Sequence-Based Localization in Wireless Sensor Networks
USC Technical Report CENG-2006-4

Kiran Yedavalli and Bhaskar Krishnamachari

Abstract

We introduce a novel sequence-based localization technique for wireless sensor networks. We show that the localization space can be divided into distinct regions that can each be uniquely identified by sequences that represent the ranking of distances from the reference nodes to that region. For n reference nodes in the localization space, combinatorially, $O(n^n)$ sequences are possible, but we show that, due to geometric constraints, the actual number of feasible location sequences is much lower, only $O(n^4)$. Using these location sequences, we develop a localization technique that is robust to random errors due to multi-path and shadowing effects of wireless channels. Through extensive systematic simulations and a representative set of real mote experiments we show that our light-weight localization technique provides comparable or better accuracy than other state-of-the-art radio signal strength-based localization techniques over a range of wireless channel and node deployment conditions.

Index Terms

Wireless Sensor Networks, Localization, Location Sequence, Arrangement of Lines

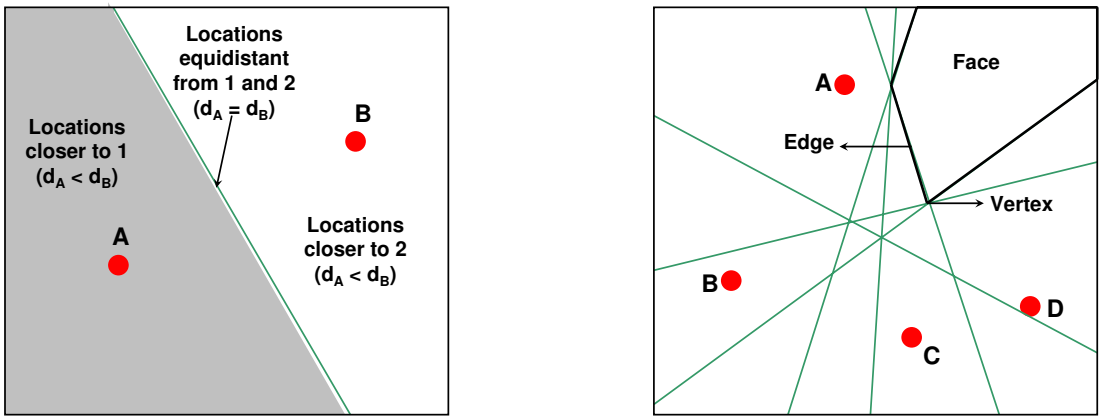
I. Introduction

ACCURATE localization is an essential part of many wireless sensor network applications. Over the years many researchers have proposed many different solutions for this problem ([1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11]). In these techniques, there is a tradeoff between the accuracy of localization and the complexity of implementation. For instance, least squares estimation techniques ([1]) require accurate RF channel parameters such as the radio path loss exponent; finger-printing based techniques (such as [8]) require extensive pre-configuration studies that depend on the features of the localization space; other techniques require specialized hardware ([5]) or a complex configuration procedure ([11]). On the other extreme, really simple techniques such as computing centroid of nearby beacons ([7]) provide low accuracy. In this paper, we present a novel sequence-based RF localization technique that is lightweight, works with any hardware and provides accurate localization without requiring accurate channel parameters or any pre-configuration.

At the heart of our proposed technique is the division of a two-dimensional (2D) localization space into distinct regions by the perpendicular bisectors of lines joining pairs of *reference nodes* (nodes with known locations). We show that each distinct region formed in this manner can be uniquely identified by a *location sequence* that represents the distance ranks of reference nodes to that region. We present an algorithm to construct the *location sequence table* that maps all these feasible location sequences to the corresponding regions, using the locations of reference nodes. This table is used to localize an *unknown node* (i.e. the node whose location has to be determined) as follows.

This project is supported by a Bosch RTC Gift Grant, NSF CAREER Award# CNS-0347621 and NSF Award# CNS-032587.

Kiran Yedavalli (kyedaval@usc.edu) and Bhaskar Krishnamachari (bkrishna@usc.edu) are with the Department of Electrical Engineering - Systems at the University of Southern California, Los Angeles.



(a) (b)

Fig. 1. (a) The perpendicular bisector of the line joining two reference nodes divides the localization space into three distinct regions. (b) Illustration of arrangement of 6 bisector lines for 4 reference nodes placed uniformly randomly in a square localization space.

The unknown node first determines its own location sequence based on the measured strength of signals between itself and the reference nodes. It then searches through the location sequence table to determine the "nearest" feasible sequence to its own measured sequence. The centroid of the corresponding region is taken to be its location.

In this paper, we focus only on RF signal-based localization since radios are used for the essential task of communication and are therefore freely available on all devices in a wireless network. Ideally, the measured distance order of reference nodes should be identical to the distance order based on true Euclidean distances. But this is not true in the real world as the RF signals are subjected to multi-path fading and noise. These non-ideal effects corrupt the location sequence measured by the unknown node. For n reference nodes in the localization space, the possible number of combinations of distance rank sequences is $O(n^n)$. However, we prove in this paper that the actual number of feasible location sequences is much lower due to geometric constraints, only $O(n^4)$. The lower dimensionality of the sequence table enables the correction of errors in the measured sequence. This is one of the reasons our proposed sequence-based localization technique performs well despite channel errors.

The rest of the paper is organized as follows: We formally define location sequences in Section II and describe the procedure of localization using them in Section III. In the same section, we derive the maximum number of feasible location sequences, illustrate the construction of the location sequence table, discuss the effect of RF channel non-idealities on unknown node location sequences and describe metrics to measure "distance" distance between sequences. In Section IV, we describe localization procedures for two different application scenarios and show their robustness to RF channel random errors through examples. In Section V, we present an exhaustive systematic performance study of our localization technique in addition to conducting a comparative study with state-of-the-art localization techniques. We present the evaluation of our technique in real mote experiments in Section VI and discuss related work in Section VII. We conclude and discuss our future work in Section VIII.

II. Location Sequences

In this section we define location sequences and illustrate them through examples.

Assume that a 2D localization space consists of n reference nodes. Consider any two reference nodes and draw a perpendicular bisector to the line joining their locations. This perpendicular bisector divides the localization space into three different regions that are distinguished by their proximity to either of the

reference node, as illustrated in Figure 1(a). Similarly, if perpendicular bisectors are drawn for all $\frac{n(n-1)}{2}$ pairs of reference nodes, they divide the localization space in to many regions of three different types - *vertices, edges* and *faces*, as shown in Figure 1(b). This subdivision of a 2D space into vertices, edges and faces by a set of lines is an *arrangement* induced by that set [12].

Now, for each region created by the arrangement induced by the set of perpendicular bisectors, determine the ordered sequence of reference nodes' ranks based on their distances from them. We define this ordered sequence of distance ranks as the *location sequence.*

***Proposition** 1:* The location sequence of a given region is unique to that region.

Proof: The proof is by contradiction. Assume that two different regions in the arrangement have the same location sequence. This implies that the distance ranks of reference nodes are the same for both the regions. This further implies that there is no bisector line that separates the two regions. The implication applies to all possible combinations of regions such as two faces, two edges, two vertices, a face and an edge, an edge and a vertex and a face and a vertex, in their own different ways. Otherwise, if there was a bisector line of two arbitrary reference nodes that separated the two regions then it would rank those reference nodes differently for the two regions. But this is a contradiction, as by definition, two different regions in the arrangement are separated by at least a single bisector line. ■

Therefore, each region created by the arrangement has a unique location sequence. Further, we make the following observations:

- All locations inside a region have the same location sequence.
- If each region in the arrangement is represented by its centroid, there is a one-to-one mapping between a location sequence and the centroid of the region it represents. For a vertex, the centroid is the vertex itself; for an edge, the centroid is its midpoint and for a face, the centroid is the centroid of the polygon that bounds it.
- The total number of unique location sequences is equal to the sum of the number of vertices, the number of edges and the number of faces created by the arrangement in the localization space.

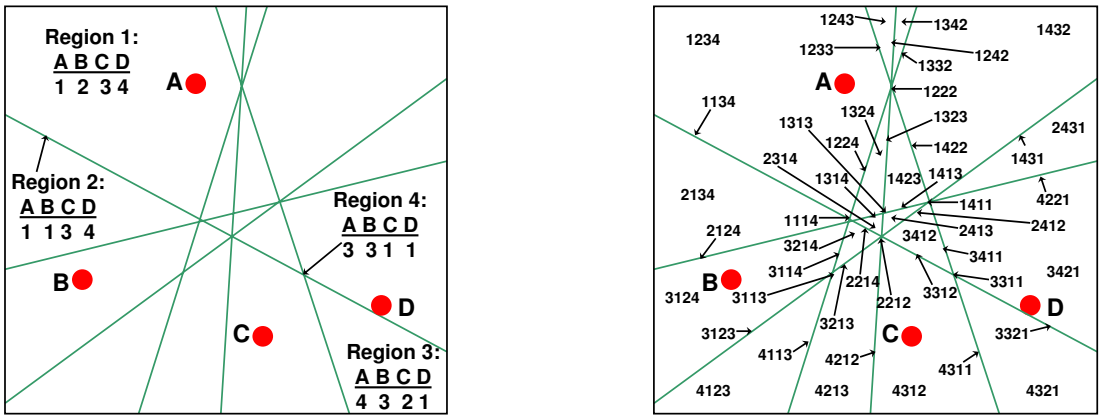
The order in which the ranks of reference nodes are written in a location sequence is determined by a pre-defined order of reference node IDs. We illustrate the above ideas through examples. Figure 2(a) shows the location sequences four different regions. In the example the pre-defined order of reference node IDs is ABCD. Region 1 is a face and its location sequence is 1234, since the distance rank of A from it is 1 (A is the closest) and the respective distance ranks of B,C and D are 2,3 and 4 (D is the farthest). Similarly, for Region 3 the location sequence is 4321 as the distance rank of A is the farthest (distance rank 4), D is the closest (distance rank 1) and B is closer than C and A. For Region 4, which is a vertex, the distance ranks of A,B and C,D are equal in pairs as it lies on the intersection of perpendicular bisectors of those pairs of reference nodes. Also, the pair C,D is closer to it than the pair A,B. Therefore, its location sequence is 3311. Similarly, for Region 2, which is an edge, the distance ranks of A and B are the same and its location sequence is 1134. Figure 2(b) shows all feasible location sequences for the topology of reference nodes of Figure 2(a).

Next, we describe how location sequences can be used for localization.

III. Localization using Location Sequences

The procedure for localization of unknown nodes using location sequences is as follows:

1) Determine all feasible location sequences in the localization space and list them in a *location sequence table*.



(a) (b)

Fig. 2. (a) Examples of location sequences for a four reference node topology. (b) All feasible location sequences for the topology of (a).

2) Determine the location sequence of the unknown node location using received signal strength (RSS) measurements of localization packets exchanged between itself and the reference nodes. The RSS based location sequence will be a corrupted version of the original location sequence.
3) Search in the location sequence table for the "nearest" location sequence to the unknown node location sequence. The centroid mapped to by that sequence is the location estimate of the unknown node.

