
Refining hop-count for localisation in wireless sensor networks

Xiaojun Zhu, Xiaobing Wu* and Guihai Chen

State Key Laboratory for Novel Software Technology,
Nanjing University,
Nanjing, Jiangsu 210093, China
Email: gxjzhu@gmail.com
Email: wuxb@nju.edu.cn
Email: gchen@nju.edu.cn
*Corresponding author

Abstract: Distance estimation is a crucial component in localisation for wireless sensor networks. Among the estimation methods, hop-count is widely used in situations where only connectivity information is available. However, hop-count is integer-valued, implying crude distance estimation. In this paper, we refine hop-count to achieve better distance estimation. This is done by estimating neighbour distance and then approximating non-neighbour distance by the length of the shortest path. To estimate neighbour distance, we propose three estimators and show that they have negligible bias. We also show that the variance of the estimators is related to node density. The final refined hop-counts are further studied by simulations. Results verify the improvement on distance estimation and show that existing localisation methods can benefit from the improvement in various scenarios.

Keywords: wireless sensor network; range-free localisation; refined hop-count; distance estimation; proximity estimation.

Reference to this paper should be made as follows: Zhu, X., Wu, X. and Chen, G. (2012) 'Refining hop-count for localisation in wireless sensor networks', *Int. J. Sensor Networks*, Vol. 12, No. 4, pp.232–243.

Biographical notes: Xiaojun Zhu received the BS degree in Computer Science from Nanjing University in 2008. He is pursuing his PhD degree in Nanjing University. His current research interests include RFID systems, smartphone systems, wireless sensor networks and vehicular networks.

Xiaobing Wu is an Assistant Professor in Nanjing University. He received his BS and MEng degrees in Computer Science from Wuhan University in 2000 and 2003, respectively. He earned his PhD degree in Computer Science from Nanjing University (NJU) in 2009. He had been with China Merchants Bank for two years before moving to NJU in 2005. His research interests are mainly in the areas of wireless networking.

Guihai Chen is a Professor at Nanjing University, China. He received his BS degree from Nanjing University in 1984, the ME degree from Southeast University in 1987, and the PhD degree from the University of Hong Kong in 1997. He has a wide range of research interests with focus on sensor networks, peer-to-peer computing, high-performance computer architecture, and combinatorics.

1 Introduction

Localisation is a fundamental issue in wireless sensor networks and attracts much research effort (Hightower and Borriello, 2001; Liu et al., 2010). Most localisation methods take distance information as input. Based on the granularity of the distance information, we can classify existing localisation methods into two categories: range-based and range-free. Range-based methods rely on accurate distance measurement, e.g. TDOA (Priyantha et al., 2000). Though they can achieve high accuracy, the requirements on extra hardware or intensive labour work make them only

applicable to small scale networks. On the other hand, range-free methods do not rely on precise distance measurements, and they use *hop-count* as the distance measurement, (e.g. Niculescu and Nath, 2003). Here hop-count is defined as the least number of hops between two nodes; therefore, it can be computed from connectivity information.

However, hop-count is integer-valued in nature, so it suffers from coarse-grained accuracy. To achieve better accuracy, researchers seek approaches to provide sub-hop resolution (Zhong and He, 2009; Xi et al., 2010). Among them, RSD (Zhong and He, 2009) uses RSSI information to

identify the relative near-far relationship between neighbours, and Virtual-hop (Xi et al., 2010) exploits neighbour distribution to refine hop-count. Both are heuristic in nature. Though they work well in the proposed scenarios, it is unclear how the network model can influence their solutions. Instead, we seek non-heuristic solutions to the same problem.

In this work, we refine hop-count by mere connectivity information. This is done by relating the geometry of radio coverage to the number of nodes within a certain area. Specifically, we consider to use the number of common neighbours to infer the distance between two neighbouring nodes. For this purpose, we propose three estimators, and show the associated bias and variance by numerical analysis. Since the estimators are only suitable for neighbours, we extend it to the non-neighbour scenario by using the shortest path distance to approximate non-neighbours' distance. Consequently, there are three resulting refined-hop-counts. They can be incorporated to both distributed and centralised localisation algorithms that are based on connectivity information. Numerical simulations confirm the improvement over traditional hop-count. We also apply our proposed method to other two scenarios, namely irregular radio coverage scenario and anisotropic node distribution scenario. Results indicate that our method is also effective in these two scenarios.

The rest of the paper is organised as follows. Section 2 reviews related works. Section 3 formulates the problem. We show the outline of our approach in Section 4. The detailed neighbour distance estimation methods are presented in Section 5 and numerically studied in Section 6. Section 7 evaluates our solutions and Section 8 concludes the paper.

2 Related works

In the following, we first review the range-free methods, and then review works devoting to sub-hop resolution.

Range-free methods require only connectivity information and are cost-effective alternatives to locate large-scale sensor networks (Niculescu and Nath, 2003; Shang et al., 2003; Biswas et al., 2006; Giorgetti et al., 2007; Wang et al., 2011). DV-hop (Niculescu and Nath, 2003) is a representative work. In DV-hop, each node computes the hop-count distance to each beacon. Each beacon floods a distance-per-hop message to the network, so that each unknown node can calculate the absolute distance to beacons and then be localised. Another set of range-free methods uses optimisation-based techniques. Examples include SDP (Biswas et al., 2006), MDS-MAP (Shang et al., 2003), SISR (Kung et al., 2009), SOM (Giorgetti et al., 2007) etc. They can use both connectivity information and range measurements. By considering anisotropic node distribution, the work of Lim and Hou (2005) finds a linear mapping that transforms proximity measurements between sensors to geographic distances. These works are orthogonal to ours and can incorporate our refined hop-count by using it as the hop-count or proximity measurement. Another interesting topic is to provide analytical error bound for a localisation method given only connectivity information. The results for DV-Hop

and MDS-MAP have been reported by Karbasi and Oh (2010) and Oh et al. (2010), respectively.

Since hop-count is integer-valued, its performance can be improved if we use more information. Amorphous (Nagpal, 1999; Bachrach et al., 2004) is the earliest to provide sub-hop resolution for hop-count. It adjusts hop-count by averaging over all neighbours hop-counts minus 0.5. However, the derivation is based on one-dimensional networks. The work of Ma et al. (2009) calculates the node-to-anchor distance by considering the hop distribution of a node's neighbours. RSD (Zhong and He, 2009) proposes *regulated signature distance* to measure the distance between neighbouring nodes, which is based the observation that from a node's point of view, a larger RSSI indicates a shorter distance. Similar observations have also been used to infer the relative locations of sensors deployed along a line (Zhu and Chen, 2011). Virtual-hop, a sub-system in CDL (Xi et al., 2010), relies on the following heuristic rule: a node nearer to the beacon is likely to have more previous-hop neighbours and less next-hop neighbours. In this work, we analytically refine hop-count based on radio coverage and node distribution.

