

Accurate Range-Free Localization for Anisotropic Wireless Sensor Networks

SHIGENG ZHANG, Central South University
XUAN LIU, Hong Kong Polytechnic University
JIANXIN WANG, Central South University
JIANNONG CAO, Hong Kong Polytechnic University
GEYONG MIN, University of Bradford

Position information plays a pivotal role in wireless sensor network (WSN) applications and protocol/algorithm design. In recent years, range-free localization algorithms have drawn much research attention due to their low cost and applicability to large scale WSNs. However, the application of range-free localization algorithms is restricted because of their dramatic accuracy degradation in practical anisotropic WSNs, which is mainly caused by large error of distance estimation. Distance estimation in the existing range-free algorithms usually rely on a unified *per hop length* (PHL) metric between nodes. But PHL between different nodes might be greatly different in anisotropic WSNs, resulting in large error in distance estimation. We find that, although PHL between different nodes might be greatly different, it exhibits significant *locality*, i.e., nearby nodes share similar PHL to anchors that know their positions in advance. Based on the locality of PHL, a novel distance estimation approach is proposed. Theoretical analyses show that the error of distance estimation in the proposed approach is only one fourth of that in state-of-the-art pattern-driven scheme (PDS). An adaptive anchor selection is devised to further improve localization accuracy by mitigating the negative effects from the anchors that are poorly distributed in geometry. By combining the locality-based distance estimation and the adaptive anchor selection, a range-free localization algorithm named Selective Multilateration (SM) is proposed. Simulation results show that SM achieves localization accuracy higher than $0.3r$, where r is the communication radius of nodes. Compared with state-of-the-art solutions, SM improves distance estimation accuracy by up to 57 percent and improves localization accuracy by up to 52 percent consequently.

Categories and Subject Descriptors: C.2.2 [Computer-Communication Networks]: Network Protocols

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Wireless sensor networks, anisotropic wireless networks, localization, range-free, anchor selection

ACM Reference Format:

Shigeng Zhang, Xuan Liu, Jianxin Wang, Jiannong Cao, and Geyong Min, 2014. Accurate Range-Free Localization for Anisotropic Wireless Sensor Networks. *ACM Trans. Sensor Netw.* xx, xx, Article 1 (March 2014), 25 pages.

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

This work is supported by the National Science Foundation of Chian, under grant 61173203, grant 61073036, and grant 61173169.

Author's addresses: S. Zhang and J. Wang, School of Information Science and Engineering, Central South University; X. Liu and J. Cao, Department of Computing, the Hong Kong Polytechnic University; G. Min, Department of Computing, University of Bradford.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2014 ACM 1550-4859/2014/03-ART1 \$15.00

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

1. INTRODUCTION

Wireless sensor networks (WSNs) have the potential to provide economical and practical solutions to many important applications, e.g., environmental monitoring [Navarro et al. 2013], object tracking [Blumrosen et al. 2013], and wild animal monitoring and controlling [Misra and Singh 2012]. Position information of nodes plays a pivotal role in many WSN applications. For instance, in the environmental monitoring applications, the collected data should be notated with the positions where they are sensed to be further analyzed to provide meaningful information. Position information of nodes is also necessary in the design of many WSN protocols and algorithms, like topology control [Zhu et al. 2012a], geographical routing [Huang et al. 2012b], and event detection [He et al. 2012]. The process of determining the positions of nodes in a WSN is usually termed as *localization*. Designing accurate and efficient localization algorithms remains challenging due to the extremely limited resources of nodes and the complex environments where the WSNs are deployed.

Range-free localization algorithms have received much research attention in the past several years [Wang et al. 2010; Xiao et al. 2010a; Wang and Xiao 2008; Li and Liu 2010; Lim and Hou 2009; Xiao et al. 2010b; Tan et al. 2010; Kubo et al. 2012; Shang et al. 2004a]. Compared to range-based localization [Wang et al. 2011; Aspnes et al. 2006] that requires special hardware to measure distances/angles among nodes, range-free localization is more suitable to large-scale WSNs due to its low requirement on node hardware and low cost. Instead of directly measuring distances or angles between nodes, range-free localization algorithms estimate the distances between nodes by using network topology and the proximity relationship among nodes, e.g., DV-Hop [Niculescu and Nath 2001] and Amorphous [Nagpal et al. 2003]. Although these algorithms achieve high localization accuracy in isotropic WSNs (i.e., WSNs formed by nodes uniformly scattered in regular regions), they face severe accuracy degradation in anisotropic WSNs [Lim and Hou 2009; Xiao et al. 2010b; Tan et al. 2010; Kubo et al. 2012; Wang and Xiao 2008]. The accuracy degradation mainly comes from large error of distance estimation in anisotropic environments. For example, existing algorithms like DV-Hop and Amorphous usually rely on a unified *per hop length* (PHL) to estimate distances between nodes. In anisotropic WSNs, PHL between different node pairs might be greatly different from each other, making algorithms relying on DV-Hop and Amorphous output largely erroneous distance estimates. The inaccurate distance estimates consequently result in inaccurate position estimates.

Improving the accuracy of range-free localization in anisotropic WSNs has been one of the main focuses in WSN localization research [Li et al. 2005; Wang and Xiao 2008; Kung et al. 2009; Yang and Liu 2010; Xiao et al. 2013; Li and Liu 2010; Wang et al. 2010; Xiao et al. 2010b; Lim and Hou 2009; Shang et al. 2004b; Xiao et al. 2010a; Tan et al. 2010; Shang et al. 2004a]. Existing solutions can be roughly classified into four categories. The first category of approaches employ robust statistical location estimates to tolerate the large error in distance estimation [Li et al. 2005; Wang and Xiao 2008; Kung et al. 2009; Yang and Liu 2010; Xiao et al. 2013]. The second category of approaches rectify the distance estimates within an additional phase and calculate positions of nodes with the rectified distances [Li and Liu 2010; Wang et al. 2010; Xiao et al. 2010b; Lim and Hou 2009; Lederer et al. 2009] in order to improve localization accuracy. The third category of approaches are more conservative that they just simply discard the distance estimates that might have large error according to heuristic rules [Shang et al. 2004b; Xiao et al. 2010a; Tan et al. 2010]. There are also some other approaches exploit the local isotropy of the network to build accurate local coordinate systems and form the global coordinate system by combining local ones [Shang et al. 2004a; Costa et al. 2006; Ji and Zha 2004]. However, most of these approaches

are *reactive*, i.e., they try to make improvement on the erroneous distance estimates obtained with traditional approaches like DV-Hop and Amorphous. They fail to tackle the problem *proactively* at the starting point, i.e., designing novel approaches to obtaining accurate distance estimation between nodes.

In this paper, a novel approach is proposed to obtain accurate distance estimation between nodes in anisotropic WSNs. The key observation is that, although in an anisotropic network PHL between different node pairs might be greatly different, it exhibits significant *locality*: For two nearby nodes, their PHL to the anchors (nodes that know their positions in advance) are similar. Based on the locality of PHL, a novel distance estimation approach is proposed that outputs more accurate distance estimates than state-of-the-art pattern-driven scheme (PDS) [Xiao et al. 2010b] does. We also devise an algorithm to adaptively select a subset of anchors for every node to perform localization, which can further improve the localization accuracy. By combining the two techniques, we propose a range-free localization algorithm that achieves high localization accuracy in anisotropic WSNs.

In summary, this paper makes the following main contributions:

- A novel distance estimation approach is proposed based on the locality of PHL. Theoretical analyses show that the expected error of distance estimation in the proposed approach is only one fourth of that in PDS [Xiao et al. 2010b], i.e., $0.1r$ in our approach versus $0.4r$ in PDS, where r is the communication radius of nodes.
- An efficient algorithm is devised to adaptively select a subset of anchors for every node to perform localization. The selection of anchors effectively improves localization accuracy by mitigating the negative effects from the anchors that are poorly distributed in geometry.
- By combining the locality-based distance estimation and the adaptive anchor selection, a range-free localization algorithm named **Selective Multilateration (SM)** is proposed. SM achieves localization accuracy higher than $0.3r$, well satisfying the basic requirements of location-dependent protocols/algorithms in WSNs, i.e., $0.4r$ as pointed out in [He et al. 2005].

The rest of this paper is organized as following. Section 2 introduces background on range-free localization. Section 3 shows the locality of PHL in anisotropic WSNs and gives theoretical explanation of the locality. The locality-based distance estimation approach is detailed in Section 4, along with the analyses on its accuracy. Section 5 describes the adaptive anchor selection algorithm. Section 6 proposes the SM algorithm built on top of the locality-based distance estimation approach and the anchor selection algorithm. Extensive simulation experiments are conducted to evaluate the performance of SM and make comparisons with state-of-the-art solutions, and the results are reported in Section 7. Section 8 reviews related work. Finally, Section 9 gives some conclusion remarks.

2. BACKGROUND

2.1. Range-Free Distance Estimation

In range-free distance estimation approaches [Niculescu and Nath 2001; Nagpal et al. 2003], the distance between two nodes is estimated as the multiplication of the PHL and the hop count between them. Let h_{ij} be the hop count between node s_i and node s_j . Then the distance d_{ij} can be estimated as

$$d_{ij} \approx h_{ij} * PHL. \quad (1)$$

PHL represents the expected distance each hop progresses. As to be revealed in Section 3.1, PHLs between different node pairs are similar in isotropic WSNs. Thus, its value

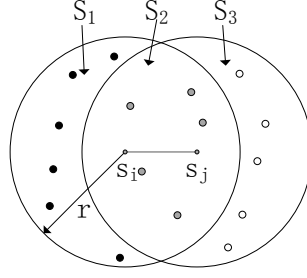


Fig. 1. Estimating proximity level between s_i and s_j : The proximity level can be inferred from $A(S_1)/A(S_2)$, where $A(S_1)$ ($A(S_2)$) is the area of region S_1 (S_2). $A(S_1)/A(S_2)$ can be approximately represented as the ratio of $|N(s_i) - N(s_j)|$ (the number of black nodes) to $|N(s_i) \cap N(s_j)|$ (the number of gray nodes).

can be estimated by sampling on anchors [Niculescu and Nath 2001]

$$PHL \approx \frac{\sum_{s_i \neq s_j, s_i, s_j \in A} d_{ij}}{\sum_{s_i \neq s_j, s_i, s_j \in A} h_{ij}}, \quad (2)$$

where A is the set of all anchors, d_{ij} is the Euclidean distance between two anchors s_i and s_j which can be calculated according to the two anchor's coordinates. This approach, usually called DV-Hop, is the basic technique employed by most range-free localization algorithms [Wang et al. 2010; Xiao et al. 2010a; Wang and Xiao 2008; Li and Liu 2010; Xiao et al. 2010b; Tan et al. 2010; Kubo et al. 2012; Li et al. 2005] to estimate distance between nodes. It can output fairly accurate distance estimates in isotropic WSNs, but faces severe accuracy degradation in anisotropic WSNs. We will discuss the reason behind this in Section 3.1.

