

A Qualitative and Quantitative Evaluation of Differential Signal Strength Fingerprinting Methods

Christos Laoudias*, Sunwoo Kim[†], Demetrios Zeinalipour-Yazti[‡], and Christos G. Panayiotou*

*KIOS Research Center, University of Cyprus, Nicosia, Cyprus

[†]Department of Electronic Engineering, Hanyang University, South Korea

[‡]Department of Computer Science, University of Cyprus, Nicosia, Cyprus

Email: laoudias@ucy.ac.cy, remero@hanyang.ac.kr, dzeina@cs.ucy.ac.cy, christosp@ucy.ac.cy

Abstract—Indoor location systems that rely on WiFi signal strength values from the existing building infrastructure deliver adequate accuracy with no installation cost at the expense of time and effort to populate the signal database. To this end, crowdsourcing has been widely explored to leverage on large volumes of user-collected data; however, mobile devices that report signal strength differently, known as device diversity, limit its applicability in practice. We consider crowdsourced location systems and present a qualitative (in terms of analytical results) and a quantitative (with respect to real-life experimental data) evaluation of several approaches that are robust to diverse devices while focusing on differential signal strength methods.

I. INTRODUCTION

Location awareness has become an integral part of the mobile user's daily life and the wide availability of location information has triggered the advent of innovative location-based services, while according to *Market Research Future* the global market is estimated to grow up to USD 80 Billion by 2023 [1]. These services will be increasingly targeting indoor application scenarios as people already spend 80-90% of their time in indoor environments.

Among the wide range of indoor location technologies, WiFi signal strength fingerprinting that may be augmented with other sensor readings is gaining traction. This solution relies on the collection of location-tagged Received Signal Strength (RSS) measurements from the surrounding WiFi Access Points (AP) inside the building, referred to as *fingerprints*, which are stored in a signal database known as *radiomap*. Even though this approach has demonstrated comparable or even better accuracy than expensive solutions that require additional hardware, it has not been widely adopted. This is due to the data collection process, which can be tedious and time consuming especially for multi-floor buildings.

To reduce the data collection effort, crowdsourcing has emerged as a viable alternative that leverages user-collected WiFi RSS readings. Even though several crowdsourcing systems have been presented in the literature (see, for example, [2] and references therein), the key challenge that hinders their wider adoption is device diversity [3]. That is, mobile devices report largely different RSS values from a WiFi

AP, even if they reside at the same location, due to multi-vendor WiFi adapters, varying antenna gain, different circuitry packaging, etc. This fact makes the RSS values collected with diverse devices incompatible with each other, thus impeding the combination of all data in a single radiomap.

Some approaches apply data transformations to remove the device dependency from RSS readings, which enables a user to start localizing accurately as soon as he/she enters the building carrying a new device. For instance, the *FreeLoc* system [4] uses relative RSS values in a Rank Based Fingerprinting (RBF) approach, which ranks the WiFi APs from stronger to weaker RSS. However, the exact RSS values are lost, which hinders fine-grain localization [3]. The Hyperbolic Location Fingerprinting (HLF) approach uses normalized logarithm RSS ratios for all AP pairs in the observed fingerprint [5].

Another data transformation approach uses RSS differences, rather than absolute RSS values, to homogenize the RSS data from multiple devices. For instance, the DIFF method takes the difference between all AP pairs [6]. The Signal Strength Differences (SSD) method subtracts the RSS value of an anchor AP from the other RSS values [7]; however, it is reported that DIFF performs better than SSD for crowdsourced indoor localization [8]. Recently, the Mean Differential Fingerprint (MDF) method proposed to subtract the mean RSS value across all APs in the original fingerprint from each RSS value [9]. It was shown theoretically that MDF delivers exactly the same location estimates as DIFF, but at significantly less computational expense. These differential methods demonstrated promising location accuracy in simulation and small-scale experimental setups; however, still their behaviour has not been fully analyzed and they have not been compared against competing solutions in real-life conditions.

To this end, the contribution of this work is twofold. First, we attempt to gain theoretical insights on differential fingerprinting for crowdsourced indoor localization. We derive analytical models for their respective probability of correct location estimation and employ these models to assess and analyze their performance theoretically for the first time. Second, we conduct extensive experiments using real-life WiFi RSS data from two publicly available datasets and compare differential methods against other approaches. The results validate our analytical models and offer valuable guidelines for the design of crowdsourced localization systems.