The closest works to ours are the geometric distance estimation methods (Kroller et al., 2006; Aslam et al., 2009). Assuming that the number of nodes in any disk follows uniform distribution, the work of Kroller et al. (2006) expressed the distance between neighbours in terms of the fraction of their common neighbours, i.e. $|N(i) \cap N(j)| / |N_i \setminus \{j\}|$ where $N(i)$ is the set of node i 's neighbours. Arguing that $|N(i) \cup N(j)|$ is more reliable than $|N(i) \cap N(j)|$, the work of Kroller et al. (2006) improved the estimation to be in terms of $|N(i) \cap N(j)| / |N(i) \cup N(j)|$. Interestingly, with a little difference in the exact expression and approximation, these two works are basically the same as our first and second estimators given in Section 5, respectively. However, we differ in three aspects. First, we assume that nodes are uniformly distributed in the whole network field, which results in Poisson distribution of the number of nodes in a certain area, while the previous works assumed uniform distribution of the number of nodes in any certain area. Consequently, variance analysis is different. Second, our derivation naturally leads to the third estimator in Section 5, which gives more stable performance improvement under various scenarios. Third, besides the quality of distance estimation, we also studied the distance estimators in the context of localisation.

3 Problem formulation

We model the network as an undirected graph $G = (V, E)$ where V is the set of nodes and E is the set of communication links. There is a link between two nodes if and only if their distance is less than or equal to a fixed distance r , named transmission range. This assumption is named circular transmission range assumption, which is common in all connectivity-based schemes, e.g. recently in the Tan et al.'s (2010) study. We assume sensor nodes are

uniformly deployed in a field with density ρ . For node i , denote $N(i)$ as the set of its neighbours. With a little abuse of notation, let $i \in N(i)$. We assume each node knows its neighbours by some neighbour discover scheme such as the one proposed by You et al. (2012). Given two nodes i and j , the hop-count distance $h(i, j)$ is defined as the least number of nodes that a message from node i must encounter before arrival at node j . Note that $h(i, j) = h(j, i)$. Let $d(i, j)$ be the physical distance between nodes i and j .

The above model is widely used in range-free localisation works (e.g. Niculescu and Nath, 2003; Lotker et al., 2004), and the hop-count is used to approximate the physical distance. To show how hop-count performs, we generate a network with 400 nodes uniformly deployed in a 10×10 square with r as the unit length. For each pair of nodes, we plot their hop-count with respect to their physical distance in Figure 1. It is evident that hop-count is discrete and each hop-count may correspond to a wide range of distances. Therefore, hop-count can only provide a crude estimation of distance. Our target in this paper is to design a new distance metric to replace hop-count such that the resulting plot has better correspondence to physical distance.

Figure 1 Hop-count vs physical distance. Four hundred nodes are deployed in a 10×10 square with $r = 1$ (see online version for colours)

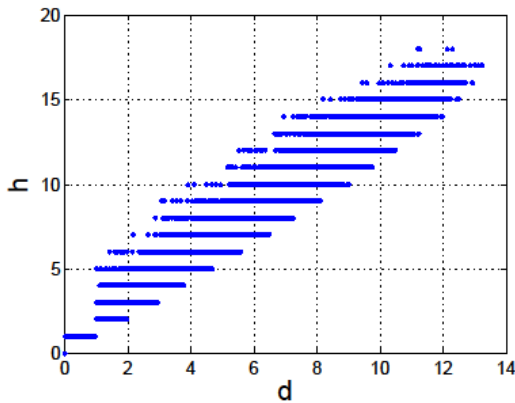


Table 1 Terminologies

Symbol	Definitions
n	Number of sensor nodes
ρ	The number of nodes per unit square.
r	Transmission range
τ	Average neighbourhood size, $\tau = \rho\pi r^2$
$d(i, j)$	Physical distance between nodes i and j
$h(i, j)$	The hop-count distance from i to j
$\tilde{l}(i, j)$	Temporary hop-count distance from i to j , only defined for neighbours
$l(i, j)$	refined hop-count distance from i to j
$N(i)$	Neighbour set of node i . Note $i \in N(i)$.

Note that there are idealised assumptions in the model including uniform node distribution and circular radio coverage. These assumptions make theoretical derivation

tractable and the results under these assumptions can provide insights into the distance estimation for sensor nodes. In the evaluation section, we also consider situations with non-uniform node distribution and radio irregularity.

4 Outline of our approach

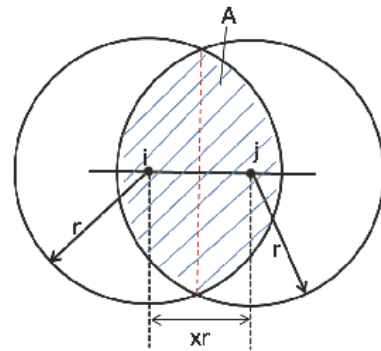
In this section, we present the basic idea about our design of refined hop-count l . To compute this refined hop-count, nodes exchange neighbour information with their neighbours and compute distances to neighbours. If two nodes are not neighbours but their distance is required, one of them floods a message to the network, and the other node computes the shortest path distance. We first introduce distance estimation for neighbouring nodes, and then extend the initial estimation to all node pairs.

4.1 Initial distance estimation

In this subsection, we focus on neighbouring nodes and provide initial distance estimation $\tilde{l}(i, j)$ between two neighbouring nodes i and j . We first relate the physical distance to area of two circular caps, and then relate the area to the number of nodes.

Consider a pair of neighbours i and j . Their radio coverage can be illustrated in Figure 2. In Figure 2, $d(i, j) = x \cdot r$ and A denotes the intersected radio coverage area. Ideally, the result we want is $l(i, j) = x$. Therefore, we need to estimate x . Our basic idea is to relate x to the geometry area of A , which is proportional to the number of nodes within A .

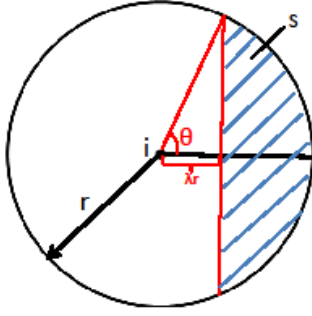
Figure 2 Radio coverage for nodes i and j . A is the common radio coverage area (see online version for colours)



For area A , it can be divided by the dashed line into two halves. We show a half in Figure 3. According to basic geometry, we have $S = \left(\theta - \frac{\sin 2\theta}{2} \right) r^2$. Note that $\theta = \arccos \lambda$, $x = 2\lambda$

and $A = 2S$, so we get $A = 2 \left(\arccos \frac{x}{2} - \frac{x}{2} \sqrt{1 - \frac{x^2}{4}} \right) r^2$. Let B

be the union of radio coverage area of nodes i and j . We have $B = 2\pi r^2 - A$. Now we have finished the geometry part of the derivation.

Figure 3 A circular cap (see online version for colours)

On the other hand, consider the nodes in area A , denote them as $N(A)$. Under our uniform assumption, $|N(A)|$ follows a Poisson distribution with expectation ρA . Similarly, $|N(B)|$ follows a Poisson distribution with expectation ρB . In summary, we have the following information:

$$A = 2 \left(\arccos \frac{x}{2} - \frac{x}{2} \sqrt{1 - \frac{x^2}{4}} \right) r^2,$$

$$B = 2\pi r^2 - A,$$

$$|N(A)| \sim \text{Pois}(\rho A),$$

$$|N(B)| \sim \text{Pois}(\rho B),$$

where $a \sim \text{Pois}(\lambda)$ means a follows Poisson distribution with parameter λ .