2.2. Sub-Hop Proximity Estimation

In recent years, there have been some researches focusing on sub-hop proximity estimation between adjacent nodes, i.e., obtaining proximity information more fine-grained than hop count to improve the accuracy of range-free localization algorithms [Wu et al. 2012; Zhu et al. 2012b; Zhong and He 2011]. Figure 1 illustrates the intuition to estimate sub-hop proximity between two adjacent nodes. s_i and s_j share some common neighbors. The ratio of common neighbors is related to the distance between them. For example, when d_{ij} is small, i.e., the two nodes are very close to each other, most of their neighbors are common neighbors. In contrast, when d_{ij} is large, they share only a few or even no common neighbors. We can infer more fine-grained proximity information than traditional hop-count by using the ratio of common neighbors.

We briefly explain how to estimate the proximity level between s_i and s_j by using Figure 1 as an example. Consider the case in which the communication radius r (i.e., one longest hop in traditional approaches) is divided into K proximity levels. The proximity level between s_i and s_j can be estimated as

$$\kappa = \lceil \frac{d_{ij}}{r/K} \rceil = \lceil K * \frac{d_{ij}}{r} \rceil. \quad (3)$$

Let $A(S_1)$ and $A(S_2)$ be the area of region S_1 and S_2 in Figure 1, respectively. The ratio of $A(S_1)$ to $A(S_2)$ is determined by only d_{ij} , i.e.,

$$\frac{A(S_1)}{A(S_2)} = f(d_{ij}) = \frac{\pi r^2}{2 \arccos \frac{d_{ij}}{2r} r^2 - d_{ij} \sqrt{r^2 - d_{ij}^2/4}} - 1. \quad (4)$$

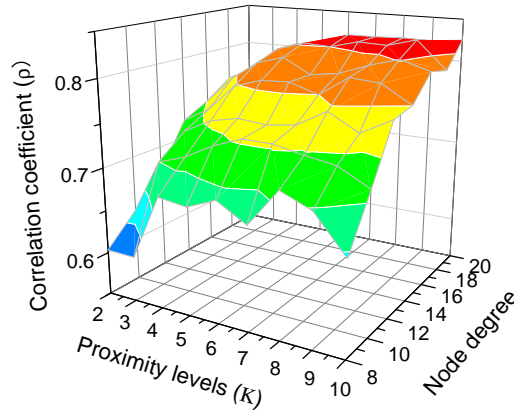


Fig. 2. The impact of the number of proximity levels (K) and node degree on the correlation coefficient between distance and proximity estimation.

Let $N(s_i)$ be the set of all the neighbors of s_i . Then for s_i , $A(S_1)/A(S_2)$ can be approximated as

$$f(d_{ij}) \approx \frac{|N(s_i) - N(s_j)|}{|N(s_i) \cap N(s_j)|}. \quad (5)$$

Then d_{ij} can be estimated as

$$d_{ij} \approx f^{-1}\left(\frac{|N(s_i) - N(s_j)|}{|N(s_i) \cap N(s_j)|}\right). \quad (6)$$

Substituting Eq. (6) into Eq. (3), s_i can obtain the proximity level between itself and s_j . In the same way, s_j can obtain its proximity level between itself and s_i . Note the proximity levels obtained by s_i and s_j might be different due to uneven distribution of nodes. In this case, s_i and s_j can exchange their estimated proximity levels and use the mean value as the final proximity level estimation.

The number of proximity levels (K) affects the distance estimation accuracy from two aspects. On one hand, with a larger K more fine-grained proximity information can be obtained and hence the distance estimation accuracy can be improved. On the other hand, more proximity levels might induce larger proximity level estimation error. For instance, let κ and κ' be the ground-truth and the estimated proximity level, respectively. The maximum proximity level estimation error would be about $r * \frac{|K-1|}{K}$, i.e., $\kappa = 1$ but $\kappa' = K$ or $\kappa = K$ but $\kappa' = 1$. When $K = 2$, the maximum estimation error would be $0.5r$. When $K = 10$, the maximum estimation error would be $0.9r$.

The value of K should be appropriately selected. Figure 2 plots the Pearson correlation coefficient (ρ) between the proximity estimation and the corresponding ground-truth distance between two nodes when K and node degree (to be defined in Section 7) change. It can be observed that ρ becomes almost stable when $K \geq 4$. Thus we set $K = 4$ through this paper. Meanwhile, we observe that ρ is always larger than 0.6 when the node degree is larger than eight, showing that the proximity estimation can effectively represent the distance between two nodes. To be consistent with the existing works, we still use the term ‘‘hop’’ to denote one proximity level in this paper. In other words, there might be κ hops between two adjacent nodes s_i and s_j , where $1 \leq \kappa \leq K$.

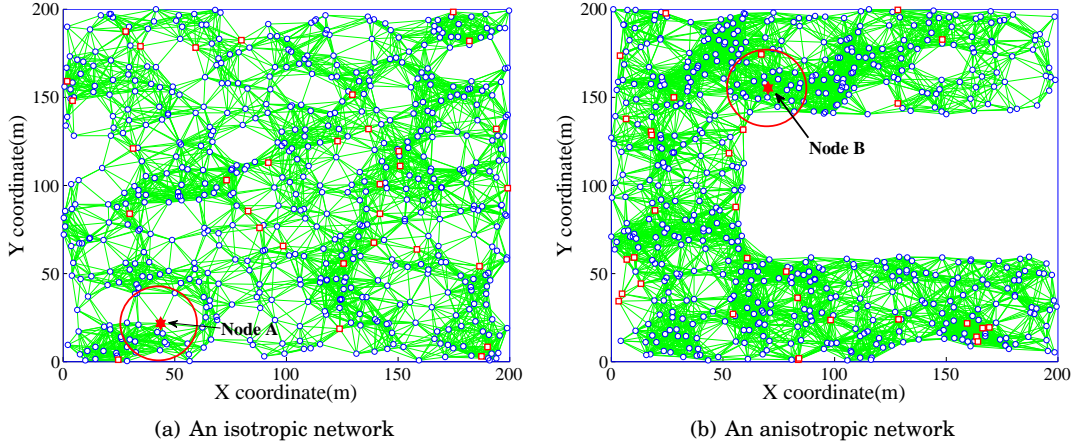


Fig. 3. (a): A typical isotropic network, and (b) a typical anisotropic network. The squares represent anchors.

3. LOCALITY OF PHL

3.1. Locality of PHL in Anisotropic WSNs

In this section we investigate the different properties of PHL in isotropic and anisotropic WSNs.

Definition 3.1 (PHL vector). Consider a network consisting of n nodes, among which m are anchors. We label the n nodes as $\{s_1, \dots, s_m, s_{m+1}, \dots, s_n\}$, where s_1, \dots, s_m are anchors. The PHL vector of s_i ($1 \leq i \leq n$) is defined as $\mathbf{V}_i = [u_1, \dots, u_m]$, where the k -th element is the PHL between s_i and s_k , i.e.,

$$u_k = \frac{d_{ik}}{h_{ik}}. \quad (7)$$

When s_i itself is an anchor, $u_i = 0$.

Definition 3.2 (Similarity between two PHL vectors). Given two nodes' PHL vectors \mathbf{V}_i and \mathbf{V}_j , we define their similarity as the cosine of the angle between them, which is calculated as

$$\text{Sim}(\mathbf{V}_i, \mathbf{V}_j) = \frac{\langle \mathbf{V}_i, \mathbf{V}_j \rangle}{|\mathbf{V}_i| |\mathbf{V}_j|}, \quad (8)$$

where $\langle \mathbf{V}_i, \mathbf{V}_j \rangle$ is the inner product of the two vectors

$$\langle \mathbf{V}_i, \mathbf{V}_j \rangle = \sum_{k=1}^m \mathbf{V}_i[k] * \mathbf{V}_j[k], \quad (9)$$

and $|\mathbf{V}_i|$ ($|\mathbf{V}_j|$) is the length of \mathbf{V}_i (\mathbf{V}_j).

The similarity between different nodes' PHL vectors exhibits different properties in isotropic and anisotropic WSNs. In isotropic WSNs like Figure 3(a), the similarity between different nodes' PHL vectors is always very high and concentrates in a very small interval. In contrast, in anisotropic WSNs like Figure 3(b), the similarity between different nodes' PHL vectors scatters in a much larger interval. As an example, we randomly select two nodes, i.e., A in Figure 3(a) and B in Figure 3(b), and plot the similarity between their PHL vectors and other nodes' PHL vectors in the same network. The similarity between A and other nodes' PHL vectors is plotted in Figure 4(a),

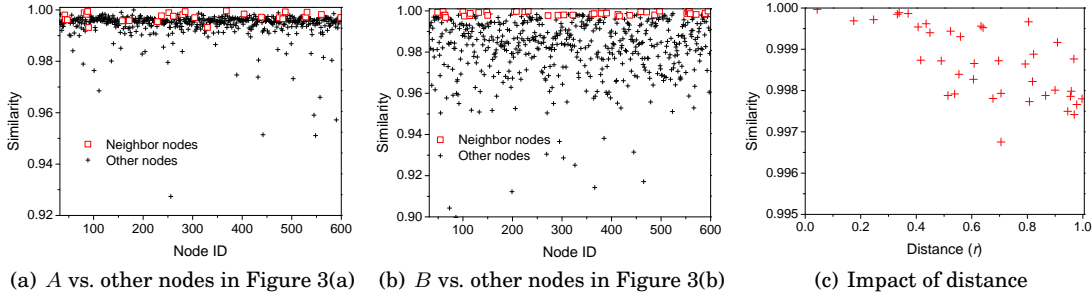


Fig. 4. Similarity of PHL vectors: (a) similarity between A and other nodes in Figure 3(a), (b) similarity between B and other nodes in Figure 3(b); (c) similarity between B and its neighbors versus the corresponding distances. The similarity between A (B) and its direct neighbors is marked as squares.

and the similarity between B and other nodes' PHL vectors is plotted in Figure 4(b). It can be observed that most (more than 81 percent) of the similarity between A and other nodes in Figure 3(a) concentrates in a very small interval $[0.993, 0.998]$, while the similarity between B and other nodes in Figure 3(b) scatters in a much larger interval $[0.95, 1]$. This is the intrinsic reason why algorithms like DV-Hop and Amorphous perform well in isotropic WSNs but encounter severe accuracy degradation in anisotropic WSNs, because they rely on a unified PHL to estimate the distances between nodes.

However, PHL exhibits significant *locality* in anisotropic WSNs, i.e., the similarity between two nearby nodes' PHL vectors is very close to 1. This means that two nearby nodes have similar PHL to anchors. In Figure 4(b) we mark the similarity between B 's PHL vector and its direct neighbors' PHL vectors with red squares. It can be observed that the similarity between B and its direct neighbors is significantly higher than that between B and other nodes. We call this property as *locality of PHL*. Actually, the locality of PHL also holds in isotropic WSNs, as shown in Figure 4(a).

