

# Error Characteristics and Calibration-free Techniques for Wireless LAN-based Location Estimation

Youngjune Gwon and Ravi Jain  
DoCoMo Communications Laboratories USA, Inc.  
181 Metro Drive, Suite 300  
San Jose, California 95110  
{gyj, jain}@docomolabs-usa.com

## ABSTRACT

Using wireless LAN technology for location estimation provides alternate means to enable location-based applications without investment in sensor network infrastructure and special hardware. However, the main drawback of wireless LAN-based location systems is *calibration* of signal strength as a function of location in spatially high-density, which consumes manual labor and needs to be carried out repeatedly. In this paper, we analyze empirical error characteristics of calibration-based location algorithms such as triangulation in various spatial densities of calibration, using commercially available wireless LAN products. Then, we propose Triangular Interpolation and eXtrapolation (TIX), a calibration-free location algorithm, and present empirical performance evaluation. TIX can achieve mean distance error within 5.4 m, which is comparable to within 4.7 m errors of the calibration-based algorithms. We also present theoretical analysis on error characteristics of the location algorithms deriving accuracy limits and quantifying the effect of RF measurement and calibration.

## Categories and Subject Descriptors

C.2.1 [Computer-communication Networks]: Network Architecture and Design – *wireless communication*; C.4 [Performance of Systems]: Measurement Techniques; G.1.1 [Numerical Analysis]: Interpolation – *interpolation formulas, extrapolation*

## General Terms

Algorithms, Experimentation, Measurement, Performance

## Keywords

RF-based location estimation, static scene analysis, wireless LAN, calibration, triangulation, interpolation, extrapolation

## 1. INTRODUCTION

Location information of a mobile user is regarded as a key requirement enabling new means to construct the future mobile applications and services. Due to poor indoor coverage of the Global Positioning System (GPS), indoor location estimation has been a challenging research issue. One approach for indoor

location is to deploy infrared or ultrasound sensor network infrastructures. However, this approach is naturally concerned with investment cost, installation, and maintenance of sensor network infrastructure and special hardware. An alternate approach leverages the existing wireless LAN to enable RF-based location estimation as proposed in the previous literature such as [1, 2, 3, 4, 5, 6, 7, 8, 9]. *Static scene analysis* collectively refers to the latter approach.

Static scene analysis employs a number of wireless access points (APs) transmitting RF beacons. Using a wireless client as location sensor, received signal characteristics of an area covered by the location system are examined. However, the main drawback of this approach is *calibration* of signal strength as a function location in spatially high-density, which consumes manual labor and needs to be carried out repeatedly. So far, few known work [6, 7] has addressed the issue of reducing calibration. Elimination of calibration, however, has not been addressed in any previous work and regarded as a very difficult research problem.

This paper discusses calibration-free techniques for wireless LAN-based indoor location estimation. In particular, we propose Triangular Interpolation and eXtrapolation (TIX) algorithm, which is a central part of our calibration-free techniques. Using commercially available IEEE 802.11b wireless LAN products, we obtained empirical error characteristics (or accuracy) of both calibration-based and calibration-free algorithms. Especially, we applied various spatial densities of calibration for calibration-based classical algorithms. TIX achieved mean distance error within 5.4 m, which is 0.7 – 1.4 m worse than densely-calibrated classical algorithms. We also present theoretical analysis on error characteristics of the location algorithms deriving accuracy limits and quantifying the effect of RF measurement and calibration.

Rest of the paper is organized as follows. Section 2 provides background of the work. Section 3 presents empirical error characteristics of calibration-based location estimation using classical algorithms. Section 4 introduces TIX algorithm, describes associated calibration-free techniques, and presents empirical error characteristics of calibration-free location estimation using TIX. Section 5 presents our theoretical analysis. In Section 6, we discuss related work and how our work differs from others. Finally, Section 7 concludes the paper.

## 2. BACKGROUND

### 2.1 Static scene analysis

Static scene analysis examines certain features such as RF propagation that can characterize various areas covered by a location system. Based on an experimental experience, Bahl *et al.*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*MobiWac'04*, October 1, 2004, Philadelphia, Pennsylvania, USA.  
Copyright 2004 ACM 1-58113-920-9/04/0010...\$5.00.

[1] claims that signal strength (or a proxy for signal strength) serves the best as a measurable RF feature for wireless LAN. Received signal strength indicator (RSSI) metric, which is measured from captured wireless LAN beacons, is our choice for static scene analysis although use of other metrics is also possible. In the remainder of this paper, we use two terms ‘RF metric’ and ‘RSSI’ interchangeably. Static scene analysis using such RF metrics consists of ‘offline’ and ‘run-time’ stages.

1. *Offline stage*: measurement and sampling RSSIs at known locations occur. The eventual purpose is to calibrate the location system by relating RSSI value to coordinates of the known locations. Collected samples are processed statistically and stored in a location database.
2. *Run-time stage*: RSSI values are measured at an unknown location. Using this run-time RSSI, the location database is searched for the closest match among RSSIs of the known locations. A location algorithm is then applied to estimate the unknown location estimate using coordinates of the matched known locations.

## 2.2 Classical location algorithms

Three classical location algorithms that are calibration-based are used in our experiments.