Our target is to estimate x based on $N(i)$ and $N(j)$. These two sets can be obtained in a local manner by broadcasting each node's neighbour list, which does not incur much overhead. Given $N(i)$ and $N(j)$, we can compute $|N(A)| = |N(i) \cap N(j)|$, or $|N(B)| = |N(i) \cup N(j)|$, or even $|N(B \setminus A)| = |N(B)| - |N(A)|$. It should be noted that $|N(A)|$ and $|N(B)|$ are not mutually independent due to the fact that area A belongs to area B . To estimate x , we can use heuristic idea, or apply rules from estimation literature. We will design three estimators in Section 5. Currently, suppose we have designed an estimator \tilde{l} . However, the obvious limitation is that it is defined only for neighbours. We then extend it to non-neighbours.

4.2 Shortest path distance as refined hop-count

For non-neighbours, we can estimate their distance by the shortest path distance between them, which completes our refined hop-count. Specifically, after obtaining $\tilde{l}(i, j)$, we construct a graph $G' = (V, E, W)$ with V being the set of nodes, E being the set of links, and $W: E \rightarrow R^+$ being the initial distance estimation as $W(i, j) = \tilde{l}(i, j)$ for $(i, j) \in E$. The refined hop-count l for any two nodes i and j is then the shortest path distance between nodes i and j in graph G' .

Here are some remarks on $l(i, j)$. First, it may not be equal to $\tilde{l}(i, j)$ even if $(i, j) \in E$. This is possible when there exists node k such that $\tilde{l}(i, k) + \tilde{l}(k, j) < \tilde{l}(i, j)$ due to the probabilistic nature of node distribution. In this case,

it is beneficial to use the shortest distance rather than the original $\tilde{l}(i, j)$, because the estimation between closer nodes is more reliable. Second, it is not necessary to compute l for all pairs. Usually, only the distances between unknown nodes and a beacon node are required (e.g. in DV-Hop; Niculescu and Nath, 2003). In this case, the distance can be computed by a DV-like flooding from the beacon node. Therefore, it incurs negligible overhead.

5 Neighbour distance estimators

We have shown the outline of our approach. In this section, we focus on deriving the distance between neighbours. We will propose three estimators, based on different formulations.

There is a basic function that we will use frequently, i.e. $\arccos \frac{x}{2} - \frac{x}{2} \sqrt{1 - \frac{x^2}{4}}$. We approximate it by a linear function of x .

Lemma 1: Let $f(x) = \arccos \frac{x}{2} - \frac{x}{2} \sqrt{1 - \frac{x^2}{4}}$ and $g(x) = \pi/2 - x$. For all $0 \leq x \leq 1$, it holds (1)

$$\frac{\pi}{3} - \frac{\sqrt{3}}{4} \leq f(x) \leq \frac{\pi}{2}; \quad (2) \quad 0 \leq f(x) - g(x) < 0.044.$$

Proof: For equation (1), since $f'(x) = -\frac{1}{2} \sqrt{4 - x^2} \leq 0$, the function $f(x)$ is monotonically decreasing. In addition, $f(0) = \frac{\pi}{2}$, and $f(1) = \frac{\pi}{3} - \frac{\sqrt{3}}{4}$.

For equation (2), let $h(x) = f(x) - g(x)$. Consider its derivative

$$h'(x) = 1 - \frac{1}{2} \sqrt{4 - x^2} \geq 0$$

which shows that $h(x)$ is an increasing function. In addition, $h(0) = 0$ and $h(1) = 1 - \frac{\sqrt{3}}{4} - \frac{\pi}{6} < 0.044$. This completes our proof.

This lemma shows that the approximation is quite good. It makes the later derivations possible. With this notation, we have $A = 2f(x)r^2$.

Let $C(i)$ be the radio coverage area of node i . Then we have $A = C(i) \cap C(j)$ and $B = C(i) \cup C(j)$.

5.1 The first estimator

Consider $C(i)$. The number of nodes within $C(i)$ is $|N(i)|$, which follows Poisson distribution with expectation $\rho C(i)$. We have the following equation

$$\frac{E[|N(A)|]}{E[|N(i)|]} = \frac{\rho A}{\rho C(i)}.$$

We will design an estimator based on this equation. We can approximate $E[|N(A)|]$ by $|N(A)|$ and $E[|N(i)|]$ by $|N(i)|$. In addition, $A = 2f(x)r^2 \approx 2g(x)r^2$ and $C(i) = \pi r^2 - x$. These lead to

$$\frac{|N(A)|}{|N(i)|} \approx \frac{2g(x)}{\pi}.$$

Recall that $N(A) = N(i) \cap N(j)$. Substituting $g(x) = \pi/2 - x$ yields

$$x \approx \frac{\pi}{2} \left(1 - \frac{|N(i) \cap N(j)|}{|N(i)|} \right)$$

which gives our first estimator

$$\tilde{l}_1(i, j) = \frac{\pi}{2} \left(1 - \frac{|N(i) \cap N(j)|}{|N(i)|} \right). \quad (1)$$

The intuition behind $\tilde{l}_1(i, j)$ is that the distance between i and j is inversely proportional to the ratio of common neighbours. Consider the extreme case where nodes i and j have the same set of neighbours, i.e. $\frac{|N(i) \cap N(j)|}{|N(i)|} = 1$.

Then $\tilde{l}_1(i, j) = 0$. It is worth mentioning that $\{i, j\} \in N(i) \cap N(j)$ since i and j are neighbouring nodes.

There is an undesired property regarding $\tilde{l}_1(i, j)$. It may be asymmetry, i.e. the distance computed by i may be different from that computed by j , since $|N(i)|$ may not be equal to $|N(j)|$ due the probabilistic nature of node distribution. This asymmetry can be eliminated by computing the average, the maximum or minimum of the two values. In this work, we adopt the average approach. Though we can solve the asymmetry issue by additional processing, it would be better if the estimator itself could give symmetric estimation. This idea leads to our second estimator.

5.2 The second estimator

Consider the area A and B . Recall that

$$\frac{E[|N(A)|]}{E[|N(B)|]} = \frac{\rho A}{\rho B}$$

which is the basis of the second estimator. Again, we can approximate $E[|N(A)|]$ by $|N(A)|$ and $E[|N(B)|]$ by $|N(B)|$. Substituting $B = 2\pi r^2 - A$ and $A = 2f(x)r^2$ gives

$$\frac{|N(A)|}{|N(B)|} = -1 + \frac{\pi}{\pi - f(x)}.$$

Since $N(A) = N(i) \cap N(j)$ and $N(B) = N(i) \cup N(j)$, approximating $f(x)$ by $g(x)$ and solving for x lead to

$$x \approx \frac{\pi}{2} \left(1 - \frac{2|N(i) \cap N(j)|}{|N(i) \cap N(j)| + |N(i) \cup N(j)|} \right).$$

Note that the term $|N(i) \cap N(j)| + |N(i) \cup N(j)|$ is actually equal to $|N(i)| + |N(j)|$. Our second estimator is

$$\tilde{l}_2(i, j) = \frac{\pi}{2} \left(1 - \frac{2|N(i) \cap N(j)|}{|N(i)| + |N(j)|} \right). \quad (2)$$

This estimator has the same intuition as the first estimator: the distance between two nodes is inversely proportionally to the percentage of common neighbours. In fact, it generalises the first estimator. To see this, suppose $|N(i)| = |N(j)|$, then $\tilde{l}_2(i, j) = \tilde{l}_1(i, j) = \tilde{l}_1(j, i)$. Our second estimator has the additional advantage that it is symmetric, i.e., the distance computed by i is always equal to that computed by j .