The locality of PHL provides an opportunity to obtain accurate distance estimates to anchors in anisotropic WSNs. For example, assume that s_i is an anchor and one of its neighbor s_j wants to estimate distances to anchors. Since all the anchors' positions are known, s_i can calculate its PHL vector and broadcast the vector to s_j . Then s_j can use s_i 's PHL vector to obtain accurate distance estimations to anchors. This locality-based distance estimation approach will be elaborated in Section 4.

3.2. Theoretical Foundation of the Locality of PHL

In this section, we analyze how the distance between two nodes affects the similarity between their PHL vectors. According to the following analyses, we will see that when two nodes are close to each other, the similarity between their PHL vectors will approach 1, in either isotropic WSNs or anisotropic WSNs.

Consider two nodes s_i and s_j . We use two vectors $\mathbf{D}_i = [d_{ik}]_{1 \leq k \leq m}$ and $\mathbf{H}_i = [h_{ik}]_{1 \leq k \leq m}$ to represent s_i 's distances and hop counts to anchors, where d_{ik} and h_{ik} represent the distance and the hop count between s_i and anchor s_k , respectively. Similarly, we use \mathbf{D}_j and \mathbf{H}_j to represent the two vectors for s_j . The PHL vectors of s_i and s_j can be written as

$$\mathbf{V}_i = \mathbf{D}_i * (\text{diag}(\mathbf{H}_i))^{-1}, \mathbf{V}_j = \mathbf{D}_j * (\text{diag}(\mathbf{H}_j))^{-1}, \quad (10)$$

where $\text{diag}(\mathbf{H})$ represents the diagonal matrix taking elements in \mathbf{H} as elements on its diagonal.

In order to facilitate the analysis, we define two vectors $\Theta = [\theta_k]_{1 \leq k \leq m}$ and $\Xi = [\xi_k]_{1 \leq k \leq m}$, where

$$\theta_k = \frac{d_{jk}}{d_{ik}} - 1, \xi_k = \frac{h_{jk}}{h_{ik}}. \quad (11)$$

Θ represents the relative difference between D_i and D_j , and Ξ represents how H_i and H_j are close to each other. With Θ and Ξ , D_j and H_j can be rewritten as

$$D_j = D_i * (\mathbf{I} + \text{diag}(\Theta)), H_j = H_i * \text{diag}(\Xi), \quad (12)$$

where \mathbf{I} is the identify matrix.

In order to calculate the similarity between V_i and V_j , we first calculate $\langle V_i, V_j \rangle$ as

$$\begin{aligned} \langle V_i, V_j \rangle &= V_i * V_j^T \\ &= D_i * (\text{diag}(H_i))^{-1} * ((\text{diag}(H_j))^{-1})^T * D_j^T \\ &= D_i * (\text{diag}(H_i))^{-1} * ((\text{diag}(H_i * \text{diag}(\Xi)))^{-1})^T * (D_i(\mathbf{I} + \text{diag}(\Theta)))^T \\ &= D_i * (\text{diag}(H_i \odot H_i \odot \Xi))^{-1} * D_i^T + D_i * (\text{diag}(H_i \odot H_i \odot \Xi))^{-1} * \text{diag}(\Theta) * D_i^T \\ &= \sum_{k=1}^m \frac{1}{\xi_k} \frac{d_{ik}^2}{h_{ik}^2} + \sum_{k=1}^m \frac{1}{\xi_k} \theta_k \frac{d_{ik}^2}{h_{ik}^2}. \end{aligned} \quad (13)$$

Here \odot means element-wise multiplication. For example, let $\mathbf{A} = [a_1, \dots, a_m]$ and $\mathbf{B} = [b_1, \dots, b_m]$. Then $\mathbf{A} \odot \mathbf{B} = [a_1 * b_1, \dots, a_m * b_m]$.

Let ξ_{max} and ξ_{min} be the maximum and the minimum elements in Ξ , respectively. Then there must exist $\xi_{min} \leq \xi_1 \leq \xi_{max}$, such that

$$\sum_{k=1}^m \frac{1}{\xi_k} \frac{d_{ik}^2}{h_{ik}^2} = \frac{1}{\xi_1} \sum_{k=1}^m \frac{d_{ik}^2}{h_{ik}^2} = \frac{1}{\xi_1} |V_i|^2. \quad (14)$$

Similarly, let θ_{max} and θ_{min} be the maximum and the minimum elements in Θ , respectively. There must exist $\theta_{min} \leq \theta_1 \leq \theta_{max}$, such that

$$\sum_{k=1}^m \frac{1}{\xi_k} \theta_k \frac{d_{ik}^2}{h_{ik}^2} = \theta_1 \sum_{k=1}^m \frac{1}{\xi_k} \frac{d_{ik}^2}{h_{ik}^2} = \theta_1 \frac{1}{\xi_1} |V_i|^2. \quad (15)$$

Substituting Eqs. (14) and (15) into Eq. (13), we get

$$\langle V_i, V_j \rangle = |V_i|^2 \frac{1}{\xi_1} (1 + \theta_1). \quad (16)$$

We calculate $|V_j|^2$ in a similar way.

$$\begin{aligned} |V_j|^2 &= V_j * V_j^T \\ &= D_i * (\mathbf{I} + \text{diag}(\Theta)) * (\text{diag}(H_i * \text{diag}(\Xi)))^{-1} * (D_i * (\mathbf{I} + \text{diag}(\Theta)) * (\text{diag}(H_i * \text{diag}(\Xi)))^{-1})^T \\ &= D_i * \mathbf{X} * D_i^T + 2D_i * \mathbf{X} * \text{diag}(\Theta) * D_i^T + D_i * \mathbf{X} * \text{diag}(\Theta \odot \Theta) * D_i^T, \end{aligned} \quad (17)$$

where

$$\mathbf{X} = (\text{diag}(H_i \odot H_i \odot \Xi \odot \Xi))^{-1}. \quad (18)$$

Note that

$$D_i * \mathbf{X} * D_i^T = \sum_{k=1}^m \frac{1}{\xi_k^2} \frac{d_{ik}^2}{h_{ik}^2}. \quad (19)$$

Thus, there must exist $\xi_{min} \leq \xi_2 \leq \xi_{max}$ such that

$$\sum_{k=1}^m \frac{1}{\xi_k^2} \frac{d_{ik}^2}{h_{ik}^2} = \frac{1}{\xi_2^2} \sum_{k=1}^m \frac{d_{ik}^2}{h_{ik}^2} = \frac{1}{\xi_2^2} |\mathbf{V}_i|^2. \quad (20)$$

Similarly, there exists $\theta_{min} \leq \theta_2 \leq \theta_{max}$ such that

$$\mathbf{D}_i * \mathbf{X} * \text{diag}(\Theta) * \mathbf{D}_i^T = \sum_{k=1}^m \theta_k \frac{1}{\xi_k^2} \frac{d_{ik}^2}{h_{ik}^2} = \theta_2 * \frac{1}{\xi_2^2} |\mathbf{V}_i|^2, \quad (21)$$

and exists $\theta_{min} \leq \theta_3 \leq \theta_{max}$ such that

$$\mathbf{D}_i * \mathbf{X} * \text{diag}(\Theta \odot \Theta) * \mathbf{D}_i^T = \sum_{k=1}^m \theta_k^2 \frac{1}{\xi_k^2} \frac{d_{ik}^2}{h_{ik}^2} = \theta_3^2 * \frac{1}{\xi_2^2} |\mathbf{V}_i|^2. \quad (22)$$

Substituting Eqs. (20)-(22) into Eq. (17), we have

$$|\mathbf{V}_j|^2 = \frac{1}{\xi_2^2} (1 + \theta_2 + \theta_3^2) |\mathbf{V}_i|^2. \quad (23)$$

Substituting Eqs. (16) and (23) into Eq. (8), we can write the similarity between \mathbf{V}_i and \mathbf{V}_j as

$$\text{Sim}(\mathbf{V}_i, \mathbf{V}_j) = \frac{\frac{1}{\xi_1} (1 + \theta_1)}{\frac{1}{\xi_2} \sqrt{1 + \theta_2 + \theta_3^2}}. \quad (24)$$

According to Eq. (24), we can explain the locality of PHL as follows. When s_i and s_j are close to each other, for a faraway anchor s_k , h_{ik} and h_{jk} are very close to each other and thus ξ_1 and ξ_2 approach 1. Similarly, when s_i and s_j are close to each other, d_{ij} is small. Thus for the faraway anchor s_k , the distance between s_i and s_j is relatively very small compared to their distances to s_k . In this case, θ_1 , θ_2 , and θ_3 will approach 0. Thus $\text{Sim}(\mathbf{V}_i, \mathbf{V}_j)$ will approach 1 if the anchors are faraway enough, which is usually the case in a large networks with only a small number of anchors. In Figure 4(c) we plot how the similarity between B and its direct neighbors when the distance between them changes. It can be clearly observed that when the distance between two nodes is small, the similarity between their PHL vectors is very close to 1. The similarity drops when the distance increases. For this reason, when a node can obtain PHL vectors from more than one neighboring anchors, it should use the nearest neighbor's PHL vector to perform distance estimation. The proximity estimation approach given in Section 2.2 can help to determine which neighbor is the nearest one.

We should point out that locality of PHL is pervasive in both isotropic and anisotropic WSNs, because it has no special requirements on the network topology or nodal distribution. Furthermore, the locality does not mean that node pairs having high similarity must be close to each other. For example, as shown in Figure 4(a), in isotropic WSNs the similarity between two disperse nodes might also be high. However, in anisotropic WSNs as shown in Figure 3(b), the similarity between adjacent nodes is significantly higher than that between disperse nodes, and thus exhibits very strong locality.

4. LOCALITY-BASED DISTANCE ESTIMATION

Based on the locality of PHL, we propose a novel distance estimation approach and analyze its accuracy.

4.1. Estimating Distances to Anchors

As being pointed out in Section 3.1, nearby nodes share similar PHL vectors. Thus, if s_i has an anchor neighbor s_j , it can use s_j 's PHL vector, i.e., \mathbf{V}_j , to estimate its distances

to other anchors. In other words, \mathbf{D}_i , the vector containing s_i 's distances to vectors, can be estimated as

$$\mathbf{D}_i \approx \mathbf{V}_j \odot \mathbf{H}_i. \quad (25)$$

For instance, if $\mathbf{V}_j = [2.1, 2.3, 3.2]$ and $\mathbf{H}_i = [5, 10, 3]$, we obtain $\mathbf{D}_i \approx [2.1 * 5, 2.3 * 10, 3.2 * 3] = [10.5, 23, 9.6]$.

In case s_i has no anchor neighbors, it waits until at least one of its neighbors has been localized and then uses that node's PHL vector to estimate distances to anchors. Different from anchors, the position of the localized node might contain error, which will consequently induce error in its PHL vector. We will discuss the error accumulation problem in Section 4.2. When s_i has more than one localized neighbors, it uses the nearest neighbor's PHL vector to perform distance estimation.

