

# Fundamental Limits of RSS Fingerprinting based Indoor Localization

Yutian Wen<sup>1</sup>, Xiaohua Tian<sup>2,3</sup>, Xinbing Wang<sup>1,3</sup>, Songwu Lu<sup>4</sup>

1. School of Electronic, Info. & Electrical Engineering, Shanghai Jiao Tong University, China

2. Dept. of Electronic Engineering, Shanghai Jiao Tong University, China

3. National Mobile Communications Research Laboratory, Southeast University, China

4. Dept. of Computer Science, University of California, Los Angeles, USA

{nogerw, xtian, xwang8}@sjtu.edu.cn, slu@cs.ucla.edu

**Abstract**—Indoor localization has been an active research field for decades, where the received signal strength (RSS) fingerprinting based methodology is widely adopted and induces many important localization techniques such as the recently proposed one building the fingerprint database with crowdsourcing. While efforts have been dedicated to improve the accuracy and efficiency of localization, the fundamental limits of RSS fingerprinting based methodology itself is still unknown in a theoretical perspective. In this paper, we present a general probabilistic model to shed light on a fundamental question: how good the RSS fingerprinting based indoor localization can achieve? Concretely, we present the probability that a user can be localized in a region with certain size, given the RSS fingerprints submitted to the system. We reveal the interaction among the localization accuracy, the reliability of location estimation and the number of measurements in the RSS fingerprinting based location determination. Moreover, we present the optimal fingerprints reporting strategy that can achieve the best accuracy for given reliability and the number of measurements, which provides a design guideline for the RSS fingerprinting based indoor localization facilitated by crowdsourcing paradigm.

## I. INTRODUCTION

Indoor localization has long been an active research field, which enables a vast range of mobile computing applications [1]. Various wireless techniques have been exploited to achieve the accurate and efficient indoor localization, where the received signal strength (RSS) fingerprinting based methodology has been a seminal idea induces many indoor localization systems with different flavors [2]. Most of the RSS fingerprinting based localization systems are implemented in the 802.11 wireless local area network (WLAN) environment, where the RSS measured for frames sent from different access points (APs) is utilized to infer the user's location. Specifically, the system first collects the RSS information from APs in the area of interest, where each piece of information is termed as a *fingerprint* and many such fingerprints result in a fingerprint database. During the localization phase, a user submits measured fingerprints to the system, which are compared with the fingerprint database so that the current location of the user can be estimated.

The fingerprint database can be built in many ways. The element in the database could be deterministic, which is just the RSS reading obtained from the wireless card's routine operation of RSS measurement [3]. The element could also be

probabilistic, which is the RSS distribution that can be used for location determination in a probabilistic manner. As the RSS itself is a coarse characterization of radio propagation, which is influenced by many environmental factors, recent research turns to finer-grained wireless feature, i.e., channel state information (CSI) [2], for a higher localization accuracy. Moreover, no matter if the RSS or the CSI is used for fingerprinting, building and updating the fingerprint database is expensive and laborious for any single entity, which spurs the recent active research on location determination with fingerprints collected with the crowdsourcing paradigm [4], [13].

While efforts have been dedicated to the RSS fingerprinting based indoor localization in order to improve the accuracy and efficiency, the fundamental limits of RSS fingerprinting based methodology itself is still unknown in a theoretical perspective. Kaemarungsi et al develop a preliminary analytical model for a localization system instance with simplified assumptions of wireless signal propagation and system design [6]. Elnahrawy et al perform a comparative study of indoor localization with experiments [5], where the results are highly dependent on the environment and the device implementation. The current lack of a theoretical insight into the RSS fingerprinting based methodology could incur the blindness for system designers: could we further improve the performance of the localization system with better implementations or this has been the best we can do with the methodology?

In this paper, we present a general probabilistic model to shed light on the fundamental question: how good the RSS fingerprinting based indoor localization methodology can achieve? Concretely, we first generalize the assumption of the widely used Log-Normal Path Loss (LNPL) model [2] to provide a more reasonable portrait of the RSS particularly in the indoor environment. We then construct a multi-dimensional probability space based on the measure theory, in order to model all possible submitted RSS fingerprints in the location determination phase. Given the expected accuracy, the localization reliability calculation is transformed into the problem of integration over an event in the multi-dimensional sample space of the probability space.

Based on the problem formulation, we present the domains of integration in the sample space for the location estimation in the one-dimensional and two-dimensional physical indoor

space, which are used to model the localization process in the corridors and ordinary rooms, respectively. We then derive the consequent reliability of the location estimation in those two cases for any given accuracy. Some interesting findings about the shape of the integration domains are presented, where skillful mathematical techniques are demonstrated. As far as we know, such relationship has never been revealed in such kind of general frameworks before.

Moreover, we provide an insight into the RSS fingerprinting based location determination, where we present the condition that there must be a function from a particular subspace of the entire sample space to the physical space. With such an insight, we present the optimal fingerprints reporting strategy that can achieve the best accuracy for given reliability and the number of measurements, which provides a design guideline for the recent proposal of RSS fingerprinting based indoor localization facilitated by crowdsourcing paradigm.

## II. RELATED WORK

Numerous techniques have been proposed in the past decades for the indoor localization, where a systematical survey of these efforts can be found in [2]. Since the work in this paper constructs a probabilistic model to explore the fundamental limits of indoor localization and sheds light on the best fingerprints reporting strategy for location determination with fingerprints database built with crowdsourcing, we are to review some related work in the following two perspectives.

### A. Probabilistic Models Used for Indoor Localization

The early indoor localization system in the context of WLAN is to infer the device's location using the technique of nearest neighbor(s) in signal space (NNSS)[3], where the idea is to compute the Euclidean distance between the measured RSSes and the recorded RSSes from APs strategically deployed at a set of locations. The system returns the location that minimizes the distance. One drawback of the nearest neighbor approach is that it does not fully utilize the opportunity of joint location determination from different APs [7], which leaves room of accuracy improvement.

In order to provide a model for fusing fingerprints from multiple APs, the probabilistic model has been used to estimate the user's location. The Nibble system utilizes Bayesian networks to infer the location of a mobile device [7], where the prior distribution probabilities about a location are obtained by performing sampling for the location over several days in the training phase. With the prior distribution and the Bayesian network, a posterior probability distribution over an estimated location given a set of fingerprints can be derived.

Besides the data fusion, the probabilistic model is also used to deal with the noisy features of the wireless channel, which incurs significant deviations of the sampled RSS fingerprints from those stored in the database thus impacts the accuracy. Youssef et al propose a joint clustering technique [8], [9], which leverages the Bayes estimation theory addressing the noisy wireless channel and reducing the computational cost of searching through the fingerprints database. Battiti et al

propose a similar model for user location error caused by the variability of RSS measurements, which is utilized by the local search heuristic technique for improving the localization accuracy [10], [11]. A comparative study on the performance of the indoor localization is presented in [5], where many probabilistic techniques are briefly surveyed and evaluated with experiments.

While probabilistic models are used in the indoor localization system in an ad hoc manner, most of them focus on inferring the best location estimation in the tactical level. Kaemarungsi et al develop a preliminary probabilistic model for a localization system based on the NNSS approach [6]. With simplified assumptions on the wireless channel behavior in the indoor environment, the essential properties of the RSS fingerprinting based methodology still remain unknown. The probabilistic model to be presented in this paper is used to analyze the fundamental limits of the general indoor localization technique based on the RSS fingerprinting methodology. We have a very general assumption of the wireless channel and no assumption on the pre-deployment efforts. Many interesting theoretical findings are to be presented, which have not been shown as of the writing of the paper.

