Accuracy Characterization of Cell Tower Localization

Jie Yang[†], Alexander Varshavsky[‡], Hongbo Liu[†], Yingying Chen[†], Marco Gruteser[♯]
[†]Department of ECE, Stevens Institute of Technology, Hoboken, NJ, USA
[‡]AT&T Labs, Florham Park, NJ, USA
[♯]WINLAB, Rutgers University, North Brunswick, NJ, USA
[†]{jyang, hliu3, yingying.chen}@stevens.edu, [‡]varshavsky@research.att.com,
[♯]gruteser@winlab.rutgers.edu

ABSTRACT
Cell tower triangulation is a popular technique for determining the location of a mobile device. However, cell tower triangulation methods require the knowledge of the actual locations of cell towers. Because the locations of cell towers are not publicly available, these methods often need to use estimated tower locations obtained through wardriving. This paper provides the first large scale study of the accuracy of two existing methods for cell tower localization using wardriving data. The results show that naively applying these methods results in very large localization errors. We analyze the causes for these errors and conclude that one can localize a cell accurately only if it falls within the area covered by the wardriving trace. We further propose a bounding technique to select the cells that fall within the area covered by the wardriving trace and identify a cell combining optimization that can further reduce the localization error by half.

Author Keywords
Cell localization, cell tower, received signal strength

ACM Classification Keywords
K.8.m Personal Computing: Miscellaneous

General Terms
Experimentation, Measurement, Performance

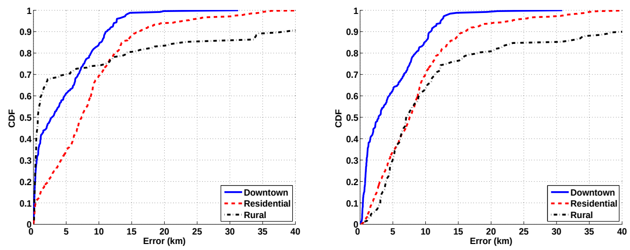
INTRODUCTION
Many mobile applications rely on cell tower triangulation to determine their position [4]. Even when Global Positioning System (GPS) is available, cell tower triangulation offers a significantly faster time to first fix and lower energy consumption. Cell tower triangulation methods [9, 5], however, require the knowledge of the actual locations of cell towers. This is also the case for many more general localization techniques such as Particle filters [7], lateration [10] and Bayesian networks [11].

Ubicomp 2010, Sep 26-Sep 29, 2010, Copenhagen, Denmark.

Although there are several public sources of cell tower locations, we found these sources to be both incomplete and inaccurate [1, 2, 3]. For instance, in the area we studied, these websites had data about less than 10% of cell towers from the service provider that we studied. Because the locations of cell towers are not readily available, a common way to estimate cell tower positions is through wardriving [5]. In wardriving, a vehicle drives within the target area, recording signals emanating from nearby cell towers (or WiFi access points) and the locations these signals were received at [8]. Using this dataset, one can estimate the locations of cell towers with algorithms such as weighted centroid [5, 6] or strongest received signal strength (strongest RSS) [12]. It has been shown for WiFi localization algorithms that the accuracy of such estimated access point locations directly affects the accuracy of mobile device localization algorithms [8]. However, we are not aware of any prior work that studied cell tower localization algorithm performance at a large scale.

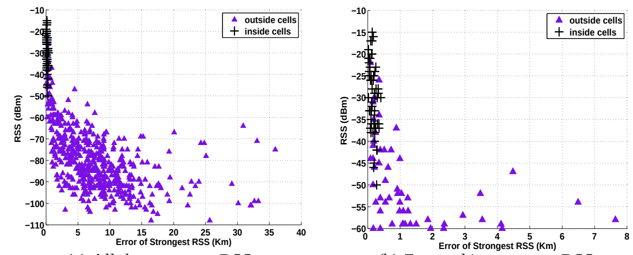
Validating these algorithms requires the knowledge of the actual locations of cell towers. Chen et al. [5] used the weighted centroid method based on the GSM wardriving data. They reported a cell localization accuracy by physically visiting 6 cell towers, which is not representative in a metropolitan area. Varshavsky et al. [12] used the strongest RSS method to localize cell towers for their study, but did not validate the cell localization accuracy. The closest work to ours is that of Kim et al. [8], in which they studied the accuracy of estimated WiFi access point (AP) locations from wardriving data and compared them with the actual AP locations. They showed that the estimated AP locations have a median error of 40 meters and that this error has a significant effect on the accuracy of estimating locations of mobile users. In contrast, we study the accuracy of cell tower localization algorithms and show that the different cellular radio characteristics such as frequency, antenna height and propagation environment, make the problem of cell tower localization different from WiFi AP localization.