4.2. Accuracy Analysis

4.2.1. When Using Anchor's PHL Vector. We first analyze the accuracy of distance estimation when s_i uses an anchor neighbor s_j 's PHL vector to perform distance estimation. Without loss of generality, we consider the error in distance estimate to anchor s_k ($s_k \neq s_j$). As shown in Figure 5, we consider the general case in which the shortest paths between s_i and s_k and between s_j and s_k are not detoured. We will deal with the cases in which the shortest paths are detoured at the end of this section.

Let d_1 and h_1 be the distance and the hop count between s_j and s_k , respectively. Similarly, denote by d_2 and h_2 the distance and the hop count between s_i and s_k , respectively. Obviously, $d_1 - r \leq d_2 \leq d_1 + r$, because s_i and s_k can communicate with each other. We derive the distance estimation error of d_2 when the number of proximity levels is K .

We first computer the error of distance estimation when $d_2 = x$. In the ideal case, the hop counts h_1 and h_2 would be

$$h_1 = \lceil \frac{d_1}{r/K} \rceil = \lceil K * \frac{d_1}{r} \rceil, h_2 = \lceil \frac{x}{r/K} \rceil = \lceil K * \frac{x}{r} \rceil. \quad (26)$$

Then the distance between s_i and s_k is estimated as

$$\hat{d}_2 = h_2 * \frac{d_1}{h_1} = \lceil K * \frac{x}{r} \rceil * \frac{d_1}{\lceil K * \frac{d_1}{r} \rceil}. \quad (27)$$

The error of distance estimation is

$$ef(x) = \left| x - \lceil K * \frac{x}{r} \rceil * \frac{d_1}{\lceil K * \frac{d_1}{r} \rceil} \right|. \quad (28)$$

In our previous work [Zhang et al. 2012], we have proved that the distance estimation error is upper bounded by $ef(x) \leq \frac{2 * h_2 * r}{h_1 - K}$. We now calculate the mean value of $ef(x)$. As shown in Figure 5, $ef(x)$ can be treated as a constant when d_2 varies in a small interval $[x, x + \Delta x]$. Denote by $Disk(k, x)$ and $Disk(j, r)$ the two disks that center at s_k and s_j with radius x and r , respectively. The size of the intersection region between the two disks will increase when d_2 increases. Denote by ΔS the increased size when d_2 increases from x to $x + \Delta x$. Then the probability that the distance estimation error is $ef(x)$ is

$$p_x = \frac{\Delta S}{\pi * r^2}, \quad (29)$$

where

$$\Delta S = \frac{2\alpha}{2\pi} \pi((x + \Delta x)^2 - x^2) \approx 2\alpha * x * \Delta x. \quad (30)$$

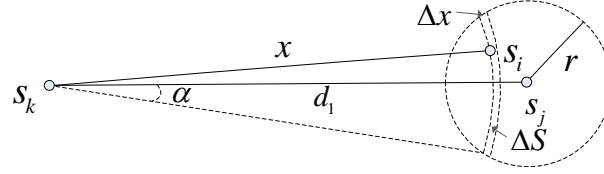


Fig. 5. The distance estimation error when s_i uses s_j 's PHL vector to estimate distance to anchor s_k .

According to the cosine law, the angle α is

$$\alpha = \arccos \frac{d_1^2 + x^2 - r^2}{2xd_1}. \quad (31)$$

Thus the expected distance estimation error of d_2 is

$$E[ef(x)] = \int_{d_1-r}^{d_1+r} ef(x) * p_x dx. \quad (32)$$

Figure 6 plots the expected distance estimation error of d_2 when d_1 varies from r to $20r$. The average distance estimation error varies from $0.2r$ to $0.5r$ when $K = 1$, i.e., when the traditional ‘‘hop count’’ metric is used. When we use the approach described in Section 2.2 to obtain sub-hop level proximity information, the expected distance estimation error drops accordingly. When $K = 4$, i.e., the communication radius is divided into up to four proximity levels, the expected distance estimation error is around $0.1r$. In general, the expected distance estimation error will be smaller than $r/2K$ when the number of proximity levels is K .

The distance estimation error is periodic and is not very relevant to the absolute value of d_1 . This can help us tackle the special case when the shortest paths between s_k and s_i and between s_k and s_j are detoured. As shown in Figure 7, because s_i and s_j are close to each other, the shortest paths between s_i and s_k and between s_j and s_k share the same detoured sub-path. Thus we can straighten the detoured sub-path and move s_k to s'_k . This straighten operation will add the same value to d_{ik} and d_{jk} . Since the distance estimation accuracy will not be affected a lot by the distance between s_j and s_k (as shown in Figure 6), the distance estimation of the detoured path d_{ik} will also be accurate.

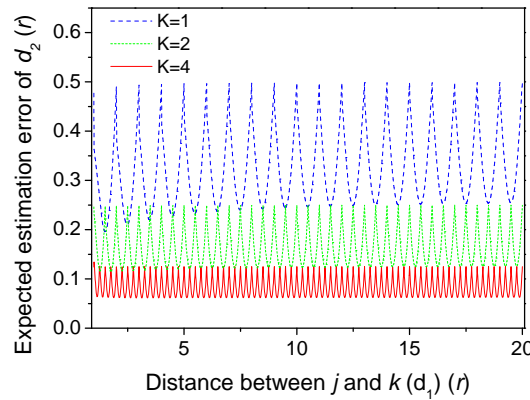


Fig. 6. Expected distance estimation error of d_2 for different proximity levels when d_1 increases.

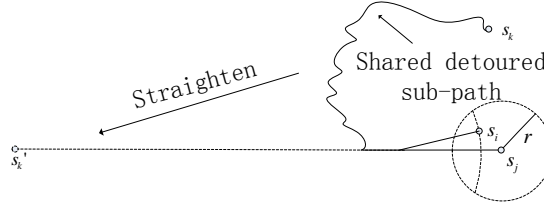


Fig. 7. Accuracy analysis when the shortest path between s_k and s_j and that between s_k and s_i are detoured.

4.2.2. When Using Node's PHL Vector. When node s_i has no anchor neighbors, it needs to use one of its localized neighbor's PHL vector to estimate its distances to anchors. Assume that s_i uses s_j 's information. The problem is that s_j knows only *erroneous estimated position* rather than *exact position*. We now analyze how s_j 's location estimation error affects node s_i 's distance estimation accuracy.

Without loss of generality, we still consider error in distance estimation to anchor s_k . Assume that the location estimation error of s_j is ε . Denote by P_j and P'_j the exact position and the estimated position of s_j , respectively. Then we have $|P_j - P'_j| = \varepsilon$. We use d_1 to denote the exact distance between s_j and s_k , and use d'_1 to denote the distance calculated with s_j 's estimated position and s_k 's exact position. Let $\Delta d_1 = d'_1 - d_1$ be the difference between d_1 and d'_1 . It is easy to know that

$$|\Delta d_1| = |d'_1 - d_1| \leq \varepsilon. \quad (33)$$

Note that the location estimation error of s_j affects only the calculation of d'_1 ; it does not affect h_1 (i.e., the hop count between node j and anchor k), because h_1 is determined by only the exact positions of the two nodes and the network topology. Thus, according to Eq. (28), s_i 's distance estimation error to anchor s_k is

$$\begin{aligned} ef'(x) &= \left| x - h_2 * \frac{d_1 + \Delta d_1}{h_1} \right| \\ &= \left| x - h_2 * \frac{d_1}{h_1} - h_2 * \frac{\Delta d_1}{h_1} \right| \\ &\leq ef(x) + |\Delta d_1| \frac{h_2}{h_1} \\ &\leq ef(x) + \frac{h_2}{h_1} \varepsilon \approx ef(x) + \varepsilon. \end{aligned} \quad (34)$$

Eq. (34) shows that s_j 's position estimation error affects s_i 's distance estimation error in an additive manner. Assume that the position estimation error approximately equals the distance estimation error. If in average there is one anchor in s_i 's l -hop (here "hop" takes its traditional meaning) neighborhood, its distance estimation error will be approximately $l * r/2K$. When the average node degree is ND , it requires the anchor ratio being larger than $1/(ND * l^2)$ to have at least one anchor in s_i 's l -hop neighborhood. For example, if we require the distance estimation error to be smaller than $0.4r$ when $K = 4$, l should be no larger than 3.2 ($l * r/2K \leq 0.4r$ implies $l \leq 3.2$). When $ND = 10$, this requires at least $1/(10 * 3.2 * 3.2) \approx 0.01$ anchors to achieve the required accuracy, i.e., anchor ratio should be larger than 1%. In realistic environments, it may require more anchors to achieve the desired accuracy because many factors other than anchor ratio might affect the localization accuracy.

5. ANCHOR SELECTION

Besides distance estimation error, the geometry distribution of anchors also affects the localization accuracy of nodes [Bu et al. 2012; Yarlagadda et al. 2002; Langley 1999; Yang and Liu 2010]. In this section, we propose an anchor selection algorithm that jointly considers distance estimation accuracy and geometry distribution of anchors to further improve the localization accuracy of nodes.

5.1. The Optimal Anchor Selection Problem

Let A be the set of all anchors, i.e., $A = \{s_1, \dots, s_m\}$. Because in the d -dimensional space a node needs to know distance measurements to at least $d+1$ anchors to uniquely determine its position, any anchor set that could be used to uniquely localize a node contains at least $d+1$ anchors. Denoting by C the set of all such candidate anchor sets, we have $C = \{A' \subseteq A \mid |A'| \geq d+1\}$. The optimal anchor set for node s_i to achieve the highest localization accuracy is

$$A^* = \arg \min_{A' \in C} |PEst(A', i) - P_i|, \quad (35)$$

where $PEst(A', i)$ is the position estimate when anchors in A' and the corresponding distance estimates are used to calculate position of s_i , and P_i is the ground-truth position of s_i . *The optimal anchor selection problem is to find the anchor set A^* .*

However, it is hard to find the exact optimal anchor set A^* , because we do not know P_i . This motivates us to resort to indirect metrics (i.e., metrics that can characterize the impact of anchors on the position error of a node without knowing its true position) to judge the quality of an anchor set. We design an efficient algorithm to select a good anchor set based on the geometric dilution of precision (GDOP) metric [Yarlagadda et al. 2002; Langley 1999].

5.2. GDOP-Threshold based Anchor Selection

5.2.1. GDOP Calculation. GDOP is widely used to measure how the geometry distribution of anchors affect the position error [Yarlagadda et al. 2002; Langley 1999]. The GDOP value of a set of anchors can be roughly interpreted as the ratio of position error to the distance estimation error. Thus, a small GDOP means that the anchors are well distributed in geometry and can generate accurate position estimation with high probability.

