Lookup Lateration: Mapping of Received Signal Strength to Position for Geo-Localization in Outdoor Urban Areas

Andrey Shestakov[1], Danila Doroshin[2], Dmitri Shmelkin[2], and Attila Kertész-Farkas[1(✉)]

[1] School of Data Analysis and Artificial Intelligence, Faculty of Computer Science, National Research University Higher School of Economics (HSE), Moscow, Russia
akerteszfarkas@hse.ru
[2] Huawei Technologies Co. Ltd., Moscow, Russia

Abstract. The accurate geo-localization of mobile devices based upon received signal strength (RSS) in an urban area is hindered by obstacles in the signal propagation path. Current localization methods have their own advantages and drawbacks. Triangular lateration (TL) is fast and scalable but employs a monotone RSS-to-distance transformation that unfortunately assumes mobile devices are on the line of sight. Radio frequency fingerprinting (RFP) methods employ a reference database, which ensures accurate localization but unfortunately hinders scalability.

Here, we propose a new, simple, and robust method called lookup lateration (LL), which incorporates the advantages of TL and RFP without their drawbacks. Like RFP, LL employs a dataset of reference locations but stores them in separate lookup tables with respect to RSS and antenna towers. A query observation is localized by identifying common locations in only associating lookup tables. Due to this decentralization, LL is two orders of magnitude faster than RFP, making it particularly scalable for large cities. Moreover, we show that analytically and experimentally, LL achieves higher localization accuracy than RFP as well. For instance, using grid size 20 m, LL achieves 9.11 m and 55.66 m, while RFP achieves 72.50 m and 242.19 m localization errors at 67% and 95%, respectively, on the Urban Hannover Scenario dataset.

Keywords: Non-line-of-sight
Outdoor mobile device geo-localization · Fingerprinting

1 Introduction and Literature Review

The localization of wireless mobile devices [8,20] has become a key issue in emergency cases [1,6], surveillance, security, family tracking, radio network planning and optimization, et cetera. GPS-based localization, while very accurate, is not

preferred because of its high energy consumption; thus, a wide range of technologies has emerged for localization based on received signal strength (RSS), time of arrival (ToA), and angle of arrival (AoA). The two latter methods, ToA and AoA, require additional equipments, such as an antenna array to time synchronization between the transmitter and the receiver, while AoA requires some array to identify the signal's angle. However, RSS measurements are ubiquitous and readily available in almost all wireless communication systems; hence, it seems plausible to gain information about mobile devices' position.

The main problem is characterized as follows: We assume that we are given a set of antennas $A = \{a_i = (a_{i_x}, a_{i_y}) \in R^2\}$ and several RSS measurements for a single mobile phone: $M_{RSS} = [r_{a_1}, r_{a_2}, ..., r_{a_{|A|}}]$, where r_{a_i} denotes RSS from antenna a_i, measured in dBm if it is observed; NaN otherwise. Antenna a_i and its location will be used interchangeably in this article. The main task is to determine the mobile phone's location, where its RSS measurements were observed. A straightforward method to carry out a simple localization is the triangular lateration (TL) that calculates the mobile phone's location by solving the following optimization problem:

$$\min_{(x,y)} \sum_{i:r_{a_i} \in M_{RSS}} \left((a_{i_x} - x)^2 + (a_{i_y} - y)^2 - (d(r_{a_i}))^2\right)^2, \tag{1}$$

where $d(.)$ denotes a distance function. For monotone functions $d(.)$ and for at least three measurements, the optimization problem becomes convex and avoids the local minima phenomenon. However, there are two main issues with approach. First, it requires a well-calibrated signal-to-distance conversion function $d(.)$. For urban areas, the COST-HATA models were developed by the COST European Union Forum based on various field experiments and research [13]. Second, it assumes mobile devices on line of sight (LoS), which is hardly met in practice in urban areas because signal propagation is hindered by obstacles, concrete constructions, churches, tunnels, underpasses, etc. In this case, the observed distance can be decomposed as $d(r_{a_i}) = L_i + e_i + NLOS_i$, where L_i is the true distance and e_i is the receiver noise assumed to be a zero mean Gaussian random variable. Note that, $NLOS_i \gg e_i$ according to field test results.

The quantity $NLOS_i$ denotes the error caused by obstacles on the signal propagation path. These obstacles interfere with the signal by either weakening it or changing its path a.k.a multipath propagation. In either case, the antenna receives weaker signals indicating mobile phones farther away than they are as illustrated in Fig. 1A. The NLOS error can be modeled by a probabilistic distributions [5] or can be considered a deterministic error [10,12]. These methods show promising results when few obstacles can be found on the landscape, though how these methods will work in a dense urban area remains to be seen.

