Optimization of Fingerprints Reporting Strategy for WLAN Indoor Localization

Xiaohua Tian, *Member, IEEE*, Wenxin Li, Yucheng Yang, Zhehui Zhang, and Xinbing Wang, *Senior Member, IEEE*

Abstract—This paper investigates how to optimize the fingerprints reporting strategy to improve localization accuracy, and how the optimal strategy theory can be utilized to streamline the design of WLAN fingerprinting localization systems. In particular, we first reveal that the fingerprints reporting problem is essentially an NP-Hard size-constrained supermodular maximization problem, and then show the inapplicability of the state-of-the-art approximation algorithms to the problem. We then propose a new algorithm and show that if the number of fingerprints measurements is large enough, then the localization accuracy is at most $1-\varepsilon$ times worse than the optimal value, with ε any given constant close to 0. Moreover, we demonstrate how the optimal 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, with an iterative algorithm developed to cross check fingerprints sampled in different locations, in order to derive the best possible result of localization. Further, we reveal the relationship between accuracy of location estimation and coverage of Wi-Fi signals in indoor spaces when planning deployment of APs. Experiment results are presented to validate our theoretical analysis.

Index Terms—Fingerpringting, localization, performance analysis

1 Introduction

Wi-Fi received signal strength (RSS) fingerprinting based approach has been very popular for indoor localization [1], [2], [3]. The basic idea of the approach is to first perform site survey for 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 observed 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 neighbor(s) in signal space (NNSS) technique, which is to compute the euclidean distance between reported RSSes and the RSSes in the database [4]. 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 [5]. The recent trend for designing the indoor localization system is to leverage crowdsourcing for data training and collaborative location estimation [1], [6], [7], [8], [9], where data from sensors embedded in smartphones are utilized [6], [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 [3].

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 [2]. 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. In contrast to the AP selection schemes for scalability and energy efficiency with clustering [18], [19], [20], [21], the new finding provides a theoretical basis for streamlining the design of fingerprinting based localization systems. However, how to find the optimal reporting strategy with reasonable computational cost, and how to exploit the theory to streamline the design of the localization system are still not concretely known.

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 reveal that the objective function of the optimal fingerprints reporting strategy is supermodular, and the fingerprints reporting problem is essentially an NP-Hard size-constrained supermodular

- *X. Tian is with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China. E-mail: xtian@sjtu.edu.cn.*
- *W. Li, Y. Yang, Z. Zhang, and X. Wang are with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: {469634196, yangyuchengalbert, qiaomai, xwang8}@sjtu.edu.cn.*

Manuscript received 7 Oct. 2016; revised 6 June 2017; accepted 11 June 2017. Date of publication 15 June 2017; date of current version 5 Jan. 2018. (Corresponding author: Xiaohua Tian.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2017.2715820

maximization problem. We then present analysis of the state-of-the-art approximation algorithms [26], [27] and show their inapplicability to the fingerprints reporting strategy optimization case. We propose a new algorithm and show that if the number of fingerprints measurements is large enough, then the localization accuracy is at most $1-\varepsilon$ times worse than the optimal value, where ε is a given constant close to 0.

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 analyze how to optimally deploy APs, which improves the performance of localization systems. 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 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.

2 Related Work

Wi-Fi Based Indoor Localization. A number of wireless techniques can be used for indoor localization such as acoustic signals, ultrasound, infrared, RFID, Bluetooth, Cellular and Wi-Fi, where a comprehensive survey can be found in [30]. The basis of such localization techniques is to measure the spatial feature of the corresponding wireless signals. Wi-Fi based indoor localization has drawn much attention in recent years due to the wide deployment of Wi-Fi APs and ubiquitous use of smartphones. The easy-access received signal strength (RSS) can be used for localization. With the RSS measurement, the radio propagation model can be applied to derive the distance between APs and the mobile device, and the device's location can be triangulated. This is known as model based approach. Another method is to collect RSSes observed in different locations to build a RSS fingerprints database, which is then compared with the user reported RSSes to estimate the user's location. This is known as fingerprinting based approach.

The time and angle extracted from the channel state information (CSI) of the wireless signal also can be used for localization. Chronos utilizes a novel algorithm to compute the time-of-flight (ToF) of the signal to derive the distance between the transmitter and receiver [31]. Kotaru et al. propose SpotFi system, which is based on ToF and angle-of-arrival (AoA) collected from each AP [32]. Xiong et al. develop an indoor location system ArrayTrack, where the basic idea is to derive the AoA information of the mobile device's signal with respect to multiple antennas. Such systems can achieve centimeter-level accuracy; however, there are only limited types of hardware on the market supporting CSI retrieval [33], [34], and those customized hardware are not widely deployed in existing buildings as regular Wi-Fi APs. We focus on fingerprinting based localization in this paper, where the advantage is the convenient deployment.

TABLE 1
Performance of Fingerprinting-Based Positioning System

Systems	Error	Description
Radar [4]	2-3m	Nearest neighbour(s) in signal space (NNSS).
Horus [5]	1.5-2.1m	Probabilistic approach.
Peer-assisted Localization [10]	1.6-3m	With acoustic ranging.
Zee [8]	1.2-2.3m	Particle filter and crowd-sourcing.
EZ [11]	2-7m	Models wireless propagation constraints.
Graph-Fusion [12]	1.52-4.53m	Particle filter and graph discretization of indoor map.
Moloc [16]	1.13-4.91m	Maximum likelihood of fingerprint and motion.
Walkie-Markie [13]	1.3-2.8m	RSSI sequence.
Unloc [14]	median 1.69m	WiFi landmark.
PTS [15]	0.4-3.8m	RSSI peak in a temporal sequence.

Wi-Fi Fingerprinting Localization. The early fingerprinting localization systems employ pure RSS fingerprints to determine the mobile device's location [4], [5], and the later systems utilize data from device's embedded sensors for both improving the localization accuracy [8], [10], [11], [12] and deriving radio map of the space covered [13], [14], [15], [16]. Some representative mechanisms are tabulated in Table 1; comprehensive summary of the fingerprinting localization systems can be found in a number of survey papers such as [17].

Fingerprints Reporting Strategies. In those fingerprinting localization systems, the fingerprints reporting strategy determines which AP's generated RSS values the mobile device should report to the localization server in the online phase. Since an indoor region could be covered by many APs, it is found that processing all the RSS data that can be observed will incur high computational complexity for the energy-constrained mobile device [18], [19]. Moreover, the number of APs providing continuous and stable RSS fingerprints for a given location is limited, there is a need to choose a subset of observable APs at a location, so that the computational complexity and bias incurred by the unstable APs can be reduced.

Clustering approach has been adopted by early large-scale localization systems to improve the scalability and accuracy, where locations in an indoor space are grouped into clusters according to the coverage of APs [18]. In particular, all locations that share q APs are put into one cluster, which makes sure that all locations are covered by at least q APs, and q APs with the largest signal strength values at each location are chosen for localization. As the order of APs with largest RSS values slightly changes with time, the q APs are treated as a non-ordered set.

Chen et al. propose an information gain based AP selection method, where the basic idea is to select those APs with most discriminative features for localization [19]. Each AP is viewed as a feature and the radio map of the space is described by features of APs, where the AP_i's feature value is the average signal strength. The discriminative power of feature AP_i is measured by the information gain when the feature value of AP_i is known. The information gain is defined as the difference between the entropy of the radio map and the conditional entropy given AP_i's feature value.

Kushki et al. propose a spatially localized positioning method in wireless local area networks [20]. The first step of the method is to perform spatial filtering over location fingerprints, where locations with similar observable fingerprints are grouped into a cluster. The system then selects APs corresponding to each cluster, where the goal is to make selected d APs to be with the smallest divergence. The divergence is measured by two metrics: Bhattacharyya distance and information potential, both of which measure the distance between two probability density functions.

Lin et al. propose a group-discrimination based AP selection scheme, which considers the dependence among APs [21]. The proposed mechanism proposes an algorithm utilizing the risk function from supporting vector machines (SVMs) to measure the localization capability of each group. The group-discrimination value of each group is derived by maximizing the margin between reference points of the space. A recursive feature elimination (RFE) mechanism is then proposed to reduce the computational overhead.

The AP selection approaches mentioned above are basically targeting at how to group APs into clusters. The recent study on fingerprints reporting strategies reveals that measuring different APs results in different levels of accuracy even within the same cluster [2]; however, the particular scheme for finding the strategy is not elaborated. While our previous work [36] propose a stimulated annealing algorithm to find an appropriate measurements sequence, a rigorous theoretical analysis of the AP selection issue is yet to be presented.