Without loss of generality, we assume that node s_i obtains distance estimates to all the m anchors. Then s_i calculates the GDOP of the m anchors as follows [Yarlagadda et al. 2002; Langley 1999]. Let (x_k, y_k) ($1 \leq k \leq m$) be the position of anchor s_k , and let (x, y) be the position of s_i . The distance between s_i and s_k could be expressed as

$$\rho_k = \sqrt{(x - x_k)^2 + (y - y_k)^2}. \quad (36)$$

We construct a $m \times 2$ matrix

$$H = \begin{bmatrix} \frac{\partial \rho_1}{\partial x} & \frac{\partial \rho_1}{\partial y} \\ \vdots & \vdots \\ \frac{\partial \rho_m}{\partial x} & \frac{\partial \rho_m}{\partial y} \end{bmatrix}. \quad (37)$$

The GDOP of the m anchors can be calculated as

$$g = \sqrt{\text{tr}((H^T H)^{-1})}. \quad (38)$$

Note that because node s_i does not know its true position, it needs to use one of its localized neighbor's position to calculate the GDOP. For example, if one of s_i 's neighbor s_j knows its position (e.g., s_j is an anchor or a localized node), the GDOP can be

calculated as

$$g = \sqrt{\text{tr}((H^T H)^{-1})} \Big|_{x=x_j, y=y_j}, \quad (39)$$

where (x_j, y_j) is s_j 's (estimated) position. When s_i has more than one localized neighbors, it uses the neighbor with whose PHL vector it estimates distance to anchors to calculate the GDOP.

5.2.2. Anchor Selection. The GDOP metric characterizes only the geometry distribution of anchors. It does not take the accuracy of distance estimate into account. In this section, we propose an efficient two-stage anchor selection algorithm that jointly considers distance estimation accuracy and geometry distribution of anchors.

The algorithm runs on every node in a distributed manner. In the first stage, the node sorts all the anchors in ascending order according to their distance estimate errors (we will discuss how to do this in our localization algorithm in Section 6). We denote the sorted anchor list as $\{s_{\pi(1)}, \dots, s_{\pi(m)}\}$, where the distance estimation error to anchor $s_{\pi(k)}$ is no larger than the distance estimation error to anchor $s_{\pi(k')}$ when $\pi(k) < \pi(k')$.

In the second stage, the node uses a GDOP-threshold based method to select anchors. To mitigate the negative effects of distance estimations with large error, it first picks anchors in the front of the sorted anchor list. In detail, it first sets $A^* = \{s_{\pi(1)}, \dots, s_{\pi(d+1)}\}$ and calculates the GDOP of A^* . If the GDOP is smaller than a specified threshold g_t , the node uses anchors A^* to calculate its position. Otherwise, it adds the next anchor in the list into A^* and repeats the procedure until the GDOP of A^* is smaller than g_t . Then the node uses anchors in the final A^* to calculate its position.

This anchor selection algorithm jointly considers both distance estimation accuracy and geometry distribution of anchors. First, because anchors are sorted in ascending order according to their distance estimation error, the selected anchors are among those who have small distance estimation errors. Second, by setting a suitable threshold g_t , we can obtain an anchor set with good geometry distribution. The threshold g_t should be selected carefully. On one hand, with a very large g_t , only a few anchors will be selected and the node cannot well exploit the redundancy of anchors to improve localization accuracy. On the other hand, with a very small g_t , almost all the anchors will be selected, which makes the location estimation suffer from large distance estimation errors of some anchors.

6. PUT THEM TOGETHER: SELECTIVE MULTILATERATION

In this section we present the *Selective Multilateration* (SM) localization algorithm that integrates the locality-based distance estimation (Section 4) and the adaptive anchor selection (Section 5).

6.1. Algorithm Framework

The framework of SM is presented in Algorithm 1. The first step is to calculate hop counts for all the nodes to the anchors. This step is the same as in DV-Hop except that hop counts in SM are calculated based on proximity estimations by using the method given in Section 2.2. Then anchors calculate their PHL vectors and broadcast the vectors to their direct neighbors. After receiving the PHL vector, a node estimates its distances to all anchors. It then selects suitable anchors and calculates its position by using multilateration. This is the reason why we name our algorithm as *selective multilateration*.

SM is an iterative algorithm in the means that localized nodes could help other nodes perform localization. As shown in Algorithm 1, when a node has no anchor neighbors, it can request PHL vector from one localized neighbor to estimate distances to anchors.

ALGORITHM 1: Framework of the SM Algorithm

```

for Every node  $s_i$  do
  Obtain proximity estimation to neighbors // Section 2.2;
  Obtain the hop counts to anchors;
  if  $s_i$  is an anchor or a localized node and some of its neighbors are not localized
  then
    Calculate its PHL vector  $V_i$  and broadcast  $V_i$  to direct neighbors;
  end
  if  $s_i$  is not localized and hears PHL vector from neighbors then
    Estimate its distances to anchors // Section 4;
    Select suitable anchors to perform localization // Section 5.2.2 and Section 6.2;
  end
end

```

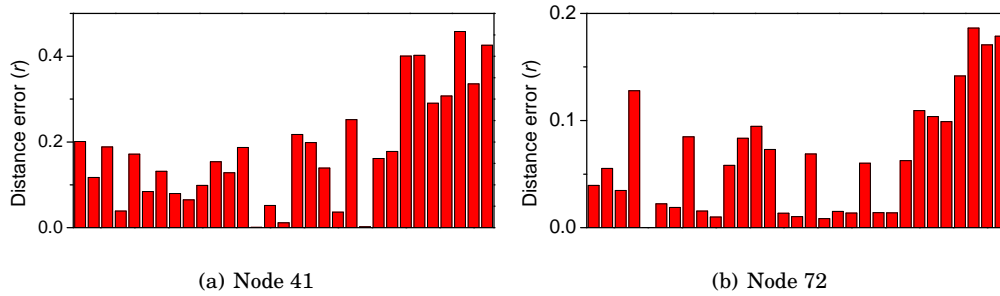


Fig. 8. Distance estimation errors (after sorting) of two randomly selected nodes in Figure 3(b).

Note that not all the localized nodes need to broadcast their PHL vectors. Only when it has at least one non-localized neighbors, a localized node needs to calculate its PHL vector and broadcast it. Note that PHL vector defined in Section 3.1 contains only PHL to anchors. Thus localized nodes' estimated positions will not be used to calculate the position of other nodes, which effectively suppresses error accumulation.

6.2. Sorting of Anchors

Because the exact distance estimation errors to anchors are unknown, in SM we use a heuristic method to sort anchors. We observe from experiments that anchors with larger hop counts usually have larger distance estimation errors. We thus sort anchors in ascending order according the node's *hop counts* to them. We argue that although this approach cannot generate exact anchor ordering as required in Section 5.2.2, in general it can sort anchors with relatively smaller distance estimation errors before those with relatively larger errors.

Figure 8 plots the sorting result of two randomly selected node in Figure 3(b). We observe that the distance estimations with large errors are put at the end of the sorting result. Because our anchor selection algorithm first picks nodes in the front of the list, the negative effect of anchors with large distance estimation errors can be effectively mitigated because they usually will not be selected. Note that this heuristic works in only our locality-based distance estimation approach. In other range-free distance estimation algorithms like DV-Hop it might not work.

6.3. Communication Overhead Analysis

In SM, every anchor needs to flood its position information to all the nodes in the network once. Every anchor also needs to broadcast its PHL vector to its direct neighbors. Besides, some localized nodes might need to broadcast their PHL vectors to direct neighbors. Let l be the total number of nodes (including anchors) that need to broadcast their PHL vectors. The total communication cost in SM is $O(nm + lm)$, where the first term represents the communication cost to flood anchors' position information, and the second term represents the communication cost incurred in broadcasting PHL vectors (each PHL vector contains m values). Because l is usually much smaller than n , the total communication cost of SM is $O(nm)$. This is as low as the communication cost of DV-Hop [Niculescu and Nath 2001] and the PDS algorithm proposed in [Xiao et al. 2010b], and is much lower than that in PDM [Lim and Hou 2009] and MDS-MAP(P) [Shang et al. 2004a].

7. SIMULATION RESULTS

We develop a simulator using JAVA programming language to evaluate the performance of SM and compare SM with two algorithms: the PDS algorithm [Xiao et al. 2010b] and the MDS-MAP(P) algorithm [Shang et al. 2004a]. PDS is state-of-the-art range-free anchor-based localization algorithm for anisotropic WSNs. It can tolerate multiple anisotropy factors of the network and achieves comparable accuracy as PDM [Lim and Hou 2009] with lower communication cost. The MDS-MAP(P) algorithm uses the divide and conquer paradigm to deal with global anisotropy of the network. It is an anchor-free connectivity-based localization algorithm designed for anisotropic WSNs. We mainly consider the performance of these algorithms in WSNs with moderate density, thus we do not compare with the algorithm proposed in [Kubo et al. 2012] because it requires very high network density.

7.1. Simulation Settings and Performance Metrics

Network topology: Two types of anisotropic topologies are used to evaluate the performance of different algorithms: the C-shape topology (see Figure 3(b)) and the O-shape topology (see Figure 9(a)). C-shape topology is the de facto topology to test the performance of range-free localization algorithms in anisotropic WSNs [Kubo et al. 2012; Wang et al. 2010; Kung et al. 2009; Lim and Hou 2009; Wang and Xiao 2008; Li and Liu 2010; Xiao et al. 2010b]. O-shape topology appears in many practical applications like volcano monitoring [Huang et al. 2012a], and PDS outperforms many other algorithms in this topology.

Performance metrics: We mainly consider the following metrics.

- *Distance estimation accuracy:* The estimation accuracy of d_{ij} is defined as $\mu(i, j) = |d_{ij} - \hat{d}_{ij}|$, where \hat{d}_{ij} is the estimate of d_{ij} . The distance estimation accuracy of an algorithm is defined as the average estimation accuracy over all the distance estimates that are used in localizing nodes.
- *Localization accuracy:* The localization accuracy of s_i is defined as $\gamma_i = |P_i - P'_i|$, where P_i and P'_i represent s_i 's ground-truth position and estimated position, respectively. The localization accuracy of an algorithm in a network is the average localization accuracy over all the nodes in the network.
- *Computational cost:* The computational cost of an algorithm is represented by the average number of anchors used to localize a node.

Controlled parameters: We tune the following parameters to investigate their impact on the performance of the considered algorithms.