To characterize the NLoS error, we analyzed the Urban Hannover Scenario dataset [15], which contains approximately 22 Gb of localized RSS measurements from Hannover. For further details, see results section or [15]. In our analysis we first selected tower ID = 183 randomly and identified mobile phones located at the fourth degree from the tower's azimuth. The data are shown in Fig. 1C, where
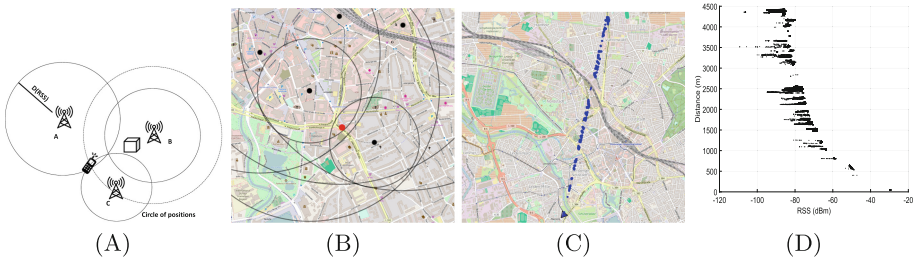
Fig. 1. (A) Triangular lateration hindered by obstacles. D(RSS) denotes the distance determined by RSS. The dashed circle denotes the true distance of a mobile phone from tower B, which appears farther because of the obstacle (represented by a box) between them. (B) Lateration in Hannover dataset using the six strongest RSSs. The red dot denotes the mobile device's true location, the black dots denote antennas, and the circles denote the circle of positions. Data (User ID = 10048 and Time = 0.1) is from [15]. (C) Location of mobiles phones (marked blue dots) from antenna ID = 183 (marked blue triangle at the bottom) at 4th degree to its azimuth (clockwise). (D) Distance vs. RSS for phones from panel (C). (Color figure online)

the blue triangle denotes the tower's location and where the blue dots indicate the mobile phones' locations. The scatter plot of the measured RSSs versus the true distance is shown in Fig. 1D. The relationship indicates an exponential-like correlation. However, the weaker the signal is, the noisier the correlation becomes. As signal strength decreases, the conversion becomes more ambiguous. For instance, an observed −90 dBm signal could indicate a mobile phone located somewhere between 2.4 km and 4.5 km. It is also worth noticing that the phones are not distributed evenly in this range; rather, they are concentrated at certain distances. For instance, no phone is located between 2.5 km and 3.0 km in this direction. This gap can be seen at the railways shown in Fig. 1C.

The radio frequency fingerprinting (RFP) [2,7] approach is designed to overcome the NLoS problem. This is a two-phase algorithm. The first step, called offline training phase or surveying, is to construct a reference dataset that consists of a collection of localized RSS measurement vectors $M_{XY} = [r_{a_1}, r_{a_2}, \ldots, r_{a_{|A|}}]$ all over the area, where x, y denotes its true position. In the second phase, localization phase, the query device's position producing M_{RSS} is estimated based on the nearest dataset member via a k-Nearest Neighbor (kNN) approach. In other words, the location x, y is determined by solving the following search problem:

$$\tilde{x}, \tilde{y} = \underset{x,y}{\operatorname{argmin}} \, d(M_{RSS}, M_{XY}), \qquad (2)$$

where the distance function $d(.,.)$ can be, say, an Euclidean distance or cosine distance [9], etc., which simply omit NaN values. One drawbacks of RFP is the lack of scalability, which means that it can be slow over large areas.

Grid systems are often utilized to mitigate data redundancy, making localization faster with less memory consumption in which measurements replaced by their grid centers. Therefore, RFP can be formulated as follows: localize query

M_{RSS} in the center of grid $\tilde{x}y$ calculated by $\tilde{x}y = \mathrm{argmin}_{xy \in \mathcal{G}} \; d(M_{RSS}, \mu_{xy})$, where μ_{xy} is the sample mean observed in grid xy, $\mathcal{G}$ denotes the set of grids and $d(.,.)$ is a suitable distance function. This reduces search time and the size of the reference datasets by a couple of magnitudes at the expense of localization accuracy, and the trade-off can be controlled with the grid's size. To make localization even faster, Campos et al. [3] have developed filtering procedures to reduce the search space.

Besides RFP, which is based on a k-NN, several classic machine learning methods have been evaluated and tested on localization problems. For instance, Wu et al. [18] have tried support vector machines, Nuño-Barrau et al. [14] have used linear discriminant analysis, and Campus et al. and Magro et al. [4,11] have used genetic algorithms for positioning measurements. Artificial neural networks are also popular; they can learn a regression function to map a M_{RSS} to its corresponding location $\tilde{x}, \tilde{y}$ [16,17,19]. No matter which methods are used, they all need to learn highly nonlinear mapping like in Fig. 1D to provide a bona fide approximation. In our opinion, this seems quite challenging in practice.