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE) and from the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

The rest of the paper is structured as follows. Section II describes the problem of crowdsourced RSS fingerprinting and formulates existing differential fingerprinting methods. In Section III, analytical expressions are derived for these methods followed by an analysis of their localization performance with respect to various system parameters. Experimental results pertaining to real-life WiFi RSS datasets are discussed in Section IV. Finally, Section V provides concluding remarks.

II. CROWDSOURCED FINGERPRINTING

A. Problem Formulation

Traditional fingerprint-based systems operate in two stages, namely the *offline* phase where the WiFi RSS values are used to create the location-tagged radiomap of the building and the *online* phase where the radiomap is exploited to compute user location. In crowdsourced systems, data come from multiple heterogeneous devices that report RSS differently, while the user-carried device during localization most likely differs from the crowdsourcing devices.

In this context, the problem of crowdsourced fingerprint localization is formulated next. In the *offline* phase, a grid of reference locations $\{L : \ell_i = (x_i, y_i), i = 1, \dots, l\}$ are used to discretize the localization area. A set of crowdsourcing devices $D^{(m)}, m = 1, \dots, M$ collect RSS values from n WiFi APs. Device m visits a subset of the reference locations $\{L^{(m)} : \ell_i = (x_i, y_i), i = 1, \dots, l^{(m)}\}$, so that $L^{(m)} \subseteq L$ and $L = \bigcup_{m=1}^M L^{(m)}$. A reference fingerprint $\mathbf{r}_i^{(m)} = [r_{i1}^{(m)}, \dots, r_{in}^{(m)}]^T$ associated with location ℓ_i is a vector of RSS readings and $r_{ij}^{(m)}$ denotes the RSS reading from the j -th AP recorded by device $D^{(m)}$. The fingerprints collected by $D^{(m)}$ form the device-specific radiomap $\mathbf{R}^{(m)} \in \mathbb{Z}_{l^{(m)} \times n}^+$ that may partially cover the area, while all devices contribute their respective radiomaps. The naive approach for building the crowdsourced RSS radiomap $\mathbf{R} \in \mathbb{Z}_{l \times n}^+$, using all available radiomaps, is by aggregating the RSS values for each AP across all contributing devices M_i at location ℓ_i , where $1 \leq M_i \leq M$, according to

$$r_{ij} = \frac{1}{M_i} \sum_{m=1}^{M_i} r_{ij}^{(m)}. \quad (1)$$

In the *online* phase, given a new fingerprint $\mathbf{s} = [s_1, \dots, s_n]^T$ measured at the unknown location ℓ by the user-carried device $D^{(m')}$, the crowdsourced radiomap $\mathbf{R}$ is employed to estimate $\hat{\ell}$. For simplicity we use the Nearest Neighbor (NN) approach that determines location as

$$\hat{\ell}(\mathbf{s}) = \arg \min_{\ell_i} d_i, \quad d_i^2 = \sum_{j=1}^n (r_{ij} - s_j)^2, \quad (2)$$

where d_i^2 is the squared Euclidean distance between the reference fingerprints $\mathbf{r}_i$ and the observed fingerprint $\mathbf{s}$.

B. Crowdsourcing with Differential Fingerprints

We assume that the RSS values (in dBm) are given by the simple log-distance radio propagation model

$$RSS = A - 10\gamma \log_{10} d + X, \quad (3)$$

where d denotes the physical distance between a WiFi AP and the device, while A provides the RSS value at distance $d = 1$ m and encapsulates device specific characteristics, such as the antenna gain and transmitter power. The path loss exponent γ depends on the propagation conditions and $X \sim \mathcal{N}(0, \sigma^2)$ represents Gaussian noise that disturbs the RSS values.

Using RSS differences removes the constant term A in (3), which makes the differential fingerprints from diverse devices compatible with each other and enables the creation of the crowdsourced radiomap [9]. In the following, we outline the main differential fingerprint approaches.

1) *DIFF Approach*: Creates the differential fingerprints by taking the difference between all pairwise AP combinations [6]. Therefore, for a single device the DIFF reference fingerprint $\tilde{\mathbf{r}}_i$ at location ℓ_i and the DIFF fingerprint to be localized $\tilde{\mathbf{s}}$ are defined as

$$\tilde{\mathbf{r}}_i = [\tilde{r}_{i12}, \dots, \tilde{r}_{i(n-1)n}]^T \text{ and } \tilde{\mathbf{s}} = [\tilde{s}_{12}, \dots, \tilde{s}_{(n-1)n}]^T, \quad (4)$$

where $\tilde{r}_{ijk} = r_{ij} - r_{ik}$ and $\tilde{s}_{jk} = s_j - s_k$, $1 \leq j < k \leq n$ denote the RSS difference between the j -th and k -th APs in the reference and localization fingerprint, respectively.

