Delta d \sim U(-\varepsilon, \varepsilon)$, where $\varepsilon = \{0.1, 0.15, 0.2, 0.25, 0.3\}d$. First, the effect of the total number of reference nodes (N_{ref}) on localization error was studied with $N_{ref} = \{6, 10, 15, 20, 25\}$. The other parameters were set to the default values ($N_{ref\_set} = 6$, $GRT_{th} = 0.95$, $N_{final\_est} = 6$). The results are shown in Fig. 10(a). As expected, localization error decreases when the number of reference nodes increases, probably because there will be more references in a close proximity. Further, increasing ranging error results in increased localization error, which is also intuitive. When N_{ref} is 15 or more, localization error of ARBL drops below 10 m, on average, even for ranging error as high as 30%.

The effect of the number of nodes in a reference set (N_{ref_set}) on localization error was studied with $N_{ref\_set} = \{5, 6, 7, 8, 9, 10\}$, while the other parameters were set to the default values ($N_{ref} = 15, GRT_{th} = 0.95, N_{final\_est} = 6$). The results are shown in Fig. 10(b). As can be seen, the localization error seems to be quite independent of the value of N_{ref_set} (slightly increasing, particularly for high ranging errors), reaching its smallest value when $N_{ref\_set} = 6–8$. This is surprising, since increasing the number of nodes in a reference set (N_{ref_set}) increases the number of reference node combinations and, therefore, the number of potentially good combinations. However, the bigger the N_{ref_set} the bigger are the distances and distance errors between the references and the unlocalized node, on average, which may partly explain the behavior.

The effect of the GRT threshold value (GRT_{th}) on localization error was studied with $GRT_{th} = \{0.65, 0.7, \dots, 0.95, 0.96, \dots, 1\}$. The other parameters were set to the default values ($N_{ref} = 15, N_{ref\_set} = 6, N_{final\_est} = 6$). In this case, only the GRT_{th} condition of ARBL was used. The results are shown in Fig. 10(c). As can be seen, changing the value of GRT_{th} does not have much effect on localization error when GRT_{th} is below 0.9. However, above it, localization error starts to slowly increase. When GRT_{th} is above 0.99, the effect on localization error is drastic. This indicates that GRT_{th} should be used at least to exclude the combinations with very bad localization geometry.

The effect of the number of estimates in final estimation (N_{final_est}) on localization error was studied with $N_{final\_est} = \{1, \dots, 20\}$. The other parameters were set to the default values ($N_{ref} = 15, N_{ref\_set} = 6, GRT_{th} = 0.95$). The results are shown in Fig. 10(d). As can be seen, the localization error reduces first when N_{final_est} increases, reaching its optimal values between 5 and 7. Localization error starts to increase again after N_{final_est} gets bigger until it is around 15. Above that, increasing the value of N_{final_est} does not affect localization error much. This may be explained by the fact that there are probably not enough combinations that are below GRT_{th}, and therefore, the actual value used is smaller.

In summary, the simulation results show that, on average, the ARBL algorithm can produce quite accurate results in various network scenarios. Almost without exceptions, it outperforms the often used nearest-neighbors approach, regardless the number of reference nodes. In most cases, it also results in smaller localization error than the EATL [18] and RNST [19] algorithms, that are based on localization geometry in reference node selection. However, there are cases when the localization error of some nodes is increased due to bad geometry and ranging error.

7. Performance evaluation by experimental measurement

In addition to the simulation, we set up a sensor network to collect experimental measurement data for localization. In this section, we first describe the test equipment and data collection process, and the data and methods used for evaluating and comparing the localization accuracy and precision of ARBL. Next, we present and analyze the results.

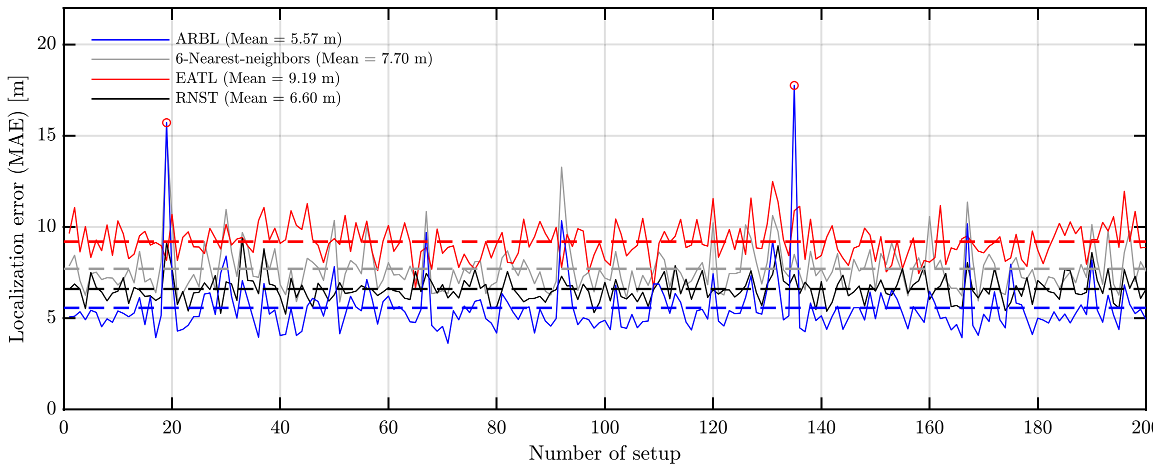


Fig. 9. Localization error (MAE) of ARBL, the best nearest-neighbors approach (NN(6)), EATL [18], and RNST [19] for different setups. The default simulation parameters for ARBL were the following: $N_{ref} = 15$, $N_{ref\_set} = 6$, $GRT_{th} = 0.95$, $N_{final\_est} = 6$, and $\Delta d \sim U(-0.2d, 0.2d)$. The red circles indicate the cases (2 out of 200) when ARBL resulted in higher MAE than the other approaches. For EATL, the parameter values were the following: $R = 50$ m, $r = 0.3R$, $threshold_a = 0.5R$. No iterations were used, and 23% of nodes were not located. For RNST, we classified reference triangles as 'almost equilateral' if $\sum_{i=1}^{3} |60° - \alpha_i| < 75°$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

