Squeeze More from Fingerprints Reporting Strategy for Indoor Localization

Zhehui Zhang*, Duowen Liu*, Sujie Zhu*, Shangjie Chen*, Xiaohua Tian*†
*School of Electronic, Info and Electrical Engineering, Shanghai Jiao Tong University, China
†National Mobile Communications Research Laboratory, Southeast University, China
{qiaomai, ldw123, zhusujie, galalz, xtian}@sjtu.edu.cn

***Abstract*—Recent study on Wi-Fi RSS fingerprinting based indoor localization reveals that reporting fingerprints with respect to different set of access points (APs) results in location estimations in different levels of accuracy; however, how to find the best strategy for fingerprints reporting with reasonable computational cost, and how to exploit the finding to streamline the design of the localization system are still unknown. In this paper, we revisit the design principles of the localization system with the opportunity provided by the theory of best fingerprints reporting strategy. We first present the localization reliability bounds under the best strategy, and develop algorithms to find the best strategy in practice. We then demonstrate how the best strategy theory can be utilized to improve accuracy of location estimation by resolving the issue of similar fingerprints for both faraway and close-by locations. An iterative algorithm is developed to cross check fingerprints sampled in different locations, in order to derive the best possible result of localization. Moreover, we reveal the relationship between accuracy of location estimation and coverage of Wi-Fi signals in a region, when planning deployment of APs. Experiment results are presented to validate our analysis and design.**

I. Introduction

Wi-Fi received signal strength (RSS) fingerprinting based approach has been very popular for indoor localization [1]. The basic idea of the approach is to first perform site survey of the indoor space that needs localization service, where the RSS reading observed with respect to each access point (AP) at each landmark is uploaded to a localization server. By aggregating uploaded RSS fingerprints, the server can build up a database associating fingerprints with corresponding landmarks, which is termed as the training or offline phase. The database can be utilized when a user wants to be localized: the user could report the RSS readings to the server, which searches the database and derives the estimation of the user's current location. This process is usually termed as location estimation or online phase.

Many indoor localization systems have been developed with the fingerprinting approach. Early systems such as Radar are based on the nearest neighbors(s) in signal space (NNSS) technique, which is to compute the Euclidean distance between reported RSSes and the RSSes in the database [2]. Later systems such as Horus utilizes probabilistic techniques to estimate the user's location, where information about the signal strength distributions is derived from the database [3]. The recent trend for designing the indoor localization system is to leverage crowdsourcing for data training and collaborative location estimation [4]–[9], where information from sensors embedded in smartphones is utilized [4], [7], [9]. Empirical studies are presented to evaluate performance of existing localization systems, where extensive experimental results are analyzed to obtain an empirical quantification of accuracy limits of RSS localization [1].

While efforts have been devoted to improving accuracy of indoor localization systems in an ad-hoc manner, recent theoretical study reveals fundamental limits of RSS fingerprinting based approach oblivious of specific implementation details [10]. An interesting theoretical result of the study is that: reporting RSSes with respect to different APs in the online phase results in different levels of accuracy in location estimation. However, how to find the best strategy with reasonable computational cost, and how to exploit the abstract theory to streamline the design of the localization system are still unknown.

In this paper, we reveal the implication behind the best strategy for fingerprints reporting, based on which we streamline design methodologies of important components in localization systems. We first analyze the theory of best strategy for fingerprints reporting, where we derive the localization reliability bounds under the best strategy. In order to harness the computational cost, we develop a simulated annealing based algorithm to find a suboptimal reporting strategy and present convergence analysis of the algorithm.

We then demonstrate how the best strategy theory can be utilized to improve accuracy of location estimation by resolving the fingerprints similarity issue, which means that fingerprints observed in faraway locations can be similar to each other due to the randomness of radio propagation. An iterative algorithm is developed to cross check fingerprints sampled in different locations, which is based on extra information provided by the best strategy theory. We further illustrate that the proposed algorithm can benefit the localization accuracy, even if locations with similar fingerprints are near to each other and the associated best strategies are the same.

Moreover, we present analyze how to optimally deploy APs, which guarantees that every point in the covered area can be localized with accuracy higher than a threshold. We reveal the relationship between accuracy of location estimation and coverage of Wi-Fi signals in a given region, when planning deployment of APs. We show that the requirement of guaranteed localization accuracy is higher than that of guaranteed signal

978-1-5090-1732-4/16/$31.00 ©2016 IEEE

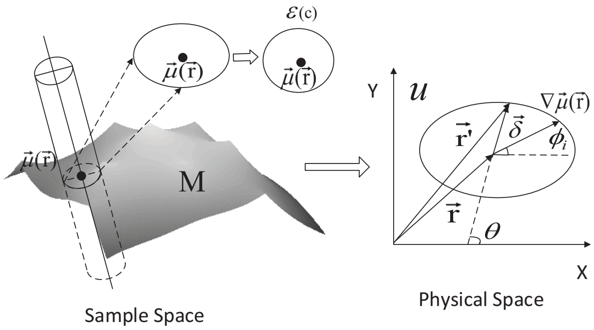


Fig. 1. Intersection in sample space

coverage. We further discuss the landmark collocation issue with AP deployment taken into account, and point out that it can be difficult to discriminate one landmark from another in the RSS sample space, if landmarks are deployed based on the current pure geometric model.

II. Best Strategy for Fingerprints Reporting

Recent theoretical studies about localization performance focus on building analytical models for localization determination stage [11]–[14]. The theory of best fingerprints reporting strategy is first presented in [10], where it is shown that reporting RSSes obtained from different APs results in accuracy of location estimations in different levels. The location estimation process of RSS fingerprinting based localization is in fact a mapping from the RSS fingerprints space to the physical space, where fingerprints space is constructed in the training stage. If we use Q to denote an area in the physical space, which is centered at the users actual location $\vec{r}$ with radius δ, then there must be an event E in the sample space, which makes the localization system to estimate the user's location in Q. Thus the probability the user is correctly localized in Q is equal to the probability that the event E happens. The event E is a set of outcomes of RSS measurements, the shape of which in the sample space turns out to be a hypercylinder [10]. It means that if the reported RSS readings fall into the hypercylinder then the localization system will estimate the user's location in the area Q. The hypercylinder intersects with the mean surface of RSSes, and the intersected surface is in the shape of an ellipse, as the example shown in Fig. 1.

