Access Point Localization Using Local Signal Strength Gradient

Dongsu Han[1], David G. Andersen[1], Michael Kaminsky[2], Konstantina Papagiannaki[2], and Srinivasan Seshan[1]

[1] Carnegie Mellon University
[2] Intel Research Pittsburgh

Abstract. Many previous studies have examined the placement of access points (APs) to improve the community's understanding of the deployment and behavioral characteristics of wireless networks. A key implicit assumption in these studies is that one can estimate the AP location accurately from wardriving-like measurements. However, existing localization algorithms exhibit high error because they over-simplify the complex nature of signal propagation. In this work, we propose a novel approach that localizes APs using directional information derived from local signal strength variations. Our algorithm only uses signal strength information, and improves localization accuracy over existing techniques. Furthermore, the algorithm is robust to the sampling biases and non-uniform shadowing, which are common in wardriving measurements.

1 Introduction

Locating the source of a radio frequency (RF) transmission is important for a wide range of purposes. These include finding rogue access points (APs), creating wardriving maps, and estimating the RF propagation properties of an area. The traditional approach to localizing an AP is to perform wardriving-like measurements (i.e., measure received signal strength and location information) and to apply a number of common techniques to process the data. Unfortunately, the types of algorithms that can be applied to such data are limited, and state-of-the-art algorithms exhibit high error and high variation. Recent studies [1] have shown that using erroneous results produced by state-of-art localization algorithms impair the performance of mobile user positioning systems like Place Lab [2] and cause inaccurate estimates of coverage and interference.

One approach to improving localization accuracy is to use more sophisticated data collection techniques. For example, Subramanian et al. [3] recently improved accuracy by using angle of arrival (AoA) information collected with a steerable beam directional antenna. While such techniques improve accuracy, the cost of hardware and human time is high. In this paper, we present a novel AP localization algorithm called *gradient* that uses only information collected from conventional wardriving, and that does not require extra hardware. Our approach uses the local signal strength distribution to estimate the direction of the AP. The gradient algorithm localizes the AP by combining directional estimates from

S.B. Moon et al. (Eds.): PAM 2009, LNCS 5448, pp. 99–108, 2009.

multiple vantage points. We show that *gradient* improves the mean accuracy by 12% over the state-of-the-art algorithm, and reduces the maximum error and standard deviation of errors by more than 33%. This paper describes the key insights and the design of the *gradient algorithm*, verifies the idea through simulation (Section 2), and compares its real-world performance against existing methods (Section 3).

2 Localization Algorithm

The goal of our work is to estimate the location of an AP given a set of received signal strength (RSS) measurements from different locations. These RSS measurements are typically obtained by passively monitoring 802.11 frames from a moving vehicle—a practice known as *wardriving*. The data contains measurement locations (x, y) and a received signal strength RSS at each location.

We first examine the the most commonly used localization algorithms, centroid, weighted centroid [2] and trilateration [4], to motivate the need for a better algorithm. Given the set of measurement points $<x_i, y_i, RSS_i>$ of an AP, centroid algorithms locate the AP at the averaged location: $\left(\sum_{i=1}^{N} w_i x_i, \sum_{i=1}^{N} w_i y_i\right)$. For centroid, $w_i = 1/N$, and for weighted centroid, $w_i = \dfrac{SNR_i}{\sum_{j=1}^{N} SNR_j}$.

Trilateration [4] estimates the distance from the signal source at a measurement point using the RSS and combines these distance estimates to infer the location of the AP. The RSS is converted to distance using the log-distance path loss model [5]. The model defines the path loss ($\overline{PL}$) from transmitter to receiver as a function of distance (d) as $\overline{PL}(\text{dB}) = \overline{PL}(d_0) + 10n \log\left(\frac{d}{d_0}\right)$, where d_0 is a reference distance and n is the path loss exponent.

Despite their simplicity, centroid-based algorithms perform as well as other existing algorithms that use signal strength [1, 3]. To understand why our gradient algorithm can outperform all of these algorithms, we first examine one representative case where centroid algorithms perform poorly.

2.1 Motivating Example

Figure 1 shows wardriving measurements collected for one AP inside an apartment. The building is near a three-way junction. The AP is located towards the front of the building. Each measurement point is shaded to indicate the received signal strength according to the gray-scale legend on the right side of the figure.