To perform the study, we obtained access to the actual locations of cell towers in the greater Los Angeles area and a wardriving trace that covers a total area of 1396.2 km^2 in the downtown, residential and rural areas of Los Angeles covered by 54 cell towers, each with multiple cell sectors. Our study of cell tower localization algorithms on this dataset lead to the following contributions: (a) We characterize the



(a) Strongest RSS (b) Weighted Centroid

Figure 1. The localization error CDF figures when localizing the observed cells using all the RSS readings in three areas.



(a) All the strongest RSS (b) Zoomed in strongest RSS

Figure 2. The strongest RSS to distance relationship for inside and outside cells in Residential area.

performance of the weighted centroid and the strongest RSS algorithms in the greater Los Angeles area and show that if used naively these algorithms have very large localization errors of more than 40km. (b) We show that one can hope to localize a cell tower accurately only if it falls within the area covered by the wardriving trace. (c) We propose a bounding technique to select the towers that fall within the area covered by the wardriving trace and study their performance. (d) Finally, we identify a cell combining optimization that can further reduce localization errors by half.

DATA DESCRIPTION

We obtained access to a wardriving trace that covers three areas in the greater Los Angeles area. The Downtown trace covers an area of 3.5km×4.2km in the downtown Los Angeles. The Residential trace covers an area of 6.3km×17km in the southern part of the Los Angeles County. The Rural trace covers an area of 35.4km×36km in the Victor Valley of San Bernardino County.

The wardriving trace was collected over a period of 2 months in February and March of 2009. The GSM signal strength measurements and their locations were recorded every 2 seconds and the speed of the car averaged about 32kmph. In total, we have 2,613,465 received signal strength (RSS) readings from 105,271 unique locations, resulting, on average, in 24.8 RSS readings from different cells per location. Each cell tower has 2, 3 or 6 cells attached to it, depending on the characteristics of the area and the coverage requirements. We know which cells belong to which cell tower and the actual location of each cell tower.

EVALUATION OF CELL LOCALIZATION ALGORITHMS

In this section, we describe two cell tower localization algorithms that have been used in the recent literature [12, 5, 6] and evaluate their performance on our wardriving trace.

Strongest RSS: Strongest RSS estimates a cell's location as the location of the measurement with the strongest observed RSS from that cell. This approach works well when a cell is located close to the road where wardriving measurements were collected, but often fails otherwise.

Weighted Centroid: The Centroid algorithm estimates the cell's location as the geometric center of the positions of the measurements where the cell was observed at. We used the Weighted Centroid method, in which the coordinates of each position are weighted by the signal strength observed at that position. The accuracy of the Weighted Centroid algorithm is sensitive to the density and homogeneity of measurements around the cell.

Evaluation: We tested the performance of both the Strongest RSS and Weighted Centroid algorithms on our wardriving traces in the three areas. Figure 1 plots the cumulative distribution function (CDF) of the difference between the estimated and the actual cell tower locations. Note that the error showed on the X-axis is in kilometers.

The figure shows that both Strongest RSS and Weighted Centroid perform very poorly. The median error in the Downtown area is 2.75km for Strongest RSS and 2.83km for Weighted Centroid. Interestingly, the Strongest RSS outperforms Weighted Centroid significantly in the rural area, achieving 0.7km median error vs. 7km, respectively. We believe this is because many of the cells in the rural area are located near roads and Strongest RSS seems to work well in this case.