Assuming M_i devices for crowdsourcing data at location ℓ_i , the differential radiomap $\tilde{\mathbf{R}} \in \mathbb{Z}_{l \times \frac{n(n-1)}{2}}$ contains the reference fingerprint $\tilde{\mathbf{r}}_i$ where $\tilde{r}_{ijk}$ is computed across all devices using (1) with $\tilde{r}_{ijk}^{(m)}$. Note that this approach increases the dimension of the original RSS fingerprints as the new fingerprints contain $\binom{n}{2} = \frac{n(n-1)}{2}$ RSS differences. Localization with the NN method is performed by replacing d_i^2 in (2) with

$$\tilde{d}_i^2 = \sum_{k=2}^n \sum_{j=1}^{k-1} (\tilde{r}_{ijk} - \tilde{s}_{jk})^2. \quad (5)$$

2) *SSD Approach*: Creates the differential fingerprints by subtracting the RSS value of an anchor AP from the other RSS values in the original fingerprint [7]. Without loss of generality, we assume that ρ is the anchor AP and we define for a single device the SSD reference fingerprint $\tilde{\mathbf{r}}_i$ at location ℓ_i and the SSD fingerprint to be localized $\tilde{\mathbf{s}}$ as

$$\tilde{\mathbf{r}}_i = [\tilde{r}_{i1}, \dots, \tilde{r}_{i(n-1)}]^T \text{ and } \tilde{\mathbf{s}} = [\tilde{s}_1, \dots, \tilde{s}_{n-1}]^T, \quad (6)$$

where $\tilde{r}_{ij} = r_{ij} - r_{i\rho}$ and $\tilde{s}_j = s_j - s_\rho$, $j = 1, \dots, n$, $j \neq \rho$ denote the RSS difference between the j -th AP and the anchor AP ρ in the reference and localization fingerprint, respectively.

Assuming M_i devices for crowdsourcing data at location ℓ_i , the differential radiomap $\tilde{\mathbf{R}} \in \mathbb{Z}_{l \times (n-1)}$ contains the reference fingerprint $\tilde{\mathbf{r}}_i$ where $\tilde{r}_{ij}$ is computed across all devices using (1) with $\tilde{r}_{ij}^{(m)}$. Note that in this case the dimension of the new fingerprints is decreased as they contain the $n - 1$ RSS differences that are independent. Localization with NN is performed by replacing d_i^2 with

$$\tilde{d}_i^2 = \sum_{j=1, j \neq \rho}^n (\tilde{r}_{ij} - \tilde{s}_j)^2. \quad (7)$$

3) *MDF Approach*: Creates the differential fingerprints by subtracting the mean RSS value pertaining to all APs from each RSS value in the original fingerprint [9], [10]. Thus, for a single device the MDF reference fingerprint $\bar{\mathbf{r}}_i$ at location ℓ_i and the MDF fingerprint to be localized $\bar{\mathbf{s}}$ are defined as

$$\bar{\mathbf{r}}_i = [\bar{r}_{i1}, \dots, \bar{r}_{in}]^T \text{ and } \bar{\mathbf{s}} = [\bar{s}_1, \dots, \bar{s}_n]^T, \quad (8)$$

where $\bar{r}_{ij} = r_{ij} - \bar{r}_i$ and $\bar{s}_j = s_j - \bar{s}$, $j = 1, \dots, n$ denote the RSS difference between the j -th AP and the mean RSS value in the reference and localization fingerprint, respectively. The mean RSS values in fingerprints $\mathbf{r}_i$ and $\mathbf{s}$ are given by $\bar{r}_i = \frac{1}{n} \sum_{j=1}^n r_{ij}$ and $\bar{s} = \frac{1}{n} \sum_{j=1}^n s_j$, respectively.