### B. Crowdsourcing based Indoor Localization

Reviewing the indoor localization schemes above, it is noted that the pre-deployment efforts are needed: there must be some fixed APs whose locations are known and there must be a large number of training data collected in different areas with respect to those fixed APs for calibration. With more and more APs deployed by different operators, the designer of the indoor localization system has to face a dilemma: the information sources are not fully utilized if just using a limited number of pre-deployed APs; however, collecting the training data from all possible APs could be laborious and expensive for any single entity, moreover, how the fingerprints from those APs in unknown locations can be utilized to achieve the most accurate location determination is a challenge.

To fully utilize existing APs and offer more convenience to users, Chintalapudi et al propose the EZ localization scheme with limited pre-deployment efforts [12]. The mobile devices record and report RSS fingerprints perceived with respect to different APs at possible unknown locations in the fingerprints collecting phase. The fingerprint is represented as the mean and standard deviation of the RSS seen from those APs. Compared with traditional schemes, EZ only needs the mobile device to occasionally obtain an absolute location at the edge of the indoor environment through GPS, and users can move around at will in the indoor space in normal course.

This almost pre-deployment free service model could spare the explicit efforts needed from indoor localization service providers for training data. The fingerprints collection can be performed in the paradigm of crowdsourcing, where any ordinary smart phone user without any professional training can collect the fingerprints in an area in passing that the user moves around the area. Rai et al present a calibration zero-effort system Zee [13], which leverages the embedded sensors

of mobile devices to track the device itself as the carrier of the device traverse an indoor environment while simultaneously performing Wi-Fi scans.

EZ and Zee emphasize on the implementation of the crowdsourcing based indoor localization, where a heuristic AP selection scheme is presented [12]. However, some fundamental issues about the crowdsourcing based indoor location determination are still unclear. With those crowdsourced RSS fingerprints, must there be a mapping from any combination of fingerprints to a location? Chintalapudi et al mentioned that RSSes from different APs are unequally effective [12], but which APs are more valuable for location determination if the locations of APs are unknown? Our work in this paper will shed light on these fundamental issues.

### III. SYSTEM MODEL

Consider an indoor space denoted by  $S$ , where the long and narrow space such as a corridor can be modeled as an one-dimensional Cartesian space with  $S \subset \mathbb{R}$ , and the ordinary space such as a room can be modeled as a two-dimensional Cartesian space with  $S \subset \mathbb{R}^2$ . We use  $\vec{r}$  to denote a location in  $S$ , where  $\vec{r} = x_1$  and  $\vec{r} = (x_1, x_2)$  in the one-dimensional and two-dimensional Cartesian coordinate system, respectively. For both fingerprints collection and location determination, the mobile device reports the RSS readings obtained by measuring signals sent from each AP. The measured result is a random variable with respect to a specific location denoted as  $\mathcal{P}(\vec{r})$ :

$$\mathcal{P}(\vec{r}) = \mu(\vec{r}) + \sigma\mathcal{Y}, \quad (1)$$

where  $\mu(\vec{r})$  represents how the mean of RSS readings varies with respect to locations.  $\mathcal{Y}$  is the normalized Gaussian additive noise with  $\mathcal{Y} \sim \mathcal{N}(0, 1)$  and  $\sigma$  is the amplitude of the noise.

Eq. (1) is a more general model derived from the Log-Normal Path Loss (LNPL) model [12], where  $\mu(\vec{r})$  is continuous. If letting  $\mu(\cdot)$  be the logarithm function, Eq. (1) degenerates to the LNPL model. We use the general model because the LNPL is an inaccurate description of the RSS, which only considers the path loss and shadowing but ignores the dominant small scale fading incurred by multipath effect in the indoor environment. As the real wireless environment is very unpredictable, we are unable to know exactly what the mean value of the measured RSSes  $\mu(\vec{r})$  is like. However, it is observed from many previous experiments [5], [6], [9], [12] that the mean of measured RSSes changes in a non-dramatic manner with the small change of locations, which makes the continuity assumption of  $\mu(\vec{r})$  reasonable. From these observation we also notice that the signal might deteriorate significantly as they propagate through solid obstructions like walls, so the continuity might be disrupted in those areas. However as we are more interested in localize users, which are not likely to appear inside a wall, the wall areas are eliminated from the map, thus defended the assumption. Consequently, we can have the following approximation:

$$\mu(\vec{r}') \approx \mu(\vec{r}) + \nabla\mu(\vec{r})(\vec{r}' - \vec{r}), \quad (2)$$

where  $\vec{r}'$  is an estimated location of the user given the actual location of the user  $\vec{r}$ . Note that all the analysis based on Eq. (1) can be applied to the scenario using the LNPL model, which itself has been widely adopted in many indoor localization systems.

In the training phase, the mobile device is randomly assigned a point in the indoor space and the device randomly chooses an AP and measures the RSS fingerprint once. As the result of each measurement is a random variable as shown in Eq. (1), all possible outcomes for the measurement form a sample space  $\Omega$ . We define a  $\sigma$ -algebra  $\mathcal{F}$  that is the collection of all events, where each event is a set containing zero or more outcomes. Eq. (1) gives an assignment of probabilities to the events  $\mathcal{P} : \Omega \rightarrow [0, 1]$ , thus we can construct a probability space  $(\Omega, \mathcal{F}, \mathcal{P})$ . Suppose that the mobile device performs the measurement  $n$  times at a given assigned point, then the Cartesian product of probability spaces induced by RSS measurements forms an  $n$ -dimensional probability space  $(\Omega^n, \mathcal{F}^n, \mathcal{P})$ , where we abuse  $\mathcal{P} : \Omega^n \rightarrow [0, 1]$  for the convenience of demonstration.

The location determination phase can be considered as a mapping  $\mathcal{M} : \Omega^n \rightarrow S$ ,  $\vec{r}' = \mathcal{M}(\vec{\sigma})$ , where  $\vec{\sigma}$  is an outcome of  $\Omega^n$ . It means that the localization system outputs an estimated location  $\vec{r}'$  for a series of measurements of RSSes from randomly chosen APs. Since RSS measurement results are independently and identically distributed, the induced sample spaces of RSS measurements are orthogonal to each other, and  $\Omega^n$  is homeomorphic to  $n$ -dimensional Cartesian space. If applying the coordinate system of  $n$ -dimensional Cartesian space to  $\Omega^n$ , we can obtain a presentation of  $\vec{\sigma}$  denoted as  $\vec{P} = [P_1, P_2, \dots, P_n]^T$ , in which  $P_i$  is the reading of each measurement result. Consequently, the original measurement  $\mathcal{P} : \Omega^n \rightarrow [0, 1]$  becomes  $f : \mathbb{R}^n \rightarrow [0, 1]$ .

We use  $Q$  to denote the area that is centered at the user's actual location with radius  $\delta$  in the physical space.  $E(\delta)$  is used to denote the event in the sample space that the localization system estimates the user's location in  $Q$ . We will use  $E$  to represent  $E(\delta)$  in the following discussion to avoid tedious mathematical presentation. The size of  $\delta$  determines the *accuracy* of the localization system. We use  $R$  to denote the probability that the user's estimated location is within the area  $Q$ , which is defined as the *reliability* of the localization system.