The above procedure opens itself to the following questions: How many feasible location sequences are there in a 2D localization space? How can we get them? How do random errors in RSS measurements affect the unknown node location sequence? What is the meaning of "nearest" location sequence and how do we measure distances between location sequences?

In the rest of this section we answer the above questions. We begin by determining the maximum number of feasible location sequences in the localization space.

A. Maximum Number of Location Sequences

For n reference nodes in the localization space, the number of possible combination sequences of distance ranks is $O(n^n)$. However, we show that the actual number of feasible location sequences is much lower, in the order of $O(n^4)$ at worst.

As stated previously, the number of feasible location sequences is equal to the sum of the number of vertices, edges and faces created by the arrangement induced by the perpendicular bisectors of reference nodes. Therefore, its upper bound can be obtained by determining the maximum number of such vertices, edges and faces, given the locations of the reference nodes. In [12], the authors show that the maximum number of vertices, edges and faces for an arrangement induced by n lines is $\frac{n(n-1)}{2}$, n^2 and $\frac{n^2}{2}+\frac{n}{2}+1$ respectively. Using these results, for $\frac{n(n-1)}{2}$ perpendicular bisectors of n reference nodes,

1) The number of vertices is at most $\frac{n^4}{8}-\frac{n^3}{4}-\frac{n^2}{8}+\frac{n}{4}$.
2) The number of edges is at most $\frac{n^4}{4}-\frac{n^3}{2}+\frac{n^2}{4}$.
3) The number of faces is at most $\frac{n^4}{8}-\frac{n^3}{4}+\frac{3n^2}{8}-\frac{n}{4}+1$.

Owing to the properties of perpendicular bisectors, it is possible to derive tighter upper bounds on the number of vertices, edges and faces.

***Theorem** 1:* Let L be the set of bisector lines for n reference nodes, $|L| = \frac{n(n-1)}{2}$. Let $A(L)$ be the arrangement induced by L. Then,

1) The number of vertices of $A(L)$ is at most $\frac{n^4}{8} - \frac{7n^3}{12} + \frac{7n^2}{8} - \frac{5n}{12}$.
2) The number of edges of $A(L)$ is at most $\frac{n^4}{4} - n^3 + \frac{7n^2}{4} - n$.
3) The number of faces of $A(L)$ is at most $\frac{n^4}{8} - \frac{5n^3}{12} + \frac{7n^2}{8} - \frac{7n}{12} + 1$.

Proof: We make use of the property that the perpendicular bisectors of the sides of a triangle intersect at a single point. Assume that $(i-1)$ reference nodes have already been added, implying that the localization space already has $\frac{(i-1)(i-2)}{2}$ bisector lines. When the i^{th} reference node is added, $(i-1)$ new bisector lines are added to the localization space.

Vertices: The first of the $(i-1)$ bisector lines intersects the already present lines in at most $\frac{(i-1)(i-2)}{2}$ new vertices. The second new line is the perpendicular bisector of a side of the triangle in which the first new line is also a perpendicular bisector. Therefore, the second new line has to pass through at least one of the vertices created by the first new line, thus creating at most $\frac{(i-1)(i-2)}{2} - 1$ new vertices. Similarly the third new line creates at most $\frac{(i-1)(i-2)}{2} - 2$ new vertices. This is illustrated in Figure 3 for $n = 4$. Finally the $(i-1)^{th}$ new line creates at most $\frac{(i-1)(i-2)}{2} - (i-2)$ new vertices. Therefore, the total number of new vertices added by the i^{th} reference node is at most

$$\frac{(i-1)(i-2)}{2} + \frac{(i-1)(i-2)}{2} - 1 + \frac{(i-1)(i-2)}{2} - 2 + \cdots + \frac{(i-1)(i-2)}{2} - (i-2) \tag{1}$$

$$= (i-1)\frac{(i-1)(i-2)}{2} - (1 + 2 + \cdots + (i-2)) = (i-1)\frac{(i-1)(i-2)}{2} - \frac{(i-2)(i-1)}{2} \tag{2}$$

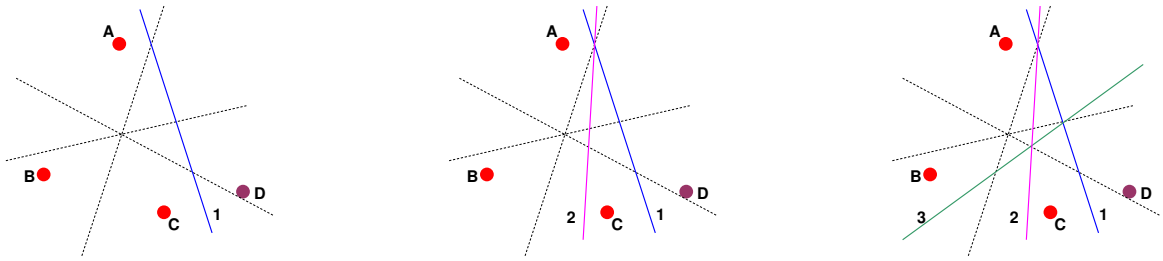
$$= \frac{(i-1)(i-2)^2}{2} \tag{3}$$

The maximum number of vertices for $n = 3$ is 1. Therefore, for n reference nodes, the maximum number of vertices is

$$1 + \sum_{i=4}^{n} \frac{(i-1)(i-2)^2}{2} = 1 + \sum_{i=4}^{n} \left[\frac{i^3}{2} - \frac{5i^2}{2} + 4i - 2\right] = \sum_{i=1}^{n} \left[\frac{i^3}{2} - \frac{5i^2}{2} + 4i - 2\right] \tag{4}$$

$$= \frac{n^4}{8} - \frac{7n^3}{12} + \frac{7n^2}{8} - \frac{5n}{12} \tag{5}$$

Edges: As explained previously, the first new line intersects the already present lines in at most $\frac{(i-1)(i-2)}{2}$ vertices and creates at most $\frac{(i-1)(i-2)}{2} + 1$ new edges on the new line and at most $\frac{(i-1)(i-2)}{2}$ new edges on the old lines which add up to $\frac{(i-1)(i-2)}{2} \cdot 2 + 1$ new edges at most. Since the second new line passes through at least one of the vertices created by the first new line, it creates at most $\frac{(i-1)(i-2)}{2} + 1$ new edges on the second new line and it creates at most $\frac{(i-1)(i-2)}{2} - 1$ new edges on the old lines including the first new line. This adds up to at most $\frac{(i-1)(i-2)}{2} \cdot 2$ new edges in the localization space. This trend is again illustrated in Figure 3 for four reference nodes in the localization space. Finally, the $(i-1)^{th}$ new line adds $\frac{(i-1)(i-2)}{2} \cdot 2 - (i-3)$ new edges to the localization space. Therefore, the total number of new edges added by the i^{th} reference node is at most



(a) (b) (c)

Fig. 3. Addition of fourth reference node D adds 3 new bisector lines to the localization space. (a) The first of the 3 new bisector lines, line 1, the perpendicular bisector of CD, creates 3 new vertices (equal to the number of pre-existing lines in the localization space), 4 new faces and 7 new edges at most. (b) The second line, line 2, the perpendicular bisector of BD, has to pass through the intersection point of the bisectors of CD and BC because, $\{BD, CD, BC\}$ form a triangle and the perpendicular bisectors of the three sides of a triangle intersect at a single point. Therefore line 2 creates 2 new vertices, 4 new faces and 6 new edges at most. (c) Similarly, line 3, the perpendicular bisector of AD has to pass through the intersection points of perpendicular bisectors of AB, BD and AC, CD as $\{AD, AB, BD\}$ and $\{AD, AC, CD\}$ are two triangles with a common side AD. Therefore, line 3 creates 1 new vertex, 4 new faces and 5 new edges at most.

$$\frac{(i-1)(i-2)}{2}\cdot 2+1+\frac{(i-1)(i-2)}{2}\cdot 2+\frac{(i-1)(i-2)}{2}\cdot 2-1+\cdots+\frac{(i-1)(i-2)}{2}\cdot 2-(i-3) \tag{6}$$

$$=2\cdot(i-1)\frac{(i-1)(i-2)}{2}+1-(1+2+\cdots+(i-3))=1+(i-1)^2(i-2)-\frac{(i-3)(i-2)}{2} \tag{7}$$

$$=i^3-\frac{9i^2}{2}+\frac{15i}{2}-4 \tag{8}$$

The maximum number of edges for $n=3$ is 6. Therefore, for n reference nodes, the maximum number of edges is

$$6+\sum_{i=4}^{n}\left[i^3-\frac{9i^2}{2}+\frac{15i}{2}-4\right]=\sum_{i=1}^{n}\left[i^3-\frac{9i^2}{2}+\frac{15i}{2}-4\right]=\frac{n^4}{4}-n^3+\frac{7n^2}{4}-n \tag{9}$$

Faces: The number of new faces created by a new line is equal to the number of edges on the new line. Therefore, the number of new faces created by the first new line among the $(i-1)$ new lines is at most $\frac{(i-1)(i-2)}{2}+1$. Since the second new line has to pass through one of the intersection points of the first line, it would also create $\frac{(i-1)(i-2)}{2}+1$ new faces and this trend continues for all the $(i-1)$ new lines as illustrated in Figure 3. Therefore, the total number of new faces added by the i^{th} reference node is at most

$$(i-1)\left(\frac{(i-1)(i-2)}{2}+1\right) \tag{10}$$

The localization space has one face when $n=1$. Therefore, for n reference nodes the maximum number of faces in the localization space is given by:

$$1+\sum_{i=2}^{n}(i-1)\left(\frac{(i-1)(i-2)}{2}+1\right)=\frac{n^4}{8}-\frac{5n^3}{12}+\frac{7n^2}{8}-\frac{7n}{12}+1 \tag{11}$$

■

***Corollary** 1:* The maximum number of unique location sequences due to n reference nodes is $\frac{n^4}{2} - 2n^3 + \frac{7n^2}{2} - 2n + 1$.

Proof: The maximum number of unique location sequences is the sum of the maximum number of vertices, edges and faces due to n reference nodes, derived in Theorem 1.

$$\left(\frac{n^4}{8} - \frac{7n^3}{12} + \frac{7n^2}{8} - \frac{5n}{12}\right) + \left(\frac{n^4}{4} - n^3 + \frac{7n^2}{4} - n\right) + \left(\frac{n^4}{8} - \frac{5n^3}{12} + \frac{7n^2}{8} - \frac{7n}{12} + 1\right) \tag{12}$$

$$= \frac{n^4}{2} - 2n^3 + \frac{7n^2}{2} - 2n + 1 \tag{13}$$

■

Next, we illustrate how to obtain all these feasible location sequences in the localization space and store them in the location sequence table.