Submodular Optimization. We are to model the problem of finding the best strategy as a size-constrained submodular minimization problem. It is known that the unconstrained submodular minimization problem can be solved in polynomial time; however, if the problem is with additional constraints, it will become extremely challenging. It is proposed to leverage the basic polyhedral theory to resolve the NP-Hard submodular minimization problem with cardinality constraint [26], where the optimal solutions under certain size constraints can be found in polynomial time. Iyer et al. propose a framework for both unconstrained and constrained submodular set function optimization problems based on discrete semidifferentials [27]; an approximation algorithm for the size-constrained submodular minimization problem is also presented. However, those recently-proposed algorithms can not be applied directly to the problem studied in this paper, and we will propose an approximation algorithm dedicated to this end.

3 System Model

The theoretical basis of best fingerprints reporting strategy is first presented in [2], where it is shown that reporting

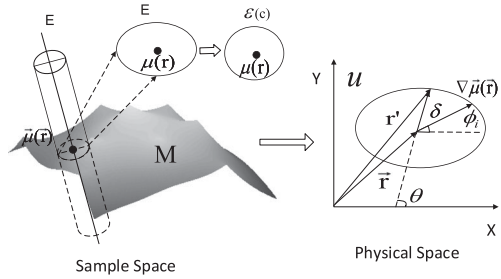


Fig. 1. Intersection in sample space.

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. If we use Q to denote an area in the physical space, which is centered at the user's actual location $\vec{r}$ with radius δ, then there must be a corresponding 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 [2]. 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 [2] 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 symbol σ_i denotes the observed RSS standard deviation with respect to AP_i.

The ellipse $\mathcal{U}$ can be transformed into the form $Q_1\rho^2 cos^2\theta + Q_2\rho^2 sin^2\theta + Q_3\rho^2 cos\theta sin\theta = 4c^2$, 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 u is in value, the more likely the user can be localized at $\vec{r}$, which leads to higher accuracy of location estimation. This means that maximizing the localization accuracy is equivalent to finding 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 $\mathbb{U} = \{AP_i\}, i = 1, \ldots, m$,

a sequence of measurements on APs can be denoted by $\mathcal{V}_l = (s_1, \ldots, s_l), s_j \in \mathbb{U}$ and l is the number of times for measurements. The symbol s_j is the index of the measurement in the sequence. Note that s_j does not necessarily 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 $\mathbb{U}^l$, where the size of the set is m^l.

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

$$\mathcal{F}(\mathcal{V}_l) = \left(\sum_{i\in\mathcal{V}_l} |Z_i|\right)^2 - \left|\sum_{i\in\mathcal{V}_l} Z_i\right|^2, \tag{1}$$

thus the optimal fingerprints reporting strategy can be denoted by $\mathcal{V}_l^*, \mathcal{V}_l^* \in \mathbb{U}^l$, where

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

While interesting insight into location estimation is revealed, the proposed best strategy for fingerprints reporting in [2] is presented in a concise form without details. Our work in this paper provides a systematical analysis of the best strategy and exploits the best strategy to streamline design of the localization system.

4 Analysis of Optimal Strategy for Fingerprints Reporting

4.1 Supermodularity of the Objective Function

Consider the localization system with m APs, each of which can be measured multiple times. Suppose that the mobile device is allowed to perform AP measurements l times, then the objective function of AP selection is equivalent to the function $g(\cdot) : 2^{[lm]} \to R$, and for any set $S \subseteq [lm]$, we have

$$g(S) = \left(\sum_{a\in S} |s_{\lfloor\frac{a-1}{l}+1\rfloor}|\right)^2 - \left|\sum_{a\in S} s_{\lfloor\frac{a-1}{l}+1\rfloor}\right|^2, \tag{3}$$

where $[lm]$ denotes the set of positive integers $\{1, 2, \ldots, lm\}$.

Intuitively, more times of measurements can lead to more accurate location estimation; moreover, it is straightforward that the first term of $g(S)$ is dominant compared with the second one. Such characteristics indicate that $g(S)$ could be a supermodular function. We provide the rigorous proof in the following.

Theorem 1. *Function* $g(S) = (\sum_{a\in S} |s_{\lfloor\frac{a-1}{l}+1\rfloor}|)^2 - |\sum_{a\in S} s_{\lfloor\frac{a-1}{l}+1\rfloor}|^2$ *is a supermodular set function defined on the set* $[lm]$.

Proof. We prove the theorem with the definition of supermodular set function [26]. We first perform the following transformation of $g(\cdot)$:

$$\begin{aligned} g(S) &= \left(\sum_{i\in S} |s_{\lfloor\frac{i-1}{l}+1\rfloor}|\right)^2 - \left|\sum_{i\in S} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2 \\ &= \left(\sum_{i\in S} p_{\sigma(i)}\right)^2 - \left|\sum_{i\in S} p_{\sigma(i)} e^{2i\phi_{\sigma(i)}}\right|^2 \\ &= \left(\sum_{i\in S} p_{\sigma(i)}^2 + \sum_{i\neq j\in S} p_{\sigma(i)}p_{\sigma(j)}\right) \\ &\quad - \left(\sum_{i\in S} p_{\sigma(i)}^2 + \sum_{i\neq j\in S} p_{\sigma(i)}p_{\sigma(j)} e^{2i(\phi_{\sigma(i)}-\phi_{\sigma(j)})}\right) \\ &= 2\sum_{i\neq j\in S} p_{\sigma(i)}p_{\sigma(j)} \sin^2(\phi_{\sigma(i)} - \phi_{\sigma(j)}). \end{aligned}$$

For the convenience of demonstration, we define a new function: $\sigma(i) = \lfloor\frac{i-1}{l}+1\rfloor (1 \le i \le lm)$; if we add a new element k to the set S, the value increase of the function introduced by the element is

$$\begin{aligned} &g(S\cup\{k\}) - g(S) \\ &\quad = 2\sum_{j\in S} p_{\sigma(k)}p_{\sigma(j)} \sin^2(\phi_{\sigma(k)} - \phi_{\sigma(j)}). \end{aligned}$$

As each term above is non-negative, if adding the same element k into a smaller set, the value of the function must be smaller compared with that if adding it into a larger set S, thus function $g(\cdot)$ is a supermodular set function.[1] □

With the supermodularity of the objective function, the AP selection problem can be intuitively modeled as a size-constrained supermodular maximization problem. In particular,

$$\max_S \{g(S) : S \subseteq [N], S = k\}, \tag{4}$$

where $g(\cdot) : 2^{[N]} \to R$ is a supermodular set function on the set $[N] = \{1, 2, \ldots, N\}$. This problem has been proved to be NP-hard [26], [27], [28], and normally the approach for finding the solution is to transform the size-constrained supermodular maximization problem into the corresponding submodular minimization problem [28]. This is because for any supermodular set function $g(\cdot)$, if we define a new function $f(\cdot) = C - g(\cdot)$, and C is a constant that is large enough such as $C = \sup g(\cdot)$, then it is easily to verify that the function $f(\cdot)$ is a submodular set function. With the newly defined function $f(\cdot)$, maximizing $g(\cdot)$ is equivalent to minimizing $f(\cdot)$.

The size-constrained submodular minimization problem is currently with no standardized approximation algorithm to the best of our knowledge [26], [27], [28]. We find two algorithms published recently [26], [27], and examine whether they could be used for resolving our problem with guaranteed performance. In the following, we will show why the two state-of-the-art approximation algorithms are unable to provide the guaranteed approximation ratio, which motivates us to develop our algorithm in Section 4.2.

Iyer et al. propose an approximation algorithm for minimizing the submodular set function [27]. Recall the definition of $f(\cdot)$ mentioned above, the submodularity of $f(\cdot)$ can still

1. A set function $f(\cdot) : 2^E \to R$ defined on set E is a supermodular set function if for any set $B \subseteq A \subseteq E$, $f(A\cup\{i\}) - f(A) \ge f(B\cup\{i\}) - f(B)$, where i is an element in set E but not in set A.

hold even without the constant C. The purpose of C is just to ensure that the value of $f(\cdot)$ is non-negative, which is required by [27]. With Iyer's approximation algorithm, the solution can be obtained g_o and the optimal solution g^* satisfies

$$\frac{C-g_o}{C-g^*} \leq 1+\varepsilon, \tag{5}$$

where ε is a constant close to 0. Since C can be a very large constant, the algorithm is unable to ensure a good approximation ratio to the optimal solution for the problem.

Nagano et al. propose a size-constrained submodular minimization algorithm by leveraging the minimum norm base [26]. We here briefly describe the basic idea of the algorithm, and then show why the algorithm can not be applied in the AP selection scenario. The submodular set function can be mapped into a N-dimensional polyhedron, with N being the number of elements in the set. An example of polyhedron corresponding to a submodular set function could be found in our technical report [29].