Assuming M_i devices for crowdsourcing data at location ℓ_i , the differential radiomap $\bar{\mathbf{R}} \in \mathbb{Z}_{l \times n}$ contains the reference fingerprint $\bar{\mathbf{r}}_i$ where $\bar{r}_{ij}$ is computed across all devices using (1) with $\bar{r}_{ij}^{(m)}$. Note that the MDF fingerprints have the same dimension as the original RSS fingerprints. Localization with NN is performed by replacing d_i^2 with

$$\bar{d}_i^2 = \sum_{j=1}^n (\bar{r}_{ij} - \bar{s}_j)^2. \quad (9)$$

III. ANALYSIS OF DIFFERENTIAL FINGERPRINTS

To gain insight in the behaviour of differential fingerprints, we derive analytically the probability of correct location estimation and then analyze their performance for crowdsourcing.

A. Probability of Correct Location Estimation

Let ℓ_1 and ℓ_2 be two neighbouring locations in the localization area. We consider two devices, i.e., $D^{(1)}$ and $D^{(2)}$, that collect RSS values at both ℓ_1 and ℓ_2 and we assume that r_{1j} and r_{2j} computed with (1) are deterministic because each has been averaged over a sufficiently large number of samples collected with both devices to filter out the noise. Therefore, the corresponding RSS fingerprints $\mathbf{r}_1$ and $\mathbf{r}_2$ in the crowdsourced radiomap are constant vectors. Moreover, for simplicity, we assume that during localization the user is located at ℓ_1 carrying device $D^{(1)}$ and observes the RSS fingerprint $\mathbf{s}^{(1)}$. The fingerprint $\mathbf{s}^{(1)}$ is a normally distributed random vector, i.e. $\mathbf{s}^{(1)} \sim \mathcal{N}(\mathbf{r}_1^{(1)}, \Sigma)$, where $\Sigma = \sigma^2 I_n$ is the covariance matrix, while σ^2 is the variance of the Gaussian noise that disturbs the RSS values and I_n is the identity matrix.

The objective is to analytically derive the probability that the NN localization method will return the correct location ℓ_1 , instead of the incorrect location ℓ_2 when RSS, DIFF, SSD, or MDF fingerprints are used, respectively. Next, we generalize our results for areas where the radiomap contains several location fingerprints.

1) *Analytical Model for RSS Fingerprints*: In the case of traditional RSS fingerprints, the NN method will return the correct location ℓ_1 only if the condition $d_1^2 \leq d_2^2$ is satisfied. In other words the unknown user location is correctly identified only if the distance between the observed fingerprint $\mathbf{s}$ and the correct location fingerprint $\mathbf{r}_1$ is smaller than the distance between the observed fingerprint and the incorrect neighbouring location fingerprint $\mathbf{r}_2$.

We follow an approach similar to [11] and starting from the above condition we can easily show that¹

$$d_1^2 \leq d_2^2 \Leftrightarrow 2 \sum_{j=1}^n \beta_j s_j + \sum_{j=1}^n \gamma_j \leq 0, \quad (10)$$

where $\beta_j = (r_{2j} - r_{1j})$ and $\gamma_j = (r_{1j}^2 - r_{2j}^2)$. For convenience, we use vector notation to rewrite (10) as

$$\boldsymbol{\beta} \mathbf{s} + \boldsymbol{\gamma} \leq 0, \quad (11)$$

where $\boldsymbol{\beta} = 2[\beta_1, \dots, \beta_n]$ and $\boldsymbol{\gamma} = \sum_{j=1}^n \gamma_j$.

The random variable $C = \boldsymbol{\beta} \mathbf{s} + \boldsymbol{\gamma}$ is normally distributed, as a linear function of the multivariate normal vector $\mathbf{s}$, i.e. $C \sim \mathcal{N}(\mu_C, \sigma_C^2)$ with

$$(\mu_C, \sigma_C^2) = (\boldsymbol{\beta} \mathbf{r}_1^{(1)} + \boldsymbol{\gamma}, \boldsymbol{\beta} \Sigma \boldsymbol{\beta}^T). \quad (12)$$

Thus, the probability of correct location estimation, when the NN method compares the observed fingerprint with just two location fingerprints in the radiomap, is given by

$$Pr\{C \leq 0\} = \frac{1}{2} + \frac{1}{2} \text{erf}\left(\frac{-\mu_C}{\sqrt{2\sigma_C^2}}\right). \quad (13)$$

In a real system where the radiomap is expected to contain several location fingerprints, depending on the size of the area and the density of the reference locations, the probability of correct location estimation $\mathcal{P}_C$ can be calculated as

$$\mathcal{P}_C = Pr\{C_2 \leq 0, \dots, C_l \leq 0\} \approx \prod_{i=2}^l Pr\{C_i \leq 0\}, \quad (14)$$

where the variable C_i corresponds to the condition $d_1^2 \leq d_i^2$, $i = 2, \dots, l$. Although this analytical model is based on the assumption that the random variables C_i are independent, which is not true, it still provides a good approximation of the probability of correct location estimation [11].