In this paper, we present a new method called lookup lateration (LL) for the rapid and accurate geo-localization of mobile phones based on RSS. The underlying idea is based on the decentralization of the reference dataset. For every antenna tower $a_i \in A$, LL builds a lookup table to store associating reference locations, that is, $\tau_{a_i}(r_{a_i}) = \{(XY) \mid r_{a_i} \in M_{XY}\}$. Therefore, the localization of a given query M_{RSS} can be carried out by simply determining the common locations in the associating lookup tables as follows: $\tilde{x}, \tilde{y} = \bigcap_{r_{a_i} \in M_{RSS}} \tau_{a_i}(r_{a_i})$. The main advantages of LL compared to RFP is that LL does not need to search the whole reference dataset, only two-six tables associated to query measurements. This results in a great acceleration; LL is around 100 times faster than RFP on the Urban Hannover Scenario dataset, and we believe LL can be even faster in very big cities.

LL resembles TL to some extent. Every lookup table can be considered an RSS-to-distance nonlinear mapping $d(.,.)$ (more precisely, a relation) w.r.t. a given antenna tower. However, instead of solving a nonconvex optimization problem, LL carries out the localization by determining common elements in the lookup tables. Therefore, LL does not involve local minima problems, but it may result in multiple locations, which need to be addressed.

2   Lateration Using RSS-to-Location Lookup Tables

In the previous section, we concluded that the relationship between received RSS and true distance is nonmonotone in dense urban areas. Using any nonmonotone, continuous function $d(.)$ in Eq. 1 would result in a nonconvex optimization problem. Here, we introduce a new procedure that we termed lookup lateration (LL).

The procedure is provided a collection of M_{XY} measurements annotated with their true locations as training data. Then lookup tables $\tau_a(r)$ contain mobile locations with respect to RSS r measured by antenna a. This procedure is

shown by Algorithm 1. To avoid redundant locations in lookup tables and reduce their size, nearby locations can be grouped using clustering algorithms, and the center of clusters can be stored in lookup tables. Because RSS measurements r are real valued and hindered by some measurement noise, we simply applied binning techniques to group measurements together. The corresponding bin b of a measurement r using bin size s is calculated $b = \lfloor r/s \rfloor$. In most of our experiments, we simply used bin $s = 1$; thus, all r were rounded down to the closest integer. For instance, -64.7 dBm is rounded down to -65. In the rest of the paper, $s = 1$, unless it is specified otherwise.

Figure 2A shows an example, where data were taken from the Urban Hannover Scenario and green dots mark the locations stored in $\tau_{88}(-50)$, while blue triangle (bottom) indicates the tower location (ID $= 88$).

Algorithm 1. Construction of lookup tables. The input is a list of localized RSS measurement $M = \{(x_i, y_i, r_{a_i})\}_{i=1}^N$. Triplet (x_i, y_i, r_{a_i}) denotes location x_i, y_i, where RSS r_{a_i} is measured from tower a_i. D is the maximal allowed diameter of a cluster.

```
1: procedure CONSTRUCTLOOKUPTABLES(M,D)
2:     for all antenna tower a do
3:         for all RSS r do
4:             S ← {(x_i, y_i) | r_{a_i} = r, a_i = a}
5:             τ_a(r) ← Clustering(S, D)    ▷
       Group mobiles nearby.
6:         end for
7:     end for
8:     return τ
9: end procedure
```

Algorithm 2. Lookup lateration. The input $M_{RSS} = \{r_{a_1}, r_{a_2}, \ldots, r_{a_n}\}$ is a list of RSS measurements in decreasing order. Returns a location estimation for M_{RSS}.

```
1:  procedure LOOKUPLATERATION(M_RSS)
2:      Remove r_{a_i} ∈ M_RSS from M_RSS if
        τ_{a_i}(r_{a_i}) is empty
3:      return some default location if M = ∅
4:      k ← 1
5:      C^k ← τ_{a_k}(r_{a_k})
6:      for k = 1 … |M_RSS| do
7:          break if C^k is ambiguous
8:          k ← k + 1
9:          C^k ← {c_i | c_i ∈ C^{k-1}, c_j ∈
        τ_{a_k}(r_{a_k}), d(c_i, c_j) ≤ T}
10:     end for
11:     C^k ← C^{k-1} if C^k is empty
12:     return mean of locations in C^k for query
        M_RSS
13: end procedure
```