Because the measurements are taken from a car, the path of the measurement points reflects the shape of the roads. *Wardriving measurements introduce strong sampling biases.* Figure 1 also shows non-uniform signal propagation. The area in front of the AP has denser and stronger (light-colored) measurements compared to the area behind the building. This is because the signal behind the building is shadowed by multiple walls, while the front of the building has fewer obstructions. Such non-uniform shadowing is also typical in wardriving measurements.

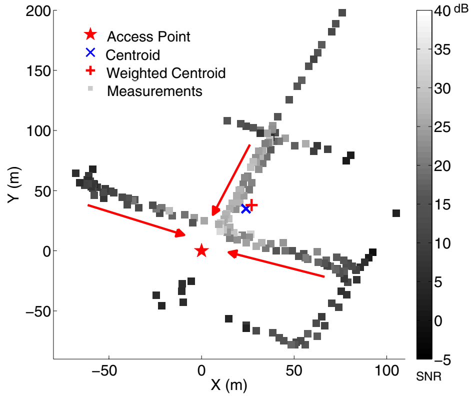


Fig. 1. Real world measurement of an AP located at the origin

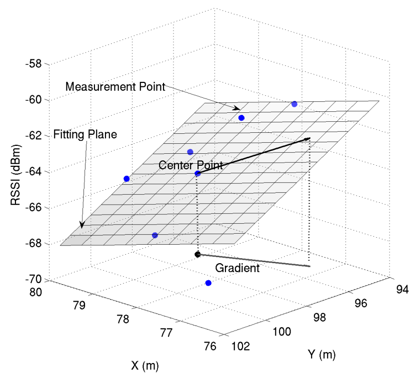


Fig. 2. Drawing the arrow

As a result, both centroid and weighted centroid give high errors of 51.8m and 52.3m (Figure 1). Centroid algorithms suffer when subject to 1) biased sampling, which is exacerbated by the layout of the road, and 2) non-uniform signal propagation due to different levels of shadowing.

Despite the skewed distribution, the signal strength distribution remains a useful hint in localizing the AP. We see a clear trend of increasing signal strength towards the AP, as denoted by the arrows in Figure 1. These arrows provide useful information about the location of AP. Based on this observation, we develop a novel AP location algorithm which we call the *gradient algorithm*.

2.2 Gradient Algorithm

The main idea in the gradient algorithm is to systematically point arrows towards the signal source. We estimate the direction of an AP from every measurement point by calculating the direction of strongest signal in the neighboring area. After this process, each measurement point has an *arrow* that points to the direction of AP. By *combining* the resulting arrows, we estimate the location of the AP. While individual arrows have some error, overall these would cancel out during this stage.

The algorithm's effectiveness arises from two properties of the signal propagation: First, despite the inherent noise in the measurement, distance from the source AP remains one of the most important determinants of RSS. Second, obstructions cause sudden attenuation of signal that weakens the correlation between distance and RSS in the global space. However, measurements near-by in space tend to have traveled through the same obstructions. Thus, when considering only local signal strength, distance is the single most important factor. The algorithm leverages these observations and estimates the direction of the AP by finding the direction of strongest signal in near-by space.

Although the effect of large-scale non-uniform shadowing may be constrained in a local area, other small-scale factors — muti-path propagation, reflection, and errors in GPS positioning, among others— affect RSS. The gradient algorithm treats them as noise and relies on averaging to mitigate their effect.

These steps of the gradient algorithm operate as follows:

Preprocessing. This phase averages signal observations to mitigate the effect of various sources of error. First, we embed the recorded GPS coordinate into a two dimensional coordinate system. Then we place a unit meter grid on the map and average the RSS and coordinates that fall into the same grid square.

Drawing the arrows. We assume free-space propagation locally, and find the direction of the AP by calculating the direction in which the RSS increases the most. We do so by fitting a plane to "near-by" measurements in the x-y-RSS space, and taking the gradient of the plane. We use minimum mean-square error (MMSE) fit to overlay a plane, and take the gradient at the center point as the direction of the AP in the x-y plane (Figure 2). We call this gradient an *arrow*. We locate "near-by" space using a square centered at the center point, and define "window size" to be half the length of the side of the square.