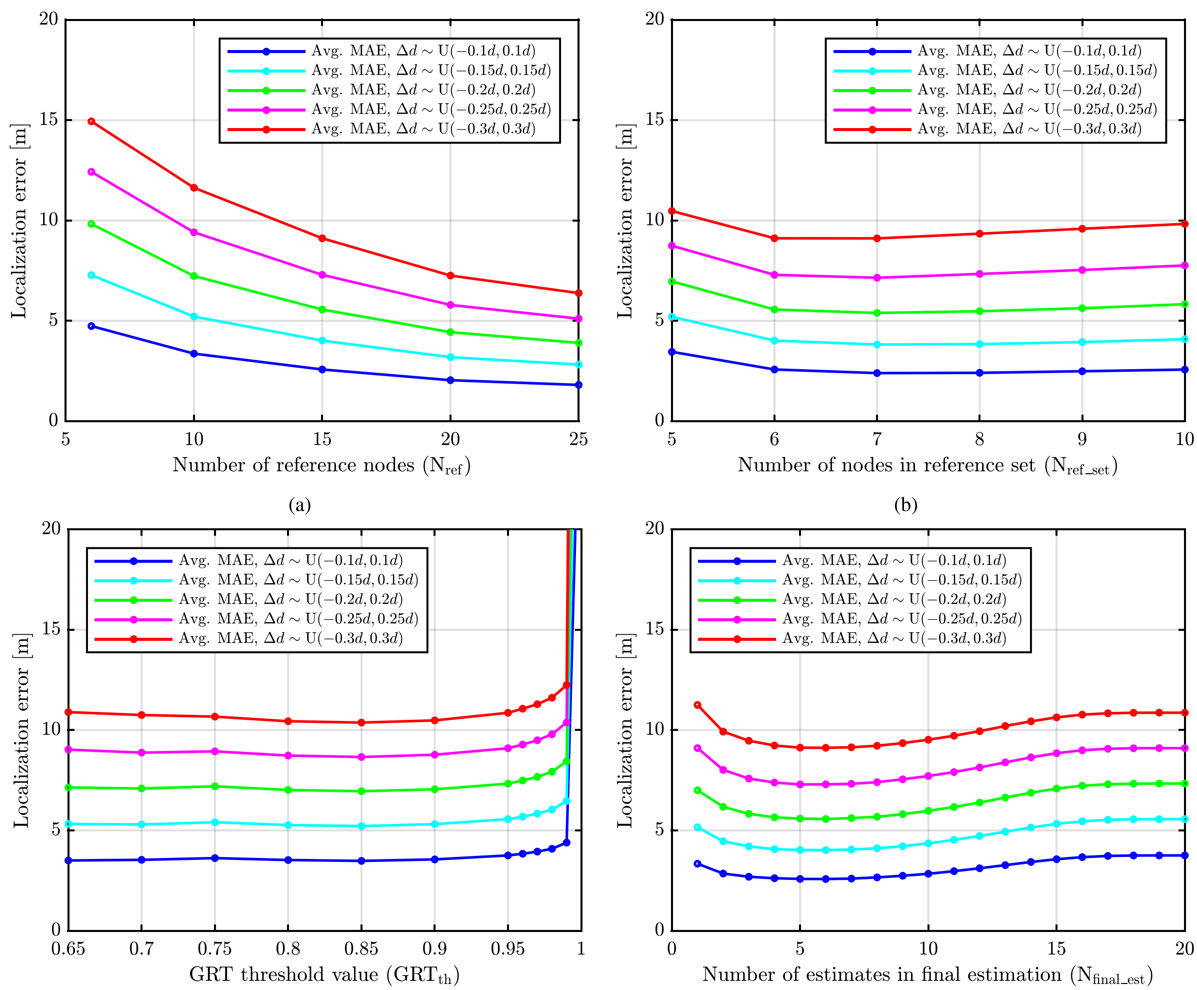


Fig. 10. (a) Localization error vs. the total number of reference nodes (N_{ref}) with $N_{ref\_set} = 6$, $GRT_{th} = 0.95$, and $N_{final\_est} = 6$. (b) Localization error vs. the number of nodes in a reference set (N_{ref_set}) with $N_{ref} = 15$, $GRT_{th} = 0.95$, and $N_{final\_est} = 6$. (c) Localization error vs. GRT threshold value (GRT_{th}) with $N_{ref} = 15$ and $N_{ref\_set} = 6$ (Note that only GRT_{th} condition of ARBL was applied in this case). (d) Localization error vs. the number of estimates in final estimation (N_{final_est}) with $N_{ref} = 15$, $N_{ref\_set} = 6$, and $GRT_{th} = 0.95$. Results averaged over 200 different setups. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

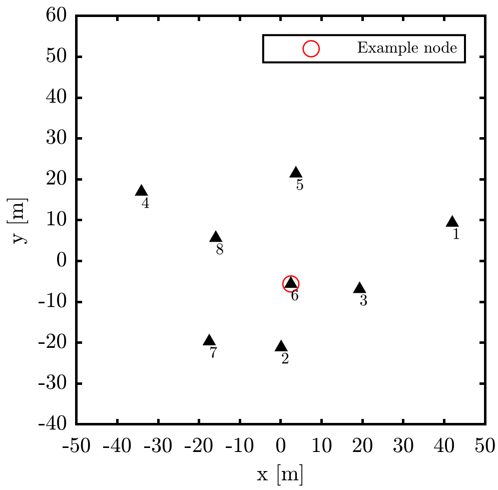


Fig. 11. Network setup of the nodes (relative locations) used in the experimental measurement. The set of the available reference nodes for any unlocalized node is all the other nodes. Note that the reference node set used in localization varies according to the given combination. An example node to be analyzed (node 6) is circled.