Now the next step is to determine the location of a given measurement $M_{RSS} = [r_{a_1}, r_{a_2}, \ldots, r_{a_{|A|}}]$. Here, the principle is that it can be carried out by determining the common locations in the corresponding tables $\tau_{a_i}(r_{a_i})$, that is, M_{RSS} is annotated by the location $\bigcap_i \tau_{a_i}(r_{a_i})$ for $r_{a_i} \neq NaN$ and $\tau_{a_i}(r_{a_i}) \neq \emptyset$. In practice, this could lead easily to an empty set or unambiguous locations. Our method, shown in Algorithm 2, is slightly different as, in a greedy manner, it takes into account that stronger signals provide more reliable information. Let $M = \{r_{a_{i_1}}, r_{a_{i_2}}, \ldots, r_{a_{i_n}}\}$ be a list of the observed RSS from M_{RSS} in decreasing order. Our algorithm starts with the set of candidate locations $C^1 = \tau_{a_1}(r_{a_1})$ provided by the strongest signal. In subsequent iterations, in the while loop at line 6, candidate location $c_i \in C^1$ is eliminated from C^1 if it does not appear as a candidate location from another antenna a_k $(k = i_2, \ldots, i_n)$ within a tolerance T. This iteration terminates if either all measurements are processed or the candidate locations in C^k have a smaller variance than a predefined threshold. The iteration also terminates when C^k is emptied. Figure 2 shows an example of how this algorithm works.
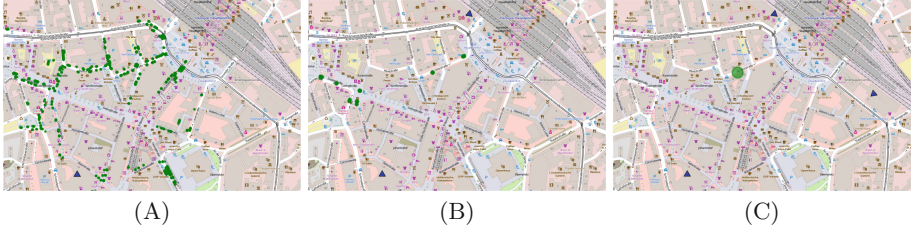
(A)   (B)   (C)

Fig. 2. Illustration of lookup lateration for query $M_{RSS} = \{88{:}{-}50;\ 27{:}{-}55;\ 135{:}{-}55\}$. Candidate locations are marked by green dots, and three antennas (ID $= 88$ [bottom], ID $= 27$ [top], ID $= 135$ [right hand side]) are marked by blue triangles. (A) Initial locations in $C^1 = \tau_{88}(-50)$ at the beginning of Algorithm 2. (B) Next, locations not present in $\tau_{27}(-55)$ are removed, and the resulting locations in C^2 are shown. (C) Points not present in $\tau_{135}(-55)$ are also removed, yielding an ambiguous location estimation in C^3 for M_{RSS}. In this example, the lookup lateration stops after three iterations. (Color figure online)

3 Error Estimation in Grid Systems

The LL method can also be used with grid systems, which can also lead to further simplifications of the algorithm. In our experiments, Algorithm 1 simply stored unique grid references in $\tau_a(r)$. Thus, clustering algorithms were not needed. Moreover, filtering condition $(d(c_i, c_j) < T)$ in line 9 of Algorithm 2 was replaced with a Kronecker delta $(\delta_{c_i, c_j} \overset{?}{=} 1)$[1], which is more quickly evaluated.

Now we can compare the error obtained by fingerprinting and lookup lateration methods under two assumptions: (1) completeness and (2) unambiguousness. By completeness, we assume that there are enough data observed in each grid. This ensures that error is not induced by data sparsity, poor design, or the localization is not carried out in, say, Paris if site survey was taken in New York, etc. By unambiguousness we assume that identical observations cannot be found in different grids. This assumption ensures that any observation can be determined unambiguously. Now we can claim the following:

Theorem 1. *Let $\mathbb{E}[\mathcal{E}_{LL}]$ and $\mathbb{E}[\mathcal{E}_{RFP}]$ be the expected error obtained with LL and RF, respectively. Under the conditions mentioned above, $\mathbb{E}[\mathcal{E}_{LL}] \leq \mathbb{E}[\mathcal{E}_{RFP}]$.*

Proof. First, let us consider the fingerprinting method. Let $\mathcal{B} \subset R^{|A|}$ be the measurement space, $\mathcal{G}$ be the grid system, P_{xy} be the density distribution of measurements belonging to grid $xy \in \mathcal{G}$, and μ_{xy} be the mean vector of P_{xy}. The distance function d in RFP localization implicitly specifies a Voronoi partition of $\mathcal{B}$: $\{\mathcal{R}_{xy}\}_{xy \in \mathcal{G}}$, where

$$\mathcal{R}_{xy} = \{r \in \mathcal{B} \mid d(r, \mu_{xy}) \leq d(r, \mu_{uv})\ \forall uv \neq xy \in \mathcal{G}\}. \tag{3}$$

---

[1] Defined by Kronecker delta $(\delta_{a,b} = 1 \iff a = b$, otherwise 0).

Second, let $M_{RSS} \in \mathcal{B}$ denote a query measurement observed at position M_{xy}, which is in grid $xy \in \mathcal{G}$.

If $M_{RSS} \in \mathcal{R}_{xy}$ and $M_{RSS} \notin \mathcal{R}_{uv}$ for any other $uv \neq xy \in \mathcal{G}$, then RFP localizes M_{RSS} in the center of grid xy. The localization error is $\epsilon_1 = d(M_{xy}, c_{xy})$, that is, the distance between M_{xy} and the center of the grid c_{xy}. Let $\mathcal{E}_1$ be this type of error and let $\mathbb{E}_{xy}[\epsilon_1]$ be the expected value of this error type in grid xy.