An interesting finding presented in [10] is: if we use another hypercylinder $\mathcal{E}(c)$ to replace the event E, where the intersection between $\mathcal{E}(c)$ and the RSS mean surface is a circle with radius c, then the corresponding area the user will be localized in the physical space is determined by the function $\rho^2(\theta) = \frac{4c^2}{\sum p_i cos^2(\theta-\phi_i)}$, which in fact is the ellipse $\mathcal{U}$ as shown in the right part of Fig. 1. In the figure, location of the user is $\vec{r}$ and the location of any point on the boundary of the ellipse $\mathcal{U}$ is $\vec{r'}$. Vector $\vec{\delta} = \vec{r'} - \vec{r}$ denotes a two-dimensional vector with the direction from the user's actual location to any point on the boundary of $\mathcal{U}$. The symbol θ denotes the angle between $\vec{\delta}$ and the horizontal axis, and ϕ_i denotes the angle between $\nabla\mu_i(\vec{r})$ and the horizontal axis, where $\nabla\mu_i(\vec{r})$ is the gradient of the mean of measured RSS with respect to AP_i at location $\vec{r}$, and $p_i = (|\nabla\mu_i(\vec{r})|/\sigma_i)^2$.

The ellipse $\mathcal{U}$ can be transformed into the form $Q_1\rho^2cos^2\theta + Q_2\rho^2sin^2\theta + Q_3\rho^2cos\theta sin\theta = 4c^2$, where $Q_1 = \sum p_i cos^2\phi_i, Q_2 = \sum p_i sin^2\phi_i, Q_3 = \sum 2p_i cos\phi_i sin\phi_i$, then the area of the ellipse is $u = 8\pi c^2/\sqrt{4Q_1Q_2 - Q_3^2}$. Since the event $\mathcal{E}(c)$ determines u in the physical space and $\mathcal{E}(c)$ is determined by reported fingerprints. The smaller the u, the more likely the user can be localized at $\vec{r}$, which leads to higher accuracy of the location estimation. This means that maximizing the localization accuracy is equivalent to find the measurement sequence minimizing u.

Specifically, if the set of all APs that can be sensed by the user's mobile device is denoted by $\mathbb{U} = \{AP_i\}, i = 1, ..., m$, a sequence of measurements on APs can be denoted by $\mathcal{V}_n = (s^1, ..., s^n), s^j \in \mathbb{U}$. The symbol s^j is the index of the measurement in the sequence, which does not mean that the measurement is performed on AP_j, since an AP can be measured more than once in the sequence. The whole set of strategies is denoted as $\mathbb{U}^n$, where the size of the set is m^n.

For the purpose of simplification, the characteristic of AP_i can be described with a complex parameter $Z_i = p_i e^{2i\phi_i}$, where p_i represents the distinctiveness of signal strength influenced by signal gradient and noise, and the corresponding direction is reflected by $2\phi_i$, double of signal gradient direction. With Z_i, minimizing u is equivalent to maximize

$$\mathcal{F}(\mathcal{V}_n) = \{(\sum_{i\in\mathcal{V}_n}|Z_i|)^2 - |\sum_{i\in\mathcal{V}_n}Z_i|^2\}. \tag{1}$$

The optimal strategy is denoted as $\mathcal{V}_n^*$, $\mathcal{V}_n^* \in \mathbb{U}^n$, where

$$\mathcal{V}_n^* = \arg\max_{\mathcal{V}_n\in\mathbb{U}^n}\mathcal{F}(\mathcal{V}_n). \tag{2}$$

While interesting insight into location estimation is revealed, the proposed best strategy for fingerprints reporting in [10] is presented in a concise form without details. Our work in this paper provides a systematical analysis on the best strategy and exploits the best strategy to streamline design of the localization system.

III. Analysis on the Best Strategy for Fingerprints Reporting

A. Reliability Bounds Analysis

Lemma 1. $\mathcal{F}(\mathcal{V}_n^*)$ is monotonically increasing with n and $\mathcal{F}(\mathcal{V}_{2n}^*) \geq 4\mathcal{F}(\mathcal{V}_n^*)$.

Proof. According to the definition of $\mathcal{V}_n^*$, we naturally have $\mathcal{F}(\mathcal{V}_n^*) \geq \mathcal{F}(\mathcal{V}_n)$. To prove that $\mathcal{F}(\mathcal{V}_n^*)$ is monotonically increasing is to prove $\mathcal{F}(\mathcal{V}_{n+1}^*) \geq \mathcal{F}(\mathcal{V}_n^*)$. Since $\mathcal{F}(\mathcal{V}_{n+1}^*) \geq \mathcal{F}(\mathcal{V}_n^* + s^j)$ with $s^j \in \mathbb{U}$, we only need to prove $\mathcal{F}(\mathcal{V}_n^* + s^i) > \mathcal{F}(\mathcal{V}_n^*)$. Let $W = \sum_{i\in\mathcal{V}_n^*}|Z_i|$ and $Z_o = \sum_{i\in\mathcal{V}_n^*}Z_i$, it is equivalent to prove

$$(W + |Z_i|)^2 - (|Z_o + Z_i|)^2 > W^2 - |Z_o|^2, \qquad \forall i \in \mathcal{V}_n. \tag{3}$$

With $|Z_o + Z_i| \leq |Z_o| + |Z_i|$, the inequation can be simplified as $(W + |Z_i|)^2 - (|Z_o| + |Z_i|)^2 \geq W^2 - |Z_o|^2$, which is equivalent to prove that $(W-|Z_o|)|Z_i| \geq 0$. It is obvious that $W \geq |Z_o|$, thus the proof of monotonic increasing is done.

To prove the second part of the lemma, our strategy is similar to that above. It is obvious that $\mathcal{F}(\mathcal{V}_{2n}^*) \geq \mathcal{F}(2\mathcal{V}_n^*)$, where $2\mathcal{V}_n^*$ means the strategy of repeating the measurement sequence in $\mathcal{V}_n^*$. As the efforts of finding $\mathcal{V}_n^*$ is highly dependent on the order of measurements, simple repeating will definitely result in $\mathcal{F}(\mathcal{V}_{2n}^*) \geq \mathcal{F}(2\mathcal{V}_n^*)$. After expanding the expression of $\mathcal{F}(2\mathcal{V}_n^*)$, it is easy to obtain $\mathcal{F}(2\mathcal{V}_n^*) = 4\mathcal{F}(\mathcal{V}_n^*)$. $\square$

Recall that $\mathcal{E}(c)$ is the probabilistic event that the user is localized within the ellipse with area u and $R(\mathcal{E}(c))$ is the probability of $\mathcal{E}(c)$, where

$$R(\mathcal{E}(c)) = 1 - e^{-\frac{u\sqrt{\mathcal{F}(\mathcal{V}_n^*)}}{16\pi}}. \tag{4}$$

We could get explicit form of Eq. (4) when u is specified.