Theorem 2. *For submodular set function $f(\cdot)=2g(E)-g(\cdot)$, the minimum norm base $\overrightarrow{b}=(b_1,b_2\ldots b_N)$ is a vector with the values of all the components being the same.*

Proof. According to the definition of minimum norm base, the components $b_1,b_2\ldots b_N$ must satisfy the following polyhedron constraints

$$\sum_{i\in S,S\not\subset E} b_i \leq 2\left(\sum_{i\in E}\left|s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|\right)^2-\left(\sum_{i\in S}\left|s_{\lfloor\frac{i-}{l}+1\rfloor}\right|\right)^2 + \left|\sum_{i\in S} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2-2\left|\sum_{i\in E} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2;$$

$$\sum_{i\in E} b_i=\left(\sum_{i\in E}\left|s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|\right)^2-\left|\sum_{i\in E} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2.$$

Now we first prove that the vector $\overrightarrow{b^*}=(b_i^*)_{1\leq i\leq N}$ satisfying the following condition is in the polyhedron P_g

$$b_1^*=b_2^*\ldots=b_N^*=\frac{\left(\sum_{i\in E}\left|s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|\right)^2-\left|\sum_{i\in E} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2}{N}.$$

It is straightforward to verify that b_i satisfies the second condition of the P_f definition, that is, $\sum_{i=1}^N b_i=f(E)$.

Note that

$$\begin{aligned}\frac{f(S)}{|S|}=&\frac{2\left(\sum_{i\in E}\left|s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|\right)^2-\left(\sum_{i\in S}\left|s_{\lfloor\frac{i-}{l}+1\rfloor}\right|\right)^2}{|S|}\\&+\frac{\left|\sum_{i\in S} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2-2\left|\sum_{i\in E} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2}{|S|}\\&\overset{(a)}{\geq}\frac{\left(\sum_{i\in E}\left|s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|\right)^2-\left|\sum_{i\in E} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2}{|S_i|}\\&\overset{(b)}{\geq}\frac{\left(\sum_{i\in E}\left|s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|\right)^2-\left|\sum_{i\in E} s_{\lfloor\frac{i-1}{l}+1\rfloor}\right|^2}{|E|}\\&=b_i.\end{aligned} \tag{6}$$

Since the function $g(\cdot)$ is continuously increasing, the inequality (a) holds as $g(S)\leq g(E)$. The inequality (b) holds because $|S|\leq|E|$. For any set S, inequality $\sum_{i\in S} b_i\leq|S|f(S)$ is established, and thus the point represented by vector $\overrightarrow{b^*}$ is in the polyhedron P_f.

According to Cauchy-inequality, we have

$$\left\|\overrightarrow{b}\right\|=\sum_{i=1}^N b_i^2\geq\frac{\left(\sum_{i=1}^N b_i\right)^2}{N}=\left\|\overrightarrow{b^*}\right\|,$$

therefore, $\overrightarrow{b^*}$ is the minimum norm base of function $g(\cdot)$. □

The theorem above indicates that the algorithm does not fit the AP selection problem under study. The theorem shows that all the components in the minimum norm base of $f(\cdot)$ are the same for the AP selection problem, which makes $|T_i|=N$ in this case. This is to select N elements in a N-norm set, which is meaningless. In the theorem above, we transform the function $g(\cdot)$ into $f(\cdot)=2g(E)-g(\cdot)$, which is a submodular set function but does not change the essence of the problem.

4.2 Algorithm for AP Selection

We propose Algorithm 1 to solve the AP selection problem. We are to first present detailed explanation of the algorithm and then theoretically analyze performance of the algorithm.

Algorithm 1. Near Optimal Strategy for Fingerprints Reporting

Require: complex number vector $(z_1,z_2\ldots z_m)$ where $z_j=p_je^{i\phi_j}(0\leq\phi_j\leq2\pi)$ represent the ith AP.
Ensure: A measurement sequence $(n_1,n_2\ldots n_l)$, where $1\leq n_i\leq l$ represent choosing the n_ith AP.

```
S ← ∅
w_i = 0 (1 ≤ i ≤ lm)
z* = 1
for i = 1 to lm do
    σ(i) ← ⌊(i−1)/l + 1⌋
end for
for i = 1 to l do
    for i ∈ [lm]\S do
        w_j ← w_j + p_σ(j) |z*| sin²(φ_σ(j) − arg(z*))
    end for
    i* ← arg max_{i∈[lm]\S} w_i
    z* ← s_σ(i*);
    S ← S ∪ {i*};
end for
return (σ(i))_{i∈S}
```

The input of the algorithm is an m-dimensional vector, where each component of the vector is a complex number. The complex number z_i of the ith dimension represents the characteristic of AP_i. The output of the algorithm is an l-dimensional vector, where each component of the vector is an integer that is greater than or equal to 1 and less than or equal to l. The output vector shows how many times each AP should be measured in the online phase, e.g., the output vector $(1,2,2,3,1,1,4)$ indicates that: if there are totally seven times of measuring opportunities, to measure AP_1 three times, AP_2 twice, AP_3 once and AP_4 once respectively can obtain the most accurate location estimation. The order of the measurements is not important, since the order does not change the value of the objective function.

In the algorithm, we first initialize the relevant intermediate variables to be used. We use set S to store the set of selected APs; w_i represents the weight of each element in the set $[lm]$. The algorithm will continue to update the weights of the elements in the following steps, which is the basis to decide whether the corresponding AP will be selected. The last variable processed in the initialization stage is z^*, which is also a complex variable and the initial value is 1.

In the fourth row of the algorithm, we map the elements in the set $[lm]$ to a corresponding AP, which is for the rigorousness of narrative. This is because the set function requires each element in the set to be unique, but the fingerprints reporting strategy studied in the paper allows an AP to be measured multiple times. After the mapping, each element in the set $[lm]$ can be considered unique, and it is straightforward that choosing element i in the set $[lm]$ is equivalent to select AP with the index $i(\bmod\ m)$.

The 6th and 7th rows of the algorithm represented by the "For" loop is the main part. In the 7th and 8th row of the algorithm, at the beginning of each loop, the algorithm updates the weight w_i of each element, increases it by $p_{\sigma(j)}|z^*|\sin^2(\phi_j - \arg(z^*))$, and z^* are selected from the last execution of each iteration.

In row 9, the algorithm selects the element with the largest weight in the remaining elements, which is indexed by i^*. Then, the algorithm copies the i^*th corresponding complex number $z_{\sigma(i^*)}$ representing AP_{i^*} to the intermediate variable z^*. Recall that finding the element with the largest weight is factually finding the complex number $z_{\sigma(i^*)}$ that can most effectively decreases the size of the ellipse as described in Fig. 1 of Section 2. Let u_k and u_{k+1} denote the areas of the ellipses that the user will be localized in the physical space when using APs in set $\{1, 2 \ldots k\}$ and $\{1, 2 \ldots k, k+1\}$ respectively, then we have

$$\begin{aligned} u_k - u_{k+1} &= \frac{1}{2}\int_0^{2\pi} (\rho_k^2(\theta) - \rho_{k+1}^2(\theta))d\theta \\ &= 2c^2 \int_0^{2\pi} \left(\frac{1}{\sum_{i=1}^{k} p_i^2 \cos^2(\theta - \phi_i)} - \frac{1}{\sum_{i=1}^{k+1} p_i^2 \cos^2(\theta - \phi_i)} \right) d\theta \\ &\propto \sum_{i=1}^{k} p_{k+1} p_i \sin^2(\phi_{k+1} - \phi_i). \end{aligned}$$

This is the fundamental reason row 9 of the algorithm is so designed.

In row 11, the algorithm merges the i^*th element into the set S, and then enter the next execution of the loop. After looping l times, we can get a set S of l elements, and finally we have the sequence $\sigma(i)$, which indicates how many times which AP should be measured in the online phase.

4.3 Algorithm Performance Analysis

We first define a new function $h(,) : 2^{[lm]} \times 2^{[lm]} \to R$, where the arguments of the function consist two sets, and the value of the function is a real number. For any two sets A, B, the value of the function of the two sets is equal to the sum of the results of the same operation between each corresponding element in the two sets

$$h(A, B) = \sum_{i \in A, j \in B} p_{\sigma(i)} p_{\sigma(j)} \sin^2(\phi_{\sigma(i)} - \phi_{\sigma(j)}). \tag{7}$$

There is close relationship between function $h(,)$ and function $g(\cdot)$, as shown in the following lemma.

Lemma 1. *For any two sets $A \in [lm]$ and $B \in [lm]$,*

$$g(A \cup B) = g(A) + g(B) - g(A \cap B) + h(A \backslash B, B \backslash A). \tag{8}$$

Proof.

$$\begin{aligned} g(A \cup B) &= g(A \backslash B) + g(B \backslash A) + g(A \cap B) + h(A \backslash B, A \cap B) \\ &\quad + h(B \backslash A, A \cap B) + h(A \backslash B, B \backslash A) \\ &= [g(A \backslash B) + g(A \cap B) + h(A \backslash B, A \cap B)] \\ &\quad + h(A \backslash B, B \backslash A) - g(A \cap B) \\ &\quad + [g(B \backslash A) + g(A \cap B) + h(B \backslash A, A \cap B)] \\ &= g(A) + g(B) + h(A \backslash B, B \backslash A) - g(A \cap B). \end{aligned} \tag{9}$$

According to the definition of $h(,)$ and $g(\cdot)$, it is straightforward to verify each step of the formula above. □

The lemma above provides a more precise description of supermodularity. If we replace the equality with the inequality of the function $h(,)$, it degenerates to the definition of supermodularity. We are to estimate the value of $h(,)$. In fact, estimating function $h(,)$ is to measure how "supermodular" the function $g(\cdot)$ is. Function $g(\cdot)$ is more supermodular with $h(,)$ being larger.