The above two estimators are both easy to understand and simple for computation. However, they only use expectation information and did not use variance information. Indeed, variance should be considered since it is widely known to be a performance metric for an estimator. Our third estimator is inspired by this observation.

5.3 The third estimator

The derivation of this estimator is much involved, but the underlying idea is simple. In estimation theory, an optimal estimator is an estimator that is unbiased and has minimum variance. Our derivation is driven by these two principles. The following derivations have been verified by symbolic computation engine WolframAlpha (WolframAlpha, n.d., <http://www.wolframalpha.com>).

Since $|N(A)|$ and $|N(B)|$ are not mutually independent, we consider a third area. Let $C = B \setminus A$. It is easy to see that $N(C) = N(B) \setminus N(A)$ so that $|N(C)|$ is independent of $|N(A)|$. For convenience, let $n_A = |N(A)|$ and $n_C = |N(C)|$. Let \hat{x} be the estimator. We assume

$$\hat{x} = a \cdot n_A + b \cdot n_C + c$$

where a, b, c are three constant parameters to be determined. They are irrelevant of n_A and n_C . It is interesting to consider several special cases. If we set $a = 0$ and $b = 0$, then this estimator degenerates into the traditional hop count – setting distance between neighbours as a constant. If we set $a = 0$, then the estimator becomes only relevant to the area C (the parameters b, c can be determined by the same procedure as follows). The case for $b = 0$ is similar. We consider the general case and will find a, b, c based on the two principles, unbiasedness and minimum variance.

To obtain unbiased estimator, we set

$$x = E[\hat{x}]$$

so that

$$\begin{aligned} x &= a \cdot \rho A + b \cdot \rho(2\pi r^2 - 2A) + c \\ &= a \cdot \rho \cdot 2f(x)r^2 + b \cdot \rho(2\pi r^2 - 4f(x)r^2) + c \\ &\approx a\rho(\pi - 2x)r^2 + b\rho \left(2\pi r^2 - 4\left(\frac{\pi}{2} - x\right)r^2 \right) + c \\ &= 2\rho r^2(2b - a)x + a\rho\pi r^2 + c. \end{aligned}$$

The above equation holds for all x . So we can obtain the following equalities. Let $\tau = \rho\pi r^2$. (It is the average neighbourhood size, i.e., $\tau = E[|N(i)|]$.)

$$\begin{cases} b = \frac{\pi}{4\tau} + \frac{a}{2}, \\ c = -a\tau. \end{cases} \quad (3)$$

Now consider the variance of the estimator.

$$\begin{aligned} \text{var}[\hat{x}] &= a^2 \cdot \text{var}[n_A] + b^2 \cdot \text{var}[n_C] \\ &= a^2 \cdot \rho A + b^2 \cdot \rho(2\pi r^2 - 2A) \\ &= a^2 \cdot \rho \cdot 2f(x)r^2 + b^2 \cdot \rho(2\pi r^2 - 2 \cdot 2f(x)r^2) \\ &\approx a^2 \rho(\pi - 2x)r^2 + b^2 \rho \left(2\pi r^2 - 4\left(\frac{\pi}{2} - x\right)r^2 \right) \\ &= \left(\tau - \frac{2\tau x}{\pi} \right) a^2 + \frac{4\tau x}{\pi} b^2. \end{aligned}$$

Substituting equation (3) into the above equation and arranging the terms yield

$$\begin{aligned} \text{var}[\hat{x}] &\approx \left(\tau - \frac{\tau \cdot x}{\pi} \right) a^2 + x \cdot a + \frac{\pi x}{4\tau} \\ &= \left(\tau - \frac{\tau x}{\pi} \right) \left(a + \frac{\pi x}{2\tau(\pi - x)} \right)^2 \\ &\quad + \frac{\pi x}{4\tau} \left(2 - \frac{\pi}{\pi - x} \right). \end{aligned}$$

It is easy to see that $\tau - \tau \cdot x / \pi > 0$ since $x < \pi$. To minimise $\text{var}[\hat{x}]$, we set

$$a = -\frac{\pi x}{2\tau(\pi - x)}.$$

Substituting this term back into equation (3) leads to

$$\begin{cases} a = -\frac{\pi x}{2\tau(\pi - x)}, \\ b = \frac{\pi}{4\tau} \left(1 - \frac{x}{\pi - x} \right), \\ c = \frac{\pi x}{2(\pi - x)}. \end{cases}$$

Our estimator becomes

$$\begin{aligned} \hat{x} &= -\frac{\pi x}{2\tau(\pi - x)} n_A + \frac{\pi}{4\tau} \left(1 - \frac{x}{\pi - x} \right) n_C \\ &\quad + \frac{\pi x}{2(\pi - x)}. \end{aligned} \quad (4)$$

From our derivation, we can see that $E[\hat{x}] \approx x$ and $\text{var}[\hat{x}] \approx \frac{\pi}{4\tau} \cdot x \left(2 - \frac{\pi}{\pi - x} \right)$ where ‘ \approx ’ is due to the approximation of $f(x)$ as $g(x)$.

We have several remarks about this estimator. Let us consider its variance $\text{var}[\hat{x}] \approx \frac{\pi}{4\tau} \cdot x \left(2 - \frac{\pi}{\pi - x} \right)$. For fixed

x , the variance is inversely proportional to τ , the average neighbourhood size. This coincides with our intuition. The building block of the estimator is to approximate geometry area by the number of nodes within the area. The denser the nodes are, the better this approximation is. On the other hand, for fixed τ , the variance is not monotonic with respect to x , which is not intuitive. Specifically, the function $x \left(2 - \frac{\pi}{\pi - x} \right)$ has a turning point at $x = \frac{1}{2}(2 - \sqrt{2})\pi \approx 0.92$, before which it increases and then decreases slightly after. To examine whether this is caused by the approximation of $f(x)$ as $g(x)$, we check the exact expression of $\text{var}[\hat{x}]$, which is

$$\frac{\pi}{2\tau} \left(\frac{x^2 f(x)}{(\pi - x)^2} + \frac{1}{4} (1 - x / (\pi - x))^2 (\pi - 2f(x)) \right).$$

As computed by WolframAlpha (<http://www.wolframalpha.com>), for fixed τ , this function also has a turning point (at $x \approx 0.9661$). Therefore, we conclude that the variance of \hat{x} is not strictly monotonic. However, since \hat{x} is not strictly unbiased, the slightly decreased variance after the turning point does not necessarily imply better estimation. Generally, it holds that smaller x gives smaller variance. This observation supports the shortest path distance methodology in the previous section, which favours shorter distance.

Now we continue our derivation. For equation (4), each node can obtain n_A and n_C by exchanging neighbour list as before. For the average neighbourhood size τ , we can use n_A and n_C to compute τ when no other information is available. Alternatively, when this estimator is incorporated to compute the distance between non-neighbours, we can compute τ during the shortest path computation by adding a field in the broadcast messages. This can reduce the bias of the estimation of τ due to small sample size.