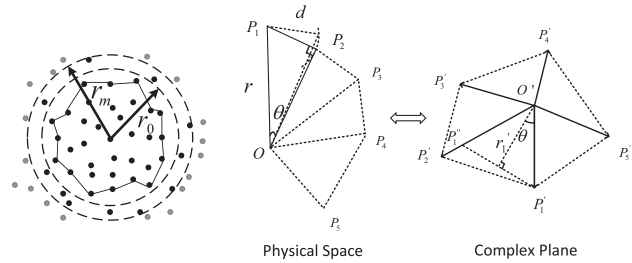
Theorem 1. *The reliability of location estimation as shown in (4) is lower bounded when* $\mathcal{F}(\mathcal{V}_n^*) = 4^{\lfloor \log_2 n \rfloor - 2}\mathcal{F}(\mathcal{V}_4^*)$, *and upper bounded when* $\mathcal{F}(\mathcal{V}_n^*) = n^2 Z_{max}^2$, *where* Z_{max} *is the largest* $|Z_i|$ *in* $\mathbb{U}$.

The proof of the the theorem is based on Lemma 1. Given n, we can find k satisfying $2^k \leq n \leq 2^{(k+1)}$. With Lemma 1, $\mathcal{F}(\mathcal{V}_n^*) > \mathcal{F}(\mathcal{V}_{2^k}^*) > 4^{k-2}\mathcal{F}(\mathcal{V}_4^*)$. Replace k with $\lfloor \log_2 n \rfloor$, the lower bound of $\mathcal{F}(\mathcal{V}_n^*)$ is $4^{\lfloor \log_2 n \rfloor - 2}\mathcal{F}(\mathcal{V}_4^*)$, which leads to the upper bound of Eq. (4). For the upper bound, we have $\mathcal{F}(\mathcal{V}_n^*) = (\sum_{i\in\mathcal{V}_n^*}|Z_i|)^2 - (|\sum_{i\in\mathcal{V}_n^*} Z_i|)^2 \leq (\sum_{i\in\mathcal{V}_n^*}|Z_i|)^2 \leq n^2 Z_{max}^2$, where Z_{max} is the largest $|Z_i|$. This leads to the lower bound of Eq. (4).

B. Reducing Searchspace of APs in Physical Space

Finding the strict optimal strategy for fingerprints reporting is computational costly. In order to harness the cost for AP selection, a natural idea is to find if we could avoid measuring APs with certain features in order to reduce searchspace of measurement strategies. Wen *et al.* mentioned that APs with certain geometric properties in the complex plane could be ignored [10]. Specifically, vectors representing different APs in the complex plane will form a convex hull, APs whose vectors fall within the convex hull can be ignored. However, such a scheme also incurs a high computational cost of $O(m^3)$, which requires to traverse all APs and all APs that could form a triangle in the complex plane at each location, which is unscalable for densely populated APs. Moreover, the implication of the theory in [10] in the physical space is still unrevealed, which hinders it from guiding designs of the practical system.

We are to demonstrate that the implication of the theory for reducing the searchspace of APs is in fact to ignore APs whose signal strength are weak from the user's perspective. We are also to propose an approach to estimate the threshold of the signal strength, so that such a theory can shed light on the practical system design. Our mechanism to eliminate APs of little avail is at the computational cost of $O(m)$.



(a) APs on the boundary (b) Illustration of the convex hull

Fig. 2. Proof of Theorem 2

Theorem 2. *Consider a physical space, where the maximum distance of two APs is* d_m. *APs with signal strength lower than* P_m *can be ignored for location estimation, where* $P_m = P_0 - 10\gamma\log(3\sqrt{2}d_m/\sqrt[4]{3})$.

Proof. We first consider the convex hull constructed by Z_i, where Z_i is the complex parameter of APs centered at the user's current location with radius $r_0 = 3d_m$, then we prove that APs with signal strength less than P_m will fall into the convex hull. In order to get the lower bound of the convex hull, we only need to consider APs on the boundary of the neighborhood shown by Fig. 2(a). Denote those two neighboring APs on the boundary with P_i and P_{i+1}, then the corresponding vectors representing them in the complex plane denoted by Z_i and Z_j can form a triangle in the complex plane.

As the example shown in the Fig. 2(b), where the distance of P_1 and P_2 is d, such that $d \leq d_m$, and the angle between P_1 and P_2 can be calculated by

$$\sin\angle P_1OP_2 = P_1P_2\frac{\sin\angle OP_2P_1}{OP_1}. \tag{5}$$

Then we can get $\theta_{max} = \arcsin\frac{d}{r-d}$. In Fig. 2(b), r_1' denotes the midperpendicular of the $\triangle P_1'O'P_1''$, which is the lower bound of inscribed circle radius.

$$r_1' = O'P_1'\cos\theta, \tag{6}$$

which indicates that r_1' decreases with θ, then we have $r_{1min}' = O'P_1'\cos(\arcsin\frac{d}{r-d})$ when $r_0 = 3d_m$, since $\angle P_1'OP_2'$ is an acute angle. We here have the lower bound of inscribed circle radius of the convex hull $r_m' = \min\{r_{1min}', r_{2min}', ...\}$, note that the minimum is determined by the outermost AP with distance less than r_0. Consequently, we have $r_m' \geq \frac{1}{r^2\sigma^2}\sqrt{1-\frac{d^2}{(r-d)^2}}$. Denoting transmitting power as P_i, we can model

$$\mu_i(\vec{r}) = P_0 - 10\gamma\log|\vec{r}| \tag{7}$$

based on LDPL model [15], which shows that $|\nabla\mu(\vec{r})| = \frac{\gamma}{|\vec{r}|}$, thus we can get $Z_i = (\frac{\gamma^2}{|r|^2\sigma_i^2})e^{2i\Phi i}$. For those APs with received signal strength $P_m \leq P_0 - 10\gamma\log(r_0\sqrt[4]{1+\frac{d^2}{r_0^2-2r_0d}})$, the length of Z_j is less than r_m', which will surely fall within the convex hull. $\square$

With the theorem above, we can directly reduce searching space by eliminate the low signal strength APs from the search space at the computational complexity cost of $O(m)$, since every AP is checked once. Moreover, to determine APs to be eliminated by signal strength can be more efficient in the perspective of physical space, since two locations a user possibly appear will share many APs with the same level of signal strength. This is more efficient than finding those APs in the complex hull.