Table 4
Collected RSSI and temperature data (1 Min samples).

Data	Description
Timestamp[a]	yyyy-mm-dd hh:mm
Node	$0 \times 810\{1,2,3,4,5,6,7,8\}$
Neighbor	$0 \times 810\{0,1,2,3,4,5,6,7,8\}$
Channel	11 – 26 (IEEE 802.15.4)
RSSI	RSSI [dBm] at $P_{TX} = +3.0$ dBm
Timestamp[b]	yyyy-mm-dd hh:mm
Node[c]	$0 \times 810\{1,2,4,5\}$
Temperature	T [°C] (SHT75)

[a]RSSI data.
[b]Temperature data.
[c]The SHT sensor of node 4 was broken.

Table 5
Parameters of the ARBL algorithm used in the experiment.

Parameter	Value
Number of unknown nodes (N_{unk})	1 (of 8)
Total number of reference nodes (N_{ref})	7
Number of nodes in reference set (N_{ref_set})	5
GRT threshold value (GRT_{th})	0.85
Number of estimates in final estimation (N_{final_est})	4

7.1. Experimental setup and data collection

For evaluating the proposed algorithm in varying conditions, we set up a WSN aimed to collect RSSI and temperature data over long-term experiments. The configuration for collecting the empirical measurement data consisted, in total, of eight WSN nodes that used Atmel ZigBit 2.4 GHz wireless modules (ATZB-24-B0) [42] with an IEEE 802.15.4-compliant AT86RF230 radio transceiver [43]. Four of the nodes were integrated with a Sensirion humidity and temperature sensor (SHT75) [44]. In addition, the setup comprised a gateway (Atmel ZigBit 2.4 GHz sink node and Raspberry Pi 3) and a database server (MongoDB). The sensor nodes were attached to lamp posts around the university campus parking lot at a height of approximately 3 m by using mounting racks. The nodes were powered by secondary batteries which were charged by solar panels and discontinuous mains power (controlled with a timer and a PECU switch). The gateway was on the terrace of the university building, and it was mains-powered and connected to the LAN via Ethernet. The network setup is illustrated in Fig. 11.

The measuring cycle for collecting RSSI and temperature data was 1 min. Once a minute, the sink node sent a broadcast message at a transmit power level of +3.0 dBm (P_{TX}) by using one of the radio channels (11–26). The radio channel was changed every minute. In addition, the temperature readings for the nodes were measured once a minute with the SHT75 sensors, where available. The nodes sent the raw data directly to the gateway, from where the data was sent to the MongoDB database server to be stored and processed further. Data processing, analysis, and computing were performed using MATLAB (R2020a) [41]. The collected and applied data is shown in Table 4.

For evaluation, the inter-node distances were measured using a laser distance meter (Leica DISTO D8), where possible. The distances that could not be measured were computed by applying trigonometry. Relative node locations were computed based on the inter-node distances by using classical multidimensional scaling (MDS). We used the MATLAB *cmdscale* function [41] to generate the configuration matrix based on a distance matrix. The relative locations (can be rotated and/or reflected) were set as the ground truth node locations against which the estimates were compared. The mean error for the reconstruction was 0.02 m (max = 0.06 m), that is, the mean absolute difference between the measured distances and the distances based on the coordinates.

To estimate the inter-node distances for localization, we used RSSI-based ranging techniques based on the ones proposed in [16]. However, range estimates can be obtained using any appropriate technique. Based on the distance estimates and the reference node coordinates, location estimates were computed using lateration.

7.2. Experimental data and methods

To evaluate the performance of the proposed localization algorithm (ARBL), we analyzed the localization error of eight unknown nodes based on real measurement data. The data used in the evaluation consisted of three separate, 2-week periods, that is, a total of 6 weeks of data. The data was collected (i) between 15 and 28 April 2019, (ii) between 27 May and 9 June 2019, and (iii) between 8 and 21 July 2019. During the first measurement period, there were no leaves on the trees of the parking lot, while during the second and third periods, the trees were in full leaf. The third period was during the summer vacation when the traffic in the parking lot was sparse, unlike during the other periods.

We compared the results given by the ARBL algorithm to the lateration with a nearest-neighbors approach (NN(k) with $k = \{3, 4, 5, 6\}$) and to the results given by lateration when using single reference node combinations, or a set of k reference nodes combinations ($k = 3, \ldots, 6$). Out of the eight nodes, one node at a time was the unknown node ($N_{unk} = 1$), and the seven others formed the set of reference nodes from which the reference nodes for lateration were taken ($N_{ref} = 7$). The k-node combinations of the reference nodes from a set of seven nodes makes a total of $\sum_{k=3}^{6} C(7,k) = 98$ combinations for an unknown node. Two of the combinations ($C = \{1,2,3\}, C = \{4,6,8\}$) were nearly collinear, which resulted in a very large localization error, and thus, had a disproportionate big effect on the overall results. Therefore, we excluded those combinations from the other computations and figures except those for the algorithms (ARBL and NN(k)). The ARBL algorithm can reject bad three-node combinations solely based on the given condition (Eq. (16)) when evaluating the reference triangles. In the experiment, we used the ARBL parameter values that were found to be more suitable for this particular network than those used in the simulations. Specifically, the number of nodes in reference set (N_{ref_set}) was 5, the GRT threshold value (GRT_{th}) was 0.85, and the number of estimates in the final estimation (N_{final_est}) was 4. The parameter values for ARBL used in the computations are shown in Table 5. The evaluation was carried out using MATLAB (R2020a) [41].