With the definition of accuracy and reliability, we can see that the probability of the event that the system correctly estimates the user's location is:

$$R(E) = \int_E f(\vec{P}) dP^n = \int_Q g(\vec{r}', \vec{r}) dr'^2, \quad (3)$$

where  $f(\vec{P})$  is the possible measurement of the sample space in the  $n$ -dimensional Cartesian coordinate system, and  $g(\vec{r}', \vec{r})$  is the probability distribution function that the user is localized at  $\vec{r}'$  given that the real location of the user is  $\vec{r}$  in the physical space. Eq. (3) indicates that the reliability can be interpreted as either the probability that the measurement fall into the event

region  $E$  or the user is localized in an area that is centered at  $\vec{r}$  and with radius  $\delta$ .

#### IV. LOCALIZATION IN ONE-DIMENSIONAL SPACE

In this section, we first calculate the localization reliability for the case of one-time measurement for a single AP, and then extend the conclusion to multiple-time measurements for multiple APs.

##### A. One-Time Measurement for Single AP

We set the origin of spacial coordinate system at the sole AP of the one-dimensional physical space. To achieve the maximal reliability, the Maximum Likelihood Estimation (MLE) is used as the estimation principle[14].

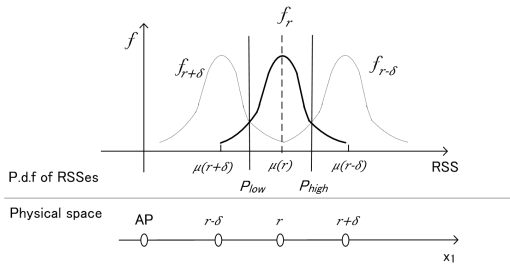


Fig. 1. Integration domain for one-dimensional localization with single measurement for single AP.

Comparing the Probability Distribution Function (p.d.f) with respect to the RSS readings of different locations, there are two thresholds  $P_{high}$  and  $P_{low}$ , where  $f_{r-\delta}(P_{high}) = f_r(P_{high})$  and  $f_{r+\delta}(P_{low}) = f_r(P_{low})$  as shown in Fig. (1). If the user's measured RSS is between the two thresholds, the user will be estimated to be in  $Q$  in the location determination phase. According to Eq. (1),

$$\begin{cases} P_{high} = \frac{\mu(r-\delta) + \mu(r)}{2}, \\ P_{low} = \frac{\mu(r+\delta) + \mu(r)}{2}. \end{cases} \quad (4)$$

Thus the reliability

$$R(\delta, r, \sigma) = \int_{P_{low}}^{P_{high}} f(P(r))d(P) = erf\left(\left|\frac{-\mu'(\vec{r})\delta}{2\sqrt{2}\sigma}\right|\right), \quad (5)$$

where  $erf(\cdot)$  is the error function defined as:  $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ .

##### B. Multiple Measurements for Multiple APs

According to the probability theory, the average of  $n$  i.i.d Gaussian variables is equivalent to a Gaussian variable with a standard deviation  $\frac{\sigma}{\sqrt{n}}$ [15]. Two measurements to a single AP can be regarded as the measurements for two identical APs locating at the same place, which is to be confirmed by our result shown in Eq. (12). If several measurements are performed on a single AP, the RSS fingerprint is set to be a new random variable with the standard deviation  $\frac{\sigma}{\sqrt{n}}$ .

By the rule of MLE, the event that the system will estimate the user to be in  $Q$  is:

$$E = \{\vec{o} \mid \prod_{i=1}^n f_r(P_i) \geq \prod_{i=1}^n f_{r+\delta}(P_i), \prod_{i=1}^n f_r(P_i) \geq \prod_{i=1}^n f_{r-\delta}(P_i)\}.$$

All outcomes in  $E$  satisfy the following inequality

$$\prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(P_i - \mu_i(r))^2}{2\sigma_i^2}} \geq \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(P_i - \mu_i(r \pm \delta))^2}{2\sigma_i^2}}$$

After simplification, it is equivalent to

$$\sum_{i=1}^n \frac{\mu_i(r \pm \delta) - \mu_i(r)}{\sigma_i} (Y_i - \frac{\mu(r \pm \delta) - \mu(r)}{2\sigma_i}) \leq 0 \quad (6)$$

where  $\mu_i(r)$  is the average RSS of AP  $i$  at  $\vec{r}$ . Normalize the RSS readings  $P_i$  to normalized noise  $Y_i = \frac{P_i - \mu_i}{\sigma_i}$ , where  $\sigma_i$  might differ among different APs, because the Gaussian noise represents the random changes in the radio propagation, and different APs certainly transmit the signal to the user through different channels.

We use  $\vec{\mu}(\vec{r})$  to denote the mean RSS outcome at  $\vec{r}$ . Now that it refers to the outcome itself, this notation does not depend on any coordinate system. Given a specific average RSS  $\vec{\mu}(\vec{r})$ , the set of  $\{Y_i\}$  forms a coordinate basis for the sample space, where the origin is  $\vec{\mu}(\vec{r})$ , and each dimension is suppressed by a factor of  $\sigma_i$ . We use this coordinate system to characterize the event  $E$  and its probability in the following several sections due to its simplicity for the small scale analysis, then we will switch to another one in section VI-A.

Apparently, the two constraint Eq. (6) are two non-parallel hyper-planes in the sample space. Vectors  $\vec{h}_1 = [\frac{1}{2\sigma_1}(-\mu_1(r) + \mu_1(r - \delta)), \frac{1}{2\sigma_2}(-\mu_2(r) + \mu_2(r - \delta)), \dots, \frac{1}{2\sigma_n}(-\mu_n(r) + \mu_n(r - \delta))]^T$ ,  $\vec{h}_2 = [\frac{1}{2\sigma_1}(-\mu_1(r) + \mu_1(r + \delta)), \frac{1}{2\sigma_2}(-\mu_2(r) + \mu_2(r + \delta)), \dots, \frac{1}{2\sigma_n}(-\mu_n(r) + \mu_n(r + \delta))]^T$  together span a plane  $\mathcal{W}$ . Restrictions to the event  $E$ , Eq. (6), can then be rewritten in the vector form:

$$\begin{cases} 2\vec{h}_1(\vec{o} - \vec{h}_1) \leq 0, \\ 2\vec{h}_2(\vec{o} - \vec{h}_2) \leq 0. \end{cases} \quad (7)$$

It is important to note that  $\vec{h}_1$  and  $\vec{h}_2$  can denote both the normal vectors to each hyperplane and the two points on each hyperplane that are closest to the origin. This fact will be helpful to deal with the two-dimensional issue. Now that the p.d.f and the constraint conditions are all normalized, we can rotate the coordinate system  $\{Y_i\}$  to another orthonormal basis  $\{\vec{e}_i\}$ ,  $i = 1, 2, \dots, n$ , where  $\vec{e}_1$  is parallel to  $\vec{h}_1$  and  $\vec{e}_2 \in \mathcal{W}$ . Consequently,  $\mathcal{W}$  is spanned by only two coordinate axes. The rest coordinate axes are therefore all orthogonal to the plane. There exists an orthonormal basis for subspace  $\mathcal{W}$ , i.e.,  $\{\vec{e}_i\}$ ,  $i = 3, \dots, n$ . Any outcome  $\vec{o}$  in the sample space can be decomposed into  $\vec{o} = \sum_i c_i \vec{e}_i$ , where coefficients  $c_i$  is determined and unique for any given vector  $\vec{o}$  and