B. Location Sequence Table Construction

Below, we present the pseudo-code for an algorithm that constructs the location sequence table given the locations of the reference nodes and the boundaries of the localization space.

***Algorithm** 1:* CONSTRUCTLOCATIONSEQUENCETABLE[1].

Input:

1) Location coordinates of reference nodes, $\{(ax_i, ay_i) \mid i = 0 \rightarrow n-1\}$.
2) Boundaries of the localization space B.

Output: Location Sequence Table.

---

0 $L = \{l_i \mid i = 0 \rightarrow (\frac{n(n-1)}{2} - 1)\} \leftarrow$ BISECTORLINES($\{(ax_i, ay_i) \mid i = 0 \rightarrow n-1\}, B$)

1 $(FL, EL, VL) \leftarrow$ CONSTRUCTARRANGEMENT(L)

► Get vertex sequences.

2 **for** $i \leftarrow 0$ **to** $(|VL| - 1)$

3 **Centroid**$[i] \leftarrow VL[i]$

4 **Sequence**$[i] \leftarrow$ GETSEQUENCE(**Centroid**$[i]$)

5 **end for**

► Get edge sequences.

6 **for** $i \leftarrow |VL|$ **to** $(|VL| + |EL| - 1)$

7 **Centroid**$[i] \leftarrow$ GETEDGECENTROID($EL[i]$)

8 **Sequence**$[i] \leftarrow$ GETSEQUENCE(**Centroid**$[i]$)

[1]C++ code files that construct the arrangement of lines and the location sequence table are available for download at http://anrg.usc.edu/downloads.html

9 **end for**

► Get face sequences.

10 **for** $i \leftarrow (|VL| + |EL|)$ **to** $(|VL| + |EL| + |FL| - 1)$

11 **Centroid**$[i] \leftarrow$ GETFACECENTROID($FL[i]$)

12 **Sequence**$[i] \leftarrow$ GETSEQUENCE(**Centroid**$[i]$)

13 **end for**

► Return the location sequence table

14 **return** {**Sequence**, **Centroid**}

- BISECTORLINES takes in the locations of the reference nodes and the boundaries of the localization space as input and returns the set L of all pair-wise perpendicular bisector lines within the boundaries of the localization space. Each line is represented by the intersection points on the left and right boundaries of the localization space.
- CONSTRUCTARRANGEMENT constructs the arrangement given a set of lines as input and returns a doubly connected edge list that consists of a vertex list (VL), an edge list (EL) and a face list (FL). Please refer to [12](Section 8.3) for a detailed description of this algorithm.
- Vertex List, VL: Contains pointers to all vertices of the arrangement induced by the set L.
- Edge List, EL: Contains pointers to all edges of the arrangement induced by the set L.
- Face List, FL: Contains pointers to all faces of the arrangement induced by the set L.
- GETEDGECENTROID takes in an edge pointer as the input and returns the centroid of the edge. The centroid of an edge (c_x, c_y) is its mid point given by:

$$(c_x, c_y) \leftarrow \left(\frac{o_x + d_x}{2}, \frac{o_y + d_y}{2}\right) \tag{14}$$

  where, (o_x, o_y) and (d_x, d_y) are the origin and destination vertices of the edge.
- GETFACECENTROID takes in a face pointer as the input and returns the centroid of the face. The centroid of a face (c_x, c_y), given its vertices $\{(x_i, y_i) | 0 \leq i \leq p-1\}$, is calculated as follows:

$$c_x \leftarrow \frac{1}{6A} \sum_{i=0}^{p-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{15}$$

$$c_y \leftarrow \frac{1}{6A} \sum_{i=0}^{p-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{16}$$

  where, p is the number of vertices that bound a given face and A is its area given by

$$A \leftarrow \frac{1}{2} \sum_{i=0}^{p-1} (x_i y_{i+1} - x_{i+1} y_i) \ ; \ (x_p, y_p) = (x_0, y_0) \tag{17}$$

- GETSEQUENCE takes in the coordinates of a point in the localization space and returns the location sequence for that point with respect to the locations of the reference nodes.

***Theorem** 2:* Algorithm 1 takes $O(n^5 \log(n))$ worst-case time and $O(n^5)$ worst case space to construct the location sequence table.

Number of Reference Nodes (n)	Number of Bisector Lines $\left(\frac{n(n-1)}{2}\right)$	Average Number of Location Sequences (Simulations)	Minimum Number of Location Sequences (Simulations)	Maximum Number of Location Sequences (Simulations)	Maximum Number of Location Sequences (Analytical)
3	3	12.3	7	13	13
4	6	44.0	23	49	49
5	10	117.3	51	141	141
6	15	274.8	217	331	331
7	21	548.4	441	653	673
8	28	988.6	840	1147	1233
9	36	1663.9	1447	1881	2089
10	45	2630.2	2321	2933	3331

TABLE I

PROGRESSION OF NUMBER OF LOCATION SEQUENCES WITH NUMBER OF REFERENCE NODES (n) IN THE LOCALIZATION SPACE. THE LAST TWO COLUMNS COMPARE THE SIMULATION AND ANALYTICAL RESULTS FOR THE MAXIMUM NUMBER OF LOCATION SEQUENCES. SIMULATION RESULTS ARE GATHERED FROM 1000 RANDOM TRIALS (WITH 100 DIFFERENT RANDOM SEEDS) IN EACH OF WHICH n REFERENCE NODES WERE PLACED UNIFORMLY AT RANDOM IN A SQUARE LOCALIZATION SPACE.

Proof: The function BISECTORLINES in line 0 takes $O(n^2)$ time and space. The algorithm CONSTRUCTARRANGEMENT that constructs the arrangement of lines takes $O(n^4)$ time, which is optimal, as proven in Theorems 8.5 and 8.6 of [12]. Since this algorithm returns the vertex list VL, the edge list EL and the face list FL, it requires $O(n^4)$ space to store all the three lists. The functions GETFACECENTROID and GETEDGECENTROID in lines 3 and 7 respectively take $O(1)$ time and space each. The function GETSEQUENCE involves sorting n reference nodes based on their distances from the centroid of the region in consideration. This takes $O(n \log n)$ time and $O(n)$ space. Since the number of faces, edges and vertices is $O(n^4)$ the worst case time requirement for lines 2-13 in the above algorithm is $O(n^5 \log(n))$ and the worst case space requirement is $O(n^5)$. Therefore, in total, Algorithm 1 takes $O(n^5 \log(n))$ worst-case time and $O(n^5)$ worst case space to construct the location sequence table. ■

Table I compares simulation results for the number of location sequences obtained using the above algorithm with analytical values from Corollary 1. The simulation results are gathered over 1000 random trials (with 100 different random seeds) in each of which n reference nodes were placed uniformly at random in the localization space. From the last two columns of the table it can be seen that the simulation results match the analytical results very closely. Note that for higher number of reference nodes the probability of occurrence of the arrangement that would produce the maximum of location sequences is less than 1 in 1000 *i.e.,* 0.001. Also, for increasing number of reference nodes, the average number of location sequences is increasingly smaller than the maximum number. Next, we discuss the effect of RF channel random errors on the unknown node location sequence.

C. Unknown Node Location Sequence

The unknown node determines its location sequence using RSS measurements of RF localization packets exchanged between itself and the reference nodes. The RSS measurements are subjected to random errors due to RF channel non-idealities such as multi-path and shadowing. In the absence of such non-idealities, the RSS measurements accurately represent the distances between the unknown node and the reference nodes. If reference nodes are ranked in a decreasing order of these RSS values then this order represents the increasing order of their separation from the unknown node.

This is not true in reality. Reference nodes that are farther from the unknown node might measure higher RSS values than reference nodes that are closer. If the reference nodes are ranked on their respective RSS measurements, the location sequence formed by these ranks will be a corrupted version of the original sequence. Corruption in unknown node location sequence results in erroneous estimation of its location. In the ideal case, when there is no corruption, the unknown node location would be the centroid of the region represented by its location sequence. However, corruption in its location sequence could erroneously estimate its location to be the centroid of some other region.

For example, if the ranks of reference nodes C and D are interchanged because of corruption due to RF channel non-idealities for Region 1 of Figure 2(a), the new location sequence would be 1243 instead of 1234. And 1243 represents a region that is adjacent to the original region as shown in Figure 2(b).

D. Feasible and Infeasible Sequences

As discussed previously, combinatorially, n reference nodes produce $O(n^n)$ location sequences. But as shown in the previous section, a localization space with n reference nodes has only $O(n^4)$ distinct regions and consequently only $O(n^4)$ ***feasible*** location sequences in the worst case. For given reference node locations, the location sequence table includes all feasible location sequences. All other sequences are *infeasible*. The non-idealities of the RF channel could corrupt a feasible location sequence either to another feasible sequence or an infeasible sequence as illustrated in Figure 4. If the corrupted sequence is infeasible, then it would be possible to detect the corruption in the sequence, whereas, if the corrupted sequence is feasible, corruption detection is not possible.

Here, we would like to emphasize the importance of low density of location sequences compared to the full sequence space. The low density of location sequences implies that many infeasible sequences are mapped to a single feasible sequence and this in turn could provide robustness to location estimation against RF channel non-idealities.

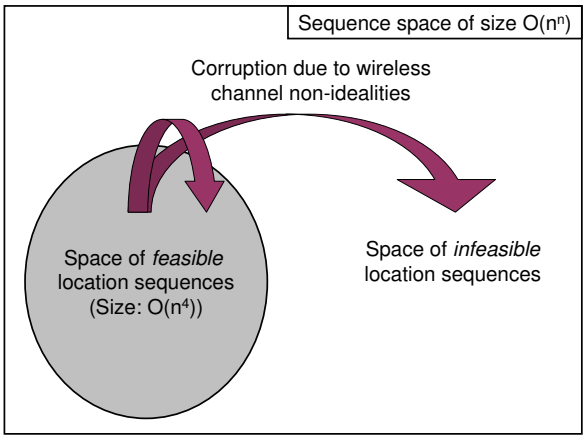


Fig. 4. RF channel non-idealities could corrupt a location sequence from the feasible space either to another sequence in the feasible space or to a sequence in the infeasible space.

Next, we present metrics to measure distance between two location sequences.

E. Distance Metrics

Given two location sequences $U = \{u_i\}$ and $V = \{v_i\}$, $1 \leq i \leq n$, where u_i's and v_i's are the ranks of reference nodes, we consider two metrics that measure the distance between them.