C. Practical Algorithm for AP selection

After reducing the searchspace, we design a simulated annealing based Algorithm 1 to find out a solution that is approximate to the best reporting strategy. By modeling the probabilistic distribution of the annealing problem as a time-inhomogeneous Markov chain, we further analyze its convergence rate and finite-time behavior.

The proposed algorithm is shown in Algorithm 1. We first define a measurements strategy as a state. Note for m APs and n measurements we have m^n states, where each state is a possible $\mathcal{V}_n$. We index each state and then use α, β to denote the index of a state, $N(\alpha)$ to denote neighboring states of α. The neighbors of a state is the set of all possible generated states from the state in one step. Although the simulated annealing algorithm has been widely used, the challenge for applying the algorithm to the AP selection problem is how to choose the generate function, so that the a good convergence rate can be obtained.

The traditional approach is that in every step, the state can generate all states with equal possibility, i.e. $|N_0(\alpha)| = m^n - 1$. However, in our analysis of the Markov chain, the convergence rate is positively related to the cardinality of $N(\alpha)$. In order to reduce the cardinality to get a faster convergence rate, we define the generate function so that one state can only generate a new state by changing only one measurement.

The *accept function* for simulated annealing algorithm [16] in the AP selection case is set to be

$$P_{\alpha\beta} = \begin{cases} 0 & \text{if } \beta \notin N(\alpha), \\ \min\{1, e^{\frac{\mathcal{F}(\mathcal{V}_n^{(\beta)}) - \mathcal{F}(\mathcal{V}_n^{(\alpha)})}{T}}\} & \text{if } \beta \in N(\alpha) \text{ and } \beta \neq \alpha, \\ 1 - \sum\limits_{\beta \in N(\alpha)} P_{\alpha\beta}(T) & \text{if } \beta = \alpha, \end{cases} \tag{8}$$

where the objective function is $\mathcal{F}(\cdot)$ in Eq. (1).

The *generate function* is set according to the algorithm where a new configuration is generated by replacing one candidate randomly, thus the probability of generating β from α is

$$g(\alpha, \beta) = 1/|N(\alpha)|, \tag{9}$$

where $|N(\alpha)| = n(m-1)$.

The *update function* to update the "temperature" in the simulated annealing algorithm at time t, denoted as T_t is set to be

$$T_t = \frac{\omega}{\log(t+k)}, t = 0, 1, 2, ..., 1 < k < \infty, \tag{10}$$

Algorithm 1 Best reporting strategy

1: Initialize $\mathcal{V}_n^{\alpha}$(Select n APs randomly).
2: **while** $T_t \geq t^*$ **do**
3: Generate new state β randomly from $N(\alpha)$.
4: Calculate $\mathcal{F}(\mathcal{V}_n^{(\beta)})$ for state β.
5: Accept state β as α with probability $P_{\alpha\beta}$.
6: Update temperature T_t.
7: **end while**
8: $\{\mathcal{V}_n\} \leftarrow \mathcal{V}_n^{\alpha}$.

where ω is the parameter tuned with AP characteristics given by $\omega > n^2 p_{max}^2$ and $p_{max} = \max_{i \in \mathbb{U}}\{|Z_i|\}$. With definitions of three main functions above we have Algorithm 1.

We now model the annealing problem as a Markov chain with transition probability matrix defined by (8). The investigation of the Markov chain can give us insights into the convergence rate of the algorithm. The first property is the strong ergodity of the Markov chain.

Theorem 3. *The Markov chain associated with Algorithm 1 is strongly ergodic.*

The proof can be found in our technical report [17].

The strong ergodity of the Markov chain associated with Algorithm 1 indicates that with a random initial state, the state probability vector at time t denoted by $v(t)$ converges to the optimum vector, the vector in which all elements are zero except global optimal ones [16]. The convergence rate is

$$||v(t) - e^*|| = O(1/t^{\min(\frac{1}{n(m-1)}, \frac{\Delta}{n^3 p_{max}})}), \tag{11}$$

where e^* is the global optimum vector, and Δ denotes the difference between the next-to-optimal and optimal value.

According to Eq. (11), the convergence rate is negatively related with the Lipschitz constant of graph underlying the Markov chain, i.e., the maximum distance between two neighboring states on the Markov chain. This indicates that if maximum distance of solutions is very large, the convergence rate will decrease since the searching tends to stuck in some local optimal solutions where accept rate of neighboring state is very low. In addition, with the increase of the difference between next-to-optimal and optimal solution, the Markov chain will converge to the optimal vector faster. Besides, when m is very large and $\frac{1}{n(m-1)}$ is less than $\frac{\Delta}{n^3 p_{max}}$, then the convergence rate is determined by $O(1/t^{\frac{1}{n(m-1)}})$, which indicates that the convergence rate has no relevance to Z values. That is to say, the performance of our algorithm can be guaranteed in indoor spaces with any AP deployment plans.

IV. Location Estimation Leveraging Best Fingerprints Reporting Strategy

The best fingerprints reporting strategy is dependent on the setting of the physical space that needs localization services. Given the indoor space with fixed AP deployment, the best strategy for each location of the space is deterministic and can be derived using the proposed algorithm. Consequently, the best strategy of a location plays a role similar to the local fingerprints stored in the database, which may distinguish one location from another. However, the best strategy is intuitively less sensitive than fingerprints in discriminating one location from another, especially in a small vicinity. Recall the complex vector characterizing an AP, which in essence represents the relative position of the AP with respect to the target location. Since the distance and angle of neighboring locations with respect to surrounding APs are almost the same, best reporting strategies for such locations should be the same. Our experiment results actually corroborate the intuition.

Such a seemingly frustrating phenomenon factually also can be leveraged to reducing errors in location estimation. Empirical studies show that the estimation errors of pure fingerprinting based localization system could be over 6m [7], [18]. The root cause of such large errors is that physically distant locations may share similar Wi-Fi signal strength, which is due to the dynamic propagation of radio signals. However, the feature of the best strategy described above provides an opportunity to eliminate such errors. Although multiple faraway locations may have similar fingerprints, their best strategies for fingerprints reporting could differ from each other, because their relative positions with respect to surrounding APs are different.

It is worth mentioning that the existing solution to deal with the fingerprints similarity is to utilize the k-nearest neighbors (KNN) algorithm [2] or the acoustic ranging estimations performed among peer smartphones; however, KNN is a machine learning approach without any theoretical basis for localization, and the acoustic ranging requires collaboration among users [7], [18].