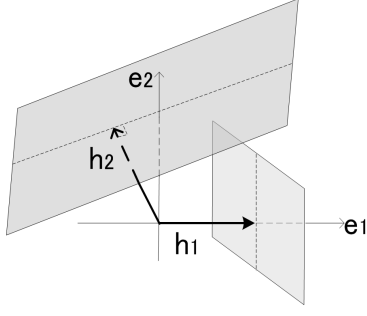


Fig. 2. Integration domain for one-dimensional localization with multiple measurements for multiple APs.

orthonormal basis  $\{\vec{e}_i\}$ . Eq. (6) can then again be rewritten in the component form of the  $\{\vec{e}_i\}$  basis:

$$\begin{cases} 2|\vec{h}_1\vec{e}_1|(c_1 - |\vec{h}_1\vec{e}_1|) \leq 0, \\ 2|\vec{h}_2\vec{e}_1|(c_1 - |\vec{h}_2\vec{e}_1|) + 2|\vec{h}_2\vec{e}_2|(c_2 - |\vec{h}_2\vec{e}_2|) \leq 0. \end{cases} \quad (8)$$

Thus the probability that the system correctly estimates the user's location can be rewritten as

$$\begin{aligned} R(E) &= \int_E f^{(n)}(\vec{r}) d\vec{e} \\ &= \int_{c_1 \leq |\vec{h}_1|} d\vec{e}_1 \int_{-\infty}^{\frac{|\vec{h}_2\vec{e}_1|^2 + |\vec{h}_2\vec{e}_2|^2 - |\vec{h}_2\vec{e}_1|c_1}{|\vec{h}_2\vec{e}_2|}} \frac{1}{2\pi} e^{-\frac{e_1^2 + e_2^2}{2}} d\vec{e}_2, \end{aligned} \quad (9)$$

$$(10)$$

where  $f^{(n)}(\cdot)$  stand for  $n$ -variable Gaussian function. As of now, we successfully reduce the multiple integral to a much simpler two dimensional one. Multivariate Gaussian integral Eq. (11) is integrated on the area indicated in Fig. (2).

By Eq. (2), the dimension of  $E$  can be further reduced, for that  $\vec{h}_1$  and  $\vec{h}_2$  will now be parallel to each other, though in different directions.

$$\begin{cases} \vec{h}_1 = [-\frac{1}{2\sigma_1}\mu'_1(r)\delta, \dots, -\frac{1}{2\sigma_n}\mu'_n(r)\delta]^T, \\ \vec{h}_2 = [\frac{1}{2\sigma_1}\mu'_1(r)\delta, \dots, \frac{1}{2\sigma_n}\mu'_n(r)\delta]^T. \end{cases}$$

Thus the two constraint conditions Eq. (6) are parallel to each other.  $R(E)$  can then be simplified as

$$R(E) = \int_{|\vec{h}_2|}^{|\vec{h}_1|} d\vec{e}_1 \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi})^2} e^{-\frac{e_1^2 + e_2^2}{2}} d\vec{e}_2 \quad (11)$$

$$\approx \text{erf}\left(\frac{\sqrt{\sum_{i=1}^n (\frac{\mu'_i(r)\delta}{2\sigma_i})^2}}{\sqrt{2}}\right) \quad (12)$$

### C. Discussions

We can see that Eq. (12) is equivalent to Eq. (5) when  $n = 1$  and there is a single AP in the room, which corroborates our analysis. Moreover, the analysis above reveals some insight into the design of the indoor localization. First, the more data are reported to the system, the more reliable the location estimation is, since  $\sqrt{\sum_i (\frac{\mu'_i(r)\delta}{2\sigma_i})^2} > \sqrt{(\frac{\mu'_i(r)\delta}{2\sigma_i})^2}$ . Second, if  $\frac{\mu'_i(r)}{\sigma_i} \geq \frac{\mu'_j(r)}{\sigma_j}, \forall i, j \leq n$ , the reliability of the result can be

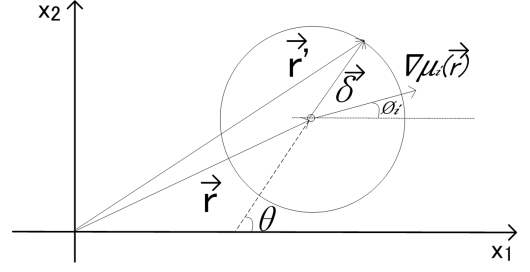


Fig. 3. Two-dimensional localization with multiple measurements for multiple APs

improved if the user reports AP  $i$ 's RSS rather than the other. This means that reporting a measurement of a cleaner channel (smaller  $\sigma$ ) is more valuable, and the sharper the signal varies around the user's location (greater  $\mu'(r)$ ), the easier it is for the system to pinpoint the user's location.

## V. LOCALIZATION IN TWO-DIMENSIONAL SPACE

### A. Multiple Measurements for Multiple APs

Compared with the case of the one-dimensional physical space, finding the integration domain for the two-dimensional localization is more challenging. Our strategy is to first prove that the integral area  $E$  for calculating the reliability is a hyper-cylinder, and then switch to a more appropriate cylindrical coordinate system to reduce the dimension of integration. After that, we will prove that the intersection between the hyper-cylinder and the cross-section plane is in the shape of an ellipse, which provides the facilitation to obtain the reliability.

Fig. (3) illustrates an example of the two-dimensional scenario, where the location of the user is  $\vec{r}$  and the location of any point on the boundary of the area  $Q$  is  $\vec{r}'$ . We use  $\vec{\delta} = \vec{r}' - \vec{r}$  to denote a two-dimensional vector with the direction from the user's actual location to any point on the boundary of  $Q$ . We use  $\theta$  to denote the angle between  $\vec{\delta}$  and the horizontal axis, and use  $\phi_i$  to denote the angle between  $\nabla\mu_i(\vec{r})$  and the horizontal axis. By Eq. (2), we have

$$\mu_i(\vec{r}') - \mu_i(\vec{r}) = \nabla\mu_i(\vec{r})\vec{\delta} = \delta|\nabla\mu_i(\vec{r})|\cos(\theta - \phi_i), \quad (13)$$

where  $\delta = |\vec{\delta}|$ .

We want to find the event that the user is located within the area  $Q$ . According to MLE, this is equivalent to find those outcomes that the probability density of the user appearing on the boundary of  $Q$  is no greater than that of the user appearing at  $\vec{r}$ , which is

$$E = \{\vec{\sigma} | \prod_{i=1}^n f_{\vec{r}}(P_i) \geq \prod_{i=1}^n f_{\vec{r}+\vec{\delta}}(P_i)\}. \quad (14)$$

All outcomes in  $E$  follow the following inequality

$$\sum_{i=1}^n \frac{\mu_i(\vec{r} + \vec{\delta}) - \mu_i(\vec{r})}{\sigma_i} (Y_i - \frac{\mu_i(\vec{r} + \vec{\delta}) - \mu_i(\vec{r})}{2\sigma_i}) \leq 0. \quad (15)$$