If $M_{RSS} \notin \mathcal{R}_{xy}$, then $\exists uv \in \mathcal{G}$ in which RFP localizes M_{RSS}. The error is $\epsilon_2 = d(M_{xy}, c_{uv})$. Let $\mathbb{E}_{xy}[\epsilon_2]$ denote the expected error of this type with respect to grid xy. Let $\mathcal{E}_2$ denote this type of error. Note that error ϵ_2 is always greater than error ϵ_1, which implies that

$$\mathbb{E}_{xy}[\epsilon_1] < \mathbb{E}_{xy}[\epsilon_2]. \tag{4}$$

If $M_{RSS} \in \mathcal{R}_{xy}$ and $M_{RSS} \in \mathcal{R}_{uv}$ for $uv \neq xy \in \mathcal{G}$, then that means either M_{RSS} is on the border of two regions $\mathcal{R}_{xy}$ and $\mathcal{R}_{uv}$ or $\mu_{xy} = \mu_{uv}$. In the first case, RFP has to choose between the two grids. Let us give some advantage to RFP and assume for the sake of simplicity, it always chooses the correct grid somehow. In the second case, we have $\mathcal{R}_{xy} = \mathcal{R}_{uv}$. In our opinion this case happens rarely in practice, so we omit this type of error, and we assume that $\mu_{ab} \neq \mu_{cd}$ for any $ab, cd \in \mathcal{G}$ in the rest of this proof.

The probability that RFP localizes M_{RSS} in its correct grid is

$$p_{xy}(\mathcal{E}_1) = \int_{\mathcal{R}_{xy}} P_{xy}(r_{a_1}, r_{a_2}, \ldots, r_{a_{|A|}}) \, dr_{a_1} \ldots dr_{a_{|A|}}, \tag{5}$$

and it is indicated by the white area under P_{xy} in Fig. 3. The probability that RFP localizes M_{RSS} in a grid incorrectly is

$$p_{xy}(\mathcal{E}_2) = \int_{\mathcal{B} \backslash \mathcal{R}_{xy}} P_{xy}(r_{a_1}, \ldots, r_{a_{|A|}}) \, dr_{a_1} \ldots dr_{a_{|A|}} = 1 - p_{xy}(\mathcal{E}_1). \tag{6}$$

and it is indicated by the gray area under P_{xy} in Fig. 3.

Last, the total expected error for the fingerprinting method can be summarized as follows:

$$\mathbb{E}[\mathcal{E}_{RFP}] = \sum_{xy \in \mathcal{G}} p(xy) \left[p_{xy}(\mathcal{E}_1) \mathbb{E}_{xy}[\epsilon_1] + p_{xy}(\mathcal{E}_2) \mathbb{E}_{xy}[\epsilon_2] \right], \tag{7}$$

where $p(xy)$ denotes a priori probability of receiving an RSS from grid xy.

Now we examine LL, and let us consider $M_{RSS} = [r_{a_1}, r_{a_2}, ..., r_{a_{|A|}}]$ measured in grid xy. Because of the first and second assumptions, we have $xy \in \tau_{a_i}(r_{a_i})$ and $\bigcap_i \tau_{a_i}(r_{a_i}) = \{xy\}$, respectively. This means when Algorithm 2 terminates, the set C^k unambiguously contains only one candidate grid, where the measurement was observed. Hence, the total expected error for lookup lateration can be summarized as follows:

$$\mathbb{E}[\mathcal{E}_{LL}] = \sum_{xy \in \mathcal{G}} p(xy) \mathbb{E}[\mathcal{E}_1], \tag{8}$$

Following from Eq. 4 we obtain $\mathbb{E}[\mathcal{E}_{LL}] \leq \mathbb{E}[\mathcal{E}_{RFP}]$, which proves our claim. ∎

To illustrate the proof of Theorem 1, let us consider a scenario in which there is one antenna tower A and two nearby grids G_{xy}, G_{uv}. Furthermore, let us assume that A receives $-53, -55, -57, -59$, and -61 dBm from G_{xy} and $-58, -60, -62, -64$, and -66 dBm from G_{uv}. The mean values are $\mu_{xy} = -57$ and $\mu_{uv} = -62$ and $\mathcal{R}_{xy} = \{x \mid x \geq -59.5\}$ and $\mathcal{R}_{uv} = \{x \mid x \leq -59.5\}$. RFP locates measurement $M = -61$ dBm in grid G_{uv} incorrectly because $M \in \mathcal{R}_{uv}$, that is, M closer to μ_{uv} than to μ_{xy}. However, LL constructs a table τ_A in which $\tau_A(-66) = \tau_A(-64) = \tau_A(-62) = \tau_A(-60) = \tau_A(-58) = \{G_{uv}\}$ and $\tau_A(-61) = \tau_A(-59) = \tau_A(-57) = \tau_A(-55) = \tau_A(-53) = \{G_{xy}\}$. Therefore, LL will find $G_{xy} \in \tau_A(-61)$ and localize M in it correctly.
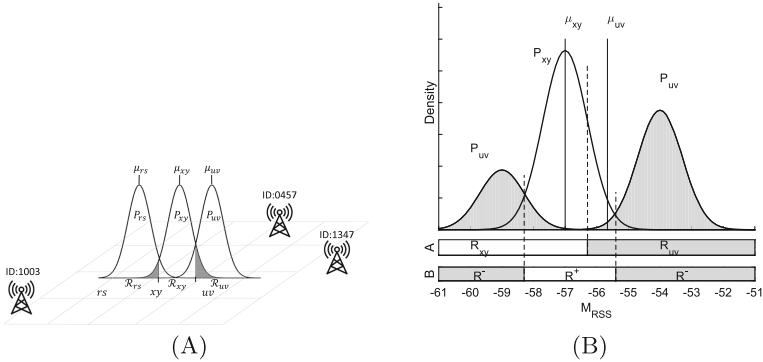