The window size is critical because it determines the area in which the signal propagation is approximated as free-space. The optimal value depends on the density of measurement points and their spatial distribution. Therefore, there is no one-size-fits-all value across different APs. We revisit this problem at the end of the section.

Combining the arrows. Once we determine the direction of the AP from each measurement point by drawing arrows as described above, we locate the AP at the position to which all of the arrows point. We position the AP at the location that minimizes the sum-squared angular error from the arrows. In calculating the sum-squared error, we weigh each error by the SNR at the measurement point. This step is similar to the technique used by Subramanian et al [3], as well as other algorithms that use AoA.

Any standard optimization tool can be used for this minimization. However, most tools do not guarantee a global minimum and they require a good initial guess from which they start searching. Because the performance of any particular optimization algorithm is not of interest, we exhaustively search the space every meter, limiting it to a 100m x 100m square centered at the maximum RSS point. When using an optimization tool, we recommend using the maximum RSS point as the initial guess.

Window size. The window size is important in accurately predicting the direction of the AP from each measurement point. If the window is too small and contains few measurements, the resulting arrow is likely to have a high error. On the other hand, making the window too large is not desirable because it breaks our free-space assumption.

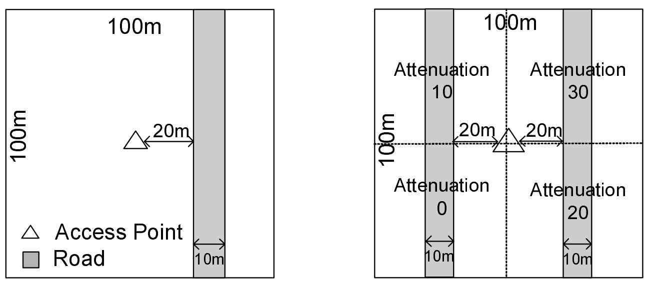
Because it is difficult to predict the best value for the window size at each measurement point, we reduce the problem to finding a single optimal value of the window size for each AP. We use the following heuristic to determine this value. First, we produce a series of estimated locations using window sizes from

1m to 50m. If the area covered by a window contains at least four measurements, we use that area to draw an arrow. At the end of the arrow drawing phase, if the arrows are drawn for fewer than 30% of all measurement points, the we do not produce an estimate for the window size.

The gradient algorithm starts using a window size of 1m and repeatedly increases it by one until five consecutive estimates converge within 5m from their average. At that point the algorithm terminates and reports the average estimate for the AP location. In Section 2.3, we verify the heuristic using simulation.

2.3 Simulation

We use simulation to verify the basic idea behind the gradient algorithm, and show that it is accurate and robust to sampling bias and non-uniform shadowing.



(a) Single road (b) Parallel road

Fig. 3. Simulation topologies: measurements are generated only on the road

Accuracy. We create a single and a parallel road topology (Figure 3) to verify the performance of the algorithm. We place an AP at the center of the maps and generate measurement points randomly along the road. The RSS at each point is calculated using the log-distance path loss model of Section 2. We set the reference distance d_0 to 1m, and subtract the path loss from signal strength at the reference point to get the RSS. For the Figure 3(a) topology, we vary the number of measurement points from 10 to 50. For (b), we generate 200 points and run another simulation that further attenuates the signal strength by the number on each quadrant to simulate non-uniform shading. Table 1 shows the average error of 20 runs using the gradient and centroid algorithms.

Gradient performs better as the density of measurement increases, and localizes the AP within 1 meter when the number of measurements exceeds 35. In contrast, the centroid algorithms suffer high error due to the biased sampling. When only non-uniform shading is introduced, gradient suffers much less than the weighted centroid, but in general needs denser measurements.

To see the effect of the path loss exponent, we vary the value of n from 2.0 to 5.0 which is the typical range in outdoor environments [5]. Figure 4 shows the average, minimum and maximum estimation error for the gradient algorithm using 10 different sampled topologies for Figure 3(a). The gradient algorithm appears unaffected by the value of the path loss exponent.

Table 1. Localization error for the topology in Figure 3(a) and (b). Units in meters.