1) *Spearman's Rank Order Correlation Coefficient* [13]: It is defined as the linear correlation coefficient of the ranks and is given by
$$\rho = 1 - \frac{6\sum_{i=1}^{n}(u_i - v_i)^2}{n(n^2-1)} \tag{18}$$
2) *Kendall's Tau* [13]: In contrast to Spearman's coefficient in which the correlation of exact ranks is calculated, this metric calculates the correlation between the relative ordering of ranks of the two sequences. It compares all the $\frac{n(n-1)}{2}$ possible pairs of ranks (u_i, v_i) and (u_j, v_j) to determine the number of matching and non-matching pairs. A pair is matching or concordant if $u_i > u_j \Rightarrow v_i > v_j$ or $u_i < u_j \Rightarrow v_i < v_j$ and non-matching or discordant if $u_i > u_j \Rightarrow v_i < v_j$ or $u_i < u_j \Rightarrow v_i > v_j$. The correlation between the two sequences is calculated as follows:
$$\tau = \frac{(n_c - n_d)}{\sqrt{n_c + n_d + n_{tu}}\sqrt{n_c + n_d + n_{tv}}} \tag{19}$$
where, n_c is the number of concordant pairs, n_d is the number of discordant pairs, n_{tu} is the number of ties in u's and n_{tv} is the number of ties in v's.

The range of both ρ and τ is $[-1, 1]$. Next, we describe the procedure to determine locations of unknown nodes using their location sequences.

F. Location Determination

The location of the unknown node is determined as follows:

1) Calculate distances between the unknown node location sequence and all location sequences in the location sequence table using the above distance metrics.
2) Choose the centroid represented by the location sequence that is closest to the unknown node location sequence as its location estimate.

Mathematically,
$$LocationEstimate = Centroid(arg \min_{1 \le i \le O(n^4)} \tau_i) \tag{20}$$
where, τ_i is the Kendall's Tau or Spearman's correlation between the unknown node location sequence and the i^{th} location sequence in the location sequence table.

Due to RF channel non-idealities, the unknown node location sequence could be a feasible sequence different from its uncorrupted version or an infeasible sequence. In any case, the above procedure maps it to the centroid of the nearest feasible location sequence in the location sequence table that represents a different region in the arrangement than the original uncorrupted version.

We measure the amount of corruption in the unknown node location sequence by calculating its distance from the uncorrupted version, using the above metrics, and denote it by T. We denote the distance between the corrupted unknown node location sequence and the nearest feasible sequence in the location sequence table by τ.

Calculating the Spearman's coefficient and Kendall's Tau between two sequences are $O(n)$ and $O(n^2)$ operations respectively. Since the location sequence table is of size $O(n^4)$, searching through it takes $O(n^5)$ and $O(n^6)$ operations respectively for the above two metrics. Later in the paper, in Section V, we compare the performance of the two distance metrics in terms of error in the unknown node location estimate.

IV. Localization Scenarios

In this section we illustrate two localization procedures for two different scenarios that are determined by the localization space size.

1) *Entire localization space is within the radio range of the unknown node*: In this case, the location sequence table remains constant for all locations of the unknown node in the localization space. Therefore, the localization procedure is as follows:
   a) Pre-construct and store the location sequence table using the locations of the reference nodes.
   b) When the unknown node initiates the localization process by broadcasting a localization packet, provide the stored location sequence table along with the RSS measurements from the reference nodes.
   c) The unknown node determines its location sequence using the RSS measurements and determines its location by searching through the provided location sequence table for the nearest feasible location sequence.

   Here, the time cost incurred by the unknown node to estimate its location is equal to the sum of the time to determine its location sequence, an $O(n \log n)$ operation, and the time to search through the location sequence table, a $O(n^6)$ operation. The amount of memory space required is of the order of $O(n^5)$ bytes.
2) *Localization space is much larger than the radio range of the unknown node*: In this case, the location sequence table changes with the location of the unknown node as a different set of reference nodes are encountered at each location. Therefore, the localization procedure is as follows:
   a) The unknown node collects the locations and RSS measurements of the reference nodes in its radio range.
   b) It constructs the location sequence table, using Algorithm 1, using the locations of the reference nodes and calculates its location sequence using the RSS measurements.
   c) It determines its location by searching for the nearest sequence in the location sequence table.

   In this case, the time cost incurred by the unknown node to estimate its location is equal to the sum of the time to calculate its location sequence, an $O(n \log n)$ operation, the time to construct the location sequence table, an $O(n^5 \log n)$ operation, and the time to search through it, a $O(n^6)$ operation. The memory requirement is $O(n^5)$ in this case also.

A wireless device that is typically used as an unknown node is of the form factor of an IPAQ [14] (that can communicate with the reference node devices, usually of the form factor of Berkeley MICA 2 motes [15]) which typically has a 300MHz processor and 128MB of RAM . In real application scenarios, a typical value for the number of reference nodes (n) is less than 15 after which there is only very marginal gain in location accuracy of the unknown node. Therefore, for a typical value of $n = 10$ reference nodes, the time and space requirements for the unknown node to construct the location sequence table are approximately 0.3 milliseconds and 32 KB respectively. And the time required to search through it is approximately 0.45 milliseconds. Thus, including the associated overhead, the total localization time taken by sequence-based localization is in milliseconds in typical application scenarios, which is very efficient. Next, we illustrate the robustness of our localization technique to RF channel non-idealities through some examples.

A. Examples

Figure 5 shows a sample layout of nine reference nodes placed in a grid and a single unknown node (P). Figure 5(a) plots the location estimate (E) for the ideal case when there are no erroneous ranks *i.e.,* the location sequence is uncorrupted or $T = 1$. In these examples we use Kendall's Tau to measure the distance between sequences. Figures 5(b), 5(c) and 5(d) show the location estimates for increasing corruption in unknown node location sequences. Even though the location estimate error increases with

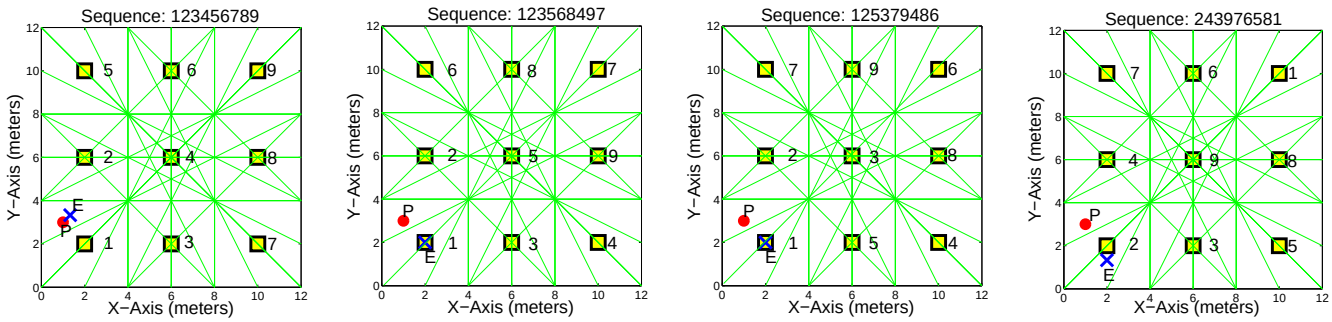


Fig. 5. *Robustness examples*: Location estimate (E) for the unknown node (P) at $(1, 3)$ for a grid layout of 9 reference nodes. The number adjacent to a reference node is its corresponding rank. The average inter-reference node distance is 6.54 meters. The location error is expressed as a percentage of the average inter-reference node distance. (a) ($T = 1$, $\tau = 1$), Estimate (E): $(1.33, 1.33)$, Location Error: 7.1% (b) ($T = 0.722$, $\tau = 0.783$), Estimate (E): $(2.0, 2.0)$, Location Error: 21.6% (c) ($T = 0.556$, $\tau = 0.667$), Estimate (E): $(2.0, 2.0)$, Location Error: 21.6% (d) ($T = 0.111$, $\tau = 0.278$), Estimate (E): $(2.0, 1.33)$, Location Error: 29.8%

increasing corruption or decreasing correlation, T, between the RSS location sequence and the true location sequence of P, it is small compared to the average inter reference node distance[2] . These examples suggest that sequence-based localization is robust to multi-path and shadowing effects of the RF channel up to some level. Intuitively, the three main reasons to which this robustness can be attributed to are:

1) The low density, $O(n^4)$, of location sequence space relative to the entire sequence space of $O(n^n)$.
2) The inherent redundancy of comparing $\frac{n(n-1)}{2}$ rank pairs in calculating the distance between two sequences using Kendall's Tau.
3) The rank order in the location sequence of the unknown node due to two reference nodes with RSS readings R_i and R_j is robust to random errors in them up to a tolerance level of $|R_i - R_j|$.

V. Evaluation

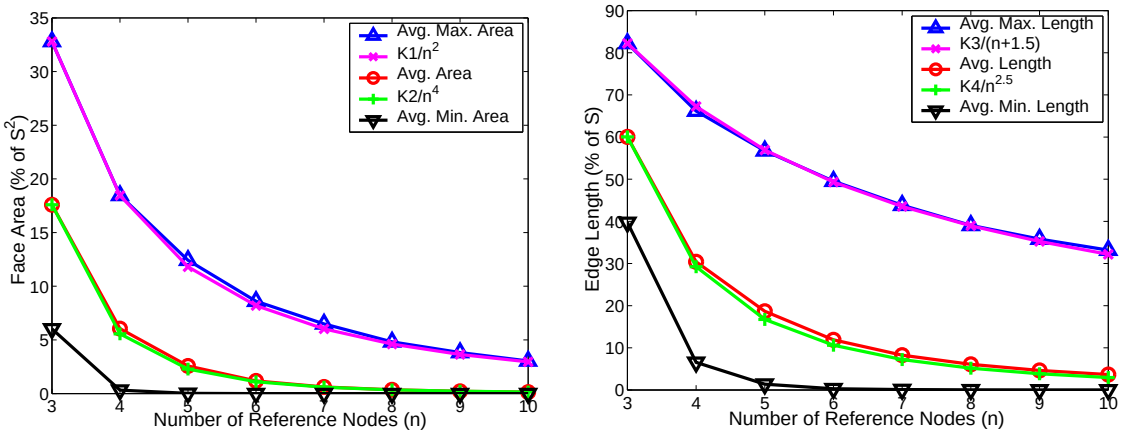
In this section, we present a complete performance evaluation of sequence-based localization (SBL). First, we discuss its inherent location error characteristics and then using simulations, we study its performance as a function of RF channel and node deployment parameters. We also present a comparative study with three other state-of-the-art localization techniques.

A. Location Error Characteristics

Each location sequence maps to the centroid of the region it represents. Representing all locations in a region by its centroid comes at the cost of error in the location estimate of the location sequence. If the region is a face, then the location error is of the order of the square-root of the area of the face and if the region is an edge then it is of the order of the length of the edge. Figure 6 plots the average, average maximum and average minimum face areas and edge lengths gathered over 1000 random trials in each of which n reference nodes were placed uniformly randomly in a square localization space of size $S \times S$ sq. meters. The main error characteristics obtained from curve fitting can be summarized as follows:

- The average face area varies proportional to $\frac{1}{n^4}$. Since the location estimate error of locations in a face region is proportional to the square-root of its area, the average location estimate error for locations in a face region reduces proportional to n^2.

[2]The average inter reference node distance is the average of distances between all pairs of reference nodes. The motivation to use this as a reference distance is described in Section V.