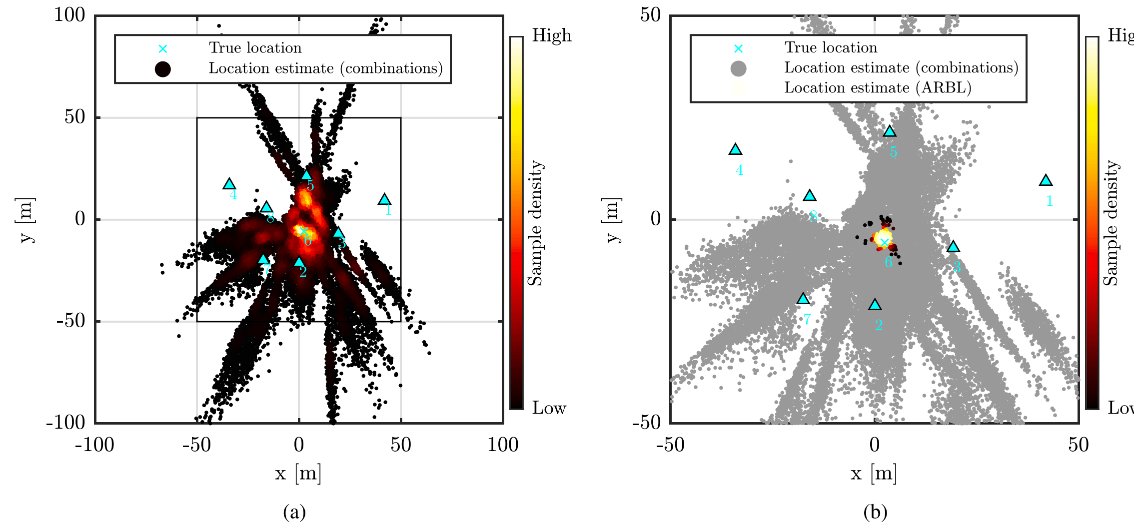


Fig. 12. (a) The location estimates (n = 97776) of node 6 for 6 weeks (1008 h) using the different reference node combinations (a total of 97). The dot of estimate $\hat{\mathbf{x}}_i$ is colored based on the density of the estimates in the circle area centered in $\hat{\mathbf{x}}_i$ and having a radius r = 2.5 m (lightness increases monotonically with density). In (b), respectively, the gray dots are the location estimates produced using the different reference node combinations, and the colored dots are the location estimates (n = 1008) produced using the proposed ARBL algorithm. The figure (b) is the magnification of the highlighted area shown in figure (a). In both figures, each dot represents a 1 h estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

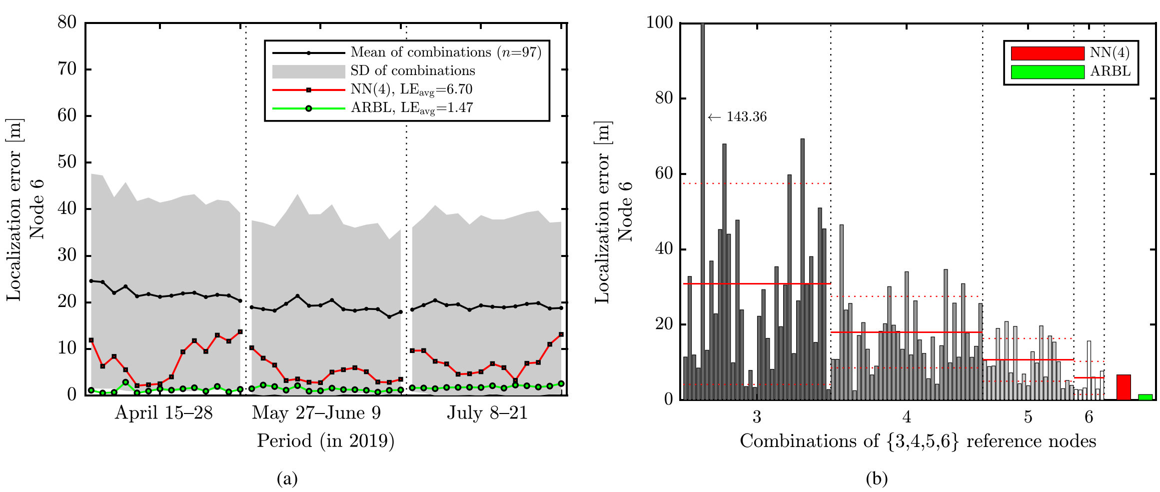


Fig. 13. (a) Localization error versus time for node 6. Combinations (mean ± SD, n = 97), NN(4), and the ARBL algorithm. The localization errors are presented as the averages of 24 h computed based on 1 h localization errors. (b) Mean localization errors (n = 1008) of each reference node combination (97), NN(4), and the ARBL algorithm computed based on 6 weeks of data for node 6 (bars). Mean ± SD of the mean values of the combinations for each number of reference nodes (lines).

7.3. Experimental results

We first take one node (node 6) as an example, and analyze its localization error. The location estimates (n = 97776) obtained using the different reference node combinations (a total of 97) and the ARBL algorithm for node 6 over the measurement period (6 weeks) are shown in Fig. 12. As can be seen from the figure (a), the location estimates vary considerably, in the order of tens of meters, when the different reference combinations are used. However, the densest area of estimates is quite close to the true location for node 6. This implies that the usage of lateration can result in very accurate estimates in the right conditions, that is, when the ranging estimates and the localization geometry are good enough. Therefore, the location estimates with the ARBL algorithm shown in Fig. 12(b) are exceptionally accurate and precise; the error is less than 2 m for node 6, on average, without any major deviations.

The localization errors of unknown node 6 over the 6 weeks of measurement data are illustrated in Fig. 13(a) and (b), both highlighting different aspects. The localization errors of the combinations (daily mean ± SD), lateration with nearest-neighbors (NN(4)) approach (daily mean), and the ARBL algorithm (daily mean) are shown in Fig. 13(a). As can be seen from the figure, the localization errors vary considerably between different reference node combinations, as well as temporally. This results from the combined effect of different ranging errors and localization geometry. What stands out in the figure is the clear benefit of the proposed algorithm. On average, it can reduce the localization error substantially compared to single combinations; the error is much smaller than the mean error of the combinations. In addition, for the ARBL algorithm, the variation in the errors is quite small for node 6. For node 6, the error of the nearest-neighbors approach (NN(4)) is also quite good, but the variation of daily mean is high.