The term x is not easy to work out, since x is exactly what we want to estimate. Fortunately, there is a natural way to approximate it. The idea is to apply iterative approximation. We begin by setting $x = 0$, and then obtain \hat{x} by equation (4). Then we set $x = \hat{x}$ and obtain a new estimation \hat{x} by equation (4). Repeat the process until convergence. We can take the converged \hat{x} as our estimator. This method, though straightforward, is difficult for analysis. It is unclear how much iterations are required. We take a different approach. Consider the converged \hat{x} . It must hold that

$$\hat{x} = -\frac{\pi \hat{x}}{2\tau(\pi - \hat{x})} n_A + \frac{\pi}{4\tau} \left(1 - \frac{\hat{x}}{\pi - \hat{x}} \right) n_C + \frac{\pi \hat{x}}{2(\pi - \hat{x})}.$$

Solving for \hat{x} yields two solutions

$$\hat{x}_1 = \frac{\pi}{4\tau} \left(n_A + n_C + \tau - \sqrt{(n_A + n_C + \tau)^2 - 4\tau n_C} \right),$$

and

$$\hat{x}_2 = \frac{\pi}{4\tau} \left(n_A + n_C + \tau + \sqrt{(n_A + n_C + \tau)^2 - 4\tau n_C} \right).$$

We can rule out \hat{x}_2 by observing that

$$\begin{aligned} E[\hat{x}_2] &\geq \frac{\pi}{4\tau} (E[n_A] + E[n_C] + \tau) \\ &= \frac{\pi}{4\tau} \left(\frac{2\tau}{\pi} f(x) + 2\tau - \frac{4\tau}{\pi} f(x) + \tau \right) \\ &= \frac{3\pi}{4} - \frac{f(x)}{2} \\ &\geq \frac{3\pi}{4} - \frac{\pi}{2} \\ &= \frac{\pi}{2} > x \end{aligned}$$

where the second inequality is due to Lemma 1. So we keep only the first solution.

Then our third estimator is

$$\tilde{l}_3(i, j) = \frac{\pi}{4\tau} \left(n_A + n_C + \tau - \sqrt{(n_A + n_C + \tau)^2 - 4\tau n_C} \right) \tag{5}$$

where τ is the average neighbourhood size, $n_A = |N(i) \cap N(j)|$ and $n_C = |N(i) \cup N(j)| - n_A$.

There are some additional remarks regarding average neighbourhood size τ . As mentioned before, τ can be computed by adding each node's neighbourhood size into the flooded message during shortest path computation. This is for homogeneous network that has a single average neighbourhood size τ . In fact, there is another natural way to compute τ . Each node can compute τ by averaging over its own neighbours' neighbourhood sizes. For example, node i has $|N(i)|$ neighbours so that node i can average over $|N(i)|$ nodes' neighbourhood sizes to obtain its local average neighbourhood size τ . In this way, not only overhead is further reduced, but also the resulting estimator can compensate for non-uniform node distribution.

6 Numerical results of the estimators

We have designed three estimators. In this section, we study their effectiveness from two perspectives: bias and variance. This is done by generating Poisson distributed random numbers and comparing the estimated parameter against the ground truth. We first introduce how we design the study and then give the results.

6.1 Methodology

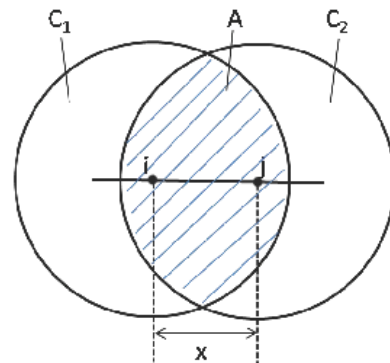
The three estimators are irrelevant of transmission range r , so we consider a unit disk scenario in Figure 4. Suppose the average neighbourhood size is τ and the distance between i and j is x . According to our assumptions, we can check that

$$\begin{cases} |N(C_1)| \sim Pois\left(\tau - \frac{2\tau}{\pi} f(x)\right), \\ |N(C_2)| \sim Pois\left(\tau - \frac{2\tau}{\pi} f(x)\right), \\ |N(A)| \sim Pois\left(\frac{2\tau}{\pi} f(x)\right). \end{cases} \tag{6}$$

More importantly, $|N(C_1)|, |N(C_2)|, |N(A)|$ are mutually independent of each other. We have $|N(i)| = |N(C_1)| + |N(A)|$, $|N(j)| = |N(C_2)| + |N(A)|$, $|N(i) \cap N(j)| = |N(A)|$ and $|N(i) \cup N(j)| = |N(C_1)| + |N(C_2)| + |N(A)|$. Several of these terms are fed into the estimators to compute the estimation for x .

We consider for $\tau = 5$ (sparse), 10(moderate), 20(dense) and $x = 0.1, 0.2, \dots, 0.9$. For each d, x pair, we randomly sample 5000 $(|N(C_1)|, |N(C_2)|, |N(A)|)$ tuples independently according to equation (6). For each tuple, we use the three estimators to estimate x .

Figure 4 Simulated scenario: the three areas C_1, A and C_2 are non-overlapping (see online version for colours)

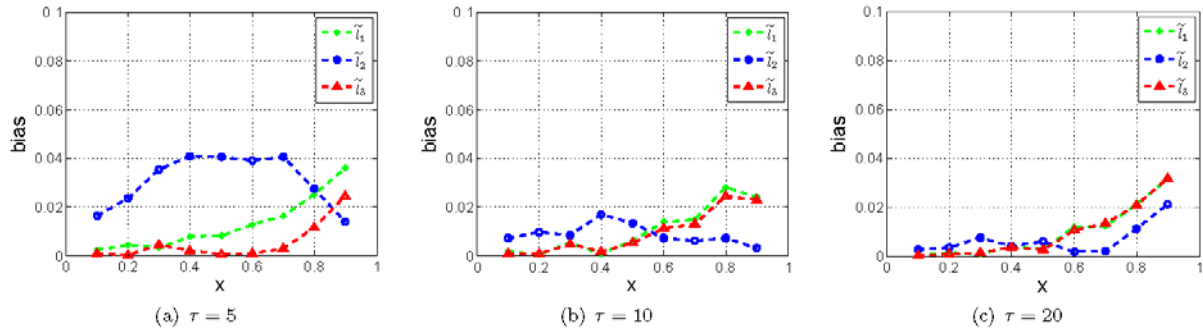
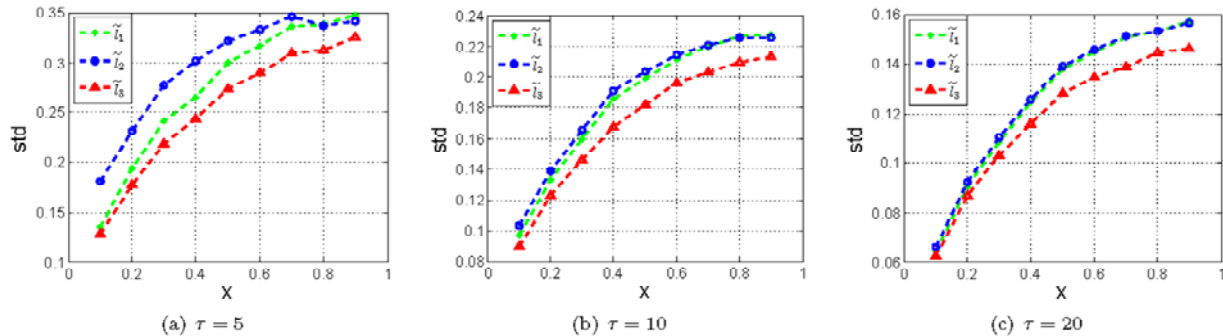


The design follows closely with our assumptions. The performance of each estimator includes the effect of approximation $f(x)$ by $g(x)$.