1. *Triangulation (TN)* is used in numerous systems such as GPS. TN requires solving a system of equations of circles centered at APs, where the radius of each circle is determined by matching the run-time RSSI values to those of the offline. Using the solution, vertices of common areas shared by the circles are found. Averaging the coordinates of the vertices that form the smallest area gives the final location estimate. More detailed description of TN is found in [4, 5].
2. *K-nearest neighbor averaging (KNN)* uses the run-time RSSI values (obtained at an unknown location) to search for K closest matches of known locations in signal space from the location database. To do so, root mean square errors between the run-time and offline RSSI values are computed and compared. Averaging the coordinates of the K-locations gives the final location estimate. KNN was used in RADAR [1].
3. *Smallest M-vertex polygon (SMP)* uses the run-time RSSI values to search for a number of candidate locations in signal space with respect to *each* AP separately. M-vertex polygons are formed by choosing at least one candidate from each AP (total of M APs). SMP chooses the smallest polygon, which has the shortest perimeter. Averaging the coordinates of vertices of the smallest polygon gives the final location estimate. SMP was used in MultiLoc [3].

## 3. EMPIRICAL ERROR CHARACTERISTICS OF CLASSICAL LOCATION ALGORITHMS

We adopted ‘mean distance error’ metric to evaluate accuracy of classical location algorithms. Mean distance error is computed by averaging Euclidean distances between the estimate  $(\hat{x}, \hat{y})$  and

the true location  $(x, y)$  in a two-dimensional Cartesian coordinate system, i.e.,  $d = \sqrt{(x-\hat{x})^2 + (y-\hat{y})^2}$ .

### 3.1 Description of experiments

Our experimental testbed spans the entire floor of a multiple-story building, which is a typical office environment having cubicles, small offices, and conference rooms. The floor map and our x-y coordinate system are illustrated in Figure 1. The unit grid for calibration measurements has a dimension of 1.422 m by 1.422 m (14/3 ft. by 14/3 ft.). The testbed spans 39.83 m by 25.60 m, for a total area of 1020 m<sup>2</sup> (10975 sq. ft.).

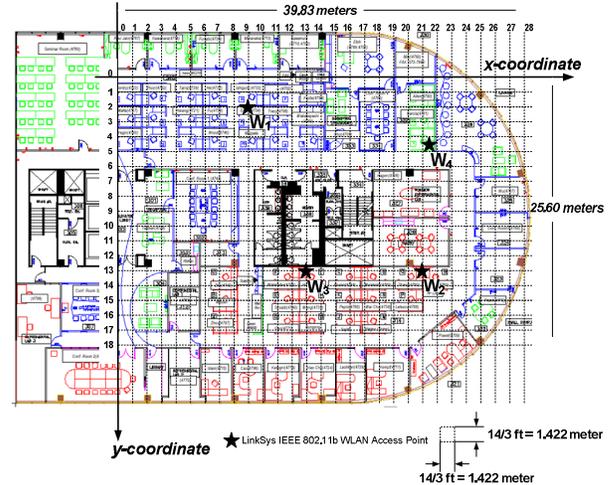


Figure 1: Map and coordinate system of experimental testbed

We placed four LinkSys™ WAP11 APs labeled  $W_1$ ,  $W_2$ ,  $W_3$ , and  $W_4$  in Figure 1. A mobile laptop used in our experiments runs Redhat Linux 2.4.18 OS with a LinkSys™ IEEE 802.11b client adaptor. We implemented a measurement module for the laptop, which is written in C based on the IEEE 802.11b device driver developed by the Linux-WLAN Project [11]. RSSI was measured in dBm. Not all grid positions were possible for measurement due to the presence of desks, walls, and other stationary components on the floor. We obtained three independent sets of data and each set included RSSI samples from 207 different locations. Each set was carried out separately at different time of a day. We obtained at least 40 RSSI samples from each of four different directions (heading north, south, west, and east) per each AP at one location.

### 3.2 Data processing and calibration

One of the three data sets was considered offline and used for calibration. Samples collected at each measurement location are first band-pass filtered, which removes the top and bottom 10 % RSSI values, and then averaged. A location database is created and stores the calibration table that relates mean RSSI values and coordinates of locations where the measurement took place.

### 3.3 Implementation of classical algorithms

We implemented TN, KNN, and SMP in MATLAB. They require RSSI measurements from at least three APs. Since four wireless APs were deployed in our testbed, the algorithms were implemented to operate with either three or four APs.

### 3.4 Performance evaluation

We present empirically observed errors for TN, KNN, and SMP using three or four APs in four different grid sizes (1x, 2x, 3x, and 4x the unit grid whose side length  $g = 1.422$  m). Total number of entries in the calibration table reduces to 119 (57 %) for grid size of 2g, 62 (30 %) for grid size of 3g, and 21 (10 %) for grid size of 4g from 207 (100 %) for the unit grid.

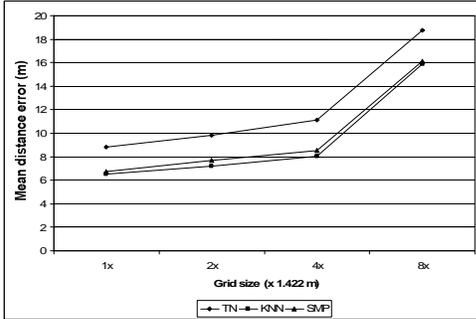


Figure 2: Mean distance errors of classical location algorithms (TN, KNN, and SMP) using three APs in different grid sizes

Figure 2 depicts mean distance errors of the classical location algorithms using three APs. KNN performs consistently better than TN and SMP, yielding the smallest mean distance errors in all four grid sizes. Mean distance errors of TN, KNN, and SMP do not increase significantly until the grid size is enlarged to 8x. This corresponds to using 10 % of the calibration table entries used for 1x. Key empirical observation is that accuracy does not degrade significantly until the calibration grid size is enlarged to some threshold (i.e., 8x). As Figure 3 depicts mean distance errors of the same algorithms using four APs, overall errors are smaller than using three APs while we observe the similar tendency of accuracy degradation as grid size increases (sparser calibration).