(A)   (B)

Fig. 3. (A) Distribution of measurements for three grids rs, xy, and uv in fingerprinting scenario. Decision boundaries between $\mathcal{R}_{rs}, \mathcal{R}_{xy}$, and $\mathcal{R}_{uv}$ are halfway between distribution means. (B) Decision boundaries in case of multi-modal density distributions. The decision boundary defined with RFP is located between μ_{xy}, μ_{uv} and the corresponding regions $\mathcal{R}_{xy}, \mathcal{R}_{xy}$ denoted on horizontal bar A. Decision boundaries defined by the NPL non-continuous and are shown on bar B.

One may argue that Theorem 1 is based on two conditions that are hard to ensure in practice. We may agree, but in our opinion, Theorem 1 shows the conceptual limitation of the fingerprinting method. However, if we easy up on the conditions and if the unambiguousness is not required, then we claim

Theorem 2. *Fingerprinting method is suboptimal in the sense of the Neyman-Pearson criterion.*

Proof. Localization using grid systems can be considered a classification problem in which grids are considered classes, and a localization method has to decide whether to classify a query M_{RSS} in a given class xy. Let $P^+(M_{RSS}) = P_{xy}(M_{RSS})$ and $P^-(M_{RSS}) = \frac{1}{Z} \sum_{uv \neq xy} P_{uv}(M_{RSS})$ be likelihood functions, where Z is an appropriate normalization factor. According to the Neyman-Pearson Lemma (NPL), the highest sensitivity can be achieved with the following likelihood-ratio rule:

$$\text{Localize } M_{RSS} \text{ in grid } xy \text{ if } \quad \frac{P^+(M_{RSS})}{P^-(M_{RSS})} \geq \eta, \tag{9}$$

where η is a trade-off parameter among false positive, false negative error, and statistical power. Let $\eta = 1$ for sake of simplicity. The region $\mathcal{R}^+ = \{M_{RSS} \mid P^+(M_{RSS}) \geq P^-(M_{RSS})\}$ in which a query $M_{RSS} \in \mathcal{R}^+$ is localized in grid xy and defined by the likelihood ratio can be noncontinuous region. However, Fig. 3B illustrates that the region $\mathcal{R}_{xy}$ defined by RFP in Eq. 3 is always continuous and different from $\mathcal{R}^+$. Therefore, RFP yields less or equal sensitivity than it could be achieved with NPL, where equality holds iff all distribution belonging to grids are symmetric and unimodal. ∎

In other words, the weakness of RFP arises from the fact that RFP does not take into account the full distribution of measurements, it uses only the mean vectors of the distributions, which can be disadvantageous for multimodal distributions. Contrary, NPL utilizes the whole distribution.

On the other hand, LL will identify all candidate grids in which the query can be found. Therefore, the set of candidate grids $C^{(k)}$ will contain the correct grid as well. If LL was programmed to localize a query by the center of the grid $\tilde{x}\tilde{y}$ for which $\tilde{x}\tilde{y} = \mathrm{argmax}_{xy \in C^{(k)}}\{P_{xy}(M_{RSS})\}$, then LL would be optimal in the sense of NPL. However, we decided to report the mean of the candidate grids. In this case, $\mathcal{E}_2$ happens, but we hope averaging the candidate locations would mitigate the amount of error ϵ_2. This also does not require us to store or model full measurement distributions.

4   Results and Discussion

We have carried out our experiments on the Urban Hannover Scenario dataset [15]. This dataset contains approximately 22 Gb of RSS measurements simulated in Downtown Hannover, along with a reference x-y location. The reference point (0-0) for the coordinate system is the lower left corner of the scenario. The data is the result of a prediction with a calibrated ray tracer using 2.5D building information. For each mobile phone, the 20 strongest RSS measurements are provided. For further details, we refer to [15]. Data was split into training and test sets randomly, and experiments were repeated ten times. Results were averaged. The variance in results was very small because of the dataset's huge size; therefore, standard deviation is not shown for the sake of simplicity. Algorithms were implemented in Python programming language and executed on a PC equipped with a 3.4 GHz CPU and a 16 GB RAM.