For the convenience of observation, we first transform the function $h(,)$ into the following:

$$\begin{aligned} h(A, B) &= \sum_{i \in A m j \in B} p_{\sigma(i)} p_{\sigma(j)} \sin^2(\phi_{\sigma(i)} - \phi_{\sigma(j)}) \\ &= \sum_{i \in A} \left[p_{\sigma(i)} \sum_{j \in B} p_{\sigma(j)} \sin^2(\phi_{\sigma(i)} - \phi_{\sigma(j)}) \right]. \end{aligned}$$

Note that

$$\begin{aligned} & p_{\sigma(i)} \sum_{j \in B} p_{\sigma(j)} \sin^2(\phi_{\sigma(i)} - \phi_{\sigma(j)}) \\ &= \frac{1}{2} \left\{ p_{\sigma(i)} \sum_{j \in B} p_{\sigma(j)} - \mathrm{Re}\left[s_{\sigma(i)} \cdot \sum_{j \in B} s_{\sigma(j)} \right] \right\}, \end{aligned} \tag{10}$$

where $\mathrm{Re}(\cdot)$ denotes the real part of a complex number. Note that the right part of the equation $\mathrm{Re}[s_{\sigma(i)} * \sum_{j \in B} s_{\sigma(j)}]$ is the real part of the product between complex number $s_{\sigma(i)}$ and the sum of all complex numbers in set B. It is obvious that the equation is less than or equal to the product of the complex number $s_{\sigma(i)}$'s norm and the norm of the sum of all complex numbers in set B; therefore,

$$\begin{aligned} h(A, B) &\geq \sum_{i \in A} \sum_{j \in B} p_{\sigma(i)} p_{\sigma(j)} - \sum_{i \in A} \left(p_{\sigma(i)} \left| \sum_{j \in B} s_{\sigma(i)} \right| \right) \\ &\geq \left(\sum_{i \in A} p_{\sigma(i)} \right) \left(\sum_{j \in B} p_{\sigma(j)} - \left| \sum_{j \in B} s_{\sigma(i)} \right| \right). \end{aligned}$$

Meanwhile, we note that

$$\sum_{j \in B} p_{\sigma(j)} - \left| \sum_{j \in B} s_{\sigma(i)} \right| = \frac{g(B)}{\sum_{j \in B} p_{\sigma(j)} + \left| \sum_{j \in B} s_{\sigma(i)} \right|},$$

thus we have

$$h(A,B) \geq \frac{\sum_{i\in A} p_{\sigma(i)}}{\sum_{j\in B} p_{\sigma(j)} + |\sum_{j\in B} s_{\sigma(i)}|} g(B) \overset{(a)}{\geq} \frac{\sum_{i\in A} p_{\sigma(i)}}{2\sum_{j\in B} p_{\sigma(j)}} g(B).$$

The inequality (a) holds because $|\sum_{j\in B} s_{\sigma(i)}| \leq \sum_{j\in B} p_{\sigma(j)}$. Similarly, we have

$$h(A,B) \geq \frac{\sum_{i\in B} p_{\sigma(i)}}{2\sum_{j\in A} p_{\sigma(j)}} g(A). \tag{11}$$

We define a new set function $\lambda(\cdot): 2^{[lm]} \rightarrow R$, then for any set A, $\lambda(A)$ equals the norm sum of all elements in set A, that is, $\lambda(A) = \sum_{j\in A} p_{\sigma(j)}$. Besides, we let $m = \min_{i\in[lm]} p_{\sigma(i)}$, $M = \max_{i\in[lm]} p_{\sigma(i)}$, which will facilitate the following analysis of the algorithm. With the conclusions of the derivations above, we have $h(A,B) \geq \frac{1}{2}\frac{\lambda(B)}{\lambda(A)} g(A)$, and symmetrically we have

$$h(A,B) \geq \frac{1}{2}\frac{\lambda(A)}{\lambda(B)} g(B).$$

Consequently, we have

$$\begin{aligned} h(A,B) &\geq \frac{1}{4}\frac{\lambda(B)}{\lambda(A)} g(A) + \frac{\lambda(A)}{\lambda(B)} g(B) \\ &= \frac{1}{2}\sqrt{g(A)g(B)} \\ &\geq \frac{1}{2}\min\{g(A), g(B)\}. \end{aligned}$$

We here introduce another concept, *curvature*, to be used in the approximation ratio analysis. Similar to the submodular function, the curvature of supermodular set function is defined as following.

Definition 1. *The curvature of a supermodular function $g(\cdot)$ is*

$$\kappa_g = 1 - \min_{j\in E} \frac{g(j)}{g(E\backslash j)}, \tag{12}$$

where $g(E\backslash j) = g(E) - g(E\backslash j)$.

Theorem 3. *We use u_{out} to denote the ellipse area corresponding to the fingerprint reporting strategy yielded by Algorithm 1, and u_{opt} to denote the ellipse area corresponding to the optimal reporting strategy, then*

$$\frac{u_{out}}{u_{opt}} \leq \frac{1}{\sqrt{1-\delta}} = 1 + \frac{\delta}{2} + O(\delta^2).$$

That is, the approximation ratio of Algorithm 1 is $1+\frac{\delta}{2}+O(\delta^2)$, and $\delta = 1 - \max\{[(1+\frac{m}{2M})e^{-(1-\kappa)} - \frac{m}{2M}], (1+\frac{m}{2Ml})e^{-(1-\kappa)l} - \frac{m}{2Ml}\}$.

Proof. Assuming that the optimal solution of the optimization problem is the set S^*. In the second "For" loop of the algorithm, the algorithm will select an element i^*, adding it into set S at the end of each execution of the loop. We assume that the set yielded from step i is S_i, which partially intersects with S^*

$$\begin{aligned} g(S^*) &= g(S^* \cup S_i) + g(S^* \cap S_i) - g(S_i) - h(S^*\backslash S_i, S_i\backslash S^*) \\ &\overset{(a)}{=} g(S^* \cup S_i) - g(S_i\backslash S^*) - h(S_i \cap S^*, S_i\backslash S^*) \\ &\quad - h(S^*\backslash S_i, S_i\backslash S^*) \\ &\overset{(b)}{=} g(S^* \cup S_i) - g(S_i\backslash S^*) - h(S^*, S_i\backslash S^*) \\ &\overset{(c)}{\leq} g(S_i) + \frac{|S^*\backslash S_i|}{1-\kappa}(g(S_{i+1}) - g(S_i)) - g(S_i\backslash S^*) \\ &\quad - h(S^*, S_i\backslash S^*) \\ &\overset{(d)}{\leq} g(S_i) + \frac{|S^*\backslash S_i|}{1-\kappa}(g(S_{i+1}) - g(S_i)) - \frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)} g(S^*). \end{aligned} \tag{13}$$

Merging $g(S^*)$ from both sides of the inequality, we have the following result

$$\left(1 + \frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right) g(S^*) \leq \frac{|S^*\backslash S_i|}{1-\kappa} g(S_{i+1}) + \left(1 - \frac{|S^*\backslash S_i|}{1-\kappa}\right) g(S_i).$$

After appropriate transposition, we can transform the inequality into

$$g(S_{i+1}) \geq \frac{1 + \frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}}{\frac{|S^*\backslash S_i|}{1-\kappa}} g(S^*) - \left(\frac{1-\kappa}{|S^*\backslash S_i|} - 1\right) g(S_i).$$

We can regard the inequality above as an iterative inequality of the set S_i, and now we do some transposition to the inequality as following:

$$\begin{aligned} &\left(1 + \frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right) g(S^*) - g(S_{i+1}) \\ &\leq \left(1 - \frac{1-\kappa}{|S^*\backslash S_i|}\right)\left[\left(1 + \frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right) g(S^*) - g(S_i)\right]. \end{aligned} \tag{14}$$

Using the formula above, we can perform the computation iteratively until reaching the case of $i = 0$. Note that $S_0 = \emptyset$, thus the following inequality holds for any i

$$\begin{aligned} &\left(1 + \frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right) g(S^*) - g(S_i) \\ &\leq \left(1 + \frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right) g(S^*) \prod_{j=0}^{i-1}\left(1 - \frac{1-\kappa}{|S^*\backslash S_j|}\right). \end{aligned} \tag{15}$$