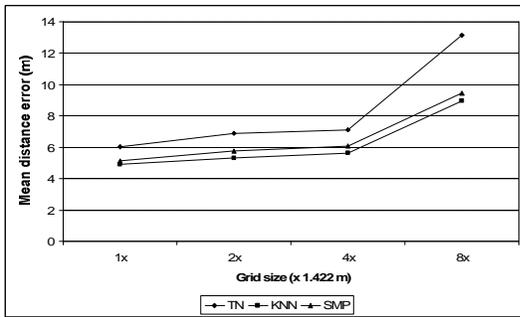


Figure 3: Mean distance errors of classical location algorithms (TN, KNN, and SMP) using four APs in different grid sizes

## 4. CALIBRATION-FREE LOCATION ESTIMATION

Calibration in RF-based static scene analysis repeatedly consumes human labors and creates significant maintenance and scalability issues. We present calibration-free location estimation techniques that eliminate laborious offline RSSI measurements.

### 4.1 Overview

An overview of our calibration-free location estimation techniques is depicted in Figure 4, comprising the following steps:

1. *Location request*: a mobile node (location client) requests location information to location system (location server) available online.
2. *Simultaneous RSSI measurement requests*.
  - a. *Mobile measurement request*: the location system requests the mobile to measure and report RSSIs from all listenable APs.
  - b. *Inter-AP measurement request*: location system requests all APs to measure and report RSSIs from all APs excluding itself.
3. *Generation of mapping curves*: location system uses the inter-AP measurements (using each *observing* AP's measurement separately) to generate multiple distance mapping curves for each AP *observed*. The curves map an RSSI value to distance from the observed AP. On a log-linear scale, RSSI vs. distance curve is roughly a straight line since the received signal strength in log scale (i.e., measured in dBm) decays linearly with increasing range.
4. *Mobile distance approximation*: location system approximates mobile's distance from each AP using an appropriate mapping curve and the mobile RSSI measurement. We use a simple heuristic called Proximity in Signal Space (PSS), which is detailed in Section 4.3, to select the appropriate curve.
5. *Location estimation using Triangular Interpolation and eXtrapolation (TIX) algorithm*: approximated distances between the mobile and APs and location coordinates of APs are input to TIX. TIX (detailed in Section 4.4) infers the final location estimate of the mobile.
6. *Location response*: location system responds to mobile's location request by informing the final location estimate.

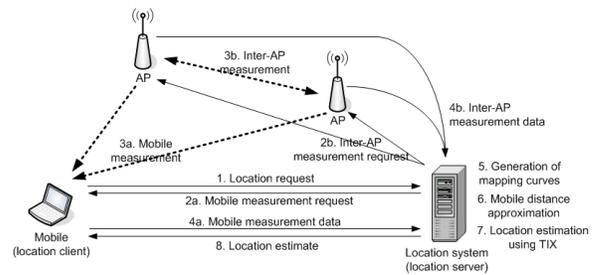


Figure 4: Calibration-free location estimation techniques

### 4.2 Generation of mapping curves

Generating equation of a line requires at least two points. We obtain the first point using inter-AP RSSI measurement and each AP's location, which is known to the system. The second point is the maximum RSSI value measured at the minimum measurable range, which is also known to the system. Since we have four APs collecting RSSIs from all other APs excluding itself, the system generates three mapping curves for each AP.

### 4.3 Mobile distance approximation

There are (N-1) mapping curves for each AP, generated by the other (N-1) APs' measurements, if there are N APs in the system. We select one most appropriate mapping curve for each AP to approximate mobile's distance from the AP based on a simple heuristic called 'Proximity in Signal Space (PSS).' We choose the mapping curves generated by using inter-AP measurement of an AP, from which the mobile received the highest RSSI value. For example, consider a mobile measurement, which results in RSSIs,  $S_2 > S_1 > S_4 > S_3$ , from AP2, AP1, AP4, and AP3, respectively. Given that  $S_2$  is the highest, it is most probable that the mobile is closest to AP2 than AP1, AP3, or AP4. Thus, we approximate the mobile distances from AP1, AP4, and AP3 all by using mapping curves generated by AP2's measurement. However, since AP2's measurement excludes itself, the mapping curve generated using AP1's RSSI measurement (AP1 gives the next highest RSSI measured by the mobile) approximates the distance between the mobile and AP2.

### 4.4 Triangular Interpolation and eXtrapolation (TIX) algorithm

TIX algorithm requires at least three APs. In case of more than three APs, TIX chooses the best three APs (i.e., highest three mobile-measured RSSIs). TIX takes two inputs: 1) distances between mobile and APs; 2) location coordinates of APs. TIX first forms a triangle, whose vertices are locations of APs. Then, TIX uses internal or external dividers of sides of the triangle to determine the final location estimate.