First, we investigated the localization accuracy of RFP and LL methods using grid sizes 5, 10, 20, and 50 m. The full dataset was split to 10% training and 90% test data. The results shown in Fig. 4 clearly tell us LL and RFP perform nearly the same when the grid size is small (Figs. 4A–B). The performance of LL remains roughly the same as the size of the grid grows (Figs. 4C–E), while the performance of RFP decreases quickly. In our opinion the dramatic drop in the performance of RFP is in accordance with Theorems 1 and 2. Large grids cover
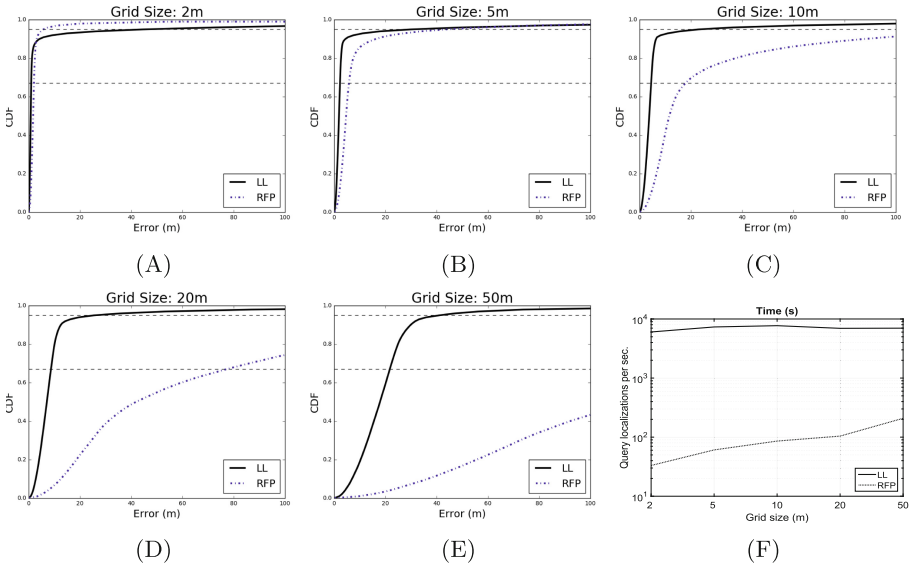
Fig. 4. Performance of RFP and LL using various grid sizes. Dashed lines corresponds to errors at 67% and 95%.

a large area containing various obstacles on the ray propagation path, which result in wide, multimodal measurement distributions w.r.t. grids. This is not taken into account by RFP. Next, we analyzed the speed of these methods, and the results are summarized in Fig. 4F. LL seems to be extremely fast, around two magnitudes faster compared to the RFP method. This improvement stems from the fact that LL processes only a few lookup tables related to a query. On the other hand, RFP iteratively processes all grids while seeking the most similar RSS vector pattern. It is also surprising, the execution time of LL does not depend on the grid size, while RFP quadratically becomes slower as the grid size decreases.

Next, we investigated how RFP and LL methods perform with improper site surveying. We calculated a site coverage defined as the ratio of grids that contain training data. For instance, 35% of coverage means that the training data belong to 35% of the total grids, and all queries would be localized in one of these grids. The coverage is driven by two factors: (i) grid size and (ii) the size of the training data. First, we took 5% of the training data and calculated the coverage and the localization error at 67% and 95% using various grid sizes. The results shown in Table 1 tell us a larger grid size results in larger coverage, and both methods yield a larger localization error at 67%. However, if we take a closer look and calculate a relative error (RE) as the ratio of the error and the grid size, we can observe opposite tendencies. The RE obtained with LL decreases as the grid size grows, and we think this is the result of increased coverage. However, the RE obtained with RFP increases in spite of increased coverage, and we explain

this by the arguments in Theorems 1 and 2. On the other hand, the RE and the overall error show opposite tendencies at 95%. LL decreases while RFP increases the overall error and RE as the sizes of grids and coverage grow.

Table 1. Localization Error (m) and Coverage with various grid sizes

Grid size (m)		1		5		10		20		50	
Cov. (%)		26.38		54.42		66.74		71.78		87.20	
		Error	RE[a]	Error	RE	Error	RE	Error	RE	Error	RE
67%	LL	0.77	*0.77*	2.44	*0.49*	4.74	*0.47*	9.11	*0.46*	22.24	*0.44*
	RFP	1.09	*1.09*	5.76	*1.15*	16.36	*1.64*	72.50	*3.63*	157.17	*3.14*
95%	LL	125.98	*125.98*	89.51	*17.9*	69.37	*6.94*	55.66	*2.78*	49.88	*0.998*
	RFP	6.32	*6.32*	45.05	*9.01*	138.62	*13.68*	242.19	*12.11*	343.94	*6.88*

The training set size was 5%. [a] Relative error (RE) is defined as the ratio of error and grid size.