(a) (b)

Fig. 6. Simulation results averaged over 1000 random trials (with 100 different random seeds) in each of which n reference nodes were placed uniformly at random in a 2D square localization area of $S \times S$ sq. meters. (a) The average maximum, average and average minimum face areas as a function of the number of reference nodes. (b) The average maximum, average and average minimum edge lengths as a function of the number of reference nodes. $K1$, $K2$ and $K3$, $K4$ are scaling constants.

- The average maximum face area varies proportional to $\frac{1}{n^2}$. Therefore, the maximum location estimate error in a face region reduces proportional to n which is slower than the reduction in average location estimate error.
- The average edge length varies proportional to $\frac{1}{n^{2.5}}$. Since, the location estimate error for locations on an edge is proportional to its length, the average location estimate error for locations on an edge reduces proportional to $n^{2.5}$ which is faster than that for locations in a face region.
- The maximum edge length varies proportional to $\frac{1}{(n+1.5)}$. Therefore, the maximum location estimate error for locations on an edge reduces proportional to n which is slower than the reduction of average location estimate error.

Apart from the above location errors, the performance of sequence-based localization is affected by random errors in RSS measurements due to multi-path and shadowing effects of the RF channel. In the rest of this section, we present results from simulation studies that capture the effect of these random errors on the performance of SBL.

B. Simulation Model

The most widely used simulation model to generate RSS samples as a function of distance in RF channels is the log-normal shadowing model [16]:

$$P_R(d) = P_T - PL(d_0) - 10\eta \log_{10} \frac{d}{d_0} + X_\sigma \tag{21}$$

where, P_R is the received signal power, P_T is the transmit power and $PL(d_0)$ is path loss for a reference distance of d_0. η is the path loss exponent and the random variation in RSS is expressed as a Gaussian random variable of zero mean and σ^2 variance, $X_\sigma = N(0, \sigma^2)$. All powers are in dBm and all distances are in meters. In this model we do not provision separately for any obstructions like walls. If obstructions are to be considered an extra constant needs to be subtracted from the right hand side of the above equation to account for the attenuation in them (the constant depends on the type and number of obstructions).

C. Simulation Parameters

The location estimate of any RF-based localization technique depends on a fundamental set of parameters which can be broadly categorized into RF channel characteristics and node deployment parameters.

1) **RF Channel Characteristics**: ([17], [16])
   a) Path loss exponent (η): Measures the power attenuation of RF signals relative to distance.
   b) Standard deviation (σ): Measures the standard deviation in RSS measurements due to log-normal shadowing.

   The values of η and σ change with the frequency of operation and the obstructions and disturbance in the environment.
2) **Node Deployment Parameters**:
   a) Number of reference nodes (n).
   b) Reference node density (β).

Table II lists the typical values and ranges for different parameters used in our simulations.

Parameter	Typical Value	Typical Range
P_T	4dBm (max.)	NA
$PL(d_0)$	55dB ($d_0 = 1$m) [18]	NA
η	4 (indoors) 4 (outdoors)	1 – 7 [17]
σ	7 (indoors) 4 (outdoors)	2 – 14 [17]
n	10	3 – 10
β	0.1 (one node in 10 sq.m)	$\{0.01, 0.04, 0.1, 1\}$

TABLE II

TYPICAL VALUES AND RANGES FOR DIFFERENT SIMULATION PARAMETERS

D. Simulation Procedure

We assume that all reference nodes are in radio range of each other and also that of the unknown node. A 48 bit arithmetic linear congruential pseudo random number generator was used and results were averaged over 100 random trials using 10 different random seeds. In each trial, n reference nodes were placed uniformly randomly in a square localization space of size $S \times S$ sq. meters and the unknown node was placed at 100 different locations on a grid of $\frac{S}{10}$ separation. In total, the results presented are averaged over 10000 different scenarios.

The performance of sequence-based localization is measured in terms of *location error* for a wide range of RF channel conditions and node deployment parameters. Location error is defined as the Euclidean distance between the location estimate and the actual location of the unknown node. The location error is averaged over 100 random trials as described previously and presented as a percentage of the average inter reference node distance (D_a). D_a is calculated as the average of the distances between all possible reference node pairs. On an average, $D_a \approx \frac{S}{2}$. The motivation to use D_a as the reference distance for location error is that it provides a normalization with respect to the reference node density.

Figure 7 plots the two distance metrics described in the previous section as a function of the number of reference nodes (n) or in other words the length of the location sequence. There is a growing difference, however small, between the two metrics with increasing length of the sequence, with Kendall's Tau performing increasingly better than Spearman's correlation in terms of the location estimate error.

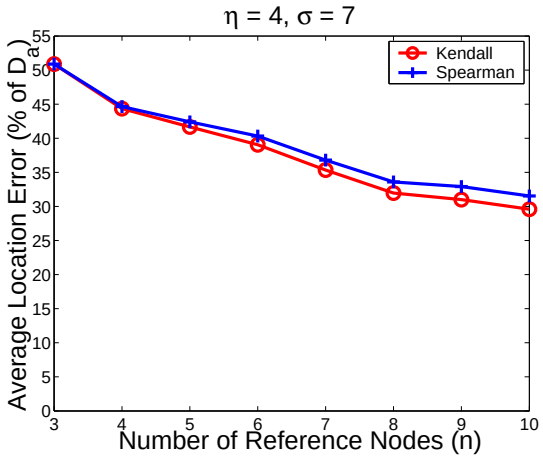
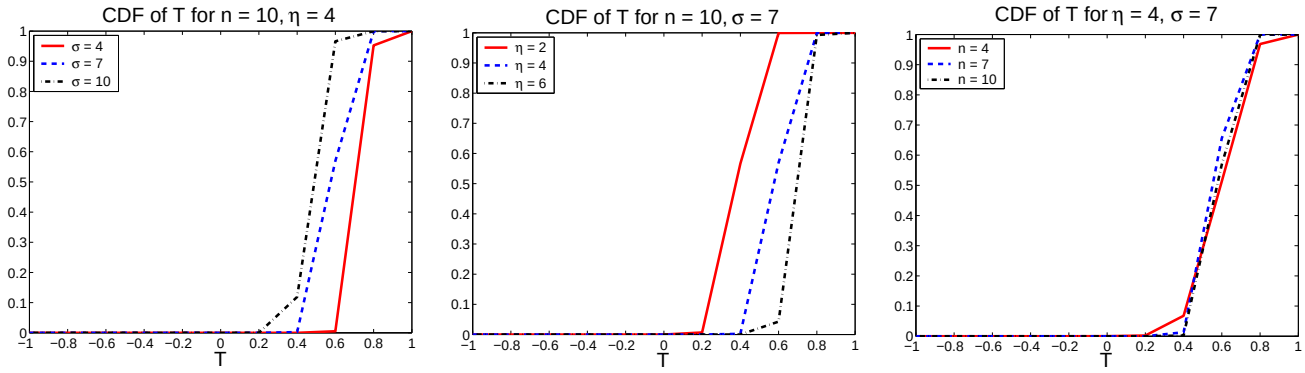


Fig. 7. Average location error as measured using Spearman's correlation and Kendall's Tau as a function of the number of reference nodes.



(a) (b) (c)

Fig. 8. *Sequence corruption*: Cumulative distribution function of *Kendall's Tau* T between the RSS location sequence and true location sequence for varying (a) standard deviation (σ) (b) path loss exponent (η) (c) number of reference nodes (n).

E. Simulation Results: Sequence Corruption

Figure 8 plots the corruption in location sequences, represented by T, due to RF channel and node deployment parameters. According to these results, the corruption in location sequences

- increases with increasing randomness in the RF channel represented by standard deviation in RSS, σ. (Figure 8(a))
- decreases with increasing path loss exponent, η. (Figure 8(b))
- is independent of the number of reference nodes in the localization space, n. (Figure 8(c))

F. Simulation Results: Performance Study

Figure 9 plots the average location error due to SBL as a function of RF channel and node deployment parameters. The main results are:

- Location error due to SBL is higher for RF channels with higher standard deviation (σ) values (Figure 9(a)). This is due to higher levels of corruption in location sequences at higher values of σ.
- Location error due to SBL is lower for RF channels with higher path loss exponent (η) values (Figure 9(b)). This is due to lower levels of corruption in location sequences at higher η values.

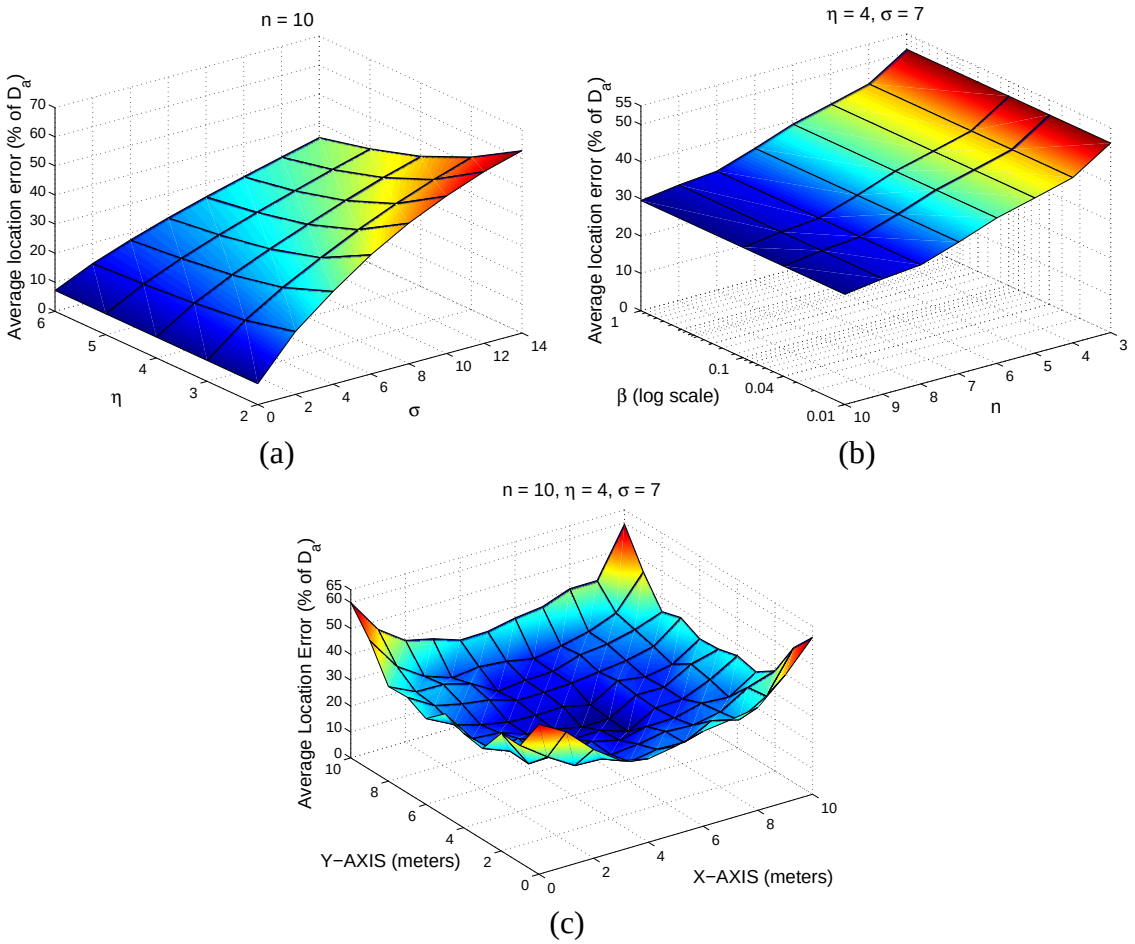


Fig. 9. *Performance*: (a) Average location error as a function of RF channel parameters - standard devation (σ) and path loss exponent (η). (b) Average location error as a function of node deployment parameters - number of reference nodes (n) and reference node density (β). (c) Average location error as a function of the location of the unknown node.