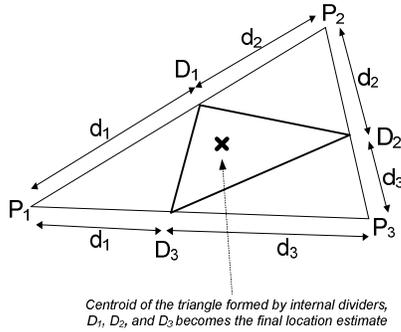


Figure 5: Triangular interpolation

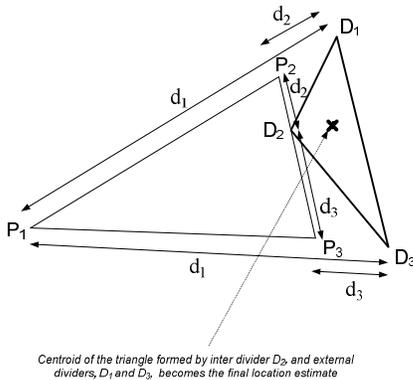


Figure 6: Triangular extrapolation

Figure 5 illustrates TIX algorithm with only internal dividers (i.e., triangular interpolation). Vertices of triangle,  $P_1$ ,  $P_2$ , and  $P_3$ , are locations of three APs, AP1, AP2, and AP3. Sides of triangle  $P_1P_2P_3$  are internally divided by points  $D_1$ ,  $D_2$ , and  $D_3$ . The centroid of triangle  $D_1D_2D_3$  gives the final location estimate.

Figure 6 illustrates TIX algorithm with at least one external divider (i.e., triangular extrapolation). In this case, two sides of triangle  $P_1P_2P_3$ , connecting  $P_1$  and  $P_2$ , and  $P_1$  and  $P_3$ , are externally divided by points  $D_1$  and  $D_3$  while the side connecting  $P_2$  and  $P_3$  is internally divided by  $D_2$ . The centroid of triangle  $D_1D_2D_3$  gives the final location estimate.

Triangular interpolation or extrapolation cases are determined based on the following:

1. *Triangular Interpolation.* If a side of triangle  $P_1P_2P_3$ , or line  $P_1P_2$ , which is the distance between AP1 and AP2, is longer than both  $d_1$  and  $d_2$ , a point along the line  $P_1P_2$  that divides in  $d_1$ -to- $d_2$  ratio is found using (4.1). We compute  $D_1$ , an internal divider of the line  $P_1P_2$ , or a side of the triangle  $P_1P_2P_3$ , by interpolation using:

$$D_1 = \frac{\frac{1}{d_1}}{\frac{1}{d_1} + \frac{1}{d_2}} P_1 + \frac{\frac{1}{d_2}}{\frac{1}{d_1} + \frac{1}{d_2}} P_2, \quad (4.1)$$

where  $D_1$  is the coordinate of the internal divider,  $D_1 = (D_{1x}, D_{1y})$ ,  $P_1$  and  $P_2$  are the coordinates of AP1 and AP2 locations,  $P_1 = (P_{1x}, P_{1y})$  and  $P_2 = (P_{2x}, P_{2y})$ .  $d_1$  is larger than  $d_2$ .  $D_1$  divides the line connecting  $P_1$  and  $P_2$  in  $d_1$ -to- $d_2$  ratio, internally.

2. *Triangular Extrapolation.* If a side of triangle  $P_1P_2P_3$ , or line  $P_1P_2$ , which is the distance between AP1 and AP2, is shorter than either  $d_1$  or  $d_2$ , a point  $D_1$  along the line connecting  $P_1$  and  $P_2$ , or locations of AP1 and AP2, which extends beyond point  $P_1$  or  $P_2$  and divides in  $d_1$ -to- $d_2$  ratio is found using (4.2). We compute  $D_1$ , an external divider of the line  $P_1P_2$ , or a side of the triangle  $P_1P_2P_3$ , beyond locations  $P_1$  or  $P_2$  extrapolating line  $P_1P_2$  using:

$$D_1 = \frac{-\frac{1}{d_1}}{\frac{1}{d_2} - \frac{1}{d_1}} P_1 + \frac{\frac{1}{d_2}}{\frac{1}{d_2} - \frac{1}{d_1}} P_2, \quad (4.2)$$

where  $D_1$  is the coordinate of the external divider,  $D_1 = (D_{1x}, D_{1y})$ ,  $P_1$  and  $P_2$  are the coordinates of AP1 and AP2 locations,  $P_1 = (P_{1x}, P_{1y})$  and  $P_2 = (P_{2x}, P_{2y})$ .  $d_1$  is larger than  $d_2$ .  $D_1$  divides the extrapolated line connecting beyond  $P_1$  and  $P_2$  in  $d_1$ -to- $d_2$  ratio.

3. Similarly for AP2 and AP3, the point that divides the line connecting AP2 and AP3 internally or externally (based on conditions given by the previous two steps), or  $D_2$ , is found by using either Step 1 or Step 2.
4. Similarly for AP1 and AP3, the point that divides the line connecting AP1 and AP3 internally or externally (based on conditions given by the previous two steps), or  $D_3$ , is found by using either Step 1 or Step 2.
5. Centroid of the triangle formed by the points  $D_1$ ,  $D_2$ , and  $D_3$ , (found through Steps 1 to 4) gives the final location estimate.

## 4.5 Description of experiments

We retained most of our experimental methods, tools, and software platform used in Section 3 and carried out empirical performance evaluation for calibration-free location estimation. Figure 7 depicts our evaluation scenario. Locations of our four APs remained unchanged and we estimated location of a mobile node in two distinct regions. First, we estimated the mobile location outside the convex hull connecting the APs (thin lines in Figure 7). Three thick dark lines in Figure 7 represent this region and 30 different locations were considered. Secondly, we estimated inside the convex hull represented as thick bright lines in Figure 7 where another 30 different locations were considered.