	Number of measurements					
	10	15	20	25	30	35
Gradient	10.0	6.5	1.8	1.2	1.1	1.0
Centroid	26.4	26.4	26.3	25.8	25.5	25.8
W. Cen.	26.1	26.2	26.1	25.7	25.3	25.6

With non-uniform attenuation			
Gradient	6.7	W. Cen.	16.0
Without attenuation			
Gradient	0.3	W. Cen.	3.0

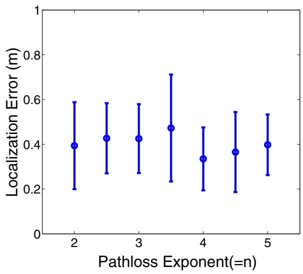


Fig. 4. Error versus path loss exponent

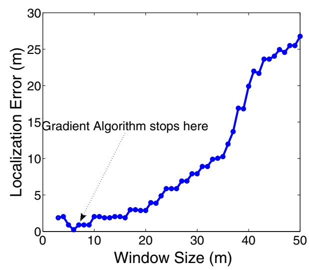


Fig. 5. Error versus window size

Effect of window size. As mentioned in Section 2.2, the window size is important for accuracy. To understand the effect, we generate 50 measurement points using Figure 3(a)'s topology and produce a series of estimates while varying the window size from 1 to 50m. The resulting curve (Figure 5) features a natural minimization point, where the gradient algorithm stops.

When the window is small, only a few measurement points are considered when drawing the arrow, thus leading to high error. As the window grows, our two assumptions break down. The free-space assumption stops holding because the large window now contains effects of non-uniform shadowing. (This effect is not present in the free-space simulation.) Second, because the global signal strength distribution is not linear, the distribution inside the large window also becomes non-linear. This results in errors due to underfitting since the gradient algorithm attempts to apply a linear fir to a non-linear distribution.

3 Evaluation

In this section, we evaluate the performance of the gradient algorithm in real environments. We first collect measurements and use them to compare the algorithm's accuracy with that of previously proposed approaches. We demonstrate that the algorithm is robust and clearly outperforms distance estimation based on signal strength.

3.1 Data Collection Framework

We collected wardriving measurements using the following setup:

Hardware and software. We used Intel-based laptops running Linux, with a Pharos GPS device, an Atheros wireless cardbus card, and an external omnidirectional antenna mounted on top of a car. In order to capture more packets, we used two laptops, with one scanning only the popular orthogonal channels 1, 6 and 11, and the other scanning channels 1 through 11. The madwifi driver and *kismet* were used to capture packets in monitor mode. The madwifi driver reports continuous received signal strength levels, which is important to our algorithm. *Kismet* records GPS coordinates, all received packets and their signal strength. From the *kismet* log, we extract an <AP MAC address, SSID, latitude, longitude, SNR, noise> tuple.

Wardriving. We took measurements in the residential Pittsburgh neighborhood of Squirrel Hill. The area consists mostly of detached homes, townhouses and small apartments. We drove at about 15mph, scanning each side of the road two to four times. Slow speed and multiple scans allowed us to get more measurements per AP. Scanning both sides of the road increases the number of measurements inside a small area, which improves the accuracy of our algorithm.

Ground truth. Determining the actual location of an AP is crucial for evaluation, but is difficult, especially in the unplanned deployment setting that we examine. We obtained ground truth locations in two ways. First, we knew some of the AP owners personally. Second, we observed that many SSIDs contained either street address or family names. Using an online phone book [6] plus the local zipcode, we mapped family names to addresses. We further verified that the resulting address was reasonably close to the observed location for the AP.

For each candidate ground truth AP, we measured the actual GPS coordinates on the street in front of the addressed building. This step was surprisingly important— address to GPS coordinate conversion services such as Google Maps often produce inaccurate mapping that were off by a few houses. We manually processed the measured coordinates by positioning the AP at the center of the house using Google Maps. This process located twenty-five APs.

3.2 Real World Experiment

We compare the accuracy of the gradient algorithm with that of centroid and trilateration using the data we collected in Section 3.1. We also compare with an algorithm that we developed called cone-fit. Cone-fit uses the same path loss model of Section 2. Given the path loss exponent as a parameter, it estimates, for all measurement $<x_i, y_i, RSS_i>$, the RSS from the path loss model and produces $<x_i, y_i, EstimatedRSS_i>$ by placing the AP at a location. It locates the AP at the location where $\sum(EstimatedRSS_i - RSS_i)^2$ is minimized. Visually, it is fitting a 3 dimensional cone to $<x, y, RSS>$ space whose skirt is shaped by the path loss exponent.