In particular, if $i = l$, we have

$$\begin{aligned} g(S_l) &\geq \left(1 - \prod_{j=0}^{l-1}\left(1 - \frac{1-\kappa}{|S^*\backslash S_j|}\right)\right)\left(1 + \frac{\lambda(S_l\backslash S^*)}{2\lambda(S^*)}\right) g(S^*) \\ &\overset{(a)}{\geq} \left(1 - \left(1 - \frac{1-\kappa}{|S^*\backslash S_l|}\right)^l\right)\left(1 + \frac{\lambda(S_l\backslash S^*)}{2\lambda(S^*)}\right) g(S^*) \\ &\overset{(b)}{\geq} \left(1 - \left(1 - \frac{1-\kappa}{|S^*\backslash S_l|}\right)^l\right)\left(1 + \frac{|S_l\backslash S^*|m}{2lM}\right) g(S^*). \end{aligned}$$

The inequality (a) can hold because increasing $|S^*\backslash S_j|$ to $|S^*\backslash S_l|$ will make the entire equation smaller in value. According to the definition of function $\lambda(\cdot)$, we have $\lambda(S_l\backslash S^*) \geq |S_l\backslash S^*|m$ and $\lambda(S^*) \leq |S^*|M = lM$, thus the inequality (b) also holds. Note that $|S_l| = |S^*| = l$, thus $|S^*\backslash S_l| = |S_l\backslash S^*|$. We use k to represent the number of elements in the two sets. One more step and we will have

the following:

$$\begin{aligned} g(S_l) &\geq \left(1-\left(1-\frac{1-\kappa}{k}\right)^l\right)\left(1+\frac{km}{2lM}\right)g(S^*) \\ &\overset{(a)}{\geq} \left(1-e^{-\frac{(1-\kappa)l}{k}}\right)\left(1+\frac{km}{2lM}\right)g(S^*) \\ &\overset{(b)}{=} (1-e^{-(1-\kappa)t})\left(1+\frac{m}{2Mt}\right)g(S^*). \end{aligned}$$

As $(1-\frac{1-\kappa}{k})^l \leq e^{-\frac{(1-\kappa)l}{k}}$, inequality (a) holds. Note that the right hand side of the inequality is actually a function with respect to k. For the convenience of demonstration, we let the variable $t=\frac{l}{k}$, and we have equation (b) with $t\in[1,l]$. Let $q(t)=(1-e^{-(1-\kappa)t})(1+\frac{m}{2Mt})$, and we find the minimum of $q(t)$ by finding t's derivative. Although t is a discrete variable according to the definition of t, we can relax it to a continuous variable when finding the extreme of the function; therefore, the analysis can be facilitated by the derivative. The first order derivative of t in function $q(\cdot)$ is

$$\frac{dq(t)}{dt}=\frac{e^{-(1-\kappa)t}[(1-\kappa)(t^2+\frac{mt}{2M})+\frac{m}{2M}]-\frac{m}{2M}}{t^2}.$$

We use $q_1(\cdot)$ to denote the numerator of the formula above, and find the derivative of function $q_1(\cdot)$

$$\frac{dq_1(t)}{dt}=(1-\kappa)te^{-(1-\kappa)t}\Big[2-\ (1-\kappa)\Big(t+\frac{m}{2M}\Big)\Big].$$

It is straightforward that $\frac{dq_1(t)}{dt}\geq 0$, $\frac{dq(t)}{dt}\geq 0$, when $0\leq t\leq \frac{2}{1-\kappa}-\frac{m}{2M}$; $\frac{dq_1(t)}{dt}\leq 0$, when $t\geq \frac{2}{1-\kappa}-\frac{m}{2M}$. This indicates that $q(t)$ have at most one zero point; therefore,

$$\begin{aligned} q(t) &\geq \min\{q(1),q(l)\} \\ &= \min\Bigg\{\left(1-e^{-(1-\kappa)}\right)\left(1+\frac{m}{2M}\right), \\ &\quad (1-e^{-(1-\kappa)l})\left(1+\frac{m}{2Ml}\right)\Bigg\} \\ &= 1-\delta, \end{aligned}$$

where $\delta=1-\max\{[(1+\frac{m}{2M})e^{-(1-\kappa)}-\frac{m}{2M}],(1+\frac{m}{2Ml})e^{-(1-\kappa)l}-\frac{m}{2Ml}\}$. Based on the results above, we have

$$\frac{u_{out}}{u_{opt}}=\frac{\frac{8\pi c^2}{\sqrt{\left(\sum_{s_i\in S_l}|s_i|\right)^2-\left|\sum_{s_i\in S_l}s_i\right|^2}}}{\frac{8\pi c^2}{\sqrt{\left(\sum_{s_i\in S^*}|s_i|\right)^2-\left|\sum_{s_i\in S^*}s_i\right|^2}}}\leq\frac{1}{\sqrt{q(t)}}.$$

Note that the Taylor expansion of function $\frac{1}{\sqrt{1+x}}$ at $x=0$ is

$$\frac{1}{\sqrt{1+x}}=1+\sum_{i=1}^{\infty}(-1)^i\frac{(2i-1)!!}{(2i)!!}x^i=1-\frac{x}{2}+O(x^2),$$

therefore,

$$\frac{u_{out}}{u_{opt}}\leq\frac{1}{\sqrt{1-\delta}}=1+\frac{\delta}{2}+O(\delta^2).$$

The approximation ratio is $1+\frac{\delta}{2}+O(\delta^2)$, and $\delta=1-\max\{[(1+\frac{m}{2M})e^{-(1-\kappa)}-\frac{m}{2M}],(1+\frac{m}{2Ml})e^{-(1-\kappa)l}-\frac{m}{2Ml}\}$. □

The algorithm analysis above indicates that every time a new element is picked out from the remaining set, the area of u of the ellipse is constantly decreasing, which means that the accuracy is getting higher and the corresponding difference from the optimal value is decreasing. If the number of selected elements has reached l, when continuing to select the new element to add in according to the selection method of the algorithm, the error will continue to be reduced. It is easy to verify that the inequality (15) derived from the analysis above also holds for the situation of $i>l$. We have

$$\begin{aligned} g(S_i) &\geq \left(1+\frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right)\left(1-\prod_{j=0}^{i-1}\left(1-\frac{1-\kappa}{|S^*\backslash S_j|}\right)\right)g(S^*) \\ &\geq \left(1+\frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right)\left(1-\left(1-\frac{1-\kappa}{l}\right)^i\right)g(S^*) \\ &\geq \left(1+\frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)}\right)(1-e^{-(1-\kappa)\frac{i}{l}})g(S^*). \end{aligned}$$

For any given constant ε close to 0, $g(S_i)\geq(1+\frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)})(1-e^{-(1-\kappa)\frac{i}{l}})g(S^*)\geq(1-\varepsilon)g(S^*)$, if $i\geq\frac{l\ln\frac{1}{\varepsilon}}{(1-\kappa)}$. That is, if the number of reported signal fingerprints is large enough to make $g(S_i)\geq(1+\frac{\lambda(S_i\backslash S^*)}{2\lambda(S^*)})(1-e^{-(1-\kappa)\frac{i}{l}})g(S^*)\geq(1-\varepsilon)g(S^*)$, then the final localization accuracy is at most $1-\varepsilon$ times worse than the optimal value.

Now we briefly analyze the time complexity of the algorithm: it is obvious that the algorithm is with polynomial complexity. In the rows 4 and 5 of the algorithm, which is the element mapping stage, there are lm loops, with each can be finished in $O(1)$, thus the complexity is $O(lm)$. In rows 7 to 11 of the algorithm, the dominant factor of the complexity is the "For" loop to update weights, which is nested in another "For" loop. The execution of each loop needs time complexity of $O(lm)$. The complexity of the algorithm's rows 6 to 11 is $O(l^2m)$. Finally, during the algorithm calculating the sequence of returned values, for all elements in set S, the corresponding AP must be found, for which the time complexity is $O(l)$. Consequently, the overall complexity of the algorithm is $O(l^2m)$.

5 Applications of the Best Strategy

5.1 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

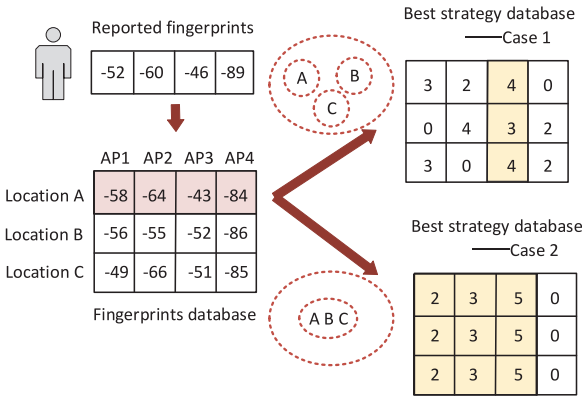


Fig. 2. Location determination facilitated by the best strategy.

same, best reporting strategies for such locations should be the same.