Exploiting the best strategy can reduce large localization errors without consuming extra resources in users' devices. We propose Algorithm 2 to implement a location estimation approach facilitated by the best strategy. We are to provide a walk-through of the algorithm using the example shown in Fig. 3. Moreover, our analysis of the algorithm will show that the localization accuracy can be improved not only for reducing large-scale errors, but also for discriminating locations in a small neighborhood, where case 1 and case 2 in Fig. 3 are used to represent such two scenarios.

The basic idea of the algorithm is to first roughly determine a set of candidate locations. The user measures APs that are included in the best strategies for all candidate locations in each iteration. The server can then reduce the estimation uncertainty according to the reported fingerprints in each iteration and finally localize the user.

As shown in Fig. 3, the user reports a fingerprint consisting RSSes with respect to AP_1 to AP_4. The server calculates the Euclidean distance between the reported fingerprint and the fingerprint associated with the location A, B and C in the database respectively, through which the server finds that all the three locations are matching the reported fingerprint. We consider the case 1, where the best strategy associated with each location is distinctive because the three locations are faraway from each other. The number in the best strategy

Algorithm 2 Location determination strategy

Get initial estimated locations $\{r_e\}$, $\{cAP_i\}^0 = \emptyset$.
Initialization(set counter $t = 1$ and calculates $\{cAP_i\}^t \leftarrow \bigcap_{\{r_e\}} \{\mathcal{V}_n^*(r)\}$).
while Radius of $\{r_e\} < r^*$ **do**
  if $|\{cAP_i\}^t - \{cAP_i\}^{t-1}| \neq 1$ **then**
    Set $\{nAP\}$ as the AP that appears most frequently.
  else
    $\{nAP\} \leftarrow \{cAP_i\}^{(t)} \backslash \{cAP_i\}^{(t-1)}$.
  end if
  Report $RSSes$ for AP in $\{nAP\}$.
  Update distance matrix $D_j = |RSSes, \mu(r_j)|$.
  Get new estimated locations $\{r_e\} \leftarrow \{r_j\}, D_j < d^*$.
  Increment counter $t = t + 1$ and recalculate candidate APs $\{cAP_i\}^t \leftarrow \bigcap_{\{r_e\}} \{\mathcal{V}_n^*(r)\}$.
end while
Return estimated location $r_0 = \frac{1}{|\{r_e\}|} \sum \{r_e\}$.

means the number of times each corresponding AP should be measured. The server finds that AP_3 is included in all three best strategies, and then asks the user to measure AP_3. After matching the reported fingerprint with respect to AP_3, the sever finds that the location C's fingerprint with respect to AP_3 in the fingerprints database is the worst match to the reported one, C can be deleted from the candidates list. Comparing the best strategies of the remaining location candidates A and B, they have AP_2 in common. The server asks the user to report another fingerprint with respect to AP_2, and the worst match is also deleted.

Now we consider the case 2 in Fig. 3, where the best strategy associated with each location is the same because the three locations are in a small neighborhood. In this case, the server asks the user to report a fingerprint with respect to AP_3, because AP_3 appears most frequently in all best strategies. In this way, the server still can cross check fingerprints reported from the user multiple times and find the best estimation. The performance evaluation presented in Section VI is to show the effectiveness of such an algorithm.

A key issue of the algorithm is when the iteration should end. The server could execute the iteration until it converges to only one candidate location. It can also end the iteration when the candidate locations are within a certain region denoted by r^*. For example, in case 2 of Fig. 3, if r^* is set to be larger than the distance between the three locations, then the iteration will end, and the location that is equally distant to each of the candidate points is returned as the estimated location. The convergence rate of the algorithm is determined by d^*, which shows in each iteration how many candidate points will be accepted to the next iteration. In most cases, d^* can be set with respect to the number of measurement times. If the measurement time is N_0 and there are l candidates, then we can set d^* so that only $\log_{N_0} l$ points will be selected to enter the next iteration, then after at most N_0 rounds, the iteration will end.

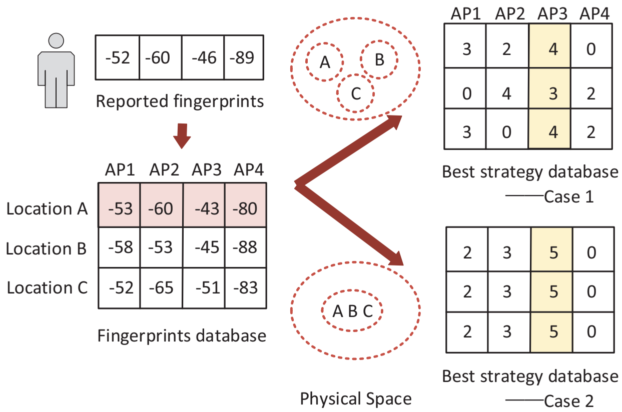


Fig. 3. Location determination facilitated by the best strategy.

V. Strategy for AP Deployment

Wi-Fi APs have been widely deployed in public indoor spaces, where coverage is an important issue. Bai *et al.* propose an AP deployment scheme based on evolving diamond pattern, which presents the minimum required number of APs to cover an area [19]. As location based service is gaining increasingly popularity, it could kill two birds with one stone if the deployment of APs take both coverage and localization into account, especially for newly constructed public buildings. Battiti *et al.* [20] propose a heuristic search method that integrates coverage requirements with the reduction of localization error, where the error is estimated relied on simple radio propagation model. While AP coverage problem is closely related to AP deployment problem for localization, the fundamental relationship between the two problems has yet been revealed.

A. AP Deployment for Localization

If we want to find out the best strategy for fingerprints reporting at location $\vec{r}$, the deployment of APs has to be given; therefore, it seems to be a paradox to determine the optimal AP deployment based on the theory of the best strategy. Our basic idea is to search over all possible AP deployment solutions, in order to find the solution that can offer the best localization performance. Such AP deployment solution is the best for localization. The crux of the idea is to evaluate the performance of localization based on the best strategy for fingerprints reporting, which is denoted by

$$R(\vec{r}) = (\sum_{i\in\mathcal{V}_n^*} |Z_i|)^2 - (|\sum_{i\in\mathcal{V}_n^*} Z_i|)^2, \tag{12}$$

where $\mathcal{V}_n^*$ is the index of measured APs determined by Eq. (2). This is because the larger $R(\vec{r})$ the smaller ellipse can be obtained (recall Fig. 1) thus the higher localization accuracy can be achieved. Here note the difference between Eq. (2) and Eq. (12). In the discussion of AP deployment problem, we assume the system always follow the best fingerprints reporting strategy, so that the system accuracy can be maximized.