The mean localization errors of each reference node combination, the nearest-neighbors approach (NN(4)), and the ARBL algorithm computed over the measurement periods (6 weeks) are presented in Fig. 13(b). The data in the figure shows the obvious variation in the localization error depending on the reference nodes used. This signifies the importance of appropriate reference node selection. On average, it seems that the bigger the number of reference nodes, the smaller the

Table 6
Absolute and Relative Localization Errors (LEs) based on 6 weeks of data.

LE [m]	Absolute LE [m] Average of nodes[a]	Relative LE [R] Average of nodes[a]
LE of combinations[b], Mean ± SD	34.50 ± 35.71	0.43 ± 0.45
Max	284.12	3.55
Min	7.63	0.10
n	96–97[c]	96–97[c]
Lateration, avg. of k-node combinations		
3 REFS $(\bar{x}, \bar{y})$, Mean ± SD	21.70 ± 5.02	0.27 ± 0.06
4 REFS $(\bar{x}, \bar{y})$, Mean ± SD	18.67 ± 3.65	0.23 ± 0.05
5 REFS $(\bar{x}, \bar{y})$, Mean ± SD	19.12 ± 3.25	0.24 ± 0.04
6 REFS $(\bar{x}, \bar{y})$, Mean ± SD	18.90 ± 3.06	0.24 ± 0.04
All $(\bar{x}, \bar{y})$, Mean ± SD	19.33 ± 3.79	0.24 ± 0.05
Lateration with nearest-neighbors approach (NN(k))		
Lateration with NN(3), Mean ± SD	19.44 ± 20.49	0.24 ± 0.26
Lateration with NN(4), Mean ± SD	17.28 ± 3.37	0.22 ± 0.04
Lateration with NN(5), Mean ± SD	18.53 ± 3.59	0.23 ± 0.04
Lateration with NN(6), Mean ± SD	20.64 ± 3.19	0.26 ± 0.04
ARBL, Mean ± SD	15.48 ± 3.19	0.19 ± 0.04

[a]Nodes 1–8.
[b]Average LE of each combination over 6 weeks of data.
[c]One or two collinear combination(s) excluded.

localization error (the mean and the standard deviation). Nevertheless, a few very good combinations can be found using only three reference nodes. Despite the large variation between the combinations, the ARBL algorithm seems to find the good combinations with a high probability and produces fairly accurate and precise location estimates. For node 6, the localization error of the ARBL algorithm is smaller than the localization error of any single combination. Also NN(4) achieves good accuracy in this case.

The results presented above are taken as an example of one node that has quite optimal yet realistic localization geometry. The average localization errors of the unknown nodes (nodes $1-8$) are summarized in Table 6, using all the combinations. The localization errors are presented as the absolute metric and relative to the estimated radio range. The estimated radio range is 80 m for all the nodes in the computations (approximately the same as the maximum link distance in the setup). We discuss only the absolute localization errors, as they are comparable to relative errors. It can be seen from the data in Table 6 that the localization error of the ARBL algorithm is 15.48 ± 3.19 m when averaged over all the nodes (1.47 ± 0.95 m for node 6). It is much smaller than the average error of combinations, 34.50 ± 35.71 m, and smaller than the error of the lateration with the best nearest-neighbor approach NN(4), 17.28 ± 3.37 m. For node 6, the localization error of the ARBL algorithm is even smaller than the mean localization error of the best single combination, 2.49 m, while it is about twice as big as the average best combination, 7.63 m, when taking all the nodes into account. On average, the ARBL algorithm reduced the mean localization error by 55% compared to the combinations' average error (15.48 m vs. 34.50 m). This shows the clear advantage of the proposed algorithm over the random selection of reference nodes. Also, the nearest-neighbors approach resulted in quite a good results; the best one was NN(4) on average. However, this particular scenario was quite challenging for both.

Furthermore, the cumulative distribution function (CDF) of the localization error for the ARBL algorithm compared to the average coordinates of the 4-node combinations, and to the nearest-neighbors approach (NN(4)) for node 6 is illustrated in Fig. 14. For node 6, the localization error of the ARBL algorithm was within 1.29 m in 50% of the cases, and within 2.35 m in 90% of the cases. It is clearly the best candidate and quite close to the theoretical optimal value, by which we assume that a node is able to pick every time the best combination (of all k-node combinations, $k = 3, \dots, 6$) with the minimum localization error. The corresponding 50th and 90th percentiles when averaged over all the nodes are shown in Table 7. In that case, the corresponding 50th and 90th percentile errors were 15.23 m and 19.71 m, respectively. Although considerably worse than for node 6, they are still better than those of nearest-neighbor approach and the average coordinates of k-node combinations. This highlights the importance of feasible reference node selection.

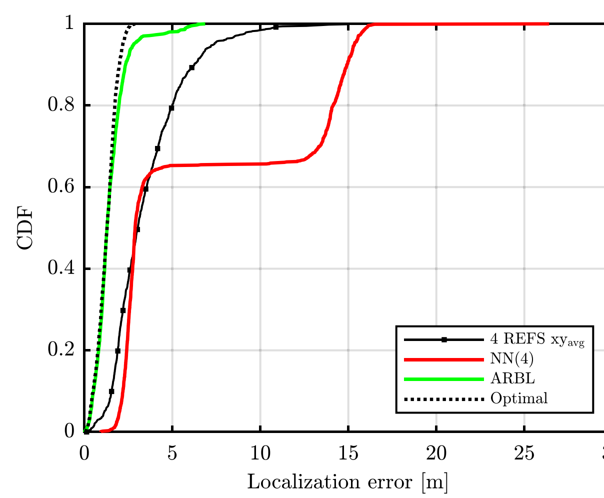


Fig. 14. Cumulative distribution function (CDF) of the localization error for node 6. The dashed line illustrates the optimal case, by which we assume that every time the reference combination with a minimum localization error was used in localization. The different behavior of NN(4) approach in is due to its location estimates are distributed in two clusters; one close to the true location, the other farther from it.