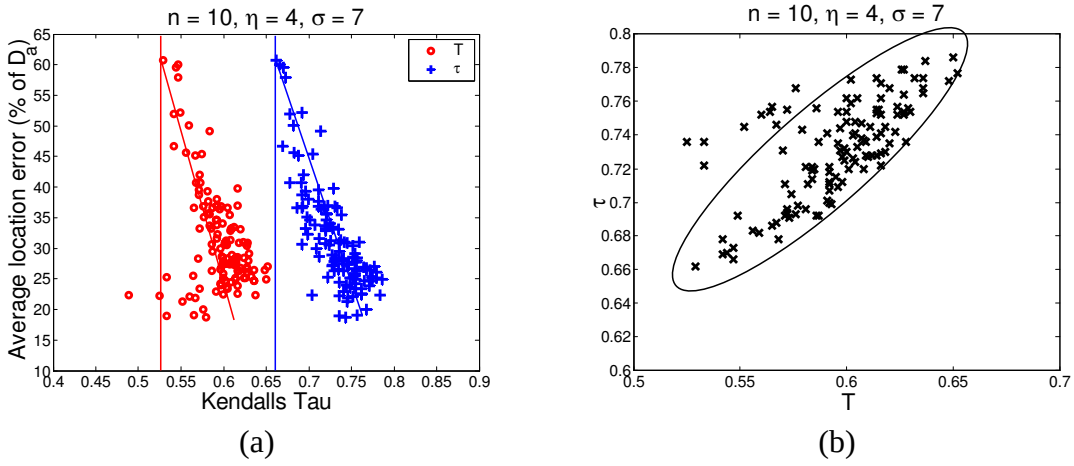


Fig. 10. (a) Average location error as a function of the sequence corruption (T) and as a function of the distance (τ) between the corrupted sequence and its nearest feasible sequence in the location sequence table. (b) Correlation between τ and T.

- Location error due to SBL reduces with increasing number of reference nodes (n) suggesting that longer sequences are more robust to RF channel non-idealities than shorter sequences. (Figure 9(b))
- Location error due to SBL reduces with increasing reference node density β. Figure 9(b) shows that it is constant with β, but the inter reference node distance D_a reduces with increasing reference node density.
- Location error due to SBL depends on the location of the unknown node. Figure 9(c) plots the average location error for all possible unknown node locations in the localization space. It shows that unknown node locations that are closer to the center of the localization space have lower location error than unknown node locations closer to the boundaries of the localization space. This can be verified from the observation (*Eg.* Figure 1(b)) that for any arrangement of bisector lines, the faces and edges towards the center of the localization space have smaller areas and lengths respectively compared to that of at its boundaries. Consequently, for unknown node locations towards the center of the localization space, the location to which the nearest feasible sequence of the corrupted sequence maps will be closer to the true location of the unknown node than for locations towards the boundaries. This results in lower location errors for unknown node locations towards the center of the localization space than for locations towards its boundaries.
- Figure 10(a) plots average location error as a function of Kendall's Tau values T and τ and Figure 10(b) plots τ as a function of T. The figures suggest that:
  - The location error is correlated to T, the corruption due to RF channel.
  - The location error is correlated to τ, the distance between the corrupted sequence and the nearest feasible sequence.
  - A correlation exists between τ and T.

  This suggests that, τ, which is a measurable quantity, as apposed to T, could be used as a quantitative indicator of the location error due to sequence-based localization. Also, owing to its correlation to T, it could also be used as an approximate indicator of the state of the RF channel.

G. Simulation Results: Comparative Study

We compare SBL with three other localization techniques - *least squares estimator*, *proximity localization* and *3-centroid*.

- *Least Squares Estimator* (LSE): It is identical to the maximum likelihood location estimator ([1], [2]) and works as follows:
  1) Measure the distance between each of the reference nodes and the unknown node using
  $$d_{mi} = 10^{\frac{P_T - PL(d_0) - \overline{P_{Ri}}}{10\eta}} \tag{22}$$
  where, d_{mi} is the measured distance and $\overline{P_{Ri}}$ is the mean received signal power between a given reference node i and the unknown node. Accurate distance measurement requires accurate estimation of the path loss exponent (η) of the environment. This requires expensive ranging techniques and/or extensive pre-configuration surveys of the localization space.
  2) For each grid point location in the localization space, determine the sum of the squares of differences in the measured distances and the true Euclidean distances of all the reference nodes from the grid point.
  $$\Sigma_{(x,y)} = \sum_{i=0}^{n-1} (d_i^{(x,y)} - d_{mi})^2 \tag{23}$$
  where, $d_i^{(x,y)}$ is the Euclidean distance between the grid location (x, y) and the reference node i.

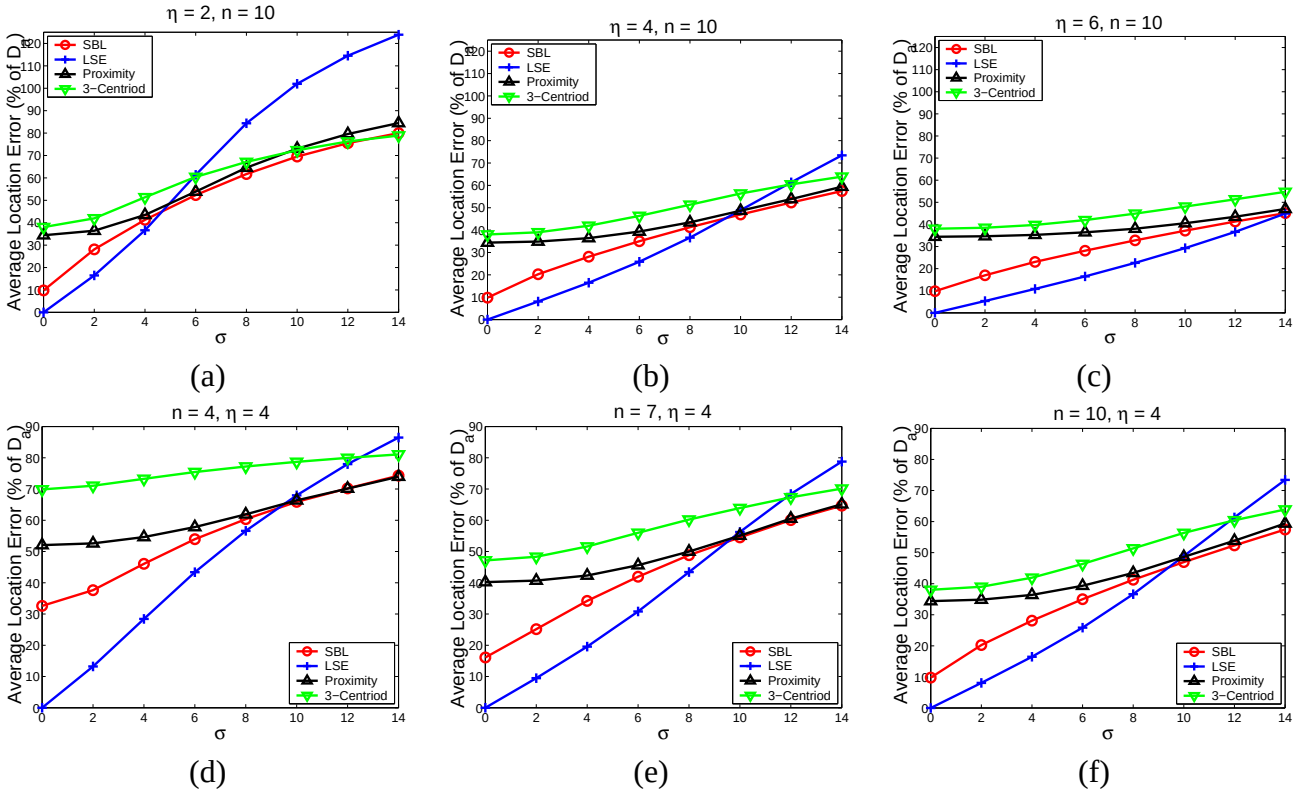


Fig. 11. *Comparison*: Average location error due to SBL, LSE, Proximity and 3-Centroid as a function of standard deviation of RSS log-normal distribution σ for different values of path loss exponent η. (a) $\eta = 2, n = 10$ (b) $\eta = 4, n = 10$ (c) $\eta = 6, n = 10$ and for different values of number of reference nodes n. (a) $n = 4, \eta = 4$ (b) $n = 7, \eta = 4$ (c) $n = 10, \eta = 4$.

3) Choose the grid point location with the least value of the above sum, $\Sigma_{(x,y)}$, as the location of the unknown node. In our study, we consider a grid resolution that is 100 times higher than the dimensions of the localization space, *i.e.*, for a $S \times S$ sq. meters localization space, we search 10000 grid points with a separation of $\frac{S}{100}$ meters between them, to determine the location of the unknown node.

- *Proximity Localization*: The location of the closest reference node by RSS value is chosen as the location of the unknown node. This is an extreme special case of SBL in which the sequence is of length 1.
- *3-Centroid*: The centroid of all the reference nodes in the radio range of the unknown node is chosen as its location ([7]). Since, in our case, all reference nodes are in the radio range of the unknown node the location error would be independent of the RF channel characteristics. In order to measure the effect of these characteristics on the centroid technique we choose the centroid of the closest three reference nodes by RSS values as the location of the unknown node.

Figure 11 plots the average location error due to SBL, LSE, Proximity and 3-Centroid as a function of the standard deviation in RSS log-normal distribution σ for different values of path loss exponents η and for different values of number of reference nodes n. The main results of the comparison are:

- SBL performs better than Proximity and 3-Centroid over a range of RF channel and node deployment parameters.
- SBL performs better than LSE for higher values of σ, whereas LSE performs better than SBL for lower values of σ. There is a crossover value of σ between the error due to SBL and LSE and this value of σ is higher for environments that have more attenuation *i.e.*, higher values of path loss exponent η. There is no significant change in the value of crossover σ with changing number of reference nodes n.