Eq. (12) indicates that the change of APs' locations results in change of $\nabla\mu(\vec{r})$. We use $\{x_i, y_i\}$ to denote the location of AP_i, and $\{\vec{x}, \vec{y}\} = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$ to denote locations of APs. Consider the localization performance at location $\vec{r}$, which is denoted by $R(\vec{r})$. Suppose the user appear in locations following the probability density function denoted by $f(\vec{r})$, the expected performance of localization in such a specific setting is $\int_S R(\vec{r})f(\vec{r})dr^2$. Consequently, the optimal strategy for deploying APs is to fix APs in the following locations:

$$\{\vec{x}, \vec{y}\}^* = \underset{\{\vec{x},\vec{y}\}}{\arg\max} \int_S R(\vec{r})f(\vec{r})dr^2, \tag{13}$$

where S is the area of the indoor space. This provides fundamental criteria to evaluate different AP deployment strategies for the purpose of localization.

The challenge for resolving the problem above is that the searching space for optimal $\{\vec{x}, \vec{y}\}$ is continuous. In practice, we are unable to search the physical space in an inch-by-inch manner, thus we can limit the locations of APs in discrete positions. Consequently, the problem can be transformed into

$$\{\vec{x}, \vec{y}\}^* = \underset{\{\vec{x},\vec{y}\}\subset\{\vec{X},\vec{Y}\}}{\arg\max} \int_S R(\vec{r})f(\vec{r})dr^2, \tag{14}$$

where $\{\vec{X}, \vec{Y}\}$ denotes the discretized physical space.

The transformed problem can also be resolved by a simulated annealing based algorithm, which is similar with the one for deriving the best strategy. However, in the current problem, a state no longer means an AP selection strategy. It means a deployment strategy, i.e., the locations for APs. Correspondingly, we present a new formulation of generating function $g(\alpha, \beta) = 1/|c|$, where c denotes the number of deployed APs. Detailed algorithm can be found in our technical report [17].

B. Localization and Coverage

In former works, access point deployment is treated as a coverage problem, however, for the performance of localization, AP deployment can be further discussed so as to meet the accuracy needs. If a specific AP deployment strategy guarantees that each point in the area can be localized correctly, then each point is surely covered by APs, which means every point can receive signals higher than the given strength. In the following discussions, we only care about the minimum noise so we use σ to denote $\min_i\{\sigma_i\}$. Here N_0 denotes measurement time and S denotes the area.

Theorem 4. *If an AP deployment plan in a region with area S satisfies the localization accuracy R^*, then every point must be covered by at least one AP within the distance $\frac{\sqrt{N_0}}{\sigma\sqrt[4]{R^*}}$, and there must be at least $\frac{2\sigma\sqrt{3R^*}}{9N_0}(S-2)$ APs need to be deployed; however, if a user is definitely covered by at least one AP at any point of the region, it is still possible that the localization accuracy can not be satisfied.*

Proof. The first part of the theorem is to prove that $\forall R^*, \forall N_0, \exists d$, such that if $\forall \vec{P}, \forall AP_i \in \mathbb{U}_n, d(\vec{P}, AP_i) > d$, then $R(\vec{p}) < \frac{N_0^2}{d^4\sigma^4}$, where $\vec{P}$ is a point in the region.

According to (12), the localization accuracy is determined by the characteristic vector of the AP, thus

$$\begin{aligned} R(\vec{P}) \leq (\sum_{i \in \mathcal{V}_n} |Z_i|)^2 =& (\sum_{i \in \mathcal{V}_n} p_i)^2 = (\sum_{i \in \mathcal{V}_n} \frac{(\nabla \mu_i)^2}{\sigma_i^2})^2 \\ =& (\sum_{i \in \mathcal{V}_n} \frac{1}{r_i^2 \sigma_i^2})^2 \leq \frac{N_0^2}{d^4 \sigma^4}. \end{aligned} \tag{15}$$

The equality in (15) holds when the location is at the center of several equidistant APs with the same RSS variance σ.

If a point can be located with required accuracy level R^*, then there must be an AP within $\frac{\sqrt{N_0}}{\sigma \sqrt[4]{R^*}}$. To guarantee that every point lies in a coverage range of some APs, the AP deployment strategy must satisfy $\exists AP_i \in \mathbb{U}_n, d(\vec{P}, AP_i) < \frac{\sqrt{N_0}}{\sigma \sqrt[4]{R^*}}$. The second port of the theorem is equivalent to finding the minimum required number of APs to cover the area given the minimum required distance d. Since the space requires indoor localization service can be divided into small square areas, the problem is reducible to circle covering problem which has been proved by Kersher [21].

With Kersher's theory that the least number of circles with radius a denoted by $\psi(a)$, satisfies

$$\pi a^2 \psi(a) > \frac{2\pi\sqrt{3}}{9}(S - 2\pi a^2), \tag{16}$$

where S denotes the area of the rectangle. Together with (15), we finally get the minimum required number is

$$\psi > \frac{2\sigma\sqrt{3R^*}}{9N_0}(S - 2). \tag{17}$$

We now give a counter example showing that the localization accuracy can not be satisfied, while the coverage is satisfied. Assume there are k APs collocated collinearly, every point will be covered if access points have enough power, however, it is impossible to localize the user if the user is standing on any point of the line. This is because the small displacement along the line is with indistinguishable changes in RSSes. If the user moves along the line, it is impossible to localize where the user is. $\square$

Eq. (17) shows the minimum number of APs required is determined by several factors. When the signal channel is clearer, i.e. σ is smaller, the number is smaller. Moreover, the decrease of measurement times N_0 brings the same impact. Moreover, when the required reliability becomes higher, the required number is larger, which increases proportionally to the square root of the reliability.

C. Discussions

We now briefly investigate the influence of our AP deployment on collation of landmarks or measurement points (MPs). Landmarks are important infrastructure for fingerprinting based indoor localization, because it is impossible to sample every point in an indoor space in the training phase, as shown in Fig. 5. Currently, the common practice in localization is to perform sampling on landmarks distributed in the space following certain pattern such as square and triangle [22]; however, current studies on the issue basically model it as a pure geometric problem, without considering the influence APs [23].

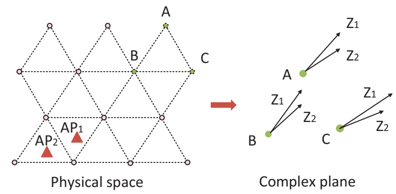


Fig. 4. Complex plane of equilateral collocation

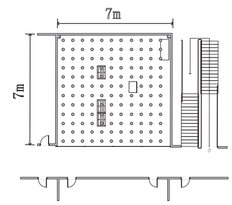


Fig. 5. Floor plan of the experiment field.