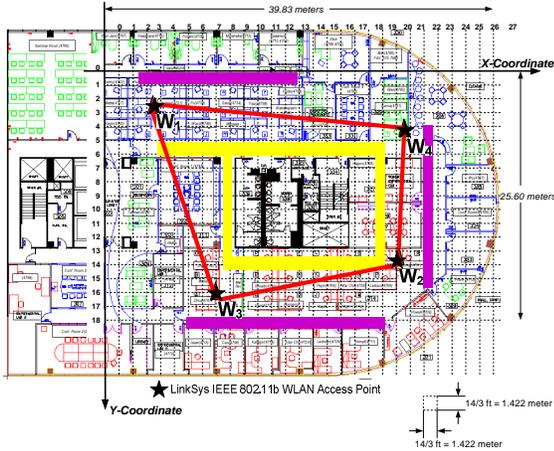


Figure 7: Performance evaluation for calibration-free location estimation techniques

## 4.6 Error characteristics of calibration-free location estimation

We computed mean distance errors of TIX. Using the same RSSI samples collected during the mobile and inter-AP measurements, we also computed mean distance errors of the following algorithms for comparison: 1) signal strength weighted centroid (SSWC) and 2) triangulation (TN).

SSWC is a simple algorithm, which estimates the user location by weighing the coordinates of locations of the APs according to their RSSI values. Consider RSSI samples,  $S_{AP1}$ ,  $S_{AP2}$ ,  $S_{AP3}$ , and  $S_{AP4}$ , obtained from four APs, AP1, AP2, AP3, and AP4, whose locations are  $L_{AP1} = (X_{AP1}, Y_{AP1})$ ,  $L_{AP2} = (X_{AP2}, Y_{AP2})$ ,  $L_{AP3} = (X_{AP3}, Y_{AP3})$ , and  $L_{AP4} = (X_{AP4}, Y_{AP4})$ . SSWC computes the location estimate,  $\hat{L}$ , as the following:

$$\hat{L} = \frac{S_{AP1}}{\Sigma} L_{AP1} + \frac{S_{AP2}}{\Sigma} L_{AP2} + \frac{S_{AP3}}{\Sigma} L_{AP3} + \frac{S_{AP4}}{\Sigma} L_{AP4}, \quad (4.3)$$

where  $S = S_{AP1} + S_{AP2} + S_{AP3} + S_{AP4}$ .

TN used here solves equations of circles by approximating mobile's distance (from APs) using our calibration-free mobile distance approximation as described in Section 4.3. Calibration-based TN used in Section 3 differs by matching the run-time RSSI values to those of the offline to approximate the distance.

Figure 8 depicts mean distance errors of SSWC, TN, and TIX. SSWC has 40 – 65 % greater mean distance error outside the

convex hull than the inside. TN and TIX errors without Proximity in Signal Space (PSS) heuristic display the similar trend, while TIX outperforms TN when PSS is applied. Furthermore, TIX's computational complexity is significantly lower than TN requiring only a number of arithmetic operations (TN requires solving the equations of circles and examining areas formed by solutions). Use of PSS heuristic alone improved the accuracy by 20 – 30 %. The best mean distance error was 4.7 m, achieved by TIX.

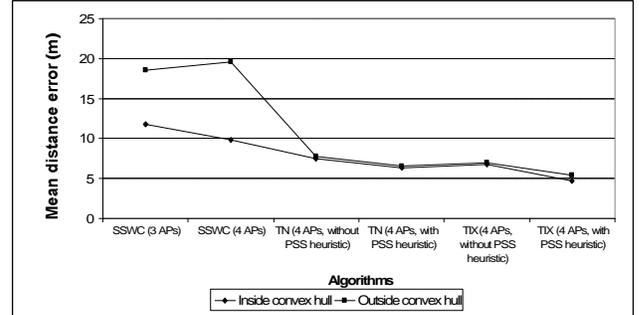


Figure 8: Mean distance errors of SSWC, TN, and TIX

## 5. THEORETICAL ANALYSIS

We present theoretical analysis on error characteristics of location algorithms primarily focusing on RF issues. We begin by quantizing the effect of calibration.

### 5.1 Impact of calibration

RF propagation is imperfect. RSSI values collected over time can vary significantly. Although regarded dynamic in general, proper statistical processing techniques can help to statically characterize RSSI.

#### 5.1.1 Primary and secondary nearest neighbors in square grid space

Figure 9 depicts a square grid for calibration with grid size  $g$ . A center grid location (designated as  $\bullet$  in Figure 9) has four closest grid positions (designated as  $\blacksquare$  in Figure 9) that are located in the next coordinate vertically or horizontally. We name these points as the primary nearest neighbors. The next closest grid positions are the secondary nearest neighbors and designated as  $\blacklozenge$  in Figure 9. The secondary nearest neighbors are located in the next coordinate diagonally from the center location.

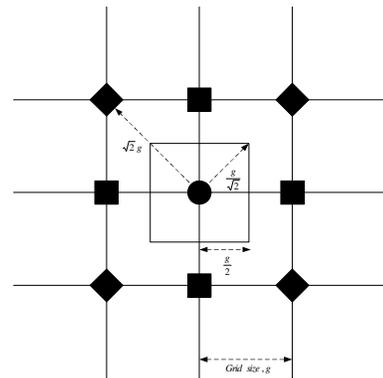


Figure 9: Nearest neighbors in square-grid space

Calibration-based algorithms are most effective when RSSI values acquired at center grid positions and their nearest neighbors exhibit clearly distinguished statistical patterns. In other words, if mean RSSI values of a center position and its primary nearest neighbors are the same, measurements at such grid alone contribute errors up to  $g$ . This inclusion of errors is known as *aliasing* in signal space, where RF properties of grid positions are blurred by physically different grid positions.