In summary, these results show that the proposed algorithm can produce quite accurate location estimates under varying ranging errors and localization geometry. These conditions are often inevitable when range-based trilateration is applied in real outdoor environments, particularly in the case of RSSI-based ranging. Moreover, the algorithm is computationally low cost, and can be implemented in resource-constrained sensor nodes by using the available ranging technique(s) that meet(s) the given requirements.

8. Discussion and conclusion

It is well-known that RSSI-based ranging is an error-prone distance estimation technique. Combined with lateration which is sensitive to

Table 7
50th and 90th Percentiles of Localization Error (LE).

LE [m]	50th Percentile[b] Average of nodes[a]	90th Percentile[b] Average of nodes[a]
Lat., avg. of k-combinations		
3 REFS $(\bar{x}, \bar{y})$	21.50	27.97
4 REFS $(\bar{x}, \bar{y})$	18.60	23.15
5 REFS $(\bar{x}, \bar{y})$	19.05	23.02
6 REFS $(\bar{x}, \bar{y})$	18.80	22.58
Lateration with NN(k)		
NN(3)	16.89	20.82
NN(4)	16.68	21.28
NN(5)	18.72	22.36
NN(6)	20.80	24.47
ARBL	15.23	19.71
Optimal[c]	3.52	6.16

[a]Nodes 1–8.
[b]The number of samples $n = 1008$.
[c]The combinations with a minimum LE.

erroneous distance estimates and localization geometry, RSSI-based ranging may lead to substantial localization errors depending on the reference nodes used, as was demonstrated in this experiment. This holds true for range-based techniques in general. In varying environmental and weather conditions, achieving location estimates that are accurate and precise is challenging. As a solution, we proposed an ARBL algorithm that is based on reference node selection. The algorithm aims to find the best reference node combinations for an unknown node at a given time and space based on the geometry of the reference triangles and ranging errors, and employs them to compute the node's location estimate. The results indicate that the localization error of the ARBL algorithm is considerably smaller than the average error of the single combinations, and that it outperforms the often used nearest-neighbors technique, as well as EATL [18] and RNST [19] algorithms with the applied simulation parameters. The algorithm can adapt to changing conditions and find the best reference node combinations for a particular case very successfully. For the practical implementation, the ARBL algorithm can be implemented in resource-constrained WSN nodes with limited processing, memory, and communication capabilities. The ARBL algorithm is practicable as it can be assumed that there are five reference nodes in most WSNs at a minimum, thus providing enough combinations for the algorithm.

The proposed algorithm is virtually range-based and independent of the ranging technique employed. Although RSSI was used for range estimation in this study (for the empirical data), distance estimates could be obtained using any other ranging technique, such as ToA or TDoA, if available. Naturally, the ranging technique affects the localization accuracy. Furthermore, the algorithm can be applied in multihop cases where an unknown node is out of the reference nodes' radio range. In that case, the distances to the reference nodes could be estimated, for example, based on the shortest paths. Alternatively, localization could be performed iteratively and cooperatively so that an unknown node will become a reference after it is localized, or by extending trilateration using the shadow edges, as in [45].

Despite the promising results, there is always room for improvement. One possible way would be to incorporate more location estimation techniques (e.g., Min–max) in the ARBL algorithm to mitigate the shortcomings of lateration, or use different trilateration techniques that are more robust to noisy distance estimates (e.g., weighted techniques). Furthermore, additional studies are needed to evaluate the performance of the algorithm in multihop situations, for example. Minimizing the costs of the algorithm (e.g., in the ranging phase) in addition to the localization error is also an aspect for future studies.

As stated in the literature (e.g., in [2,5,11,46]), one algorithm is not suitable or the best option for all possible scenarios and conditions. The choice of the algorithm depends mostly on the specific application scenario, thus requiring a combination of techniques to handle various situations [2]. However, developing techniques and algorithms that can adapt to different scenarios and changing conditions without requiring separate algorithms for each is important.

In conclusion, in this paper we set out to find ways to improve the quality of range-based localization for low-cost WSN nodes in varying outdoor conditions. We first analyzed the effect of ranging error and localization geometry on localization error based on simulations. As a solution, we proposed an ARBL algorithm that is based on reference node selection, and employs multiple reference node combinations to find the best ones to compute the final location estimate. The evaluation results show that the proposed algorithm reduced the localization error considerably. These promising findings indicate that achieving reasonable localization accuracy using range-based localization for low-cost, resource-constrained WSN nodes is possible by employing applicable techniques and information. The findings also provide new insights into anchor- and range-based localization that may be useful for future studies.

CRediT authorship contribution statement

Jari Luomala: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Ismo Hakala:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the Finnish Cultural Foundation, Central Ostrobothnia Regional fund.