6.2 Bias

Define *bias* as the absolute gap between estimated x and the ground-truth x where the estimated x is the averaged estimation for 5000 tuples. Bias is unavoidable due to the approximation of $f(x)$ as $g(x)$. However, as Lemma 1 shows, this approximation has an error of at most 0.044, which is quite small. Figure 5 shows the bias of three estimators with respect to τ and x .

We have two non-intuitive observations. First, the biases of all three estimators are within 0.044, which is the approximation error of $f(x)$ as $g(x)$. This result is not obvious, since each estimator involves several occurrences of $f(x)$. Such bias is acceptable. Second, the bias trends of \tilde{l}_1 and \tilde{l}_3 are similar, which are different from \tilde{l}_2 for $\tau = 5, 10$. This is not expected since the derivations of \tilde{l}_1 and \tilde{l}_2 are the same while the derivation of \tilde{l}_3 follows another idea. Nevertheless, it is safe to say that bias generally grows with respect to x .

Figure 5 Bias of three estimators. τ is the average neighborhood size. All three cases have negligible bias (see online version for colours)**Figure 6** Standard deviation of three estimators. (see online version for colours)

6.3 Variance

Figure 6 shows the standard deviation (square root of the variance) of each estimator. All three estimators have smaller variance for larger τ , which coincides with intuition. The denser the network is, the better the approximation of geometry area as the number of nodes is. Among the three estimators, \tilde{l}_3 has consistent smaller variance than the other two estimators. This is due to the fact that \tilde{l}_3 is derived following the minimum variance principle. \tilde{l}_1 has smaller variance than \tilde{l}_2 , and the gap vanishes when the network becomes dense ($\tau = 20$).

7 Evaluations

In this section, we combine the distance estimator for neighbours with shortest path distance estimation for non-neighbours, leading to three refined hop-counts l_1, l_2 and l_3 . Besides comparison among these three, we also compare them with the traditional hop-count h . We first study the factors influencing refined hop-counts, which is necessary due to the incorporation of shortest path computation. Then we study how localisation methods can benefit from refined hop-counts. At last, we relax assumptions and observe their impacts. In the following, we set the transmission range $i = 1$.

7.1 Factors influencing l

By considering our derivation, we can pick out two factors: (a) physical distance. It can influence both the estimator for

neighbours and the shortest path computation; (b) node density ρ . It is the source of randomness in the estimator for neighbours. For low density network, the number of neighbours is very few, which causes high bias for the estimating of area by number of nodes. We conduct two set of simulations to study the two factors.

In the first scenario, we fix the network field as 10×10 square and vary the distance between two nodes. We compute the refined hop-count between the two nodes for $d = 0.5, 1.0, \dots, 5.0$. For each distance d , we fix two nodes' locations in the network such that they are relatively in the centre of the field to avoid edge effect, and their distance is d . Then we uniformly generate 398 nodes. For each distance we perform 50 iterations. We show the estimation error Δ ($\Delta = l - d$ where $l = l_1, l_2, l_3$) and the standard deviation of each estimator in Figure 7. We can see that all three estimators perform reasonably well, with error at most ± 0.2 and standard deviation at most 0.5. The third estimator has smaller standard deviation in general.

In the second scenario, we fix two nodes in the network with distance 3 (we can see that 2–5 is roughly the same from the above simulation), and vary the node density by varying the total number of nodes. We study for $n = 200, 300, \dots, 600$, which correspond to node density $\rho = 2, 3, \dots, 6$. For each number n , we perform 50 iterations. Figure 8 shows that both the estimation error and the standard deviation decrease with the increase of node density, which is expected. The third estimator still provides smaller standard deviation. Additionally, the estimation error for all three estimators becomes negligible after $\rho \geq 4$.

Figure 7 Refined hop-count for varying distance. (see online version for colours)

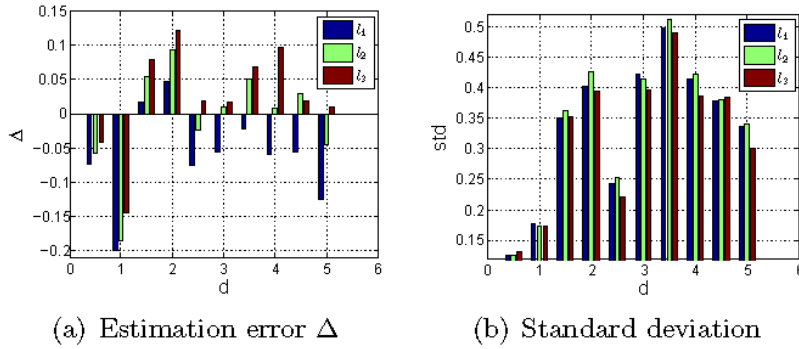
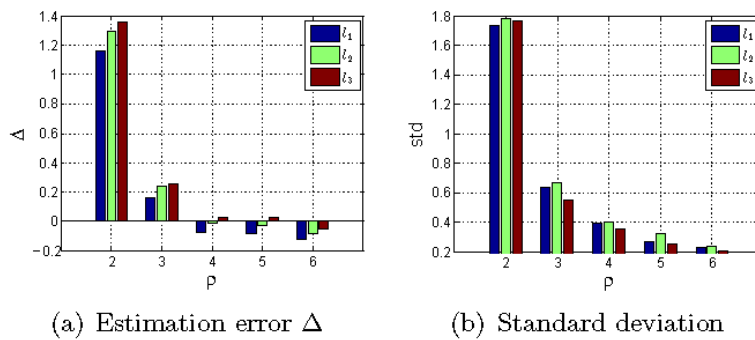


Figure 8 Refined hop-count for varying density. The distance between two nodes is fixed as 3. Note that in our settings a network with density ρ is equivalent to a network with 100ρ nodes, or a network where each node has $\pi\rho$ neighbours (including itself) on average (see online version for colours)



7.2 The effectiveness of hop-count refining method on localisation methods

Though our main focus is in refining hop-count, in this subsection, we also study its impact on localisation methods. It is worth mentioning that localisation error is an effect of both distance estimation error and the error of localisation method itself (e.g. the number of anchor nodes).