Such a seemingly frustrating phenomenon factually also can be leveraged to reduce errors in location estimation. Empirical studies show that the estimation errors of pure fingerprinting based localization system could be over 6m [9], [37], [38]. 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 mitigate the impact of 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 [9], [37].

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. 2. 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 cases 1 and 2 in Fig. 2 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. 2, 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 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.

Algorithm 2. Location Determination Strategy

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

Now we consider the case 2 in Fig. 2, 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 6 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. 2, 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.

5.2 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 [39]. 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. [40] 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.

5.2.1 AP Deployment for Localization

If we want to find 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

$$R(\vec{r}) = \left(\sum_{i\in\mathcal{V}_n^*} |Z_i|\right)^2 - \left(\left|\sum_{i\in\mathcal{V}_n^*} Z_i\right|\right)^2, \tag{16}$$

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 Eqs. (2) and (16). In the discussion of AP deployment problem, we assume that the system always follows the best fingerprints reporting strategy, so that the system accuracy can be maximized.

Eq. (16) 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 $\{\vec{x}, \vec{y}\} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ to denote locations of APs. Consider the localization performance at location $\vec{r}$, which is denoted by $R(\vec{r})$. Suppose the user appears 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:

$$\{\vec{x}, \vec{y}\}^* = \arg\max_{\{\vec{x},\vec{y}\}} \int_S R(\vec{r})f(\vec{r})dr^2, \tag{17}$$

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

$$\{\vec{x}, \vec{y}\}^* = \arg\max_{\{\vec{x},\vec{y}\}\subset\{\vec{X},\vec{Y}\}} \int_S R(\vec{r})f(\vec{r})dr^2, \tag{18}$$

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

The transformed problem can be resolved by a simulated annealing based algorithm as shown in Algorithm 3.

Algorithm 3. AP Deployment Strategy

1: Initialize $\{\vec{x}, \vec{y}\}$(Select c locations randomly).
2: **while** $T_t \geq t^*$ **do**
3: Generate new strategy by randomly change one location in $\{\vec{x}, \vec{y}\}$.
4: Calculate expected reliability H for the new and old state.
5: Accept the new state according to H_{new} and H_{old} with probability $\min\{1, e^{\frac{\mathcal{H}_{new}-\mathcal{H}_{new}}{T_t}}\}$.
6: Update temperature T_t.
7: **end while**
8: $\{\vec{x}, \vec{y}\}^* \leftarrow \{\vec{x}, \vec{y}\}$.

5.2.2 Localization and Coverage

In previous work in the literature, access point deployment problem is considered 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.

Theorem 4. *If an AP deployment plan in a region with area D satisfies the localization accuracy R^*, then every point of the region 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 $\frac{2\sigma\sqrt{3R^*}}{9N_0}(D-2)$ 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 $\forall\vec{P}, \forall AP_i \in \mathbb{U}_l, d(\vec{P}, AP_i) > d$, then $R(\vec{p}) < \frac{N_0^2}{d^4\sigma^4}$, where $\vec{P}$ is a point in the region and $\sigma = \min_i\{\sigma_i\}$.

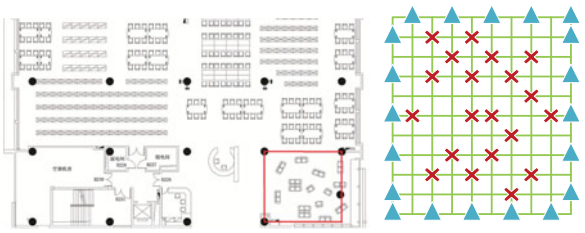
According to (16), the localization accuracy is determined by the characteristic vector of the AP, thus

$$\begin{aligned} R(\vec{P}) \leq \left(\sum_{i\in\mathcal{V}_l} |Z_i|\right)^2 &= \left(\sum_{i\in\mathcal{V}_l} p_i\right)^2 = \left(\sum_{i\in\mathcal{V}_l} \frac{(\nabla\mu_i)^2}{\sigma_i^2}\right)^2 \\ &= \left(\sum_{i\in\mathcal{V}_l} \frac{1}{r_i^2\sigma_i^2}\right)^2 \leq \frac{N_0^2}{d^4\sigma^4}. \end{aligned} \tag{19}$$

The equality in (19) holds when the location is at the center of several equidistant APs with the same RSS variance σ.

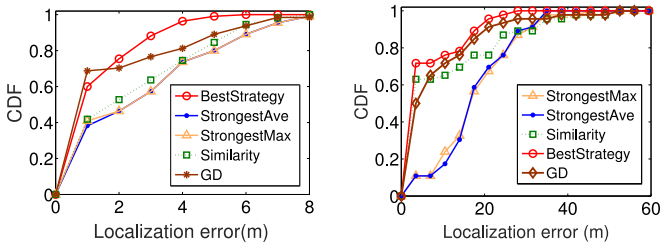
If a point can be localized with required accuracy level R^*, then there must be an AP within $\frac{\sqrt{N_0}}{\sigma\sqrt{[4]R^*}}$. To guarantee that every point lies in a coverage range of some APs, the AP deployment strategy must satisfy $\exists AP_i \in \mathbb{U}_l, d(\vec{P}, AP_i) < \frac{\sqrt{N_0}}{\sigma\sqrt{[4]R^*}}$.

The second part 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



(a) Floor plan. (b) AP deployment.

Fig. 3. Experiment field.



(a) Resutls of local experiments. (b) Results of trace data experiments.

Fig. 4. Results of localization.

small square areas, the problem is reducible to circle covering problem which has been proved by Kersher [41].

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

$$\pi a^2 \psi(a) > \frac{2\pi\sqrt{3}}{9}(D - 2\pi a^2), \tag{20}$$

where D denotes the area of the rectangle. Together with (19), we finally get the minimum required number is

$$\psi > \frac{2\sigma\sqrt{3R^*}}{9N_0}(D-2). \tag{21}$$

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. □

Eq. (21) 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.

6 Performance Evaluation

In this section, we evaluate the proposed algorithms with both local and trace data experiments. We conduct local experiments in an area of our university's library as shown in Fig. 3a. The area in the red frame is grided into 10×11 cells and the edge length of each cell is 70 cm (Note that the black nodes represent pillars.). The APs used in the experiment are uniformly distributed along the edges of the region first, as those blue triangles shown in Fig. 3b, and then deployed according to our proposed AP deployment strategy, as those red crosses. We also conduct experiments with the trace data collected by the EVARILOS testbed [46]. The data are collected in an unmanned utility room with many metal objects termed as "Zwijnaarde", where there is almost no outside interference and no persons are present in the environment. Detailed descriptions of the testbed and data could be found in [47], [48], [49].

Performance of AP Selection Strategies in CDF. We first verify that the proposed best fingerprints reporting strategy indeed can improve accuracy of localization. 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, where the detailed description of the scheme could be found in [7].

The three strategies mentioned above are to select each AP based on the AP's own importance, where the APs' group effect is not taken into account. We note that a group-discrimination based (GD) AP selection mechanism is proposed recently [21], where the positioning capabilities of a group of APs are investigated. The basic idea of the GD mechanism is to find the best combination of APs, and use the fingerprints generated by those APs as the feature of each location. We also compare our proposed BestStrategy with the GD mechanism. In the experiments, we select the best combination consisting of 6 APs according to the GD algorithm in [21], and use the fingerprints of those 6 APs to distinguish one location from another in the offline phase. We also use the fingerprints from the same 6 APs to perform the online phase to estimate the device location.

Fig. 4a shows results of the local experiments, where the horizontal axis represents the localization error observed and the vertical axis represents the corresponding cumulative distribution function (CDF) value. In the experiment, we deploy 20 APs in the area with our best deployment strategy as shown in Fig. 3. Under such setting, we perform localization in 100 randomly-selected points. The results show that the best fingerprints reporting strategy yielded from our algorithm outperform other strategies. Fig. 4b shows experimental results with the trace data from the EVARILOS testbed. Since the deployment of APs and landmarks could not be changed, we just evaluate the AP selection strategy. We use the data observed at a part of the landmarks as the training set and that observed at the rest of the landmarks as the test set to perform localization. The results corroborate our local experiment results. Our proposed algorithm outperforms others because other strategies just exploit the statistical properties of the fingerprints, but our strategy is based on the intrinsic mechanism of fingerprinting localization.