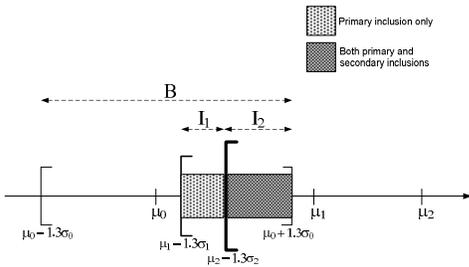
### 5.1.2 Gaussian approximation on nearest neighbor error inclusions

RF measurements occur during particular periods of time and collect a finite set of RSSI samples. Benefit of Gaussian approximation is that the (marginal) distribution of an arbitrary (or unknown) stochastic process at any finite set of points can be specified, and is Gaussian [6]. In fact, we observed that RSSI samples obtained empirically at our measurement locations were roughly normally-distributed.

**Table 1: Confidence intervals vs. ranges**

Confidence Interval (CI)	Range
0.800	$\pm 1.2816\sigma$
0.900	$\pm 1.6449\sigma$
0.950	$\pm 1.9600\sigma$
0.990	$\pm 2.5758\sigma$
0.999	$\pm 3.2905\sigma$

The Gaussian approximation considers the first and second order statistics (e.g., mean and variance/standard deviation). Consider the center location in Figure 9 has mean RSSI value and standard deviation pair,  $(\mu_0, \sigma_0)$ . Equivalently, the primary and secondary nearest neighbors have  $(\mu_1, \sigma_1)$  and  $(\mu_2, \sigma_2)$ . In the Gaussian approximation, range in multiples of the standard deviation from the mean value corresponding to Confidence Interval (CI) can be determined using error function (erf) [6] as summarized in Table 1.



**Figure 10: Error inclusion with grid size  $g$  in calibration**

A simple band-pass filtering was applied for our data processing method (see Section 3.2), which excludes the highest and lowest 10% of RSSI samples. This corresponds to the first entry in Table 1 (highlighted), approximately  $\pm 1.3\sigma$  variation from the mean received signal strength is expected according to the Gaussian approximation.

In Figure 10, we explain the impact of calibration with grid size  $g$  and the band-pass filtering at sample acquisition, which corresponds to the inclusion of errors from the primary and

secondary nearest neighbors. There are two regions of inclusion where only error from the primary nearest neighbors and from both the primary and secondary nearest neighbors are considered.

Using the same notation of Figure 10, total error contribution from the primary and secondary nearest neighbors,  $e$ , in calibration with grid size  $g$  based on the Gaussian approximation is:

$$e = g \cdot \frac{I_1}{B} + \left( \frac{1+\sqrt{2}}{2} \right) g \cdot \frac{I_2}{B}, \quad (5.1)$$

where  $B$  is Gaussian range of the center position,  $I_1$  is the range of the region where only the primary nearest neighbors contribute to the error,  $I_2$  is the range of region where both the primary and secondary nearest neighbors contribute to the error. As depicted in Figures 10 and 11, the primary nearest neighbors contributes distance error of  $g$  with probability of  $I_1/B$ . The secondary nearest neighbors contributes  $\sqrt{2}g$  with probability of  $I_2/B$ . There is also another component of error, which is due to the center location itself. The maximum self-error contribution is  $g/\sqrt{2}$  with probability of  $1 - I_1/B - I_2/B$ .

Combining the self-error contribution with (5.1), the total distance error due to the effect of calibration with grid size  $g$  is:

$$E = \frac{g}{\sqrt{2}} \cdot \left( 1 - \frac{I_1}{B} - \frac{I_2}{B} \right) + \left( \frac{\sqrt{2}+2}{4} \right) g \cdot \frac{I_1}{B} + \left( \frac{3\sqrt{2}+2}{6} \right) g \cdot \frac{I_2}{B}. \quad (5.2)$$

## 5.2 Accuracy of classical location algorithms

### 5.2.1 Triangulation (TN)

In Figure 11, the final triangle (smallest area formed by solutions of equations of circles) in TN is depicted.  $C$  is the centroid of the triangle, thus the final location estimate. Distances from  $C$  to each vertex,  $d_1$ ,  $d_2$ , and  $d_3$ , represent maximum distance errors of TN. Using the property of a triangle that the sum of two sides is larger than the length of the longest side, we observe:

$$d_2 + d_3 > S_1, \quad (5.3)$$

$$d_1 + d_3 > S_2, \quad (5.4)$$

$$d_1 + d_2 > S_3. \quad (5.5)$$

Combining (5.3), (5.4), and (5.5) yields:

$$\frac{d_1 + d_2 + d_3}{3} > \frac{1}{2} \cdot \frac{S_1 + S_2 + S_3}{3}. \quad (5.6)$$

Right side of (5.6) is mean maximum error of TN. Thus, we find the lower bound for mean maximum error of TN:

$$\text{Mean Maximum Error of TN} > \frac{1}{2} \cdot \bar{S}, \quad (5.7)$$

where  $\bar{S}$  is mean length of sides of the final triangle in TN.