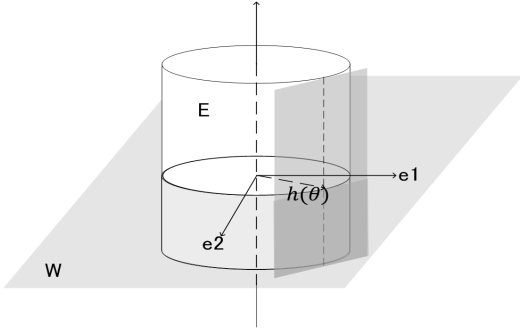


Fig. 4. Integral area of 2-dimensional space

Substituting Eq. (13) into Eq. (15), we have the specific description of  $E$ :

$$\sum_{i=1}^n \frac{\delta|\nabla\mu_i(\vec{r})|}{\sigma_i} \cos(\theta - \phi_i) (Y_i - \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \cos(\theta - \phi_i)) \leq 0. \quad (16)$$

Constraint condition Eq. (16) should hold true for any  $\theta$ . Thus there will be an infinite set of hyper-planes surrounding event  $E$  in the sample space, as shown in Fig. 4. We define the  $n$ -dimensional normal vector of the hyper-plane to be a function of  $\theta$ :

$$\vec{h}(\theta) = [\frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \cos(\theta - \phi_1), \dots, \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \cos(\theta - \phi_n)]^T$$

**Theorem 1.** *The orbit of  $\{\vec{h}(\theta)\}$  and the origin are coplanar, i.e., on the same 2-dimensional plane in the sample space.*

*Proof:* This is equivalent to prove that there exists a rank  $n-2$  complementary subspace  $\bar{\mathcal{W}}$  of  $\mathcal{W}$ , where  $\mathcal{W}$  is spanned by  $\{\vec{h}(\theta)\}$ . Formally,  $\forall \vec{g} \in \bar{\mathcal{W}}, \forall \theta, \vec{h}(\theta) \cdot \vec{g} \equiv 0$ , that is

$$[\frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \{ \cos\theta \cos\phi_1 + \sin\theta \sin\phi_1 \}, \dots, \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \{ \cos\theta \cos\phi_n + \sin\theta \sin\phi_n \}] [g_1, \dots, g_n]^T \equiv 0.$$

Consequently, we need to prove

$$\begin{cases} \cos(\theta) [\frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \cos\phi_1, \dots, \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \cos\phi_n] [g_1, \dots, g_n]^T \equiv 0, \\ \sin(\theta) [\frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \sin\phi_1, \dots, \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \sin\phi_n] [g_1, \dots, g_n]^T \equiv 0. \end{cases}$$

The equations above hold true for all  $\theta$ , thus

$$\begin{cases} [\frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \cos\phi_1, \dots, \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \cos\phi_n] [g_1, \dots, g_n]^T \equiv 0, \\ [\frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \sin\phi_1, \dots, \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \sin\phi_n] [g_1, \dots, g_n]^T \equiv 0. \end{cases} \quad (17)$$

Adding  $n-2$  lines of zero row vectors under the row vector in Eq. (17) makes an  $n \times n$  square matrix  $\mathcal{H}$ :

$$\mathcal{H} = \begin{pmatrix} \frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \cos(\phi_1) & \dots & \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \cos(\phi_n) \\ \frac{\delta|\nabla\mu_1(\vec{r})|}{2\sigma_1} \sin(\phi_1) & \dots & \frac{\delta|\nabla\mu_n(\vec{r})|}{2\sigma_n} \sin(\phi_n) \\ 0 & \dots & 0 \end{pmatrix}.$$

Then  $\bar{\mathcal{W}}$  is the solution space to the linear formula:  $\mathcal{H}\vec{g} = 0$ , where  $\text{rank}(\mathcal{H}) \leq 2$ , so  $\text{rank}(\bar{\mathcal{W}}) \geq n-2$ ; therefore,

$\text{rank}(\mathcal{W}) = n - \text{rank}(\bar{\mathcal{W}}) \leq 2$ . The straight line connecting  $\vec{h}(\theta)$  and  $\vec{h}(-\theta)$  will always come across the origin, thus the origin is also in plane  $\mathcal{W}$ . ■

Eq. (13) indicates that  $\mathcal{W}$  is a tangent plane of surface  $M$  at  $\vec{\mu}(\vec{r})$ , where  $M = \{\vec{\mu}(\vec{r}) | \vec{r} \in S\}$  is the mean surface of RSS readings. Repeat the technique we used in Section IV-B, we will be again able to reduce the multivariate integral in the whole event to a two variable integral on a subset of plane  $\mathcal{W}$ .

The next is to determine the domain of probability integration. By definition, it is the area inside the envelop of  $\{\vec{h}(\theta)\}$ , denoted  $\mathcal{D}$ .

**Theorem 2.** *The orbit of  $\{\vec{h}(\theta)\}$  is an ellipse.*

*Proof:* Theorem (1) states that the orbit of  $\{\vec{h}(\theta)\}$  is in a plane. To prove Theorem (2) is equivalent to prove that  $\forall \theta, \exists \psi$ ,

$$\vec{h}(\theta) = \vec{U} \cos(\psi) + \vec{V} \sin(\psi), \quad (18)$$

where  $\vec{U}$  and  $\vec{V}$  are constant vectors and  $\vec{U} \cdot \vec{V} = 0$ . If there exists such a constant  $\alpha = \theta - \psi$  satisfying Eq. (18), then Theorem (2) is proven.

$$\vec{h}(\theta) = \sum_{i=1}^n \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \{ \cos(\psi + \alpha - \phi_i) \} Y_i \quad (19)$$

$$= \sum_{i=1}^n \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \{ \cos(\alpha - \phi_i) \cos\psi - \sin(\alpha - \phi_i) \sin\psi \} Y_i. \quad (20)$$

If  $\vec{U}$  and  $\vec{V}$  is assigned as following then Eq. (19) is satisfied.

$$\begin{cases} \vec{U} = \sum_i \{ \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \cos(\alpha - \phi_i) Y_i \}, \\ \vec{V} = \sum_i \{ -\frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \sin(\alpha - \phi_i) Y_i \}. \end{cases} \quad (21)$$

Thus  $\vec{U} \cdot \vec{V} = 0$  is equivalent to

$$f(\alpha) = \sum_{i=1}^n \left( \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \right)^2 \sin(2\alpha - 2\phi_i) = 0. \quad (22)$$

Apparently, formula (22) has four different solutions of  $\alpha$  in the interval between 0 and  $2\pi$ , because  $f(\alpha) = -f(\alpha + \pi/2)$  and  $f(\alpha)$  is a continuous function. There will be four zero points within each  $2\pi$  period. The four solutions actually correspond to four different assignments of vector  $\vec{U}$  to the semi-major axes and semi-minor axes. However, as we are only interested in the length of the semi-major axis and semi-minor axis, all four kinds of assignments are the same. We will use  $\vec{U}$  as the semi-major axis in the following sections.

$$\sum_i \left( \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \right)^2 (\sin(2\alpha) \cos(2\phi_i) - \cos(2\alpha) \sin(2\phi_i)) = 0, \quad (23)$$

where

$$\tan 2\alpha = \frac{\sum_{i=1}^n \left( \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \right)^2 \sin(2\phi_i)}{\sum_{i=1}^n \left( \frac{\delta|\nabla\mu_i(\vec{r})|}{2\sigma_i} \right)^2 \cos(2\phi_i)}.$$

Thus the module of  $\vec{U}$  and  $\vec{V}$  are determined.