It is shown that our proposed mechanism still outperforms the GD mechanism. This is because the process of finding the best combination in [21] is based on empirical

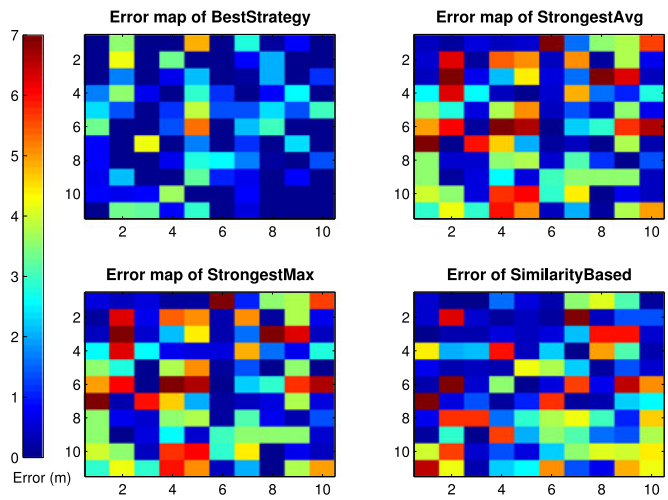


Fig. 5. Error map of local experiment.

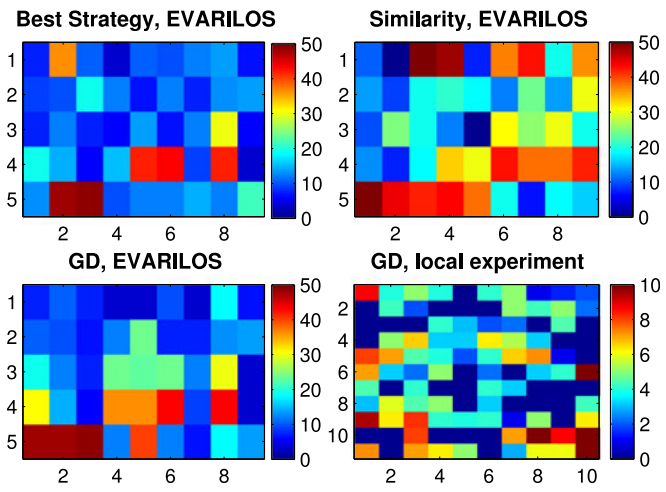


Fig. 6. Error map of selected algorithms.

metric. According to our theoretical analysis, the number of times for measurements with respect to each AP is also important, which however is not taken into account in GD [21]. It is noted that the performance of the StrongestMax and StrongestAvg are almost the same in both the EVARILOS data experiment and the local experiment. This is because a large proportion of APs selected by the two strategies are the same. If we use the localization accuracy of CDF 80 percent as the benchmark to evaluate the algorithms' performance. It shows that at 80 percent of the time in our local experiments, the localization error is within 2.2 m and 3.5 m with the BestStrategy and the second best GD strategy, respectively. In the trace data experiments, those numbers become 13 m and 17 m. That is, the BestStrategy outperforms the GD strategy by around 37 and 24 percent in the local and trace data experiments, respectively.

Performance of AP Selection Strategies in Error Map. Fig. 5 illustrates how the localization errors are distributed in the local experiment field. The experiment results with respect to the three strategies all confirm the results in [7]. Moreover, the results also indicate that the best strategy indeed incurs the smallest localization errors. Fig. 6 illustrates how the localization errors with selected mechanisms are distributed in the EVARILOS experiment field and the local experiment field. We select the BestStrategy, GD and SimilarityBased strategies. The experiment field of the EVARILOS testbed is a 56 m $\times$ 15 m rectangle space, and each cell in the figure represents a 6 m $\times$ 3.2 m space. All the experimental results show that the proposed BestStrategy outperforms other strategies.

It is worth mentioning that our experimental results are just based on pure fingerprints comparison; other assisting mechanisms such as motion sensors and acoustic ranging as described in Section 2 are not included. This is because the purpose of the paper is to investigate the fingerprints reporting strategy. Moreover, the results of trace data experiments are not as good as that of local experiments, this is due to the limitation of the trace data. It could be found in the data set that there are only a few RSS readings recorded on some landmarks, and at some landmarks, the RSS readings are exactly the same. This is perhaps because the data are collected by the automatic robot as described in [46].

AP Deployment Strategy. We conduct experiments to examine the effect of our propose AP deployment strategy. In the experiment field as shown in Fig. 3, we deploy 20 APs in two ways, where the first is to deploy them uniformly along the four edges, and the second is to follow the proposed AP deployment strategy. We then perform localization in the area for 100 times at 100 randomly-selected points. The results of localization are illustrated in Fig. 7, where the curve representing uniform deployment is labeled with "SquarePattern". It is straightforward that the proposed deployment strategy can help improving the localization performance.

Energy Consumption. We conduct 1-hour experiment with a HUAWEI MT7-TL00 and a Nexus 5 smartphone to measure power consumption of RSS fingerprints reporting strategies. We examine two kinds of strategies, where the first one selects 6 APs using our proposed algorithm, and the second one selects all 20 APs. For each strategy, we let the smartphone report the observed RSS fingerprints every 500 ms. Considering that the best strategy may vary in different locations as explained in Section 5.1 when the user is moving, we recompute the best strategy every 2 s to make sure the best strategy for each location is updated in time. This process lasts for 1 hour, during which the power consumption results are recorded using both the PowerTutor [50] and the Trepn Power Profiler APP [51] every 5 minutes. The results are shown in Fig. 8. Since the two kinds of smartphones tailored the Android OS in different ways, their basic energy consumptions are different; moreover, the energy consumption models adopted by the two APPs are different [50], thus the results by the two different APPs are not the same. However, it is clear that the energy consumption results of the two strategies in different scenarios are almost the same, which means that the energy consumption incurred by the computation of our strategy could be negligible.

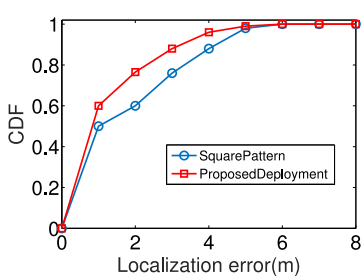


Fig. 7. Localization error CDF for purposed deployment strategy.

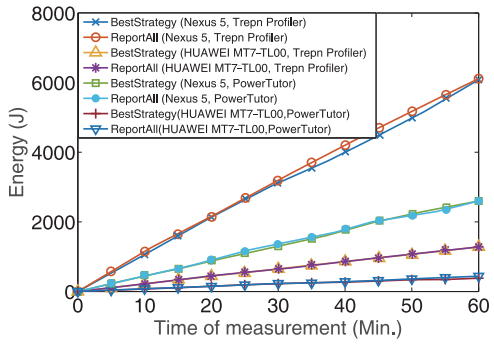


Fig. 8. Energy consumption.

7 Conclusions and Future Work

This paper have investigated how to optimize the fingerprints reporting strategy to improve localization accuracy, and how the optimal strategy theory can be utilized to streamline the design of WLAN fingerprinting localization systems. In particular, we have revealed that the fingerprints reporting problem is essentially an NP-Hard size-constrained supermodular maximization problem. We have proposed a new algorithm with a theoretical analysis of the corresponding approximation ratio. Moreover, we have demonstrated how the optimal 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, with an iterative algorithm developed to cross check fingerprints sampled in different locations, in order to derive the best possible result of localization. Further, we have revealed the relationship between accuracy of location estimation and coverage of Wi-Fi signals in indoor spaces when planning deployment of APs. We have presented comprehensive experiment results to validate our theoretical analysis. Our future work will focus on examining the influence of environmental change on the performance of fingerprinting based localization systems. We will pay particular attention to the human body shadowing effect with the corresponding aggravated multipath effect, which are supposed to change the noise level of the space dramatically.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61572319, U1405251, 61532012, 61325012, 61271219, 61428205). A preliminary version of this article appeared in the *Proceedings of the IEEE International Conference on Sensing, Communication, and Networking* (IEEE SECON 2016).

References

[1] C. Wu, Z. Yang, and Y. Liu, "Smartphones based crowdsourcing for indoor localization," *IEEE Trans. Mobile Comput.*, vol. 14, no. 2, pp. 444–457, Feb. 2015.

[2] Y. Wen, X. Tian, X. Wang, and S. Lu, "Fundamental limits of RSS fingerprinting based indoor localization," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 2479–2487.

[3] E. Elnahrawy, X. Li, and R. P. Martin, "The limits of localization using signal strength: A comparative study," in *Proc. IEEE Annu. IEEE Commun. Soc. Sensor Ad Hoc Commun. Netw.*, 2004, pp. 406–414.

[4] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE Conf. Comput. Commun.*, 2000, pp. 775–784.

[5] M. Youssef and A. Agrawala, "The horus WLAN location determination system," in *Proc. Int. Conf. Mobile Syst. Appl. Services*, 2005, pp. 205–218.

[6] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 269–280.

[7] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2010, pp. 173–184.

[8] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 293–304.

[9] H. Liu, et al., "Push the limit of WiFi based localization for smartphones," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 305–316.

[10] H. Liu, et al., "Push the limit of WiFi based localization for smartphones," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 305–316.