Consider a certain landmark collocation, we use $\mathcal{K}_i$ to denote the neighborhood of the ith landmark. We use $P(\mathcal{K}_i|\vec{r} = MP_i)$ to denote the probability that a user is localized in the neighborhood $\mathcal{K}_i$ when the real location is MP_i. If the user appears at each reference point with equal probability, the PDF of its location $P(r \in \mathcal{K}_i)$ is $1/|\mathcal{K}_i|$. We use E_r to represent the expected reliability, where

$$E_r = \sum_i P(\mathcal{K}_i|\vec{r} = MP_i)P(r \in \mathcal{K}_i). \tag{18}$$

Maximization of the accuracy is equivalent to minimization of the error rate. Since the observed RSS is Gaussian distributed [10], $f(\mu(\vec{r})|\mu(MP_i)) = \prod_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp - \frac{(\mu_k(\vec{r}) - \mu_k(MP_i))^2}{2\sigma_k^2}$, thus

$$P_e = \sum_{j \neq i} P(\mathcal{K}_j|\vec{r} = MP_i) = \sum_{j \neq i} \int_E f(\mu(\vec{r}))d(\mu(\vec{r})), \tag{19}$$

where the integration zone E is the event that the reported RSSes make the user to be localized into K_j rather than $\mathcal{K}_i$. Such a zone can be calculated by $f(\mu(\vec{r})|\mu(MP_j)) > f(\mu(\vec{r})|\mu(MP_i))$, according to the theory of maximum likelihood estimation.

Eq. (19) indicates that the localization error is positively related to the distance of neighboring MPs in the RSS sample space. Thus the optimal MP collocation should be to maximize the distance of the nearest neighbors in the sample space, instead of maximizing that in the physical space as shown in [23], where the equilateral triangle pattern is proposed as the best collocation pattern. Considering the AP deployment shown in Fig. 4, MPs A, B, and C have similar complex vectors in the complex plane since the signal gradient is small and the direction is almost collinear. The similarity of the complex vectors indicates that it is very difficult to discriminate one MP from another, which makes localization performance low; therefore, collocating a large number of MPs or let crowdsourcing workers to collect fingerprints in the region is meaningless.

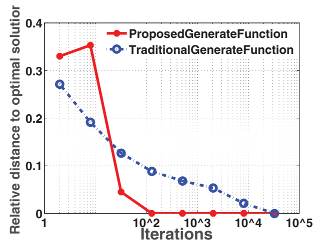


Fig. 6. Comparison of convergence rate.

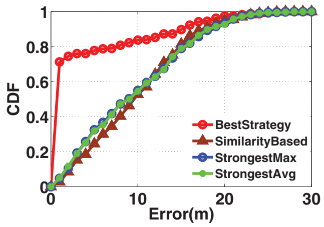


Fig. 7. Simulation results of localization.

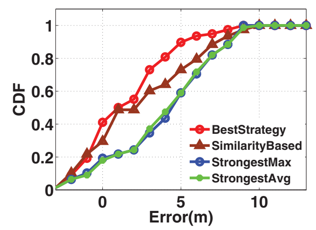


Fig. 8. Experiment results of localization.

Fig. 9. Experiment results of localization.

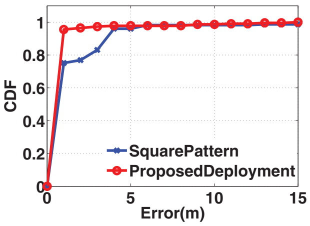


Fig. 10. Localization error CDF for purposed deployment strategy.

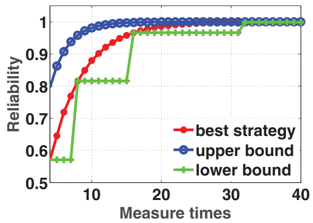


Fig. 11. Reliability bound analysis.

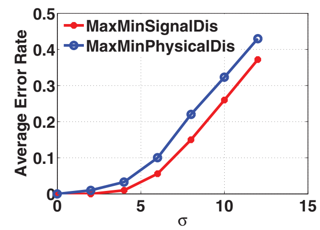


Fig. 12. Collocation of landmarks.

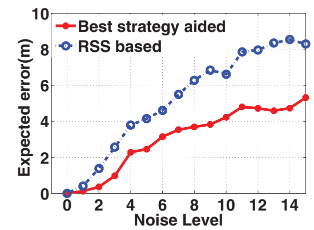


Fig. 13. Location estimation facilitated by the best strategy.

VI. Performance Evaluation

In this section, we present the performance evaluation of proposed algorithms. We conduct an real world experiment for indoor localization. The floor plan of the experiment field is shown in Fig 5. We set 11×11 measure points, laid in grid pattern. We also perform simulations to examine large-scale scenarios due to space limitation of our lab. We adopt the LDPL radio propagation model in the simulation, where the settings are the same as that in [24] with path loss exponent $\gamma = 3.5$ and $P_0 = 0dBm$. We generate the users' locations with Monte Carlo method.

A. Convergence Rate for Finding the Best Strategy

We evaluate the convergence rate of our proposed Algorithm 1, which is applied in our experiment with the number of APs $m = 12$ and the number of measurements at each location $n = 25$. Fig. 6 presents convergence rates of traditional simulated annealing algorithm and our algorithm. It is clear that the distance between the solution can be obtained to the optimal solution decreases dramatically using our algorithm, which means the convergence rate of our proposed algorithm is faster than the traditional one. This is because the generate function specifically designed for the localization case in our algorithm is more efficient, which corroborates our theoretical analysis.

B. Performance of the Best Strategy

After we have obtained the best strategy for measuring APs, we here illustrate the best strategy indeed can improve accuracy of localization. The best strategy in essence specifies which APs to measure so that corresponding reported fingerprints can result in highest accuracy. We compare the localization results with the best strategy and that with other strategies. If we measure every AP's signal strength at each location of a space, then APs with strongest average signal strength over different locations can be obtained, and the APs with strongest signal strength all over the space can also be obtained. *StrongestAvg* strategy is to always measure APs with strongest average signal strength, and *StrongestMax* is to always measure APs with the strongest signal strength. *SimilarityBased* strategy calculates the similarity matrix of APs and always measures APs with the lowest similarities. Such three strategies have been evaluated in [5]. Fig. 7 and Fig. 8 illustrates the cumulative distribution function (CDF) with respect to the localization error in the simulation and field experiment, respectively. Fig. 9 illustrates how the localization errors are distributed in the experiment field. Our simulation and experiment results with respect the three strategies all confirm the results in [5]. Moreover, the results also indicate that the best strategy indeed incur the smallest localization errors.