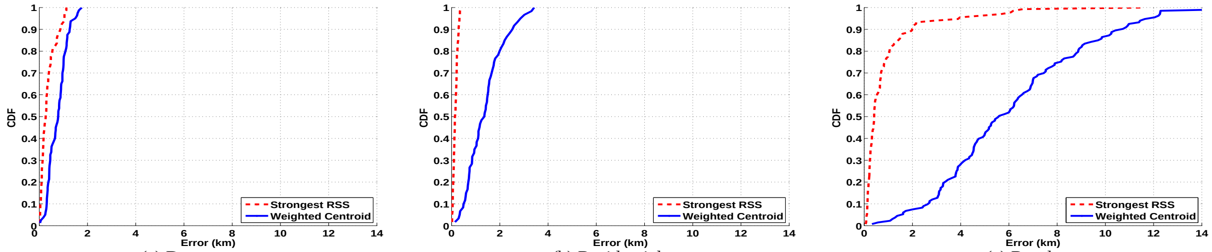
Our conclusion is that blindly applying these algorithms to estimate cell tower positions results in very large errors. In the next section, we investigate the causes for these large errors.

BOUNDING TECHNIQUE

To investigate what causes the large errors shown in Figure 1, we looked at the relationship between the strongest signal strength at which a cell could be heard and the localization error of the Strongest RSS algorithm for that cell. The results for the residential area are plotted in Figure 2. The figure shows that once the strongest observed RSS drops below roughly -60dBm, the localization error of the Strongest RSS algorithm increases significantly. Moreover, once the RSS drops below roughly -60dBm, the correlation between the strongest received signal strength and the distance to the cell also becomes weak, meaning that it is hard to predict the distance to the cell by relying just on the received signal strength.

Our conclusion is that both the Strongest RSS and the Weighted Centroid cannot estimate accurate locations for all the cells they observe because the estimated locations produced by both these algorithms are, by design, located within the wardriving area. Therefore, cells that are located outside the wardriving area cause large localization errors.

Figure 2 shows the cells that fall inside the Residential area



(a) Downtown (b) Residential (c) Rural

Figure 3. The localization CDF of different algorithms by using all the RSS readings in three areas.

as crosses (denoted as *inside cells*) and cells that fall outside the area as triangles (denoted as *outside cells*). The key observation is that the outside cells are the cause for large localization errors for the Strongest RSS algorithm and most of them have the strongest RSSs lower than -60dBm. Unfortunately, filtering cells using a simple threshold on the strongest RSS does not work well. First, taking a closer look at Figure 2 (b) reveals that some outside cells have the strongest RSS values that are higher than -60dBm, but still cause large errors for the Strongest RSS algorithm. Second, applying the threshold of -60dBm cuts out some of the inside cells as well.

The challenge is then to determine which cells are *inside cells* by just looking at the wardriving trace without having access to the actual locations of the cells. To address this challenge, we propose the *bounding* technique, which has three steps: *RSS Thresholding*, *Boundary Filtering*, and *Tower-based Regrouping*.

RSS Thresholding: Based on our findings presented in Figure 2, we filter out all cells whose strongest RSS is lower than a certain cutoff threshold. The purpose of applying this filtering is to eliminate as many outside cells as possible, while still keeping most of the inside cells. We use the threshold of -60dBm in this paper.

Boundary Filtering: The boundary filtering technique is based on the observation that the outside cells will have their strongest RSS values on the boundary or the perimeter of the wardriving area. This is because the nearest point from the outside cell to the wardriving area is the wardriving boundary. We tested this hypothesis by plotting the locations of the outside cells relative to the wardriving area and validated that this is indeed true.

Unfortunately, applying both the RSS Thresholding and the Boundary Filtering may still leave some of the outside cells in and filter some of the inside cells out due to RSS fluctuations. For instance, certain outside cells may have measurements of strongest RSS readings reside within the wardriving area and pass both the RSS Thresholding and the Boundary Filtering steps.