$$\begin{cases} |\vec{U}| = \sqrt{\sum_{i=1}^n \left\{ \frac{\delta |\nabla \mu_i(\vec{r})|}{2\sigma_i} \cos(\alpha - \phi_i) \right\}^2}, \\ |\vec{V}| = \sqrt{\sum_{i=1}^n \left\{ \frac{\delta |\nabla \mu_i(\vec{r})|}{2\sigma_i} \sin(\alpha - \phi_i) \right\}^2}. \end{cases} \quad (24)$$

Consequently, the reliability of the location estimation in the two-dimensional space is

$$R(E) = \int_{\frac{e_1^2}{|\vec{U}|^2} + \frac{e_2^2}{|\vec{V}|^2} = 1} \frac{1}{2\pi} e^{-\frac{e_1^2 + e_2^2}{2}} de_1 de_2 \quad (25)$$

$$= |\vec{U}||\vec{V}| \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - e^{-\frac{\cos^2 \psi |\vec{U}|^2 + \sin^2 \psi |\vec{V}|^2}{2}}}{\cos^2 \psi |\vec{U}|^2 + \sin^2 \psi |\vec{V}|^2} d\psi. \quad (26)$$

## B. Discussions

Most conclusions in the one-dimensional situation still hold true in the two-dimensional case. More data will yield higher reliability; however, there are some distinguishable properties in the two-dimensional case worth of mentioning. First, if the  $\nabla \mu_i(\vec{r})$  for all APs are the same, which means that  $\phi_i = \phi_j, \forall i, j$ , then it is impossible to determine the user's location. This is because  $|\vec{V}| = 0$  in this case thus  $R(E) = 0$ . It means that there should be at least two APs the direction of whose  $\nabla \mu_i(\vec{r})$  is different from each other. Second, if the user observes that  $\nabla \mu_i(\vec{r})$  and  $\nabla \mu_j(\vec{r})$  for two APs  $i$  and  $j$  are either in the same direction or in the opposite direction, then it is just like in the one-dimensional case, thus if  $\frac{|\nabla \mu_i(\vec{r})|}{\sigma_i} \geq \frac{|\nabla \mu_j(\vec{r})|}{\sigma_j}, \forall i, j \leq n$ , reporting the RSS reading from AP  $i$  is more effective than reporting that of AP  $j$  for location determination.

## VI. BEST STRATEGY FOR LOCATION DETERMINATION

The analysis above shows that the utilities for reporting RSS fingerprints from different APs are different in the location determination process. A natural question to ask is: which fingerprints should the user report to the system so that the most accurate location estimation can be obtained? Before revealing the answer to such a question, we first reveal the fundamentals of the location determination.

### A. Fundamentals of Location Determination

The fundamental issue of location determination is that: can every outcome in the sample space be mapped into a location in the physical space. The mean of RSSes  $\vec{\mu}(\cdot)$  is a continuous mapping from the physical space to the mean surface of RSSes  $M$ . According to Eq. (2), each small area around  $\vec{\mu}(\vec{r})$  can be approximated as a plane. Recall that the event  $E$  in the two-dimensional case is a hyper-cylinder. According to Theorem (2), the intersection of the hyper-cylinder and  $M$  forms an orbit, which is the same 2-dimensional plane as  $\vec{\mu}(\vec{r})$ ; therefore, if we shrink  $\delta$  to zero, then the hyper-cylinder will shrink to an  $n - 2$  dimensional body  $\bar{\mathcal{W}}$  and intersect with  $M$  at  $\vec{\mu}(\vec{r})$ , as illustrated in the analysis above.  $\bar{\mathcal{W}}$  is the event that the user is estimated to be most likely appearing at  $\vec{r}$ , because

$E$  is the event that the system estimates the user's location in the area with a radius no more than  $\delta$ . Thus

$$\vec{r} = \mathcal{M}(\bar{\mathcal{W}}), \quad (27)$$

where  $\mathcal{M} : \Omega^{n-2} \rightarrow S$  is a mapping from the set of outcomes  $\bar{\mathcal{W}}$  to the user's most likely location  $\vec{r}$ .

However, it is worth noting that we in fact abuse the notation  $\vec{r}$  here, since the location obtained from the mapping  $\mathcal{M}$  is not necessarily the actual location of the user. To see this, recall that  $\bar{\mathcal{W}}$  is a  $(n - 2)$ -dimensional body perpendicular to the tangent plane of  $M$  at  $\vec{\mu}(\vec{r})$  in the  $\{\frac{P_i}{\sigma_i}\}$  coordinate system, and  $M$  is a surface with curvature, thus it may happen that  $\bar{\mathcal{W}}_1$  and  $\bar{\mathcal{W}}_2$  that induced by two tangent planes of  $M$  intersects at an outcome, and this outcome can be mapped to two different points on  $M$ . That is, the same set of RSSes can result in different localizations.

If this happens, we also want to use MLE to derive which location the user is more likely to appear. By the definition of MLE, the intersection outcome should be mapped to the point  $\vec{r}'$  on  $M$ , if the inequality  $f(\vec{r}') > f(\vec{r}'')$  is satisfied, where:

$$\begin{aligned} f(\vec{r}') &= \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(P_i - \mu_i(\vec{r}'))^2 / 2\sigma_i^2}, \\ f(\vec{r}'') &= \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(P_i - \mu_i(\vec{r}''))^2 / 2\sigma_i^2}, \end{aligned}$$

and  $\vec{r}'$  and  $\vec{r}''$  can be any points in the physical space. If  $f(\vec{r}') > f(\vec{r}'')$ , then  $(\frac{P_i - \mu_i(\vec{r}')}{\sigma_i})^2 < (\frac{P_i - \mu_i(\vec{r}'')}{\sigma_i})^2$ . This means that in the  $\{\frac{P_i}{\sigma_i}\}$  coordinate system, the Euclidean distance from the outcome  $\vec{o} = [\frac{P_1}{\sigma_1}, \frac{P_2}{\sigma_2}, \dots, \frac{P_n}{\sigma_n}]^T$  to the point  $[\frac{\mu_1(\vec{r}')}{\sigma_1}, \frac{\mu_2(\vec{r}')}{\sigma_2}, \dots, \frac{\mu_n(\vec{r}')}{\sigma_n}]^T$  should be less than that to  $[\frac{\mu_1(\vec{r}'')}{\sigma_1}, \frac{\mu_2(\vec{r}'')}{\sigma_2}, \dots, \frac{\mu_n(\vec{r}'')}{\sigma_n}]^T$ . It should be noticed that the latter two arrays are the  $\{\frac{P_i}{\sigma_i}\}$  representation of the corresponding mean RSS outcomes  $\vec{\mu}(\vec{r}')$  and  $\vec{\mu}(\vec{r}'')$ , respectively, which are two arbitrary points on the mean RSS surface  $M$ ; therefore, we should map the intersection outcome to the point with shortest Euclidean distance to  $M$  in the  $\{\frac{P_i}{\sigma_i}\}$  coordinate system.

The analysis above indicates that if the user reports to the system the outcomes that are closer to  $M$ , it is more likely the user can be localized to the actual location. Consider the mean RSS surface  $M$ , no matter how great the curvature of  $M$  is, we can always find a very small space in  $\Omega^n$  which is around  $\vec{\mu}(\vec{r})$  on  $M$ , so that the part of  $M$  in such a small space can be approximated to its tangent plane  $\mathcal{W}$  at  $\vec{\mu}(\vec{r})$ . Each  $\bar{\mathcal{W}}$  for the given  $\vec{\mu}(\vec{r})$  is parallel to that for others. If we move  $\vec{r}$  around on  $S$ , the point  $\vec{\mu}_i(\vec{r})$  moves around correspondingly on  $\mathcal{W}$ . This is equivalent to say that  $\bar{\mathcal{W}}$  scans the entire small space in  $\Omega^n$ .