References

[1] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Comput. Netw. 52 (2008) 2292–2330, https://doi.org/10.1016/j.comnet.2008.04.002.
[2] T.J. Chowdhury, C. Elkin, V. Devabhaktuni, D.B. Rawat, J. Oluoch, Advances on localization techniques for wireless sensor networks: A survey, Comput. Netw. 110 (2016) 284–305, https://doi.org/10.1016/j.comnet.2016.10.006.
[3] W. Dargie, C. Poellabauer, Fundamentals of Wireless Sensor Networks: Theory and Practice, in: Wiley Series on Wireless Communications and Mobile Computing, John Wiley & Sons Ltd., 2010.
[4] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, A survey on sensor networks, IEEE Commun. Mag. 40 (8) (2002) 102–114, https://doi.org/10.1109/MCOM.2002.1024422.
[5] A. Boukerche, H.A.B.F. Oliveira, E.F. Nakamura, A.A.F. Loureiro, Localization systems for wireless sensor networks, IEEE Wirel. Commun. 14 (6) (2007) 6–12, https://doi.org/10.1109/MWC.2007.4407221.
[6] A. Mesmoudi, M. Feham, N. Labraoui, Wireless sensor networks localization algorithms: A comprehensive survey, Int. J. Comput. Netw. Commun. (IJCNC) 5 (6) (2013) 45–64, https://doi.org/10.5121/ijcnc.2013.5603.
[7] N.A. Alrajeh, M. Bashir, B. Shams, Localization techniques in wireless sensor networks, Int. J. Distrib. Sens. Netw. 2013 (2013) 9, https://doi.org/10.1155/2013/304628.
[8] G. Han, H. Xu, T.Q. Duong, J. Jiang, T. Hara, Localization algorithms of wireless sensor networks: a survey, Telecommun. Syst. 52 (2013) 2419–2436, https://doi.org/10.1007/s11235-011-9564-7.
[9] L. Cheng, C. Wu, Y. Zhang, H. Wu, M. Li, C. Maple, A survey of localization in wireless sensor network, Int. J. Distrib. Sens. Netw. 2012 (2012) 12, https://doi.org/10.1155/2012/962523.
[10] A.K. Paul, T. Sato, Localization in wireless sensor networks: A survey on algorithms, measurement techniques, applications and challenges, J. Sens. Actuator Netw. 6 (4) (2017) https://doi.org/10.3390/jsan6040024.
[11] K. Langendoen, N. Reijers, Distributed localization in wireless sensor networks: a quantitative comparison, Comput. Netw. 43 (4) (2003) 499–518.

[12] A. Savvides, H. Park, M.B. Srivastava, The bits and flops of the N-hop multilateration primitive for node localization problems, in: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, WSNA'02, Atlanta, GA, USA, 2002, pp. 112–121, https://doi.org/10.1145/570738.570755.

[13] K. Bannister, G. Giorgetti, S.K. Gupta, Wireless sensor networking for "hot" applications: Effects of temperature on signal strength, data collection and localization, in: The Fifth Workshop on Embedded Networked Sensors, HotEmNets '08, Charlottesville, VA, USA, 2008.

[14] H. Wennerström, F. Hermans, O. Rensfelt, C. Rohner, L. Nordén, A long-term study of correlations between meteorological conditions and 802.15.4 link performance, in: 2013 IEEE International Conference on Sensing, Communications and Networking, SECON, New Orleans, LA, USA, 2013, https://doi.org/10.1109/SAHCN.2013.6644981.

[15] J. Luomala, I. Hakala, Effects of temperature and humidity on radio signal strength in outdoor wireless sensor networks, in: M. Ganzha, L. Maciaszek, M. Paprzycki (Eds.), Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, Vol. 5, FedCSIS 2015, Łódź, Poland, in: ACSIS, 2015, pp. 1247–1255, https://doi.org/10.15439/2015F241.

[16] J. Luomala, I. Hakala, Analysis and evaluation of adaptive RSSI-based ranging in outdoor wireless sensor networks, Ad Hoc Netw. 87 (2019) 100–112, https://doi.org/10.1016/j.adhoc.2018.10.004.

[17] Z. Yang, Y. Liu, Quality of trilateration: Confidence-based iterative localization, IEEE Trans. Parallel Distrib. Syst. 21 (5) (2010) 631–640, https://doi.org/10.1109/TPDS.2009.90.

[18] Y. Fan, X. Qi, B. Yu, L. Liu, A distributed anchor node selection algorithm based on error analysis for trilateration localization, Math. Probl. Eng. 2018 (2018) 12, https://doi.org/10.1155/2018/7295702.

[19] G. Han, D. Choi, W. Lim, Reference node placement and selection algorithm based on trilateration for indoor sensor networks, Wirel. Commun. Mob. Comput. 2009 (9) (2009) 1017–1027, https://doi.org/10.1002/wcm.651.

[20] G. Mao, B. Fidan, B.D. Anderson, Wireless sensor network localization techniques, Comput. Netw. 51 (10) (2007) 2529–2553, https://doi.org/10.1016/j.comnet.2006.11.018.

[21] A. Pal, Localization algorithms in wireless sensor networks: Current approaches and future challenges, Netw. Protoc. Algorithms 2 (1) (2010) 45–73, https://doi.org/10.5296/npa.v2i1.279.

[22] S. Tian, X. Zhang, X. Wang, P. Sun, H. Zhang, A selective anchor node localization algorithm for wireless sensor networks, in: 2007 International Conference on Convergence Information Technology, ICCIT 2007, Gyeongju, South Korea, 2007, pp. 358–362, https://doi.org/10.1109/ICCIT.2007.145.

[23] H. Nomura, H. Ichikawa, Y. Kawakita, Reference node selection for range-based localization using hierarchical clustering, in: 2018 IEEE 4th World Forum on Internet of Things, WF-IoT, Singapore, Singapore, 2018, pp. 140–143, https://doi.org/10.1109/WF-IoT.2018.8355228.

[24] X. Yan, Q. Luo, Y. Yang, S. Liu, H. Li, C. Hu, ITL-MEPOSA: Improved trilateration localization with minimum uncertainty propagation and optimized selection of anchor nodes for wireless sensor networks, IEEE Access 7 (2019) 53136–53146, https://doi.org/10.1109/ACCESS.2019.2911032.

[25] D. Lieckfeldt, J. You, D. Timmermann, An algorithm for distributed beacon selection, in: Sixth Annual IEEE International Conference on Pervasive Computing and Communications, PerCom,Hong Kong, China, 2008, pp. 318–323, https://doi.org/10.1109/PERCOM.2008.78.