### 5.2.2 $K$ -Nearest Neighbors (KNN)

We consider lower bound for mean maximum error of KNN, with  $K = 3$ . Figure 12 depicts KNN in a grid space with grid size  $g$ . Averaging distances from  $C$ , which is the centroid of the triangle and hence the final estimate of KNN, yields:

$$\text{Mean Maximum Error of KNN} > \frac{g}{3} \cdot \left( \frac{\sqrt{2}}{3} + \frac{\sqrt{5}}{3} + \frac{\sqrt{5}}{3} \right) = \frac{\sqrt{2} + 2\sqrt{5}}{9} g \quad (5.8)$$

### 5.2.3 Smallest M-Polygon (SMP)

We derive the lower bound for mean maximum error of SMP, for  $M = 3$  (i.e., the polygon is triangle). Using similar principle from 5.2.1, we obtain:

$$\text{Mean Maximum Error of SMP} > \frac{1}{2} \cdot \left( \frac{1}{3} \bar{P} \right) = \frac{1}{6} \bar{P}, \quad (5.9)$$

where  $\bar{P}$  is mean perimeter length of the smallest polygons.

## 5.3 Accuracy of TIX

We applied the same analysis used for TN in Section 5.2.1. Given  $\bar{S}$ , mean length of sides of triangle formed by TIX, lower bound for the mean maximum error of TIX is:

$$\text{Mean Maximum Error of TIX} > \frac{1}{2} \cdot \bar{S} \quad (5.10)$$

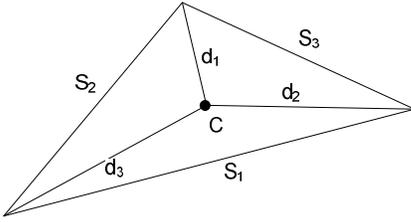


Figure 11: Final triangle and error approximation in TN

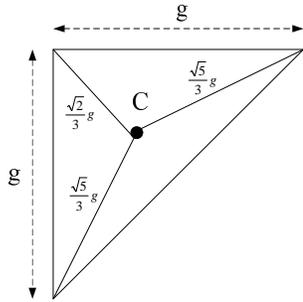


Figure 12: Error approximation in KNN

## 5.4 Comparison between empirical and theoretical errors

In Section 3.4, we observed that mean distance errors of TN, KNN, and SMP did not increase significantly until the grid size is enlarged to a certain threshold (i.e., 8x). To explain this, we first show how calibration in different grid sizes affects the distance error under the presence of the primary and secondary nearest neighbor inclusions. Figure 13 illustrates maximum possible distance errors contributed from self and nearest neighbor inclusions using the analysis in Section 5.1. We computed Equations (5.1) and (5.2) using the empirical values. As depicted in Figure 13, error contribution from the nearest neighbors decreases as grid size increases. On the other hand, the self error contribution increases linearly as grid size increases. The eventual

outcome is that the total error combining the both rises sharply after some threshold grid size (between 4x and 8x).

We present theoretical errors for TN, KNN, and SMP in Figure 14, which combines the analyses of Sections 5.1 and 5.2. Theoretical errors are smaller than our empirical results by 20 – 30 %. We observe similar behaviors between the empirical and theoretical error characteristics: 1) Errors increase gradually as grid size increases until some threshold (near 8x); 2) TN results with the highest error while KNN results with the smallest error; 3) At 8x, errors for TN and SMP become similar.

Theoretical errors for TIX are displayed in Figure 15. We computed theoretical errors in two estimation regions (inside and outside convex hull formed by APs) separately using TIX error Equation (5.10). Comparing to empirical errors that we presented in Section 4.6, theoretical errors are smaller by 12 – 17 %. Theoretical approximation for TIX reflects the empirical results better because there is no effect of calibration to be considered.

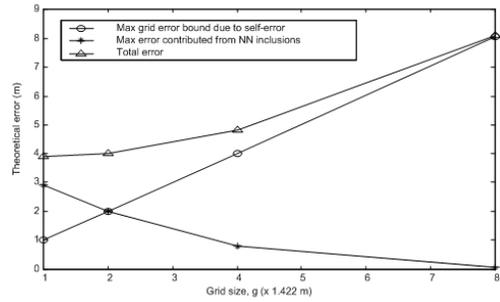


Figure 13: Maximum possible distance errors contributed from self and nearest neighbor inclusions in different grid sizes

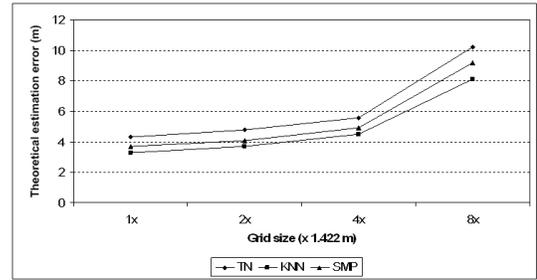


Figure 14: Theoretical errors for TN, KNN, and SMP

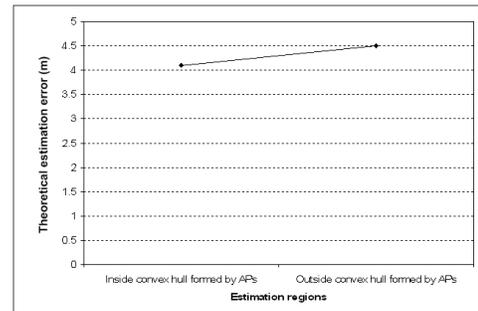


Figure 15: Theoretical errors for TIX

## 6. RELATED WORK

Early work in wireless LAN-based location estimation included the RADAR [1] system. RADAR used KNN location algorithm in a static scene analysis framework. RADAR showed that fairly accurate indoor location could be estimated without a separate sensor network infrastructure. In their following work [2], RADAR was enhanced by a Viterbi-like algorithm.