Every point within the small space is on a unique  $\bar{\mathcal{W}}$ , and there must be a mapping from  $\Omega^n$  to  $S$  for every point in the sample space around  $M$ . It is interesting to find that if the space is very small, the tangent plane approximation is more accurate thus mapping the outcome to the surface is almost the same to find the Euclidean distance. If the outcome is far

from  $M$ , it may happen that there is no such a  $\bar{W}$  so that the outcome can be mapped to a location.

In conclusion, if  $M$  is a plane, there must exist a function from  $\Omega^n$  to  $S$ ; if  $M$  is with curvature, the nearer the reported outcomes to  $M$ , the more likely the system will return a reliable location estimation with accuracy  $\delta$ .

### B. Best Strategy

With revealing the fundamentals of location determination, we now derive which APs users should measure so that they can be localized with the highest accuracy with the given reliability. Let  $\mathbb{U} = \{AP_i\}, i = 1, \dots, m$  be the set of all APs that can be sensed by the user's mobile device. A measurement strategy is defined as a sequence of measurements on APs and is denoted by  $\mathcal{V}_n = (s^1, \dots, s^n), s^j \in \mathbb{U}$ . Note that the superscript of  $s^j$  is the index of the measurement in the sequence, and it does not mean the measurement is performed on  $AP_j$ . One 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  $m^n$ . The optimal strategy is denoted as  $\mathcal{V}_n^*, \mathcal{V}_n^* \in \mathbb{U}^n$ .

Recall that the event  $E$  is a hyper-cylinder in the sample space and the intersection between the hyper-cylinder and  $M$  is an ellipse centered at  $\bar{\mu}(\bar{r})$ . We now consider another event  $\mathcal{E}(c)$ , which is also a hyper-cylinder in the sample space; however, we let the intersection between such a hyper-cylinder and  $M$  be a circle centered at  $\bar{\mu}(\bar{r})$  and with radius  $c$ , where  $\bar{r}$  is the actual location of the user. In another perspective,  $\mathcal{E}(c)$  denotes the event that the outcomes for localizing a user at  $\bar{r}$  fall in the newly defined hyper-cylinder. Thus the reliability of the location estimation is in fact the probability of the event  $\mathcal{E}(c)$ , which is similar to the previous analysis:

$$R(\mathcal{E}(c)) = \int_0^{2\pi} \int_0^c \frac{1}{2\pi} e^{-\frac{\rho^2 \cos^2 \psi + \rho^2 \sin^2 \psi}{2}} \rho d\rho d\psi \quad (28)$$

$$= 1 - e^{-c^2/2}. \quad (29)$$

Let us switch our attention to the physical space. We consider the vicinity of  $\bar{r}$ , where each point on the boundary of the vicinity represents an outcome in  $\Omega^n$ . The vicinity is denoted as  $\mathcal{U}$  and it must satisfy that the outcomes for localizing those location points on the boundary of  $\mathcal{U}$  just fall on the circle on  $M$ . The point on the circle is denoted as  $\mu(\bar{r})$ . Thus

$$\sum_{i=1}^n \frac{(\mu_i(\bar{r}) - \mu_i(\bar{r}'))^2}{(2\sigma_i)^2} = c^2. \quad (30)$$

Put  $\bar{r}$  and  $\bar{r}'$  in the polar coordinate system with the origin at  $\bar{r}$ , then Eq. (30) can be transformed into  $\sum_{i=1}^n (\rho(\theta) |\nabla \mu_i(\bar{r})| \cos(\theta - \phi_i))^2 / (2\sigma_i)^2 = c^2$ , thus we have

$$\rho^2(\theta) = \frac{4c^2}{\sum p_i \cos^2(\theta - \phi_i)} \quad (31)$$

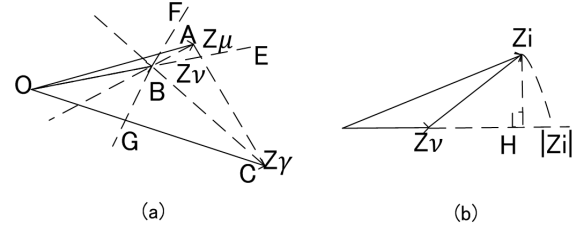


Fig. 5. Proof of Theorem 3

where  $p_i = (|\nabla \mu_i(\bar{r})|/\sigma_i)^2$ . Examining Eq. (31), and let

$$Q_1 = \sum p_i \cos^2 \phi_i,$$

$$Q_2 = \sum p_i \sin^2 \phi_i,$$

$$Q_3 = \sum 2p_i \cos \phi_i \sin \phi_i.$$

We have  $Q_1 \rho^2 \cos^2 \theta + Q_2 \rho^2 \sin^2 \theta + Q_3 \rho^2 \cos \theta \sin \theta = 4c^2$ , which means that  $\mathcal{U}$  is in fact an ellipse.

Define a complex parameter  $Z_i$  characterizing  $AP_i$ , where  $Z_i = p_i e^{2i\phi_i}$ ,  $\sum Z_i = \sum p_i e^{2i\phi_i}$ , and  $\sum Z_i^* = \sum p_i e^{-2i\phi_i}$ . The area of  $\mathcal{U}$  is denoted by  $u$ , where  $u = 8\pi c^2 / \sqrt{4Q_1 Q_2 - Q_3^2}$ . The area of the ellipse  $u$  profiles the accuracy of the localization. Recall that the event  $\mathcal{E}(c)$  determines the area of  $\mathcal{U}$  in the physical and  $\mathcal{E}(c)$  is determined by which outcomes the user has submitted. This means that which RSS fingerprints the user submitted determines the localization accuracy. To maximize the accuracy is equivalent to minimize  $u$ , thus the best strategy for the user is to adopt the measurement sequence  $\mathcal{V}_n^*$ , where

$$\mathcal{V}_n^* = \arg \max_{\mathcal{V}_n \in \mathbb{U}^n} \{ (\sum_{i \in \mathcal{V}_n} |Z_i|)^2 - |\sum_{i \in \mathcal{V}_n} Z_i|^2 \}. \quad (32)$$

It is indicated by Eq. (32) that the location determination system needs to search over the entire strategy profile  $\mathbb{U}^n$  to find the optimal strategy. In the following discussion, we are to prove that we can narrow down the searching space by eliminating APs with small  $|Z_i|$  from the set of all visible APs  $\mathbb{U}$ , and then search the strategy subspace induced by the rest of APs.

**Theorem 3.** Given that the user can choose to measure  $AP_\nu$ ,  $AP_\mu$  and  $AP_\gamma$ , where the measurement for each AP is denoted by  $Z_\nu$ ,  $Z_\mu$ ,  $Z_\gamma$ . If  $Z_\nu$  falls inside  $\triangle OZ_\mu Z_\gamma$  in the complex plane, then  $Z_\nu \notin \mathcal{V}_n^*$ .

*Proof:* Figure (5a) shows an arbitrary  $Z_\nu$  falls inside  $\triangle OZ_\mu Z_\gamma$  in the complex plane.  $O$  is the origin. We use  $A$  and  $C$  to represent the corresponding point of  $Z_\mu$  and  $Z_\gamma$ . We are to prove that any measurement such as  $Z_\nu$  that falls in the area of the triangle must not be an element of  $\mathcal{V}_n^*$  using contradiction.