We compare our refined hop-count l with the traditional hop-count h . For other methods, RSD (Zhong and He, 2009) requires the knowledge of RSSI, which is not available under our model. It is worth mentioning that l_1 can be considered as the estimator in the work of Kroller et al. (2006) extended by shortest path computation and l_2 can be considered as the estimator in the work of Aslam et al. (2009) extended by shortest path computation. For localisation method, we choose the popular DV-hop method, due to the fact that it is fully distributed and easy to implement. The widely used evaluation metric for a localisation method is the median localisation error with localisation error being the physical distance between the estimated position and the ground-truth, and the median being taken over unknown nodes. We treat traditional hop-count method as baseline, and compute the *percentage error reduction* γ defined as

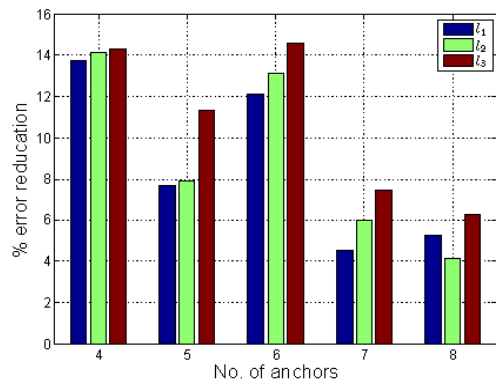
$$\gamma = \frac{e_h - e_l}{e_h}$$

where e_h is the median localisation error for traditional hop-count method, and e_l is the median localisation error for our

refined hop-count. $\gamma \geq 0$ implies that refined hop-count l incurs less error. The higher γ is, the less error l incurs.

In the first experiment, we fix node density as 4 and vary the number of anchors as 4, 5, ..., 8. Figure 9 shows that all three refined hop-counts improve over original hop-count consistently. The improvement can be up to 15%. All three refined hop-counts have comparable performance, with l_3 better than the other two.

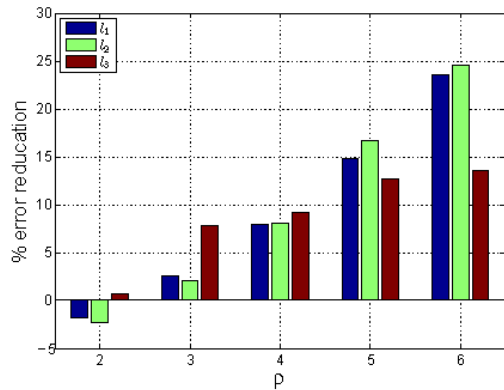
Figure 9 Error reduction for varying the number of anchors. The higher the bar is, the better the hop-count is. The node density is 4 and the result for each anchor number is averaged over 50 networks (see online version for colours)



In the second experiment, we vary node density as 2, 3, ..., 6. Figure 10 shows that the improvement magnitude increases with the increase of node density for all three refined hop-counts. For low density scenario ($\rho = 2$), l_1 and

l_2 give no improvement, but the error reduction can be up to 25% when node density is 6. The refined hop-count l_3 provides consistent improvement.

Figure 10 Error reduction for varying densities. The anchor number is 6 and the result for each density is averaged over 50 networks (see online version for colours)



7.3 Beyond the assumed scenario

We consider radio irregularity and non-uniform node distribution in this subsection.

To model radio irregularity, we use the DOI model proposed by He et al. (2003). For each sensor node, its radio coverage is irregular as follows: all nodes at a distance less than r_{low} can receive the signal; all nodes at a distance larger than r_{high} cannot receive the signal, only a portion of nodes at a distance between r_{low} and r_{high} can receive the signal. The irregularity of radio coverage is determined by DOI value, defined as the maximum radio range variations per unit degree change. We show one radio pattern with DOI = 0.2 in Figure 11. In simulations, all nodes lie in the coverage range can receive the signal from the sender. We set $r_{low} = 0.8r$ and $r_{high} = 1.2r$ in the simulation. As before, we set r as the unit length. We set the number of anchors as 6 and the number of nodes as 400. We vary DOI from 0.05 to 0.25 with a spacing of 0.05. For each DOI, we run 50 iterations.

Figure 11 One radio coverage scenario when DOI=0.1. Radio boundary is illustrated by the solid irregular curve. The inner circle has radius 0.8 and outer circle has radius 1.2. (see online version for colours)

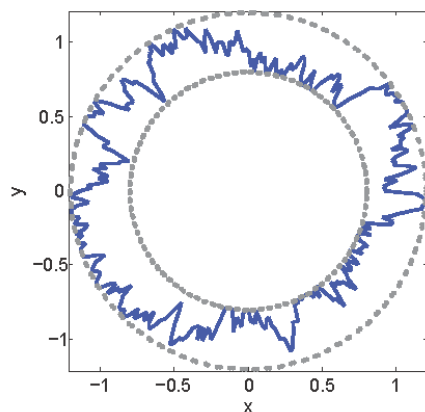
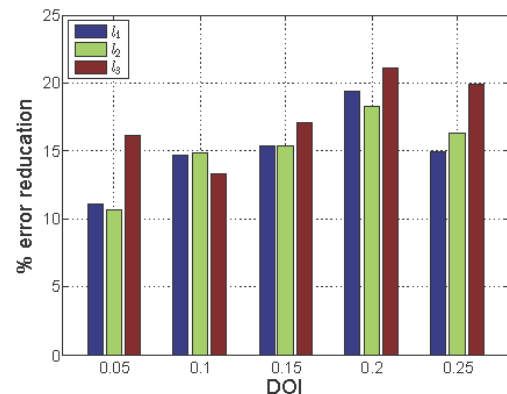


Figure 12 shows that our refined hop-counts perform stable over different DOIs and consistently better than the traditional hop-count h . This is due to fact that l computes a fine-grained hop-count between neighbours and use the shortest path distance, which reduces the effect of irregularity. The hop-count l_3 performs better than the other two.

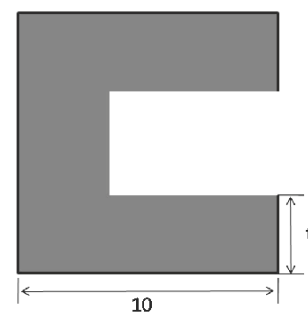
Figure 12 Median localization error for varying DOI. The anchor number is 6 and the result for each DOI is averaged over 50 networks (see online version for colours)



In the second scenario, we consider non-uniform node distribution. It is worth mentioning that there are works specifically tackling the anisotropic node distribution problem (Lim and Hou, 2005; Li and Liu, 2010). We do not target this goal specifically, and we believe their solution can be improved by replacing hop-count with our refined hop-count.

To model non-uniform distribution, we place nodes in a C-shape area as illustrated in Figure 13. Node density is 4. The parameter t controls the size of the hole. When $t = 5$, the hole disappears.

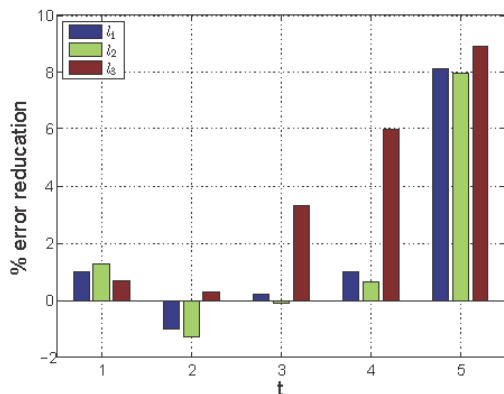
Figure 13 C-shaped area. It is obtained by cutting off a $(10-t) \times (10-2t)$ rectangle from a 10×10 square. The parameter t determines the irregularity of the shape



We consider $t = 1, \dots, 5$. Figure 14 shows the result. We find that refined hop-count l_3 always reduces localisation error, while the other two (l_1 and l_2) may increase error for some case ($t = 2$).