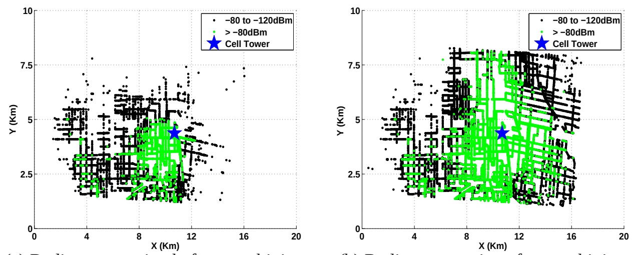
Tower-based Regrouping: The Tower-based Regrouping technique takes advantage of the relationship between a cell tower and its cells. We discovered two ways to identify cells belonging to the same cell tower. First, we found that clustering cells geographically based on their estimated positions can identify which cells belong to the same cell tower when the distance between cell towers is large. This technique works well in Residential and Rural areas, but often fails in the Downtown area due to the closer distance between cell towers. Second, we discovered that the prefix of a cell ID up to the last digit is the same for all cells belonging to the same cell tower, at least in the dataset that we studied. We used the latter technique in this paper.

The basic idea of Tower-based Regrouping is that although there may be some outside cells passing both RSS Thresholding and the Boundary Filtering steps, the other cells belonging to the same cell tower will most likely fail these two tests. Thus, if most of the cells belonging to the same cell tower are filtered out, the Tower-based Regrouping step filters out the outlier outside cell as well. Similarly, if most inside cells passed the RSS Thresholding and the Boundary Filtering steps, the outlier that was eliminated is added back. We found this technique to be very successful at determining inside cells.

Evaluation: We tested our bounding technique on the wardriving traces in the three areas and found that it eliminated all outside cells and kept all inside cells when the RSS threshold was between -67 and -58 dBm. In comparison, the basic RSS threshold technique using -60dBm, cuts out 20% of inside cells and leaves in 12% of outside cells in the downtown area. We also tested the performance of the Strongest RSS and the Weighted Centroid algorithms on the inside cells identified by the bounding technique. Figure 3 shows the CDF of the localization error for the Strongest RSS and Weighted Centroid algorithms in the Downtown, Residential and Rural areas. The results show that the performance of both algorithms has dramatically improved, with Strongest RSS significantly outperforming Weighted Centroid. The median error of the Strongest RSS algorithm in the Residential area is 139m vs. 1357m for the Weighted Centroid. In the next section, we show how the cell to cell tower relationship can help improve the localization accuracy even further.

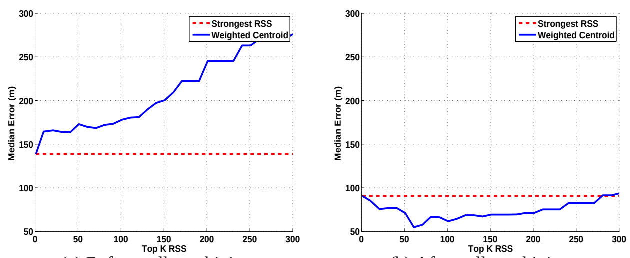
CELL COMBINING OPTIMIZATION

So far, we estimated the positions of cells belonging to the same cell tower independently from each other, even though they share the same physical location. In this Section, we show that merging wardriving traces of cells that share a common cell tower into a single trace and estimating the position of the cell tower itself can improve localization results significantly.



(a) Radio propagation before combining (b) Radio propagation after combining

Figure 4. Radio propagation of cells



(a) Before cell combining (b) After cell combining

Figure 5. The median error before and after cell combining in Residential area when choosing top K RSS readings.

During merging of traces, if a measurement contained readings from several cells belonging to the same cell tower we chose the reading with the strongest RSS. The effect of merging traces of cells belonging to the same cell tower into a single trace is illustrated in Figure 4. The figure shows that merging traces removes the measurement bias due to the directionality of a cell and results in a homogeneous coverage of the area around the cell tower. However, the improvements in the localization accuracy for the Strongest RSS and Weighted Centroid algorithms come from different sources. Strongest RSS performs better because in a merged trace the locations of all cells belonging to the same cell tower are estimated at the position where any one of the cells was heard the strongest. Weighted Centroid, on the other hand, performs better because the homogenous coverage of the area around the cell tower allows it to calculate the cell location by utilizing positions of measurements in all directions around the cell tower.