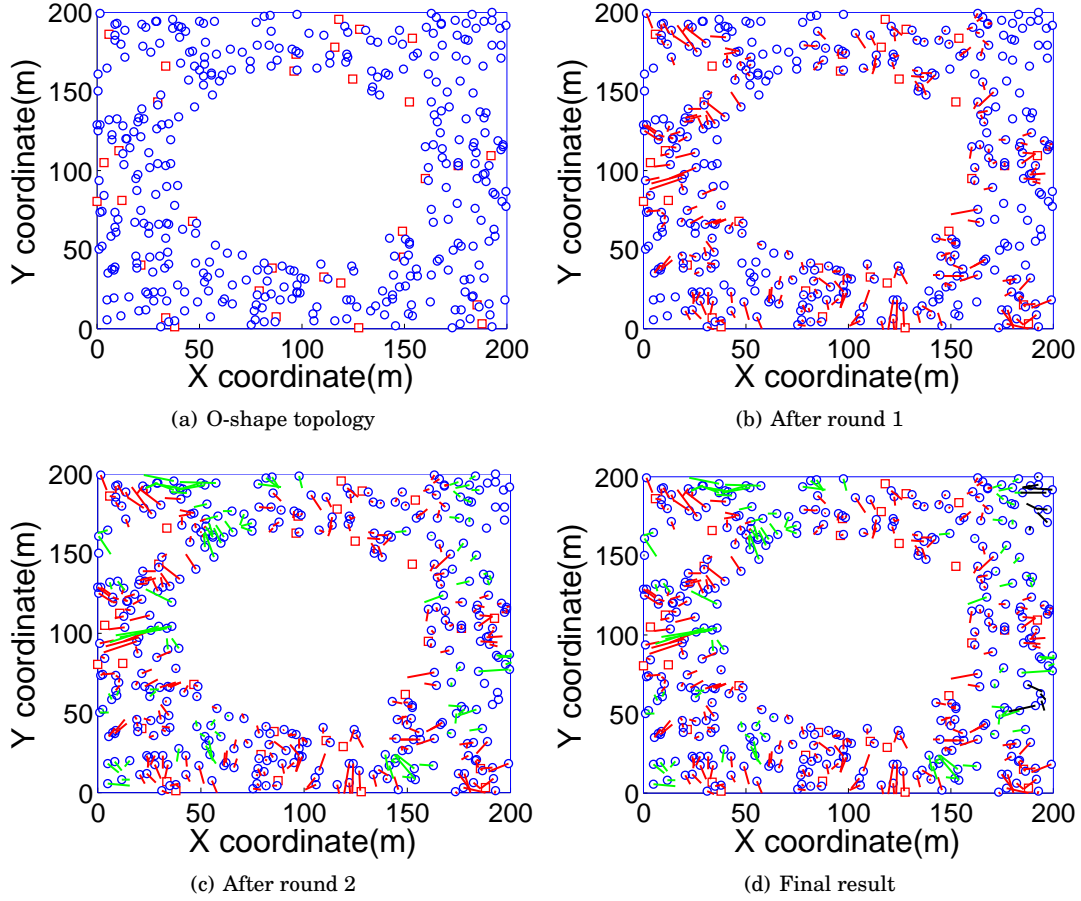


Fig. 9. Example execution of SM: (a) 400 nodes including 32 anchors (red squares) are deployed in an O-shape network ($ND \approx 15$, $AR = 0.08$), (b) results after the first round: 256 nodes are localized, the average distance estimation accuracy is $0.16r$, and the average localization accuracy is $0.27r$ (red lines), (c) results after the second round: 98 nodes are localized, the average distance estimation accuracy is $0.22r$, and the average localization accuracy is $0.37r$ (green lines), (d) results after the third round: the remaining 14 nodes are localized, the average distance estimation accuracy is $0.28r$, and the average localization accuracy is $0.4r$ (black lines). The overall localization accuracy is $0.3r$.

- *Node degree (ND)*: The node degree is defined as the average number of neighbor nodes in the network, i.e., $ND = \frac{1}{n} \sum_{i=1}^n |N(s_i)|$, where $N(s_i)$ denotes the set of s_i 's neighbors.
- *Anchor ratio (AR)*: The anchor ratio is defined as the ratio of the number of anchors to the total number of nodes in the network, i.e., $AR = m/n$.
- *Communication irregularity*: We adopt the widely used Degree of Irregularity (DOI) model [Kuhn et al. 2008; Xiao et al. 2010b] to investigate the impact of communication irregularity. Within the DOI model, the probability that two nodes can communicate with each other is controlled by a parameter d ($0 \leq d \leq 1$):

$$P(i, j) = \begin{cases} 1, & \frac{d_{ij}}{r} \leq 1 - d \\ \frac{d_{ij} - r(1-d)}{2rd}, & 1 - d < \frac{d_{ij}}{r} < 1 + d \\ 0, & \frac{d_{ij}}{r} \geq 1 + d \end{cases} . \quad (40)$$

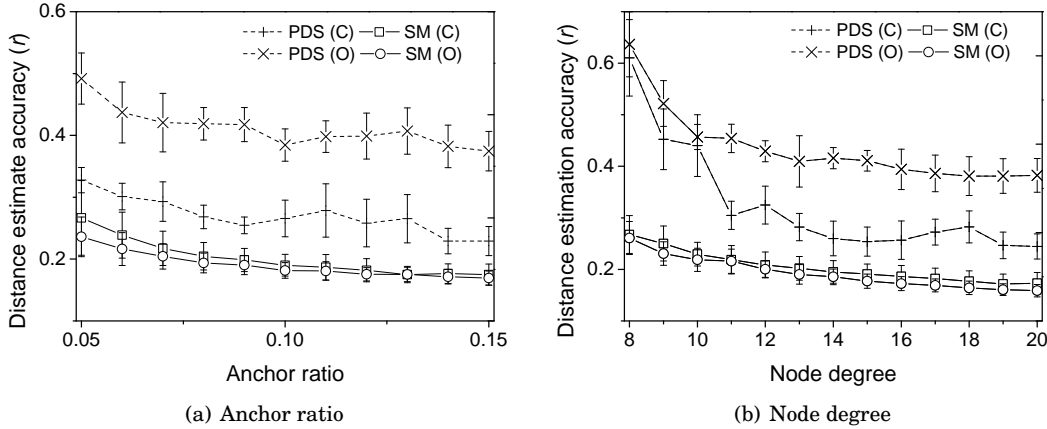


Fig. 10. Distance estimation accuracy versus (a) anchor ratio, and (b) node degree.

General settings and data collection: In default, we deploy 400 nodes in a $10r \times 10r$ square area, where r is set to $20m$ according to the previous studies [Wang and Xiao 2008; Lim and Hou 2009; Zhang et al. 2012]. The size of the void area in the C-shape topology (Figure 3(b)) and in the O-shape topology (Figure 9(a)) are $20r^2$ and $9\pi r^2$, respectively. This results in $ND \approx 14$ and $ND \approx 15$ in the C-shape topology and the O-shape topology, respectively. The default anchor number is 32, resulting in $AR = 0.08$. For every parameter setting, we randomly generate 100 network instances and report the average data over them.

7.2. A Glance at SM's Execution

We use the network plotted in Figure 9(a) to illustrate the execution flow of SM. In this figure, anchors are marked as squares and nodes are marked as circles. In the first round, every anchor calculates its PHL vector and broadcasts the vector to direct neighbors. Those nodes having at least an anchor neighbors can use SM to localize themselves. Figure 9(b) plots the result after the first round, in which 256 nodes are localized with $0.16r$ distance estimation accuracy and $0.27r$ localization accuracy. In the second round, nodes localized in the first round help their neighbors that are not localized in the first round to perform localization. In the second round 98 nodes are localized with localization accuracy $0.37r$, as shown in Figure 9(c). This procedure repeats until all the nodes are localized. In this example, all the nodes are localized after the third round, and the overall localization accuracy is $0.3r$.

7.3. Distance Estimation Accuracy

7.3.1. Impact of Anchor Ratio. Figure 10(a) plots the distance estimation accuracy in SM and PDS when AR changes from 0.05 and 0.15. It can be observed that the distance estimation accuracy in both SM and PDS improves when there are more anchors. For SM, the case with more anchors means that most nodes can use anchor's PHL vector to perform distance estimation, which effectively suppresses error accumulation caused by nodes' position error (Section 4.2). For PDS, the case with more anchors means that more nodes fall into the concentric ring (CR) pattern or the centrifugal gradient (CG) pattern, in which PDS can output fairly accurate distance estimation.

Another observation is that PDS performs worse in O-shape topology than in C-shape topology, while SM performs nearly the same in the two topologies. The reason

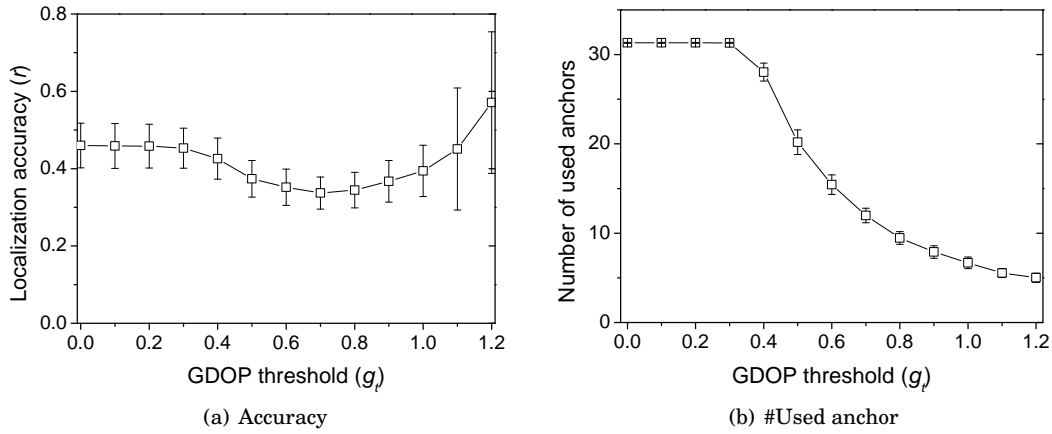


Fig. 11. GDOP threshold (g_t) versus (a) localization accuracy and (b) average number of anchors used to perform localization.

is that in O-shape topology, the shortest path between nodes might bend like a curve, but PDS requires the shortest path being a series of line segments to rectify the obtained distance estimation. Thus, the accuracy of PDS degrades in O-shape topologies. Compared with PDS, SM improves the accuracy of distance estimation by up to 34 percent in C-shape topology and up to 57 percent in O-shape topology, respectively.

7.3.2. Impact of Node Degree. Figure 10(b) plots the distance estimation accuracy in SM and PDS when ND varies. It can be observed that both algorithms achieve higher distance estimation accuracy with higher node degree. SM can obtain more accurate proximity estimation between adjacent nodes when the degree of nodes increases, thus its distance estimation accuracy improves. For PDS, because it uses a monte carlo method to smooth the estimation of the last hop which achieves higher accuracy when node degree increases, its accuracy also improves. SM always achieves much higher distance estimation accuracy than PDS, by up to 57 percent in the C-shape topology and up to 59 percent in the O-shape topology, respectively.

7.4. Localization Accuracy

7.4.1. Impact of the GDOP Threshold. The GDOP threshold (g_t) affects which anchors would be selected to perform localization for a node, and thus should be carefully set. Figure 11(a) plots the localization accuracy of SM in C-shape topology (the results for O-shape networks are similar) when g_t varies from 0 to 1.2. It can be observed that when g_t is very small (≤ 0.4) or very large (≥ 1), the localization accuracy degrades. When g_t is very small, as shown in Figure 11(b), almost all anchors are selected. In this case, those anchors to which the distance estimations contains large error will lower down the node's localization accuracy. On the other hand, when g_t is very large, only a few anchors would be selected (Figure 11(b)). In this case, the node cannot exploit the redundancy of anchors to improve localization accuracy. With extensive simulations, we observe that in most cases the highest accuracy is achieved when $g_t = 0.7$, and use this value in the following experiments.