Note that unlike gradient, trilateration and cone-fit are parameterized. Trilateration takes as inputs the path loss exponent and the signal strength at the reference point, and cone-fit takes the path loss exponent as input. We give them

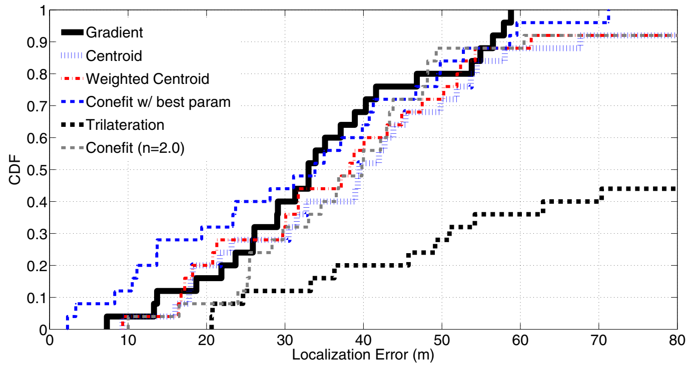


Fig. 6. Comparison of gradient, trilateration, centroid, and cone-fit

full advantage by calculating the best parameter given the actual known location of each AP. Figure 6 shows the localization error of these algorithms.

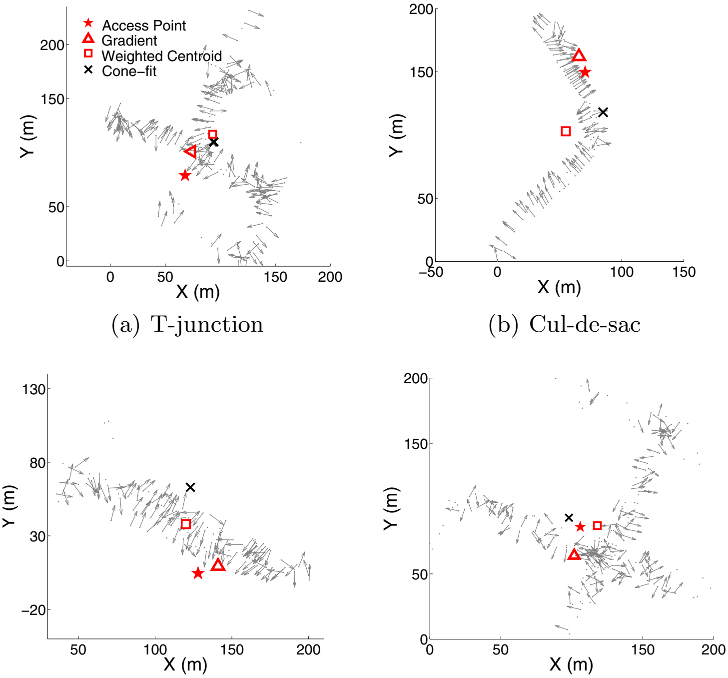
The mean error and standard deviation of gradient were 34m and 14m respectively. Gradient improves the mean error by 12% over weighted centroid, and performs better than centroids and trilateration. Moreover, it has the lowest worst case error, giving a factor of 1.4 improvement in standard deviation over weighted centroid. The maximum error of gradient was 59m versus weighted centroid's 88m. Cone-fit's average is 7% better than gradient only when the best path loss exponent was given from the actual AP's location, but when the parameter is fixed (e.g., n=2.0), it was worse than gradient. Its performance was highly variable depending on the exponent, and degrades under non-uniform shadowing and biased sampling.

Case study. To understand the characteristics of the gradient algorithm, we describe some example scenarios from the actual measurement (Figure 7). The figures show arrows at measurement points and the localization results of various algorithms[1]. Figures 7(a),7(b), and 7(c) show some skewed distributions that frequently occur in wardriving measurements. Note that a cut through the AP can be made in the x-y plane such that one side of the cut contains the vast majority of measurement points. Gradient performs well in these cases, unlike other algorithms. The more uniform distribution in Figure 7(d) provided the smallest error (of all 25 APs) for weighted centroid, because the AP was located at the corner of a junction. All algorithms perform relatively well in this scenario. While all algorithms perform well in under a uniform distribution of measurement points, only gradient performs well in all shown cases.