If  $Z_\nu \in \mathcal{V}_n^*$ , let  $T = \sum_{\mathcal{V}_n^* \setminus \{Z_\nu\}} |Z_i|$ ,  $G = \sum_{\mathcal{V}_n^* \setminus \{Z_\nu\}} Z_i$ , where  $\mathcal{V}_n^* \setminus \{Z_\nu\}$  stands for the difference sequence of  $\mathcal{V}_n^*$  eliminating an arbitrary measurement to  $AZ_\nu$ . If there are multiple measurements to  $AP_\nu$ , it makes no difference to eliminate any one of them, since the measurement order does not matter.  $\mathcal{V}_n^* \setminus \{Z_\nu\} \cup \{Z_i\}$  stands for the strategy  $\mathcal{V}_n^* \setminus \{Z_\nu\}$



plus a measurement to  $AP_i$ .  $T$  is a real number while  $G$  could be a complex number. According to Eq. (32), we should have  $u(\mathcal{V}_n^*) \leq u(\mathcal{V}_n), \forall \mathcal{V}_n \in \mathbb{U}$ , where  $u(\mathcal{V}_n)$  is the area of the ellipse in the physical space given the chosen  $\mathcal{V}_n$ . Then the following equations should hold true for both  $\mu$  and  $\gamma$ :

$$\begin{cases} \left(\frac{8\pi c^2}{u(\mathcal{V}_n^*)}\right)^2 \geq \left(\frac{8\pi c^2}{u(\mathcal{V}_n^* \setminus \{Z_\nu\} \cup \{Z_\mu\})}\right)^2, \\ \left(\frac{8\pi c^2}{u(\mathcal{V}_n^*)}\right)^2 \geq \left(\frac{8\pi c^2}{u(\mathcal{V}_n^* \setminus \{Z_\nu\} \cup \{Z_\gamma\})}\right)^2. \end{cases} \quad (33)$$

This is equivalent to prove  $(T + |Z_\nu|)^2 - |G + Z_\nu|^2 \geq (T + |Z_i|)^2 - |G + Z_i|^2$ , for  $i = \mu$  and  $\gamma$ . According to Eq. (32), we should prove that

$$\frac{G}{T}(Z_i - Z_\nu) \geq |Z_i| - |Z_\nu|. \quad (34)$$

Let  $\theta_i$  be the angle between  $Z_\nu$  and  $Z_i - Z_\nu$ , then  $|Z_i| - |Z_\nu| > |Z_i - Z_\nu| \cos(\theta_i)$  as shown in Fig. (5b). This inequality still holds for the case  $|Z_i| < |Z_\nu|$  or  $\theta_i > \frac{\pi}{2}$ , where the proof is straight and thus skipped due to the limitation of space. It is straightforward that  $|G| < T$ , therefore  $|\frac{G}{T}| < 1$ . Thus if  $Z_\nu$  were to be an element of  $\mathcal{V}_n^*$ , the following two equations should both be true:

$$e(Z_\mu - Z_\nu) > |Z_\gamma - Z_\nu| \cos(\theta_\mu), \quad (35)$$

$$e(Z_\gamma - Z_\nu) > |Z_\gamma - Z_\nu| \cos(\theta_\gamma), \quad (36)$$

where  $e$  is a unit vector.

In Fig. (5a),  $OB$  and  $OE$  are collinear. We draw two lines  $BF$  and  $BG$  so that  $\angle EBA = \angle ABF$ ,  $\angle EBC = \angle CBG$ . Eq. (35) indicates the range of direction for  $e$  is from  $BE$  to  $BF$  (counterclockwise); Eq. (36) indicates the range of direction for  $e$  is from  $BE$  to  $BG$  (clockwise);  $\angle ABC < \pi$ , which means that it is impossible for the two scopes to overlap, which means there is no such  $e$  that makes Eq. (35) and Eq. (36) true at the same time, and the inequalities (33) can not hold simultaneously. Consequently, Theorem (3) is proved by the contradiction. ■

Theorem (3) can be understood as following: If we use a point on the convex plane to represent the measurement  $Z_i$ , then there will be many points on the plane representing all possible measurements. Only those points on the convex hull of all points are possible candidates of the best strategy. It is worth noting that the parameters used for determining the best strategy can be derived by analyzing the fingerprints collected for each AP in the database. There is no need for information about the location of APs is required, and no need for explicit efforts from users either.

## VII. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a general probabilistic model to shed light on a fundamental question: how good the RSS fingerprinting based indoor localization can achieve? Concretely, we have presented the probability that a user can be localized in a region with certain size, given the RSS fingerprints submitted to the system. We have revealed the interaction among the localization accuracy, the reliability of location estimation and the number of measurements in the

RSS fingerprinting based location determination. Moreover, we have presented the optimal fingerprints reporting strategy that can achieve the best accuracy for given reliability and the number of measurements, which provides a design guideline for the RSS fingerprinting based indoor localization facilitated by crowdsourcing paradigm. Our future work is to consider the scenario that the knowledge of the mean surface of outcomes are not completely known.

## VIII. ACKNOWLEDGEMENT

This work is supported by NSF China (No. 61325012, 61271219, 61221001, 61202373, 61102052, 61102051, 61428205, 91438115, U1405251); China Ministry of Education Doctor Program (No. 20130073110025); National Mobile Communications Research Laboratory, Southeast University (No. 2012D13, 2014D07); Shanghai Basic Research Key Project (12JC1405200, 11JC1405100); Shanghai International Cooperation Project: (No. 13510711300); National Basic Research 973 Program of China (No. 2015CB352403); Jiangsu Future Network Research Project No. BY2013095-1-10.

## REFERENCES

- [1] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach and L. E. Kavraki, "Practical robust localization over large-scale 802.11 wireless networks," in *Proc. ACM MobiCom*, 2004, pp. 70–84.
- [2] Z. Yang, Z. Zhou and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp.1–32, 2013.
- [3] P. Bahl and V. N. Padmanabhan, "Radar: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, 2000, pp. 775–784.
- [4] K. Chintalapudi, A. Padmanabha Iyer and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. ACM MobiCom*, 2010, pp. 173–184.
- [5] 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.
- [6] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," in *Proc. IEEE INFOCOM*, 2004, pp. 400–408.
- [7] P. Castro, P. Chiu, T. Kremenek and R. Muntz, "A probabilistic room location service for wireless networked environments," in *Proc. ACM UbiComp*, 2001, pp. 18–34.
- [8] 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.
- [9] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proc. ACM MobiSys*, 2005, pp. 205–218.
- [10] R. Battiti, M. Brunato and A. Delai, "Optimal wireless access point placement for location-dependent services," Technical Report, 2003, Online: <http://eprints.biblio.unitn.it/489/1/DIT-03-052-withCover.pdf>.
- [11] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless lans," *Computer Networks*, vol. 47, no. 6, pp. 825–845, 2005.
- [12] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. ACM MobiCom*, 2010, pp. 173–184.
- [13] 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.
- [14] M. A. Youssef and A. Agrawala, "On the optimality of wlan location determination systems," 2003. Online: <http://www.cs.umd.edu/~moustafa/papers/cnds04.pdf>.
- [15] S. M. Ross, "Introduction to probability and statistics for engineers and scientists," Academic Press, 2009.