Figure 5 shows the median error of Strongest RSS and Weighted Centroid in the Residential area before and after applying the cell combining optimization on top of the bounding technique while varying the number of strongest K RSS readings used by the algorithms. We make three observations. First, before the cell combining, Strongest RSS always performs better than Weighted Centroid. This is caused by the non-homogeneity of measurements around the cell, which has a big adverse effect on Weighted Centroid. Second, the localization results after the cell combining significantly outperform those before the cell combining. Weighted Centroid benefits more from the integrated signal measurement map and achieves a very low medium error around 55m when $K = 60$. Third, there is a tradeoff between RSS quality and RSS quantity. Figure 5 (b) shows that the localization error of Weighted Centroid presents a decreasing trend first and then increases as K increases. The lowest medium error of 55m is achieved when $K = 60$.

Algorithm	Downtown		Residential		Rural	
	50th	90th	50th	90th	50th	90th
Strongest RSS (before)	267	877	139	291	414	1997
Strongest RSS (after)	163	399	91	194	229	711
Improvement	39%	55%	35%	33%	45%	64%
Weighted Centroid (before)	785	1263	1357	2521	5536	10730
Weighted Centroid (after)	152	427	55	161	167	698
Improvement	81%	66%	89%	92%	97%	93%

Table 1. The improvement on the median error and 90th percentile error for three areas. Note that the unit is in meters.

Finally, we present the localization accuracy improvement after applying the cell combining optimization in Table 1. In all three areas, we observed over 30% accuracy improvement in the median error and above 33% improvement in the 90th percentile error, indicating that the cell combining optimization can improve the cell localization accuracy significantly.

CONCLUSION

Accurately estimating locations of cell towers is important for many existing mobile phone localization algorithms. In this work, we conducted the first large scale study of the accuracy of the popular Strongest RSS and Weighted Centroid algorithms based on a large wardriving trace that covers downtown, residential and rural areas around greater Los Angeles. We showed that naively applying these algorithms results in very large localization errors. We analyzed the causes for these errors and concluded that one can hope to localize a cell accurately only if it falls within the area covered by the wardriving trace. We further proposed a bounding technique to select the cells that fall within the area covered by the wardriving trace and studied its performance. Finally, we identified a cell combining optimization that can further reduce localization errors by half.

Acknowledgements: This research was supported in part by NSF grant CNS-0954020.

REFERENCES

1. http://www.antennasearch.com/.
2. Mobiledia. http://www.cellreception.com/.
3. Opencellid. www.opencellid.org.
4. Skyhook Wireless. http://www.skyhookwireless.com/.
5. M. Y. Chen et al. Practical metropolitan-scale positioning for gsm phones. In *Ubicomp*, pages 225–242, 2006.
6. Y.-C. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm. Accuracy characterization for metropolitan-scale Wi-Fi localization. In *ACM Mobisys*, pages 233–245, 2005.
7. J. Hightower and G. Borriello. Particle filters for location estimation in ubiquitous computing computing: A case study". In *Ubicomp*, pages 88–106, 2004.
8. M. Kim et al. Risks of using AP locations discovered through war driving. In *Pervasive*, pages 67–82, 2006.
9. A. LaMarca et al. Place lab: Device positioning using radio beacons in the wild. In *Pervasive*, pages 116–133, 2005.
10. K. Langendoen and N. Reijers. Distributed localization in wireless sensor networks: a quantitative comparison. *Comput. Networks*, 43(4):499–518, 2003.
11. D. Madigan, E. Einahrawy, R. Martin, W.-H. Ju, P. Krishnan, and A. Krishnakumar. Bayesian indoor positioning systems. In *IEEE INFOCOM*, pages 324–331, 2005.
12. A. Varshavsky, D. Pankratov, J. Krumm, and E. Lara. Calibree: Calibration-free localization using relative distance estimations. In *Pervasive*, pages 146–161, 2008.