- For lower values of σ, the location error due to SBL decreases faster than location error due to LSE for increasing values of n. This can be seen in Figures 11(a)(b)(c) in which the difference between the location error due to SBL and LSE reduces with increasing values of n.
- LSE is out performed by all other localization techniques after some value of σ and this value is the lowest for SBL.

It should be noted that, in the above simulations LSE operates at a considerable advantage over other techniques as the exact value of the path loss exponent η is known. This advantage vanishes in real world scenarios where the value of η is very difficult to estimate accurately owing to its dependence on the area features such as walls, furniture, etc. Thus, LSE may not perform as well in real world scenarios.

VI. Real World Experiments

The performance of sequence based localization in real systems is studied through two experiments, representing different RF channel and node deployment parameters, conducted using Berkeley MICA 2 motes [15]. The first experiment was conducted in a parking lot which represents a relatively obstruction free RF channel and the second experiment was conducted in an office building with many rooms and furniture that represents a typical indoor environment. For comparison, the locations of the unknown nodes were also estimated using the three localization techniques - *least squares estimator (LSE), proximity localization, 3-centroid* - described in the previous Section.

A. Outdoor Experiment: Parking lot

The RF channel in an outdoor parking lot represents a class of relatively obstruction free channels. Eleven MICA 2 motes were placed randomly on the ground as shown in Figure 12. All motes were in line of sight of each other and all of them were programmed to broadcast a single packet without interfering with each other[3]. The motes recorded the RSS values of the received packets and stored them in their EEPROMs which were later used off-line for location estimation.

The locations of all the motes were estimated and compared with their true locations. Since all motes were in radio range of each other each mote had ten reference nodes. For the LSE method, to estimate the distances between the motes, the RSS model described by Equation 21 in Section V-B was used as there were no obstructions between motes in this experiment. The performance of the LSE technique depends on the value of the path loss exponent η, for the area in which the experiment was conducted. For this experiment we used the true distances and the corresponding RSS values between the reference nodes and the unknown node to estimate the value of η. Figure 12(a) plots RSS values as a function of distance. Linear regression analysis applied to the RSS vs distance data gives its slope as -2.9, implying that $\eta = 2.9$. We used this value of η to evaluate the LSE technique.

Figure 12(b) compares the true mote locations with SBL location estimates for all the motes. The Figure also shows the arrangement induced by the perpendicular bisectors between all pairs of reference nodes. Figure 12(c) plots the error at each mote location as a percentage of the average inter-reference nodes distance (D_a), due to all the four techniques. Evidently, SBL performs better than Proximity and 3-Centroid in ten out of eleven cases and it performs better than LSE in all the eleven cases.

Figure 12(d) plots the sequence corruption (T) at each mote location and the distance (τ) between the corrupted sequence and the nearest feasible sequence in the location sequence table for all the 11 nodes.

[3]We had actually measured RSS of 100 packets in one minute and observed that their standard deviation was less than $0.5dBm$. Therefore, we decided to use only a single packet for localization. In real application scenarios this would help in conserving energy at the mote and reducing the delay in localization without affecting its accuracy.

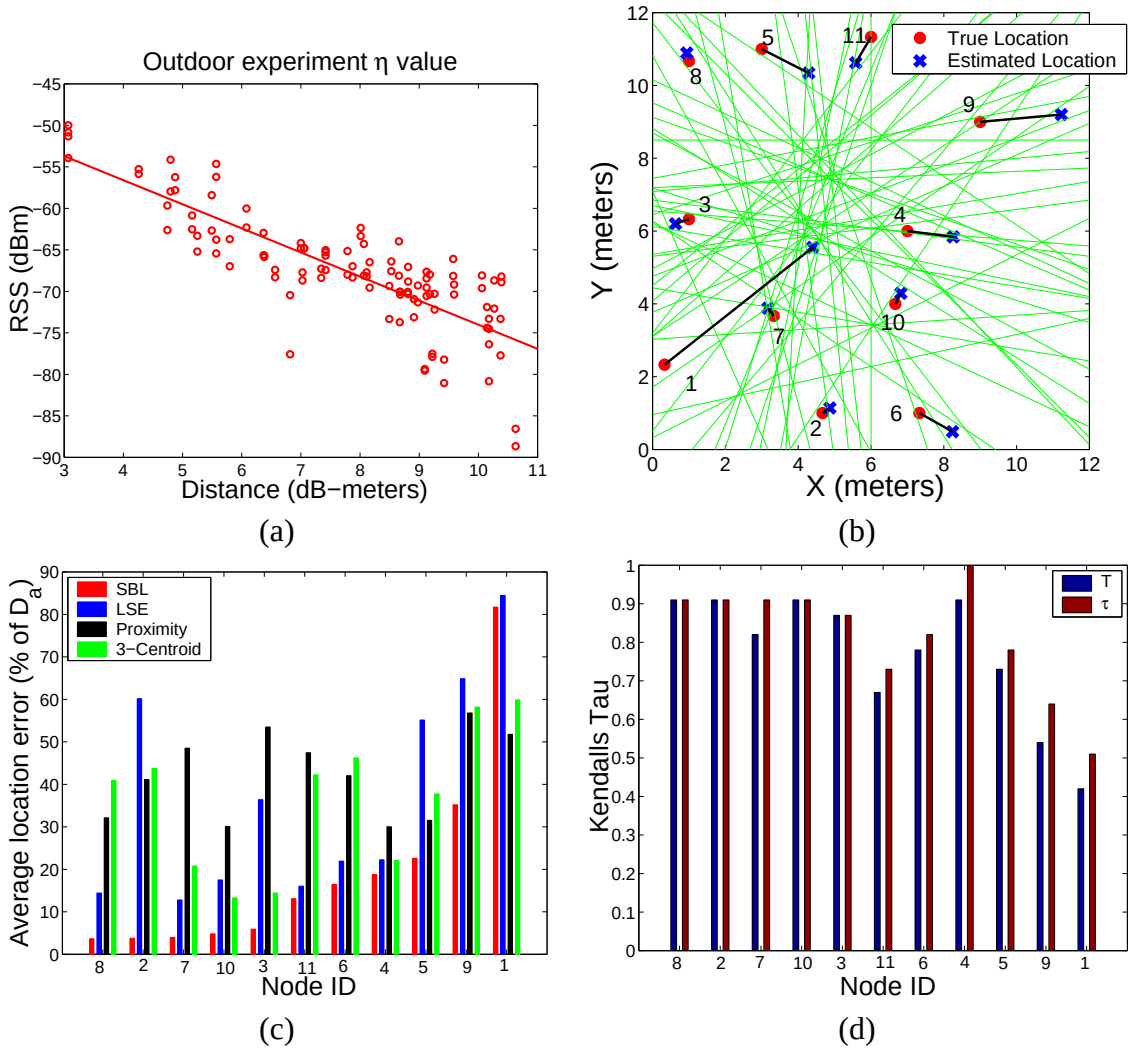


Fig. 12. *Outdoor experiment*: 11 MICA 2 motes, placed randomly in a 144 sq.meters area, were used as reference nodes as well as unknown nodes. Consequently, each unknown node had 10 reference nodes ($D_a = 6.36$ meters). (a) Path loss exponent calculation, $\eta = 2.9$. (b) Comparison between true locations and SBL location estimates. (c) Location error due to SBL, LSE, Proximity and 3-Centroid (the nodes are ordered in increasing error of SBL). (d) Corruption measure T and error indicator τ.

The correlation between T and τ can be clearly seen from the Figure. Comparing Figure 12(c) and Figure 12(d), broad correlations between T and location error and between τ and location error can be observed for SBL. For example, the location error is highest for node IDs 1 and 9, in that order, and τ is the lowest for the same node IDs in the same order. Also, the location error is almost equal for nodes 8,2,7 and 10. This trend is also reflected in the values of τ for those nodes.

B. Indoor Experiment: Office building

Office buildings with features such as rooms, corridors, furniture and other obstructions represent a distinct class of RF channels. Twelve MICA 2 motes (reference nodes) were placed on the ground randomly in a corner of the Electrical Engineering building at USC spanning different rooms and corridors. Figure 13 shows a schematic of the experimental setup. In this experiment, an unknown node was placed at five different locations and these locations were estimated using all the twelve motes as reference nodes. As in the outdoor experiment, the unknown node was programmed to broadcast a single packet from each location and the reference nodes recorded the RSS values of this packet in their respective EEPROMs which were later used off-line for location estimation.

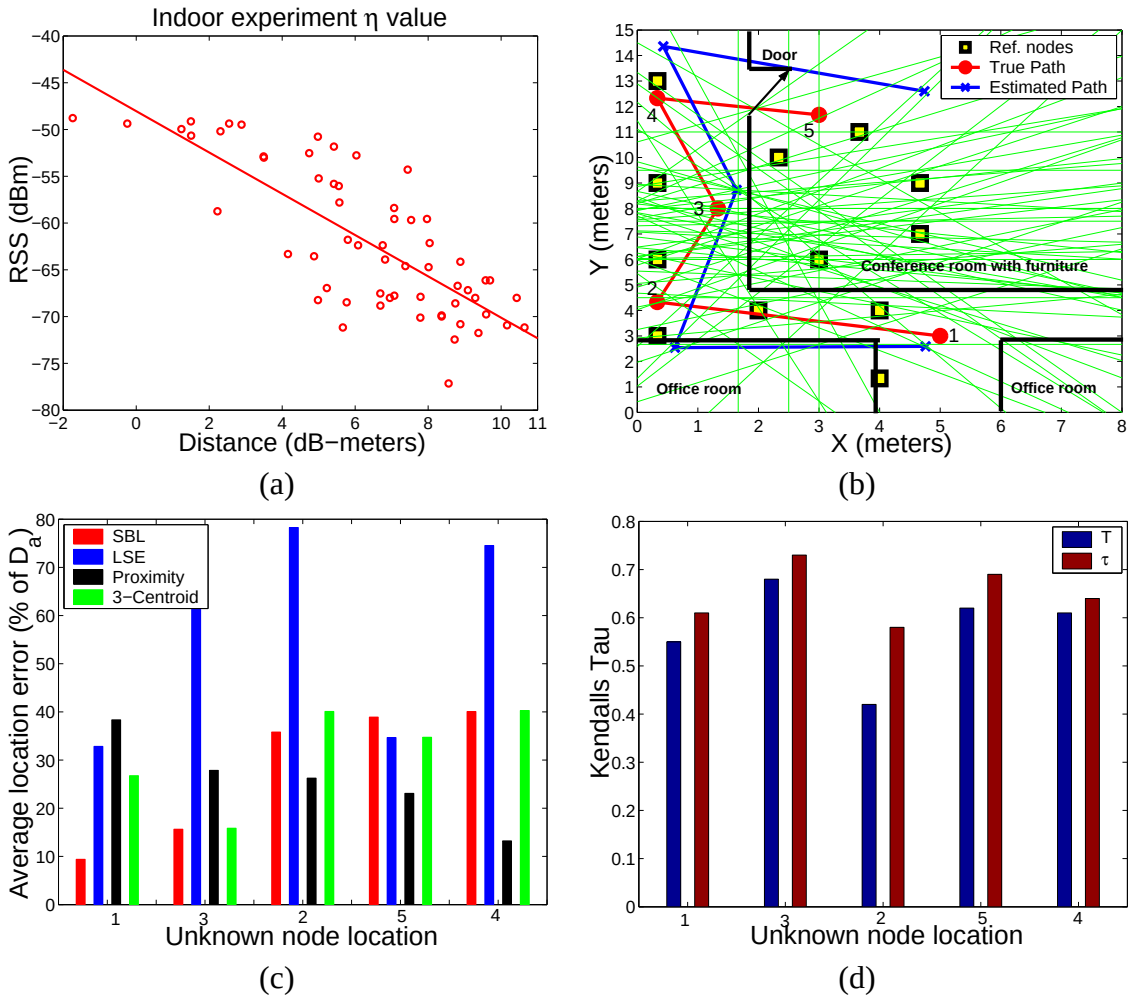


Fig. 13. *Indoor experiment*: 12 MICA 2 motes, placed randomly in a 120 sq.meters area, were used as reference nodes. The location of the unknown node was estimated for 5 different locations using the 12 reference nodes ($D_a = 5.07$ meters). (a) Path loss exponent calculation, $\eta = 2.2$. (b) Comparison between true path and SBL estimated path. (c) Location error due to SBL, LSE, Proximity and 3-Centroid (the nodes are ordered in increasing error of SBL). (d) Corruption measure T and error indicator τ.