7.4.2. Impact of Anchor Ratio. Figure 12(a) and Figure 12(b) plot the localization accuracy of SM and PDS in C-shape topology and in O-shape topology when AR varies from 0.05 to 0.15, respectively. It can be observed that the accuracy of SM and PDS

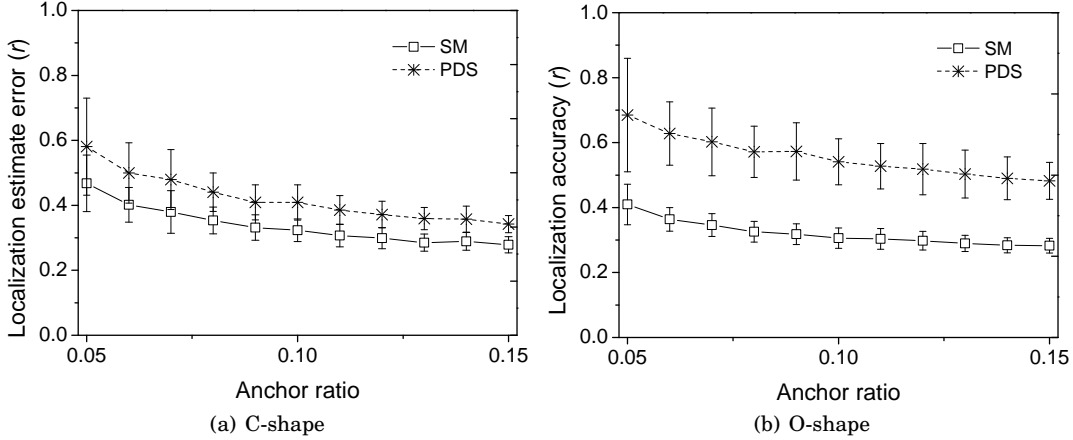


Fig. 12. Localization accuracy versus anchor ratio: (a) In C-shape topology; (b) In O-shape topology.

both improves when AR increases. The reason is that, as analyzed in Section 7.3.1, in SM and PDS the distance estimation accuracy improves with more anchors, and thus their localization accuracy improves when AR increases. Note that we do not plot the localization accuracy of MDS-MAP(P) because MDS-MAP(P) is actually an anchor-free algorithm that can run without anchors. In MDS-MAP(P) anchors are only used to translate the obtained relative coordinate system to the absolute coordinate system determined by anchors; they are not helpful to improving localization accuracy. In contrast, SM and PDS rely on anchors to determine positions of nodes.

In most cases, the localization accuracy of SM is higher than $0.4r$ and even higher than $0.3r$ when $AR \geq 0.1$. As have been pointed out in [He et al. 2005], $0.4r$ is the critical accuracy below which many location-based protocols cannot function properly. Thus SM can satisfy the needs of many such protocols, e.g., geographical routing [Huang et al. 2012b] and event detection [He et al. 2012]. PDS also achieves fairly high localization accuracy in C-shape topology (higher than $0.4r$ when $AR > 0.08$), but its accuracy in O-shape network is always lower than $0.4r$. In average, compared with PDS, SM improves localization accuracy by 20 percent in C-shape topology and by 42 percent in O-shape topology, respectively.

7.4.3. Impact of Node Degree. Figure 13(a) and Figure 13(b) plot the localization accuracy of the three algorithms in C-shape topology and O-shape topology when ND increases, respectively. It can be observed that when ND increases, the accuracy of SM and MDS-MAP(P) significantly improves while the accuracy of PDS improves only slightly. For SM, with a larger ND it can obtain more accurate distance estimations (see Section 7.3.1) and thus achieves higher accuracy. For MDS-MAP(P), with a larger ND it can build more accurate local coordinate systems and thus improve its localization accuracy. In contrast, in the PDS algorithm ND mainly impacts the estimation accuracy of the last hop, which limits the improvement in its overall localization accuracy. MDS-MAP(P) achieves comparable localization accuracy as PDS when ND is larger than 18, but SM always outperforms both of them. Compared with PDS, SM improves localization accuracy by up to 30 percent in C-shape topology and up to 52 percent in O-shape topology, respectively.

7.4.4. Impact of Communication Irregularity. MDS-MAP(P) requires symmetrical communication between nodes, and thus it cannot deal with communication irregularity. Fig-

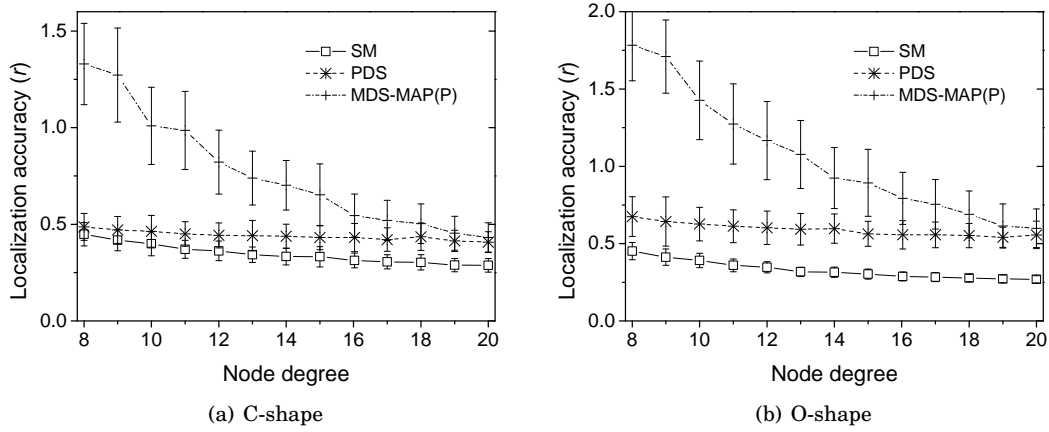


Fig. 13. Localization accuracy versus node degree: (a) In C-shape topology; (b) In O-shape topology.

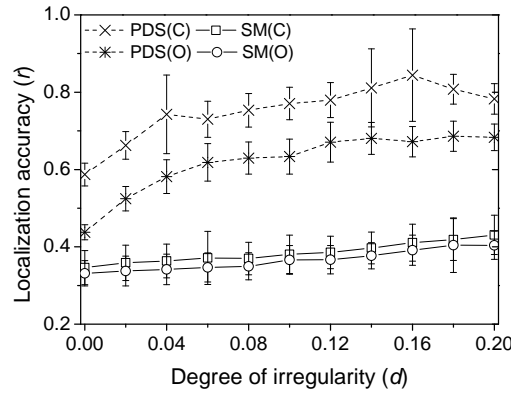


Fig. 14. Localization accuracy versus communication irregularity in SM and PDS.

Figure 14 plots the localization accuracy of SM and PDS under different communication irregularity. It can be observed that SM is more robust to the communication irregularity than PDS. When d is smaller than 0.1, the localization accuracy of SM is almost not affected. Even when d is as large as 0.2, SM still achieves high accuracy, i.e., $0.43r$ in C-shape topology and $0.4r$ in O-shape topology. In contrast, the accuracy of PDS quickly degrades when d increases. When d increases from 0 to 0.2, the accuracy of PDS drops from $0.44r$ to $0.68r$ in C-shape topology, and drops from $0.58r$ to $0.78r$ in O-shape topology. The reason of the dramatic accuracy degradation of PDS is that it builds its distance estimation algorithm on top of Amorphous [Nagpal et al. 2003], whose accuracy severely degrades when the communication is highly irregular. In contrast, SM uses a novel locality-based distance estimation approach, which is affected by only local differences between nodes and thus is more robust to communication irregularity.

7.5. Computational Overhead

Figure 15 plots how the average number of anchors every node uses to perform localization changes when the total number of anchors increases. It can be observed that

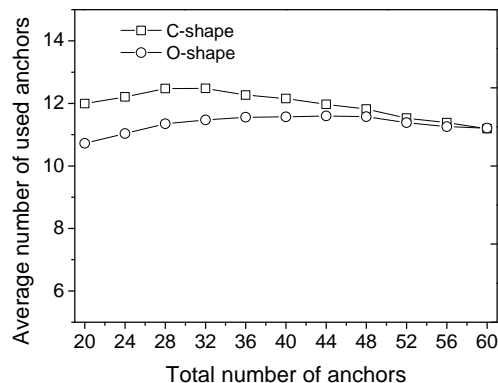


Fig. 15. Average number of anchors every node uses in localization versus the total number of anchors.

no matter how many anchors there are in the network, the average number of anchors every node used to perform localization is only slightly larger than ten, which is nearly independent to the total number of anchors. In contrast, in traditional range-free localization algorithms, all the anchors are used to perform localization. Thus the computational overhead of SM is fairly light, making it suitable to large scale WSNs.

8. RELATED WORK

Early researches on range-free localization target isotropic WSNs, e.g., DV-Hop [Niculescu and Nath 2001], Amorphous [Nagpal et al. 2003], APIT [He et al. 2005] and MDS-MAP [Shang et al. 2004a]. The two pioneering works, DV-Hop and Amorphous, provide basic technique to estimate distance between nodes and thus are the basis of many following researches including [Li and Liu 2010; Wang and Xiao 2008; Wang et al. 2010; Xiao et al. 2010b; Shang et al. 2004b; Tan et al. 2010]. However, DV-Hop and Amorphous heavily rely on the isotropy of the network to obtain accurate distance estimation. In anisotropic WSNs, their accuracy severely degrade and thus result in poor localization accuracy.

Some robust position estimators [Wang and Xiao 2008; Kung et al. 2009; Li et al. 2005] are proposed to tolerate erroneous distance estimations in anisotropic WSNs. The *i*-Multihop estimator proposed in [Wang and Xiao 2008] can recognize and filter out largely erroneous distance estimations and achieve higher accuracy than traditional minimum mean square error (MMSE) estimator. The snap-inducing shaped residuals (SISR) estimator developed in [Kung et al. 2009] is applicable to cases in which distance estimations are either very accurate or contain very large error. The least median square (LSM) estimator [Li et al. 2005] improved localization accuracy by minimizing median square error instead of mean square error. Our locality-based distance estimation approach can be combined with these estimators to exploit their advantage to further improve localization accuracy.

Rectifying the obtained distance estimations before using them to calculate node positions is another direction to cope with distance estimation errors in anisotropic WSNs [Li and Liu 2010; Wang et al. 2010; Xiao et al. 2010b; Tan et al. 2010; Lim and Hou 2009]. The Proximity Distance Mapping (PDM) algorithm replaces PHL with a proximity-distance mapping matrix to estimate distance between nodes and anchors. Although improving distance estimation accuracy, the PDM algorithm incurs much higher communication cost than DV-Hop. The *R*Ended Path (REP) algorithm [Li and Liu 2010] exploited geometric feature of the network to rectify the distance estimation

of detoured paths. In [Wang et al. 2010] and [Tan et al. 2010] the authors proposed two methods to identify turning points of detoured paths, according to which the distance estimation of detoured paths can be rectified. In the PDS algorithm [Xiao et al. 2010b], three different patterns are identified and nodes in different patterns use different method to rectify the obtained distance. These algorithms first adopt DV-Hop or Amorphous to obtain distance between nodes and then perform rectification. Our locality-based distance estimation approach could be integrated into these algorithms as a replacement of DV-Hop or Amorphous.