In summary, we find that our refined hop-counts can still help reduce localisation error in scenarios beyond our assumptions. Among the three hop-counts, l_3 gives steady performance improvement. This is caused by its minimum variance design.

Figure 14 Error reduction for different C-shape. Note that the C-shape disappears when $t = 5$ (see online version for colours)



8 Conclusion

In this paper, we derive refined hop-counts for localisation in two-dimensional wireless sensor networks. In our derivation, we use the circular radio coverage assumption to find a relationship between distance and the area of a geometry region, and then use the uniform node distribution to relate the number of nodes to the area of geometric region. Applying this idea leads to three distance estimators for neighbours. To compute the estimators, sensor nodes only need to exchange messages locally. Non-neighbours can use a DV-like protocol to compute the shortest path between them, and their distance is approximated as the length of the shortest path. We use simulations to study the factors influencing refined hop-counts and find that they work better in a network with higher density (above 5 per unit square). We also find that they improve the performance of localisation methods, both under our assumed model and a relaxed model scenario.

Acknowledgements

The authors would like to thank all the reviewers for their insightful comments. This project was supported in part by China NSF grants (60825205, 61073152, 61133006, 60903179, 61021062) and China 973 project (2012CB316200).

References

Aslam, F., Schindelbauer, C. and Vater, A. (2009) 'Improving geometric distance estimation for sensor networks and unit disk graphs', *Proceedings of the International Conference on Ultra Modern Telecommunications (ICUMT 2009)*, 12–14 October, St. Petersburg, Russia, pp.1–5.

Bachrach, J., Nagpal, R., Salib, M. and Shrobe, H. (2004) 'Experimental results for and theoretical analysis of a self-organizing global coordinate system for ad hoc sensor networks', *Telecommunication Systems*, Vol. 26, pp.213–233.

Biswas, P., Lian, T.C., Wang, T.C. and Ye, Y. (2006) 'Semidefinite programming based algorithms for sensor network localization', *ACM Transactions on Sensor*, Vol. 2, pp.188–220.

Giorgetti, G., Gupta, S. and Manes, G. (2007) 'Wireless localization using self-organizing maps', *Proceedings of the 6th International Conference on Information Processing in Sensor Networks (IPSN'07)*, 25–27 April, Cambridge, Massachusetts, USA, pp.293–302.

He, T., Huang, C., Blum, B.M., Stankovic, J.A. and Abdelzaher, T. (2003) 'Range-free localization schemes for large scale sensor networks', *Proceedings of the Ninth Annual International Conference on Mobile Computing and Networking (MOBICOM'03)*, 14–19 September, San Diego, CA, USA, pp.81–95.

Hightower, J. and Borriello, G. (2001) 'Location systems for ubiquitous computing', *Computer*, Vol. 34, pp.57–66.

Karbasi, A. and Oh, S. (2010) 'Distributed sensor network localization from local connectivity: performance analysis for the hop-terrain algorithm', *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'10)*, 14–18 June, New York, New York, USA, pp.61–70.

Kroller, A., Fekete, S.P., Buschmann, C. and Fischer, S. (2006) 'Geometric distance estimation for sensor networks and unit disk graphs', *Proceedings of the 16th International Fall Workshop on Computational Geometry*. Available online at: <http://maven.smith.edu/~streinu/FwCG/e-proc/kroller.pdf>

Kung, H., Lin, C-K., Lin, T-H. and Vlah, D. (2009) 'Localization with snap-inducing shaped residuals (SISR): coping with errors in measurement', *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking (MOBICOM'09)*, 20–25 September, Beijing, China, pp.333–344.

Li, M. and Liu, Y. (2010) 'Rendered path: Range-free localization in anisotropic sensor networks with holes', *IEEE/ACM Transactions on Networking*, Vol. 18, pp.320–332.

Lim, H. and Hou, J. (2005) 'Localization for anisotropic sensor networks', *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05)*, 13–17 March, Miami, FL, USA, pp.138–149.

Liu, Y., Yang, Z., Wang, X. and Jian, L. (2010) 'Location, localization, and localizability', *Journal of Computer Science and Technology*, Vol. 25, pp.274–297.

Lotker, Z., de Albeniz, M. M. and Perennes, S. (2004) 'Range-free ranking in sensors networks and its applications to localization', *Proceedings of the Third International Conference on Ad-Hoc, Mobile, and Wireless Networks (ADHOC-NOW'04)*, 22–24 July, Vancouver, Canada.

Ma, D., Wang, B., Lim, H.B. and Er, M.J. (2009) 'Hop-count based node-to-anchor distance estimation in wireless sensor networks', *Proceedings of the 6th IEEE Consumer Communications and Networking Conference (CCNC'09)*, Las Vegas, NV, USA, pp.920–924.

Nagpal, R. (1999) 'Organizing a global coordinate system from local information on an amorphous computer', AI Memo 1666, MIT.

Niculescu, D. and Nath, B. (2003) 'Dv based positioning in ad hoc networks', *Telecommunication Systems*, Vol. 22, pp.267–280.

Oh, S., Montanari, A. and Karbasi, A. (2010) 'Sensor network localization from local connectivity: Performance analysis for the MDS-map algorithm', *Proceedings of the 2010 IEEE Information Theory Workshop (ITW'10)*, 6–8 January, Cairo, Egypt.

- Priyantha, N.B., Chakraborty, A. and Balakrishnan, H. (2000) 'The cricket location-support system', *Proceedings of the sixth annual international conference on Mobile computing and networking (MobiCom'00)*, 6–11 August, Boston, MA, USA, pp.32–43.
- Shang, Y., Ruml, W., Zhang, Y. and Fromherz, M. (2003) 'Localization from mere connectivity', *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'03)*, 1–3 June, Annapolis, Maryland, USA, pp.201–212.
- Tan, G., Jiang, H., Zhang, S. and Kermerrec, A-M. (2010) 'Connectivity-based and anchor-free localization in large-scale 2d/3d sensor networks', *Proceedings of the 11th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'10)*, 20–24 September, Chicago, IL, USA, pp.191–200.
- Wang, X., Luo, J., Liu, Y., Li, S. and Dong, D. (2011) 'Component-based localization in sparse wireless networks', *IEEE/ACM Transactions on Networking*, Vol. 19, pp.540–548.
- Xi, W., He, Y., Liu, Y., Zhao, J., Mo, L., Yang, Z., Wang, J. and Li, X. (2010) 'Locating sensors in the wild: pursuit of ranging quality', *Proceedings of the 8th International Conference on Embedded Networked Sensor Systems (SenSys'10)*, 3–5 November, Zurich, Switzerland, pp.295–308.
- You, L., Zhu, X. and Chen, G. (2012) 'Neighbor discovery in peer-to-peer wireless networks with multi-channel MPR capability', *Proceedings of IEEE International Conference on Communications (ICC'12)*, 10–15 June, Ottawa, Canada.
- Zhong, Z. and He, T. (2009) 'Achieving range-free localization beyond connectivity', *Proceedings of the 7th International Conference on Embedded Networked Sensor Systems (SenSys'09)*, 4–6 November, Berkeley, California, USA, pp.281–294.
- Zhu, X. and Chen, G. (2011) 'Spatial ordering derivation for one-dimensional wireless sensor networks', *IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA'11)*, 26–28 May, Busan, Korea, pp.171–176.