4 Related Work and Discussion

Localization systems fall into three rough categories based upon the information they use [4]: received signal strength, time-based, and angle-of-arrival (AoA).

[1] Trilateration is not shown because its errors were too large.



(a) T-junction (b) Cul-de-sac

(c) Single road (d) Uniform distribution

Fig. 7. Case study. Arrows are trimmed to improve readability.

RADAR [7] uses RSS to infer the distance between the signal source and the receiver. However, a site-survey is required to compensate for the non-uniform shading. Similarly, time-based approaches [8, 9] require accurate time synchronization or use an additional medium such as ultrasound to address timing issues.

AoA systems measure the incident angle of the signal at the receiver. Subramanian et al. [3] proposed an AP localization algorithm that uses AoA information measured with a steerable beam directional antenna. This state-of-the-art algorithm significantly reduces the localization error. However, the evaluation is based on APs that were modified to transmit a large number of packets and placed on the same channel to speed up the data collection. Simulations of normal-speed data collection, commensurate to a real-world situation, showed that the median localization error was comparable to that of gradient.

Of the three types of information, RSS is the only information collected in conventional wardriving measurements. Furthermore, few techniques used in user localization systems can be applied to AP localization in ad-hoc environments. Therefore, centroid and weighted centroid have been commonly used [2, 1]. Our work takes a novel approach in that we measure only RSS, but derive AoA equivalent information.

Finally, particle filters [10] provide a new approach to combine information obtained from each measurement regardless of its type. It provides a probabilistic mechanism to combine data instead of the deterministic ones used by most localization techniques. An interesting avenue of future work would be to combine our work with a particle filter, to estimate locations using both the RSS and the derived AoA information from our work.

5 Conclusion

This paper presents a novel access point localization algorithm called *gradient* and its performance evaluation. Gradient is based on the idea that the signal strength indirectly reflects the direction where the signal comes from. The algorithm does not require extra hardware, and data can be collected by normal wardriving.

Our evaluation shows that gradient overcomes the sampling bias that is inherent in wardriving measurements. Furthermore, it is more accurate and has smaller variation than any other algorithm that relies on the distribution of signal strength in space. Gradient achieves a median AP localization error of 33 meters and reduces the maximum error by 33% compared to the second best, weighted centroid approach.

References

[1] Kim, M., Fielding, J.J., Kotz, D.: Risks of using AP locations discovered through war driving. In: Fishkin, K.P., Schiele, B., Nixon, P., Quigley, A. (eds.) PERVASIVE 2006. LNCS, vol. 3968, pp. 67–82. Springer, Heidelberg (2006)
[2] Cheng, Y., Chawathe, Y., LaMarca, A., Krumm, J.: Accuracy characterization for metropolitan-scale Wi-Fi localization. In: Proc. ACM MobiSys., Seattle, WA (June 2005)
[3] Subramanian, A., Deshpande, P., Gaojgao, J., Das, S.: Drive-by localization of roadside WiFi networks. In: Proc. IEEE INFOCOM, Phoenix, AZ (April 2008)
[4] Savvides, A., Han, C.C., Strivastava, M.B.: Dynamic fine-grained localization in ad-hoc networks of sensors. In: Proc. ACM Mobicom, Rome, Italy (July 2001)
[5] Rappaport, T.: Wireless Communications — Principles and Practice. Prentice-Hall, Englewood Cliffs (1996)
[6] WhitePages.com - White Pages Phone Directory, `http://www.whitepages.com`
[7] Bahl, P., Padmanabhan, V.: RADAR: An In-Building RF-based User Location and Tracking System. In: Proc. IEEE INFOCOM, Tel-Aviv, Israel (March 2000)
[8] Harter, A., Hopper, A., Steggles, P., Ward, A., Webster, P.: The anatomy of a context-aware application. In: Proc. ACM/IEEE MOBICOM, Seattle, WA (August 1999)
[9] Priyantha, N., Chakraborty, A., Balakrishnan, H.: The Cricket location-support system. In: Proc. ACM Mobicom, Boston, MA (August 2000)
[10] Fox, D., Hightower, J., Liao, L., Schulz, D., Borriello, G.: Bayesian filtering for location estimation. IEEE Pervasive Computing 02(3) (2003)