Different anchor combinations will impact the localization accuracy. In [Shang et al. 2004b] the authors proposed a simple heuristic algorithm that uses four nearest anchors to perform localization. The algorithm cannot fully utilize the redundancy in anchors to improve localization accuracy when the deployment region has complex shape. In [Yang and Liu 2010] the authors proposed the *quality of trilateration* (QoT) metric to characterize the quality of a trilateration result. However, the QoT metric cares only trilateration and incurs high computational cost when used in multilateration. Compared with them, our GDOP-based anchor selection algorithm jointly considers geometry distribution and distance estimation accuracy of anchors and can be used in multilateration.

9. CONCLUSION

In this paper, we revealed the locality of PHL in anisotropic WSNs, and proposed a novel distance estimation approach that achieves much accurate distance estimation accuracy than state-of-the-art solutions. We devise an adaptive anchor selection algorithm that jointly considers geometry distribution and distance estimation accuracy of anchors to improve localization accuracy. The SM localization algorithm built on top of the locality-based distance estimation and the adaptive anchor selection achieves localization accuracy higher than $0.3r$, which can well satisfy the requirements of typical location-dependent protocols/algorithms in WSNs.

Our locality-based distance estimation approach can be used as the replacement of DV-Hop or Amorphous in many range-free localization algorithms [Li and Liu 2010; Wang et al. 2010; Wang and Xiao 2008; Xiao et al. 2010b; Tan et al. 2010] to provide more accurate distance estimates. In the future, we plan to investigate how to integrate this approach with existing distance rectifying techniques to further improve accuracy of range-free localization algorithms in both isotropic and anisotropic WSNs. The performance of the proposed

ACKNOWLEDGMENTS

REFERENCES

- James Aspnes, Tolga Eren, David Kiyoshi Goldenberg, A. Stephen Morse, Walter Whiteley, Yang Richard Yang, Brian D. O. Anderson, and Peter N. Belhumeur. 2006. A Theory of Network Localization. *IEEE Transactions on Mobile Computing* 5, 12 (2006), 1663–1678.
- Gaddi Blumrosen, Bracha Hod, Tal Anker, Danny Dolev, and Boris Rubinsky. 2013. Enhancing RSSI-based tracking accuracy in wireless sensor networks. *ACM Transactions on Sensor Networks* 9, 3 (2013), 1–29.
- Kai Bu, Qingjun Xiao, Zhixin Sun, and Bin Xiao. 2012. Toward collinearity-aware and conflict-friendly localization for wireless sensor networks. *Computer Communications* 35, 13 (2012), 1549–1560.
- Jose A Costa, Neal Patwari, and Alfred O Hero III. 2006. Distributed weighted-multidimensional scaling for node localization in sensor networks. *ACM Transactions on Sensor Networks (TOSN)* 2, 1 (2006), 39–64.
- Shibo He, Jiming Chen, and Youxian Sun. 2012. Coverage and connectivity in duty-cycled wireless sensor networks for event monitoring. *IEEE Transactions on Parallel and Distributed Systems* 23, 3 (2012), 475–482.

- Tian He, Chengdu Huang, Brian M Blum, John A Stankovic, and Tarek F Abdelzaher. 2005. Range-free localization and its impact on large scale sensor networks. *ACM Transactions on Embedded Computing Systems* 4, 4 (2005), 877–906.
- Pei Huang, Chen Wang, and Li Xiao. 2012b. Improving end-to-end routing performance of greedy forwarding in sensor networks. *IEEE Transactions on Parallel and Distributed Systems* 23, 3 (2012), 556–563.
- Renjie Huang, Wen-Zhan Song, Mingsen Xu, Nina Peterson, Behrooz Shirazi, and Richard LaHusen. 2012a. Real-World Sensor Network for Long-Term Volcano Monitoring: Design and Findings. *IEEE Transactions on Parallel and Distributed Systems* 23, 2 (2012), 321–329.
- Xiang Ji and Hongyuan Zha. 2004. Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling. In *Proc. of the 23rd IEEE International Conference on Computer Communications (Infocom)*, Vol. 4. IEEE, 2652–2661.
- Takeshi Kubo, Atsushi Tagami, Teruyuki Hasegawa, Toru Hasegawa, and Jean C. Walrand. 2012. Range-free localization using grid graph extraction. In *Proc. of 20th IEEE International Conference on Network Protocols (ICNP)*. 1–11.
- Fabian Kuhn, Roger Wattenhofer, and Aaron Zollinger. 2008. Ad hoc networks beyond unit disk graphs. *Wireless Networks* 14, 5 (2008), 715–729.
- H. T. Kung, Chit-Kwan Lin, Tsung-Han Lin, and Dario Vlah. 2009. Localization with snap-inducing shaped residuals (SISR): coping with errors in measurement. In *Proc. of the 15th Annual International Conference on Mobile Computing and Networking (Mobicom)*. 333–344.
- R.B. Langley. 1999. Dilution of precision. *GPS world* 10, 5 (1999), 52–59.
- Sol Lederer, Yue Wang, and Jie Gao. 2009. Connectivity-based localization of large-scale sensor networks with complex shape. *ACM Transactions on Sensor Networks (TOSN)* 5, 4 (2009), 31.
- Mo Li and Yunhao Liu. 2010. Rendered path: range-free localization in anisotropic sensor networks with holes. *IEEE/ACM Transactions on Networking* 18, 1 (2010), 320–332.
- Zang Li, Wade Trappe, Yanyong Zhang, and Badri Nath. 2005. Robust statistical methods for securing wireless localization in sensor networks. In *Proc. of the 4th International Symposium on Information Processing in Sensor Networks (IPSN)*. 91–98.
- Hyuk Lim and Jennifer C. Hou. 2009. Distributed Localization for Anisotropic Sensor Networks. *ACM Transactions on Sensor Networks* 5, 2 (2009), 11:1–11:26.
- Sudip Misra and Sweta Singh. 2012. Localized policy-based target tracking using wireless sensor networks. *ACM Transactions on Sensor Networks* 8, 3 (2012), 1:27.
- Radhika Nagpal, Howard Shrobe, and Jonathan Bachrach. 2003. Organizing a Global Coordinate System from Local Information on an Ad Hoc Sensor Networks. In *Proc. of the 2nd International Workshop on Information Processing in Sensor Networks (IPSN)*. 333–348.
- Miguel Navarro, Tyler W Davis, Yao Liang, and Xu Liang. 2013. ASWP: a long-term WSN deployment for environmental monitoring. In *Proc. of the 12th International Conference on Information Processing in Sensor Networks (IPSN)*. 351–352.
- Dragos Niculescu and Badri Nath. 2001. Ad-Hoc Positioning System (APS). In *Proc. of 2001 IEEE Global Telecommunications Conference (Globecom)*. 2926–2931.
- Yi Shang, Wheeler Ruml, Ying Zhang, and Markus P. J. Fromherz. 2004a. Localization from Connectivity in Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems* 15, 11 (2004), 961–974.
- Yi Shang, HongChi Shi, and Ahmed A.Ahmed. 2004b. Performance study of localization methods for ad-hoc sensor networks. In *Proc. of the 1st IEEE Conference on Mobile Ad-hoc and Sensor Systems (MASS)*. 184–193.
- Guang Tan, Hongbo Jiang, Shengkai Zhang, and Anne-Marie Kermarrec. 2010. Connectivity-based and anchor-free localization in large-scale 2d/3d sensor networks. In *Proc. of the 11th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. 191–200.
- Chen Wang and Li Xiao. 2008. Sensor localization in concave environments. *ACM Transactions on Sensor Networks* 4, 1 (2008), 3:1–3:31.
- Xiaoping Wang, Jun Luo, Yunhao Liu, Shanshan Li, and Dezun Dong. 2011. Component-based localization in sparse wireless networks. *IEEE/ACM Transactions on Networking* 19, 2 (2011), 540–548.
- Yun Wang, Kai Li, and Jie Wu. 2010. Distance Estimation by Constructing The Virtual Ruler in Anisotropic Sensor Networks. In *Proc. of 29th IEEE International Conference on Computer Communications (Infocom)*. 1172–1180.
- Guang Wu, Shu Wang, Bang Wang, Yan Dong, and Shu Yan. 2012. A novel range-free localization based on regulated neighborhood distance for wireless ad hoc and sensor networks. *Computer Networks* 56, 16 (2012), 3581–3593.

- Bin Xiao, Lin Chen, Qingjun Xiao, and Minglu Li. 2010a. Reliable Anchor-Based Sensor Localization in Irregular Areas. *IEEE Transactions on Mobile Computing* 9, 1 (2010), 92–102.
- Qingjun Xiao, Kai Bu, Zhijun Wang, and Bin Xiao. 2013. Robust localization against outliers in wireless sensor networks. *ACM Transactions on Sensor Networks* 9, 2 (2013), 1–24.
- Qingjun Xiao, Bin Xiao, Jiannong Cao, and Jianping Wang. 2010b. Multihop Range-Free Localization in Anisotropic Wireless Sensor Networks: A Pattern-Driven Scheme. *IEEE Transactions on Mobile Computing* 9, 11 (2010), 1592–1607.
- Zheng Yang and Yunhao Liu. 2010. Quality of Trilateration: Confidence-Based Iterative Localization. *IEEE Transactions on Parallel and Distributed Systems* 21, 5 (2010), 631–640.
- R. Yarlagadda, I. Ali, N. Al-Dhahir, and J. Hershey. 2002. Gps gdop metric. In *IEE Proceedings: Radar, Sonar and Navigation*. 259–264.
- Shigeng Zhang, Jianxin Wang, Xuan Liu, and Jiannong Cao. 2012. Range-free selective multilateration for anisotropic wireless sensor networks. In *Proc. of the 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*. 299–307.
- Ziguo Zhong and Tian He. 2011. RSD: A metric for achieving range-free localization beyond connectivity. *IEEE Transactions on Parallel and Distributed Systems* 22, 11 (2011), 1943–1951.
- Xiaojun Zhu, Xiaobing Wu, and Guihai Chen. 2012b. Refining hop-count for localisation in wireless sensor networks. *International Journal of Sensor Networks* 12, 4 (2012), 232–243.
- Ying Zhu, Minsu Huang, Siyuan Chen, and Yu Wang. 2012a. Energy-efficient topology control in cooperative ad hoc networks. *IEEE Transactions on Parallel and Distributed Systems* 23, 8 (2012), 1480–1491.

Received March 2014; revised March 2014; accepted March 2014