[11] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2010, pp. 173–184.

[12] S. Hilsenbeck, D. Bobkov, G. Schroth, R. Huitl, and E. Steinbach, "Graph-based data fusion of pedometer and WiFi measurements for mobile indoor positioning," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 147–158.

[13] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie-Markie: Indoor pathway mapping made easy," in *Proc. USENIX Conf. Netw. Syst. Des. Implementation*, 2013, pp. 85–98.

[14] H. Wang, et al., "No need to war-drive: Unsupervised indoor localization," in *Proc. Int. Conf. Mobile Syst. Appl. Services*, 2012, pp. 197–210.

[15] Y. Kim, H. Shin, and H. Cha, "Smartphone-based Wi-Fi pedestrian-tracking system tolerating the RSS variance problem," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2012, pp. 11–19

[16] W. Sun, J. Liu, C. Wu, Z. Yang, X. Zhang, and Y. Liu, "MoLoc: On distinguishing fingerprint twins," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, Jul. 2013, pp. 226–235.

[17] S. He and S. H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 466–490, Jan.–Mar. 2016.

[18] M. A. Youssef, A. Agrawala, and A. U. Shankar, "WLAN localization determination via clustering and probability distributions," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2003, pp. 143–150.

[19] Y. Chen, Q. Yang, J. Yin, and X. Chai, "Power-efficient access-point selection for indoor location estimation," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 7, pp. 877–888, Jul. 2006.

[20] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 689–705, Jun. 2007.

[21] T. Lin, S. Fang, W. Tseng, C. Lee, and J. Hsieh, "A group-discrimination-based access point selection for WLAN fingerprinting localization," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 3967–3976, Oct. 2014.

[22] A. Krishnakumar and P. Krishnan, "On the accuracy of signal strength-based estimation techniques," in *Proc. IEEE Conf. Comput. Commun.*, 2005, pp. 642–650.

[23] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," in *Proc. IEEE Conf. Comput. Commun.*, 2004, pp. 1012–1022.

[24] M. A. Youssef and A. Agrawala, "On the optimality of WLAN location determination systems," in *Proc. Commun. Netw. Distrib. Syst. Model. Simul. Conf.*, 2003, pp. 1–6.

[25] M. A. Youssef, A. Agrawala, and A. Udaya Shankar, "WLAN location determination via clustering and probability distributions," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2003, pp. 143–150.

[26] K. Nagano, Y. Kawahara, and K. Aihara, "Size-constrained submodular minimization through minimum norm base," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 977–984.

[27] B. Iyer, S. Jegelka, and J. Bilmes, "Fast semidifferential-based submodular function optimization," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. III-855–III-863.

[28] Z. Svitkina and L. Fleischer, "Submodular approximation: Sampling-based algorithms and lower bounds," May 2010. [Online]. Available: https://arxiv.org/pdf/0805.1071.pdf

[29] Technical report, 2016. [Online]. Available: https://www.dropbox.com/s/nuu7i6ggd9pmjqy/IEEE-TMC-Best
[30] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surveys*, vol. 46, no. 2, 2013, Art. no. 25.
[31] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a signle WiFi access point," in *Proc. USENIX Conf. Netw. Syst. Des. Implementation*, 2016, pp. 165–178.
[32] M. Kotaru, K. Josi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proc. ACM SIGCOMM*, 2015, pp. 269–282
[33] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. USENIX Conf. Netw. Syst. Des. Implementation*, 2013, pp. 71–84
[34] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, pp. 53–53, Jan. 2011.
[35] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing," in *Proc. IEEE Conf. Decision Control*, 1985, pp. 761–767.
[36] Z. Zhang, D. Liu, S. Zhu, S. Chen, and X. Tian, "Squeeze more from fingerprints reporting strategy for indoor localization," in *Proc. Annu. IEEE Int. Conf. Sensing Commun. Netw.* 2016, pp. 1–9.
[37] H. Liu, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Accurate WiFi based localization for smartphones using peer assistance," *IEEE Trans. Mobile Comput.*, vol. 13, no. 10, pp. 2199–2214, Oct. 2014.
[38] X. Zheng, H. Liu, J. Yang, Y. Chen, R. P. Martin, and X. Li, "A study of localization accuracy using multiple frequencies and powers," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 8, pp. 1955–1965, Aug. 2014.
[39] X. Bai, Z. Yun, D. Xuan, T. H. Lai, and W. Jia, "Deploying four-connectivity and full-coverage wireless sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, 2008, pp. 1–9.
[40] R. Battiti, M. Brunato, and A. Delai, "Optimal wireless access point placement for location-dependent services," Università di Trento, Tech. Rep. DIT-03-052, pp. 1–12, (2003). [Online]. Available: http://alpha.science.unitn.it/~brunato/pubblicazioni/DIT-03-052.pdf
[41] R. Kershner, "The number of circles covering a set," *Amer. J. Math.*, vol. 61, pp. 665–671, 1939.
[42] D. Zhang, et al. "Fine-grained localization for multiple transceiver-free objects by using RF-Based technologies," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 6, pp. 1464–1475, Jun. 2014.
[43] Y. Chen, J.-A. Francisco, W. Trappe, and R. P. Martin, "A practical approach to landmark deployment for indoor localization," in *Proc. IEEE Annu. IEEE Commun. Soc. Sensor Ad Hoc Commun. Netw.*, 2006, pp. 365–373.
[44] K. Sheng, et al., "The collocation of measurement points in large open indoor environment," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 2488–2496.
[45] M. Angjelichinoski, D. Denkovski, V. Atanasovski, and L. Gavrilovska, "Cramér–rao lower bounds of RSS-based localization with anchor position uncertainty," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2807–2834, May 2015.
[46] EVARILOS testbed, 2016. [Online], Available: http://evarilos.intec.ugent.be/
[47] T. V. Haute, et al. "Platform for benchmarking of RF-based indoor localization solutions," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 126–133, Sep. 2015.
[48] T. V. Haute, et al., "Comparability of RF-based indoor localization solutions in heterogeneous environments: An experimental study," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 23, no. 1/2, pp. 92–114, 2016.
[49] F. Lemic, A. Behboodi, V. Handziski, and V. Wolisz, "Experimental decomposition of the performance of fingerprinting-based localization algorithms," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, 2016, pp. 695–704.
[50] *A Power Monitor for Android-Based Mobile Platforms*, 2011. [Online]. Available: http://ziyang.eecs.umich.edu/projects/powertutor/index.html
[51] Qualcomm developer network, "Trepn power profiler," 2017. [Online]. Available: https://developer.qualcomm.com/software/trepn-power-profiler

Xiaohua Tian (S'07-M'11) received the BE and ME degrees in communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2003 and 2006, respectively, and the PhD degree from the Department of Electrical and Computer Engineering (ECE), Illinois Institute of Technology (IIT), Chicago, in Dec. 2010. Since Mar. 2011, he has been in the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, and now is an associate professor with the title of SMC-B Scholar. He serves as the column editor of the *IEEE Network Magazine* and the guest editor of the *International Journal of Sensor Networks* (2012). He also serves as a TPC member for IEEE INFOCOM 2014-2017, a best demo/poster award committee member of IEEE INFOCOM 2014, a TPC co-chair for IEEE ICCC 2014-2016, a TPC co-chair for the 9th International Conference on Wireless Algorithms, Systems and Applications (WASA 2014), a TPC member for IEEE GLOBECOM 2011-2016, and a TPC member for IEEE ICC 2013-2016, respectively. He is a member of the IEEE.

Wenxin Li received the BS (with Hons.) degree in automation from Shanghai Jiao Tong University, in 2016. He is currently working toward the PhD degree in computer science from Ohio State University. His research interests include scheduling algorithm design, coded caching, and indoor localization.

Yucheng Yang is currently working toward the BE degree in electronics engineering from Shanghai Jiao Tong University, China, and is expected to graduate in 2018. His research interest lies in mobile networks and systems.

Zhehui Zhang received the BS degree in computer science and technology from Shanghai Jiao Tong University, in June 2016. She is currently working toward the PhD degree in computer science from the University of California, Los Angeles. Her academic interests include the areas of networked systems, mobile computing, and security.

Xinbing Wang received the BS (with Hons.) degree from the Department of Automation, Shanghai Jiaotong University, Shanghai, China, in 1998, the MS degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2001, and the PhD degree, major from the Department of Electrical and Computer Engineering, minor from the Department of Mathematics, North Carolina State University, Raleigh, in 2006. Currently, he is a professor in the Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai, China. He has been an associate editor of the *IEEE/ACM Transactions on Networking* and the *IEEE Transactions on Mobile Computing*, and a member of the Technical Program Committees of several conferences including ACM MobiCom 2012, 2014, ACM MobiHoc 2012-2017, and IEEE INFOCOM 2009-2018. He is a senior member of the IEEE.