In [3], an empirical work on a RADAR-like location system was presented. The authors adopted KNN location algorithm, which was used by RADAR, and experimented with various choices of  $K$ . They validated that choosing  $K > 3$  (RADAR chose  $K = 3$ ) did not improve the accuracy but degraded it by 5 – 20 %.

Pandya *et al.* [4] presented the MultiLoc system, which utilizes information from multiple RF technologies to improve accuracy. The MultiLoc system employs two simple ‘sensor fusion’ techniques to illustrate the benefit of using multiple technologies. In [5], more sophisticated algorithm for sensor fusion, namely SELFLOC, was presented. The authors also showed that multiple estimation algorithms could be combined to further improve accuracy using SELFLOC.

Techniques described in [6] were one of the first location algorithms with reduced calibration. Triangulation Mapping Interpolation (TMI) algorithm could achieve smaller distance error than RADAR although TMI required smaller calibration table constructed from sparser grid space. Mapping and interpolation schemes based on signal processing algorithms were used. Krumm and Platt [7] presented a more experiment-oriented work in minimizing calibration for wireless LAN-based location systems. They showed a series of empirical estimation errors while varying amount of calibration with KNN algorithm. They also described an interpolation function for smoothing estimates with reduced amount of calibration.

Ladd *et al.* [8] developed a wireless LAN-based location system using Bayesian reasoning and a hidden Markov model. While it achieved fine accuracy (in 1 – 2 m errors), very dense calibration is required. They acknowledged the extraordinary effort in calibration and suggested the use of robotics-based approach for it. Although Nibble [9] system was based on RF-based static scene analysis, it did not use any geometric location algorithms such as triangulation, but took a purely probabilistic approach (based on Bayesian network).

There are a number of other classes of location technologies. Ad hoc networking-based location systems, which require no calibration, typically impose other types of requirements such as certain degree of distribution of ad hoc users over an area covered by the location system. APS [10] is based on mixture of ad hoc routing, and range (e.g., ToA, TDoA) and angle (AoA) measurement technologies.

Our work differs from others in several aspects. First, we consider zero calibration. Location algorithms in [6,7] considered reduced amount of calibration, but never eliminated it. TIX algorithm combined with our proposed calibration-free techniques only requires at least three APs in communicative range (25 – 40 m), yet achieves comparable accuracy performance to those of calibration-based algorithms. We did not only empirically demonstrate how the impact of calibration, grid size, and use of

location algorithms resulted in error characteristics of RF-based static scene analysis, but also qualitatively explained with theoretical analysis on the same parameters. To the best of our knowledge, such theoretical analysis is unprecedented in any previous literature of similar arts.

## 7. CONCLUSIONS

We presented empirical error characteristics of the classical location algorithms in various spatial densities of calibration, and calibration-free TIX algorithm. TIX achieved mean distance error within 5.4 m with zero calibration, which is comparable to within 4.7 m errors of laborious calibration-based algorithms. Increasing calibration grid size for coarser granularity had gradual effect on estimation errors unless it was increased to a certain threshold. We presented theoretical analysis on error characteristics of location algorithms and validated key sources of errors in both calibration-based and calibration-free algorithms.

## 8. ACKNOWLEDGMENT

Authors acknowledge Raghu Dendukuri of DoCoMo USA Labs for his contribution to implementation.

## 9. REFERENCES

- [1] Bahl, P., and Padmannabhan, V., N. RADAR: An In-Building RF-based Location and Tracking System. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '00)*. Tel-Aviv, Israel, March 2000.
- [2] Bahl, P., and Padmannabhan, V., N. *Enhancements to the RADAR User Location and Tracking System*. Technical Report MSR-TR-2000-12, Microsoft Research, February 2000.
- [3] Prasithsangaree, P., Krishnamurthy, P., and Chrysanthis, P., K. On Indoor Position Location with Wireless LANs. In *Proceedings of the IEEE Personal, Indoor and Mobile Radio Communications (PIMRC '02)*. Lisbon, Portugal, September 2002.
- [4] Pandya, D., Jain, R., and Lupu, E. Indoor Location Using Multiple Wireless Technologies. In *Proceedings of the IEEE Personal, Indoor and Mobile Radio Communications (PIMRC '03)*. Beijing, China, September 2003.
- [5] Gwon, Y., Jain, R., and Kawahara, T. Robust Indoor Location Estimation of Stationary and Mobile Users. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '04)*. Hong Kong, China, March 2004.
- [6] Samilagic, A., and Kogan, D. Location Sensing and Privacy in a Context-aware Computing Environment. *IEEE Wireless Communications, Vol. 9 No.5*. October 2002.
- [7] Krumm, J., and Platt, J. *Minimizing Calibration Effort for an Indoor 802.11 Device Location Measurement System*. Technical Report MSR-TR-2003-82, Microsoft Research, November 2003.
- [8] Ladd, A., M., et al. Robotics-Based Location Sensing Using Wireless Ethernet. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking (MobiCom '02)*. Atlanta, Georgia, September 2002.
- [9] Castro, P., et al. A Probabilistic Location Service for Wireless Network Environments (Nibble). In *Proceedings of Ubiquitous Computing*, Atlanta, Georgia, September 2001.
- [10] Niculescu, D., and Nath, B. Error Characteristics of Ad Hoc Positioning Systems (APS). In *Proceedings of the Fifth ACM International Symposium on Mobile Ad Hoc Networking and Computing*. Tokyo, Japan, May 2004.
- [11] The linux-wlan™ Project. <http://www.linux-wlan.co>