[26] D. Lieckfeldt, J. You, D. Timmermann, Distributed selection of references for localization in wireless sensor networks, in: Proceedings of the 5th Workshop on Positioning, Navigation and Communication 2008, WPNC'08, Hannover, Germany, 2008, pp. 31–36, https://doi.org/10.1109/WPNC.2008.4510354.

[27] A. Bel, J.L. Vicario, G. Seco-Granados, Node selection for cooperative localization: Efficient energy vs. Accuracy trade-off, in: 2010 5th International Symposium on Wireless Pervasive Computing, ISWPC, Modena, Italy, 2010, pp. 307–312, https://doi.org/10.1109/ISWPC.2010.5483734.

[28] A. Bel, J.L. Vicario, G. Seco-Granados, Real-time path loss and node selection for cooperative localization in wireless sensor networks, in: 2010 IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops, Istanbul, Turkey, 2010, pp. 283–288, https://doi.org/10.1109/PIMRCW.2010.5670380.

[29] A. Bel, J.L. Vicario, G. Seco-Granados, Localization algorithm with on-line path loss estimation and node selection, Sensors 11 (7) (2011) 6905–6925, https://doi.org/10.3390/s110706905.

[30] S. Hadzic, J. Rodriguez, Utility based node selection scheme for cooperative localization, in: 2011 International Conference on Indoor Positioning and Indoor Navigation, uimaraes, Portugal, 2011, https://doi.org/10.1109/IPIN.2011.6071914.

[31] S. Hadzic, J. Bastos, J. Rodrigues, Reference node selection for cooperative positioning using coalition formation games, in: 2012 9th Workshop on Positioning, Navigation and Communication, Dresden, Germany, 2012, pp. 105–108, https://doi.org/10.1109/WPNC.2012.6268747.

[32] N. Patwari, J.N. Ash, S. Kyperountas, A.O. Hero III, R.L. Moses, N.S. Correal, Locating the nodes: Cooperative localization in wireless sensor networks, IEEE Signal Process. Mag. 22 (4) (2005) 54–69, https://doi.org/10.1109/MSP.2005.1458287.

[33] H. Karl, A. Willig, Protocols and Architectures for Wireless Sensor Networks, John Wiley & Sons Ltd., 2006.

[34] A. Savvides, W.L. Garber, R.L. Moses, M.B. Srivastava, An analysis of error inducing parameters in multihop sensor node localization, IEEE Trans. Mob. Comput. 4 (6) (2005) 567–577.

[35] J. Zhao, W. Xi, Y. He, Y. Liu, X.-Y. Li, L. Mo, Z. Yang, Localization of wireless sensor networks in the wild: Pursuit of ranging quality, IEEE/ACM Trans. Netw. 21 (1) (2013) 311–323, https://doi.org/10.1109/TNET.2012.2200906.

[36] G. Mao, B.D. Anderson, B. s Fidan, Path loss exponent estimation for wireless sensor network localization, Comput. Netw. 51 (10) (2007) 2467–2483, https://doi.org/10.1016/j.comnet.2006.11.007.

[37] C. Bettstetter, On the minimum node degree and connectivity of a wireless multihop network, in: MOBIHOC'02, Lausanne, Switzerland, 2002, pp. 80–91.

[38] D. Bartlett, Essentials of Positioning and Location Technology, Cambridge University Press, 2013.

[39] N. Levanon, Lowest GDOP in 2-D scenarios, IEE Proc.-Radar, Sonar Navig. 147 (3) (2000) 149–155.

[40] I. Hakala, J. Luomala, Peer-to-peer cooperative GNSS-based localization for stationary reference nodes in wireless sensor networks, in: 2017 International Conference on Localization and GNSS, ICL-GNSS, Nottingham, UK, 2017, https://doi.org/10.1109/ICL-GNSS.2017.8376241.

[41] MATLAB - The language of technical computing, URL http://www.mathworks.se/products/matlab/.

[42] Zigbit 2.4 GHz wireless modules - ATZB-24-A2/B0 datasheet, 2009, Atmel, http://www.atmel.com.

[43] Low power 2.4 GHz transceiver for ZigBee, IEEE 802.15.4, 6loWPAN, RF4ce and ISM applications - AT86rf230 datasheet, 2009, Atmel, http://www.atmel.com.

[44] Datasheet SHT7x (SHT71, SHT75) - humidity and temperature sensor IC, version 5, 2011, Sensirion, http://www.sensirion.com.

[45] G. Oliva, S. Panzieri, F. Pascucci, R. Setola, Sensor networks localization: Extending trilateration via shadow edges, IEEE Trans. Automat. Control 60 (10) (2015) 2752–2755, https://doi.org/10.1109/TAC.2015.2404253.

[46] J. Luomala, I. Hakala, Towards adaptive localization in wireless sensor networks, in: 2012 Ubiquitous Positioning, Indoor Navigation, and Location Based Service, UPINLBS, Helsinki, Finland, 2012, https://doi.org/10.1109/UPINLBS.2012.6409778.

Jari Luomala is a researcher at the University of Jyväskylä, Kokkola University Consortium Chydenius. He received the M.Sc. degree in mathematical information technology from the University of Jyväskylä, Finland, in 2011. He is currently pursuing the Ph.D. degree in software and communications engineering at the University of Jyväskylä. His current research interests include localization algorithms and techniques for wireless sensor networks.

Ismo Hakala is a professor of computer science at the Kokkola University Consortium Chydenius (KUC), University of Jyväskylä, Finland. He received his M.Sc. (1989), Ph. Lic. (1992), and Ph.D. (1997) degrees in mathematics at the University of Oulu, Finland. He joined at the University of Jyväskylä in 1999. He is currently the head of Information Technology Unit at the KUC and leads the research group which is focusing on intelligent and autonomous sensor systems with application on environmental monitoring, healthcare and smart cities. His current research interests lie in wireless sensor networks with emphasis on data processing, localization algorithms, network management, middleware, and the design and performance evaluation of wireless sensor systems.