5 Conclusions

In this article, we have presented a new method called lookup lateration for localization problems in densely populated urban areas using received signal strength (RSS) data. Our method combines the advantages from both triangular lateration and fingerprinting.

Lookup lateration outperforms triangular lateration and fingerprinting-based methods in localization accuracy as well. Triangular lateration does not take into account NLOS objects; hence, its performance in urban areas is always limited. We also showed conceptual limitations of the fingerprinting method in Theorems 1 and 2. Briefly, the problem is that RSS observations are aggregated in the same grid position, and their variances are not taken into account. As a consequence, fingerprinting is prone to annotate observations with incorrect grid positions. This situation cannot happen with lookup lateration because it stores all grid positions for any observed RSS values separately. This fact makes LL very robust to larger grid sizes.

Finally, it is very easy to migrate from RFP to LL. Since LL does not need any additional data compared to RFP, lookup tables can be constructed from the site surveying data of RFP. Lookup tables can also be an intelligent reorganization of fingerprinting data, which preserves measurement variance implicitly. Our method is very simple, easy to implement in distributed systems, and inexpensive to maintain, which, we hope, can make it appealing for large-scale industrial applications.

References

1. Directive 2002/58/EC on privacy and electronic communications
2. Bahl, P., Padmanabhan, V.N.: Radar: an in-building RF-based user location and tracking system. In: Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2000, vol. 2, pp. 775–784. IEEE (2000)
3. Campos, R.S., Lovisolo, L.: A fast database correlation algorithm for localization of wireless network mobile nodes using coverage prediction and round trip delay. In: VTC Spring. IEEE (2009)
4. Campos, R.S., Lovisolo, L.: Mobile station location using genetic algorithm optimized radio frequency fingerprinting. In: Proceedings of ITS 2010 International Telecommunications Symposium (2010)
5. Cong, L., Zhuang, W.: Nonline-of-sight error mitigation in mobile location. IEEE Trans. Wirel. Commun. **4**(2), 560–573 (2005)
6. FCC: Revision of the commission rule to ensure compatibility with enhanced 911 emergency calling system. Technical report RM-8143, Federal Communications Commission (FCC), Washington, DC (2015)
7. Gentile, C., Alsindi, N., Raulefs, R., Teolis, C.: Geolocation Techniques: Principles and Applications. Springer, Heidelberg (2012)
8. Gezici, S.: A survey on wireless position estimation. Wirel. Pers. Commun. **44**(3), 263–282 (2008)
9. He, S., Chan, S.H.G.: Tilejunction: mitigating signal noise for fingerprint-based indoor localization. IEEE Trans. Mobile Comput. **15**(6), 1554–1568 (2016)
10. Ma, C., Klukas, R., Lachapelle, G.: A nonline-of-sight error-mitigation method for TOA measurements. IEEE Trans. Veh. Technol. **56**(2), 641–651 (2007)
11. Magro, M.J., Debono, C.J.: A genetic algorithm approach to user location estimation in UMTS networks. In: The International Conference on "Computer as a Tool", EUROCON 2007, pp. 1136–1139. IEEE (2007)
12. Marco, A., Casas, R., Asensio, A., Coarasa, V., Blasco, R., Ibarz, A.: Least median of squares for non-line-of-sight error mitigation in GSM localization. In: 2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–5, September 2008
13. Neskovic, A., Neskovic, N., Paunovic, G.: Modern approaches in modeling of mobile radio systems propagation environment. IEEE Commun. Surv. Tutorials **3**(3), 2–12 (2000)
14. Nuño-Barrau, G., Páz-Borrallo, J.M.: A new location estimation system for wireless networks based on linear discriminant functions and hidden Markov models. EURASIP J. Appl. Signal Process. **2006**, 159–159 (2006)
15. Rose, D.M., Jansen, T., Werthmann, T., Türke, U., Kürner, T.: The IC 1004 urban hannover scenario - 3D pathloss predictions and realistic traffic and mobility patterns (2013)
16. Takenga, C., Xi, C., Kyamakya, K.: A hybrid neural network-data base correlation positioning in GSM network. In: 2006 10th IEEE Singapore International Conference on Communication Systems, pp. 1–5, October 2006
17. Takenga, C.M., Kyamakya, K.: Location fingerprinting in GSM network and impact of data pre-processing (2006)
18. Wu, C.L., Fu, L.C., Lian, F.L.: WLAN location determination in e-home via support vector classification. In: 2004 IEEE International Conference on Networking, Sensing and Control, vol. 2, pp. 1026–1031. IEEE (2004)

19. Yu, L., Laaraiedh, M., Avrillon, S., Uguen, B.: Fingerprinting localization based on neural networks and ultra-wideband signals. In: 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 184–189, December 2011
20. Zekavat, R., Buehrer, R.M.: Handbook of Position Location: Theory, Practice and Advances, 1st edn. Wiley/IEEE Press, Hoboken/Piscataway (2011)