Unlike in the outdoor experiment not all motes were in line of sight of each other even though they were in each other's radio range. A subset of the motes had obstructions in between them in the form of walls. As for the outdoor experiment, for the LSE method, the value of η was calculated using linear regression analysis for RSS vs. distance values between the reference nodes and the unknown node. Figure 13(a) shows the data. In this case, the value of η is 2.2.

Figure 13(b) compares the SBL location estimates of the five unknown node locations with their true locations. It can be seen that the path of the location estimates closely follows the true path of the unknown node. Figure 13(c) plots the location estimate error due to SBL, LSE, Proximity and 3-Centroid techniques for each unknown node location. It can be observed that SBL performs better than LSE and 3-centroid in four out of the five cases and better than Proximity in two out of five cases. A possible reason why proximity is performing well is the relatively dense distribution of the reference nodes.

Figure 13(d) plots the sequence corruption (T) at each mote location and the distance (τ) between the corrupted sequence and the nearest feasible sequence in the location sequence table for all the 5 unknown node locations. Comparing this Figure and Figure 12(d) shows that sequences are more corrupted in the indoor experiment than the outdoor experiment, which was expected. Also, as in the outdoor experiment there is a clear correlation between T and τ for the indoor experiment also. But the correlations between

T and location error and between τ and location error are not as clear as that in the outdoor experiment.

C. Discussion

Experimental results show that localization techniques are more accurate for relatively clutter free RF channel environments (outdoors with line of sight) than RF channels with many obstructions (indoor environment). Also, the performance of LSE in real world scenarios is worse than in simulations, as was conjectured in Section V-G. This is mainly because the radio propagation model of Equation 21 is an approximate model and the location estimate accuracy for the LSE technique depends heavily on the accuracy of η estimate. The RSS measurements in the experiments depend on antenna orientations, antenna height and transmitter/receiver non-determinism. For simulations, these issues can be captured within the log-normal random term in Equation 21.

VII. Related Work

In an earlier work [19], we presented a novel localization algorithm called *Ecolocation* that uses *location constraints* for robust localization. A location constraint is a relationship between the distances of two reference nodes from the unknown node that determines its proximity to either of the reference nodes, as shown in Figure 1(a). Location constraints can be graphically represented by perpendicular bisectors between reference nodes (Section II) and each location sequence can be written as a set of location constraints. Thus, the location constraint set is also unique to each region in the arrangement.

In this localization algorithm, the unknown node determines its set of location constraints using RSS measurements and estimates its location by searching through grid points in the localization space to determine the grid point with the highest number of matched location constraints. In [19] we show that this is a $O(\frac{n^2S^2}{r^2})$ time operation, where S is the side of the square localization space and r is the resolution of grid points. Thus, the localization algorithm using location constraints is dependent on the localization space size, the resolution of the grid points and the number of reference nodes. In contrast, the localization algorithm using location sequences which depends only on the number of reference nodes, albeit at higher time cost of $O(n^6)$. In fact, constraint based localization results tend to sequence-based localization ones for very high values of grid point resolution r. The cost differences suggest that, for smaller localization spaces and lower location accuracy requirements, constraint based localization is better compared to sequence-based one, whereas, the reverse is true for bigger localization spaces and higher location accuracy requirements.

In related works, Chakrabarty *et al* in [9] and Ray *et al* in [10] use identity codes to determine the location of sensor nodes in grid and non-grid sensor fields respectively. In this, each grid point or region in the localization space is identified by a unique set of reference node IDs whose signals can reach the point or region and this unique set is an identity-code for that point or region. The two main drawbacks of this approach are that (i) in order to uniquely identify all unknown node locations in the localization space the reference nodes need to be placed carefully according to rules determined by an optimization algorithm (ii) and that for acceptable location accuracies, the number of reference nodes required is prohibitively expensive and for sparse networks of reference nodes the accuracy is coarse-grained, in the order of radio range. For example, the number of reference nodes required to uniquely identify the location of an unknown node using identity-codes is $O(p^m)$, where m is the number of dimensions of the localization space and p is the number of grid points per dimension [9].

In another related work, authors in [6] propose a RF-based localization technique in which the unknown node location is determined by the intersection of all triangles, formed by reference nodes, that are likely to bound it. The unknown node determines its existence inside a triangle by comparing its measured RSS

values to that of its neighbors to detect a trend in RSS values in any particular direction. This technique depends on the weak assumption that signal strength decreases monotonically with distance, which is not true in real world scenarios.

VIII. Conclusion and Future Work

In this paper we presented a simple and novel localization technique based on location sequences called Sequence-Based Localization (SBL). In Sequence Based Localization location sequences are used to uniquely identify distinct regions in the localization space. The location of the unknown node is estimated by first determining its location sequence using RSS measurements of RF signals between the unknown node and the reference nodes. And then searching through a pre-determined list of all feasible location sequences in the localization space, called the location sequence table, to find the region represented by the "nearest" one. In this chapter, we derived expressions for the maximum number of location sequence and presented an algorithm to construct the location sequence table. We described distance metrics that measure the distance between location sequences and used them to determine the corruption in location sequences due to RF channel non-idealities. We identified an approximate indicator of the extent of location estimation error using the same distance metrics. Through examples we demonstrated the robustness of sequence-based localization to RF channel non-idealities. Through exhaustive simulations and systematic real mote experiments we evaluated the performance of our localization system and presented a comparison with other state-of-the-art localization techniques for different RF channel and node deployment parameters. Results showed that SBL performs well and better than other state-of-the-art localization techniques in both indoor and outdoor environments.

As part of future work we would like to incorporate location probability into the location sequence table. Owing to the features and topology of objects and obstructions in the localization space, unknown nodes are more likely to be in some locations than others. This could be incorporated into sequence-based localization by weighing feasible location sequences in the location sequence table in proportion to the location likelihoods of the regions they represent.

Acknowledgments

We wish to thank Abtin Keshavarzian of Bosch Research, Palo Alto, CA and Prof. Isaac Cohen of USC CS Department for valuable discussions on the subject.

References

[1] N. Patwari and A. H. III, "Using Proximity and Quantized RSS for Sensor Localization in Wireless Networks," in *WSNA*, San Diego, CA, September 2003.

[2] K. Yedavalli, "Location Determination Using IEEE 802.llb," Master's thesis, The University of Colorado at Boulder, December 2002.

[3] A. Savvides, H. Park, and M. Srivastava, "The Bits and Flops of the N-hop Multilateration Primitive For Node Localization Problems," in *WSNA*, Atlanta, Georgia, September 2002.

[4] A. Savvides, C. Han, and M. Srivastava, "Dynamic Fine Grained Localization in Ad-Hoc Sensor Networks," in *Proceedings of the Fifth International Conference on Mobile Computing and Networking, Mobicom 2001*, Rome, Italy, July 2001, pp. 166–179.

[5] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-Support System," in *ACM MOBICOM*, Boston, MA, August 2000.

[6] T. He, B. B. C. Huang, J. Stankovic, and T. Abdelzaher, "Range–Free Localization Schemes for Large Scale Sensor Networks," in *Mobicom*, San Diego, CA, September 2003.

[7] N. Bulusu, J. Heidemann, and D. Estrin, "Gps-less low-cost outdoor localization for very small devices," *IEEE Personal Communications Magazine*, October 2000.

[8] Victor Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System," in *IEEE INFOCOM*, Tel Aviv, Israel, 2000.

[9] Krishnendu Chakrabarty and S. Sitharama Iyengar and Hairong Qi and Eungchun Cho, "Grid Coverage for Surveillance and Target Location in Distributed Sensor Networks," *IEEE Transactions on Computers*, vol. 51, no. 12, pp. 1448–1453, December 2002.

[10] S. Ray, D. Starobinski, A. Trachtenberg, and R. Ungrangsi, “Robust Location Detection with Sensor Networks,” *IEEE JSAC Special Issue on Fundamental Performance Limits of Wireless Sensor Networks*, vol. 22, no. 6, pp. 1016–1025, August 2004.

[11] M. Maroti, P. Volgyesi, S. Dora, B. Kusy, A. Nadas, A. Ledeczi, G. Balogh, and K. Molnar, “Radio interferometric geolocation,” in *SenSys ’05: Proceedings of the 3rd international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2005, pp. 1–12.

[12] M. de Berg, M. van Krevald, M. Overmars, and O. Schwarzkopf, *Computational Geometry - Algorithms and Applications, Second Edition*, 2nd ed. Springer, 2000.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 1992.

[14] *http://welcome.hp.com/country/us/en/prodserv/handheld.html*.

[15] *http://www.xbow.com/Products/productsdetails.aspx?sid=72*.

[16] T. S. Rappaport, *Wireless Communications, Principles & Practice*. Prentice Hall, 1999.

[17] H. Hashemi, “The indoor radio propagation channel,” in *Proceedings of the IEEE*, vol. 81, no. 7. IEEE, July 1993, pp. 943–968.

[18] Marco Zuniga and Bhaskar Krishnamachari, “Analyzing the Transitional Region in Low Power Wireless Links,” in *First IEEE International Conference on Sensor and Ad hoc Communications and Networks (SECON)*, Santa Clara, CA, October 2004.

[19] K. Yedavalli, B. Krishnamachari, S. Ravula, and B. Srinivasan, “Ecolocation: A Sequence Based Technique for RF Localization in Wireless Sensor Networks,” in *The Fourth International Conference on Information Processing in Sensor Networks (IPSN 2005)*, Los Angeles, CA, April 2005.