C. Deployment of APs

We perform simulations to evaluate how to deploy APs for localization purpose. Here we set 400 candidate positions in the space and then deploy APs based on two strategies. The first strategy is to deploy APs following square pattern and the second strategy is to adopt our proposed algorithm. We use Monte Carlo experiments to find the CDF of localization accuracy, and the results are shown in Fig. 10. The result corroborate with our findings.

D. Localization Reliability Bounds under the Best Strategy

Under the same setting for evaluating the convergence rate, we show the theoretical localization reliability bounds and the reliability we observed in the experiment in Fig. 10. The results corroborate our analysis, where increasing the number of measurements can improve the reliability, and the practical reliability is bounded by our theoretical bounds.

E. Collocation of Landmarks

We perform simulations to evaluate how to collocate landmarks for the training phase, for which we set 400 landmarks in the space. We first discretize the space into grids, and then select landmarks according to two rules. The first rule is based on our proposed landmark selection algorithm [17] and the second rule is based on the optimal geometric model [23]. After landmarks are collocated, we measure the average error rate for localization in the space with respect to varying noise level, and the results are shown in Fig. 12. Our algorithm is to maximize the minimum distance of landmarks' in the RSS sample space, which incurs lower level of error rate compared with the geometric model, which maximize the minimum distance of landmarks' in the physical space. This corroborates our theoretical analysis.

F. Location Estimation facilitated by the Best Strategy

We perform simulations to evaluate the performance of the proposed location determination algorithm, which is facilitated by the best strategy. We apply our AP deployment and landmark collocation schemes in the simulation, and compare our proposed algorithm with traditional RSS based kNN approach adopted in [2], where k is set to be 2. We model RSSes from different APs with different noise factor σ_i, which is randomly generated from 1 to 4. As shown in Fig. 13, the proposed localization strategy outperforms the traditional RSS based one with lower expected localization errors. Specifically, when the noise level increase, the best strategy aided method shows higher localization accuracy, since the user will choose APs with discriminant signal strength characteristics.

VII. Conclusions

In this paper, we have revisited the design principles of the localization system with the opportunity provided by the theory of best fingerprints reporting strategy. We first have presented the localization reliability bounds under the best strategy, and developed algorithms to find the best strategy in practice. We then have demonstrated how the best strategy theory can be utilized to improve accuracy of location estimation by resolving the issue of similar fingerprints for both faraway and close-by locations. An iterative algorithm has been developed to cross check fingerprints sampled in different locations, in order to derive the best possible result of localization. Moreover, we have revealed the relationship between accuracy of location estimation and coverage of Wi-Fi signals in a region, when planning deployment of APs. Experiment results have been presented to validate our analysis and design.

VIII. Acknowledgement

This work is supported by National Natural Science Foundation of China (No. 61572319, U1405251, 61532012, 61325012, 61271219, 61428205); National Mobile Communications Research Laboratory, Southeast University (No. 2014D07).

References

[1] E. Elnahrawy, X. Li, and R. P. Martin, "The limits of localization using signal strength: A comparative study," in *Proc. IEEE SECON*, 2004, pp. 406–414.

[2] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proc. IEEE INFOCOM*, vol. 2, 2000, pp. 775–784.

[3] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proc. ACM MobiSys*, 2005, pp. 205–218.

[4] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *Proc. ACM MobiCom*, 2012, pp. 269–280.

[5] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. ACM MobiCom*, 2010, pp. 173–184.

[6] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: zero-effort crowdsourcing for indoor localization," in *Proc. ACM MobiCom*, 2012, pp. 293–304.

[7] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the limit of wifi based localization for smartphones," in *Proc. ACM MobiCom*, 2012, pp. 305–316.

[8] C. Wu, Z. Yang, and Y. Liu, "Smartphones based crowdsourcing for indoor localization," *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, pp. 444–457, 2015.

[9] Y. Shu, K. G. Shin, T. He, and J. Chen, "Last-mile navigation using smartphones," in *Proc. ACM MobiCom*. ACM, 2015, pp. 512–524.

[10] Y. Wen, X. Tian, X. Wang, and S. Lu, "Fundamental limits of rss fingerprinting based indoor localization," in *Proc. IEEE INFOCOM*, 2015.

[11] A. Krishnakumar and P. Krishnan, "On the accuracy of signal strength-based estimation techniques," in *Proc. IEEE INFOCOM*, vol. 1, 2005, pp. 642–650.

[12] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," in *Proc. IEEE INFOCOM*, vol. 2, 2004, pp. 1012–1022.

[13] M. A. Youssef and A. Agrawala, "On the optimality of wlan location determination systems," *Communication Networks and Distributed Systems Modeling and Simulation Conference*, 2003.

[14] M. A. Youssef, A. Agrawala, and A. Udaya Shankar, "Wlan location determination via clustering and probability distributions," in *Proc. IEEE PerCom*, 2003, pp. 143–150.

[15] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996, vol. 2.

[16] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing," in *Proc. IEEE DAC*, 1985, pp. 761–767.

[17] "Squeeze more from fingerprints reporting strategy for indoor localization." [Online]. Available: https://www.dropbox.com/s/z0qqb3ga0r0i36g/techrepo.pdf

[18] H. Liu, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Accurate wifi based localization for smartphones using peer assistance," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2199–2214, 2014.

[19] X. Bai, Z. Yun, D. Xuan, T. H. Lai, and W. Jia, "Deploying four-connectivity and full-coverage wireless sensor networks," in *Proc. IEEE INFOCOM*, 2008.

[20] R. Battiti, M. Brunato, and A. Delai, "Optimal wireless access point placement for location-dependent services," 2003.

[21] R. Kershner, "The number of circles covering a set," *American Journal of mathematics*, pp. 665–671, 1939.

[22] Y. Chen, J.-A. Francisco, W. Trappe, and R. P. Martin, "A practical approach to landmark deployment for indoor localization," in *Proc. IEEE SECON*, vol. 1, 2006, pp. 365–373.

[23] K. Sheng, Z. Gu, X. Mao, X. Tian, W. Wu, X. Gan, and X. Wang, "The collocation of measurement points in large open indoor environment," in *Proc. IEEE INFOCOM*, 2015.

[24] M. Angjelichinoski, D. Denkovski, V. Atanasovski, and L. Gavrilovska, "Cramér–rao lower bounds of rss-based localization with anchor position uncertainty," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2807–2834, 2015.