2) *Analytical Model for MDF and DIFF Fingerprints*: In our previous work, it was shown that the distances $\bar{d}_i^2$ between the MDF fingerprints $\bar{\mathbf{r}}_i$ and $\bar{\mathbf{s}}$ are proportional to the distances $\tilde{d}_i^2$ between the DIFF fingerprints $\tilde{\mathbf{r}}_i$ and $\tilde{\mathbf{s}}$, i.e., $\bar{d}_i^2 = \frac{1}{n} \tilde{d}_i^2$ [9]. Thus, the ordering of the candidate locations $\ell_i \in \bar{L}$ is not affected when either MDF or DIFF fingerprints are considered. Consequently, the user locations estimated with NN are exactly the same and we derive the analytical model only for MDF.

Starting from the condition $\bar{d}_1^2 \leq \bar{d}_2^2$ and using *Theorem 1* in [9], which provides the relation of the distance $\bar{d}_i^2$ between MDF fingerprints with respect to the distance d_i^2 between the corresponding RSS fingerprints, it can be shown that

$$\bar{d}_1^2 \leq \bar{d}_2^2 \Leftrightarrow 2 \sum_{j=1}^n \beta_j (s_j - \bar{s}) + \sum_{j=1}^n \delta_j \leq 0, \quad (15)$$

where $\beta_j = (r_{2j} - r_{1j})$ and $\delta_j = (r_{1j}^2 - \bar{r}_1^2 - r_{2j}^2 + \bar{r}_2^2)$.

Using vector notation, (15) is equivalent to

$$\boldsymbol{\beta} \mathbf{J} \mathbf{s} + \boldsymbol{\delta} \leq 0, \quad (16)$$

¹For brevity we drop the superscript in the localization fingerprint $\mathbf{s}^{(1)}$.

where $\beta = 2[\beta_1, \dots, \beta_n]$, $J = I - \frac{1}{n}\mathbf{e}^T\mathbf{e}$, $\mathbf{e} = [1, \dots, 1]$ and $\delta = \sum_{j=1}^n \delta_j$. The random variable $Q = \beta J s + \delta$ is normally distributed², i.e. $Q \sim \mathcal{N}(\mu_Q, \sigma_Q^2)$ with mean and variance

$$(\mu_Q, \sigma_Q^2) = (\beta J \mathbf{r}_1^{(1)} + \delta, \beta J \Sigma J^T \beta^T). \quad (17)$$

The probability of correct location estimation, using the NN localization method with MDF fingerprints, in the two-location case and the multiple location setup respectively, are given by

$$Pr\{Q \leq 0\} = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{-\mu_Q}{\sqrt{2\sigma_Q^2}}\right) \text{ and } \mathcal{P}_Q \approx \prod_{i=2}^l Pr\{Q_i \leq 0\}. \quad (18)$$

3) *Analytical Model for SSD Fingerprints*: Starting from the condition $\check{d}_1^2 \leq \check{d}_2^2$ and using *Theorem 2* in [10], which provides the relation of the distance $\check{d}_i^2$ between SSD fingerprints with respect to the distance d_i^2 between the corresponding RSS fingerprints, it can be shown that

$$\check{d}_1^2 \leq \check{d}_2^2 \Leftrightarrow 2 \sum_{j=1}^n (\beta_j - \beta_\rho)(s_j - s_\rho) + \sum_{j=1}^n \epsilon_j \leq 0, \quad (19)$$

where $\beta_j = (r_{2j} - r_{1j})$, $\beta_\rho = (r_{2\rho} - r_{1\rho})$ and $\epsilon_j = (r_{1j}^2 + r_{1\rho}^2 - r_{2j}^2 - r_{2\rho}^2 - 2r_{1\rho}\bar{r}_1 + 2r_{2\rho}\bar{r}_2)$, with ρ being the anchor AP. Again, using vector notation we may rewrite (19) as

$$\boldsymbol{\eta} M \mathbf{s} + \boldsymbol{\epsilon} \leq 0, \quad (20)$$

where $\boldsymbol{\eta} = 2[\beta_1 - \beta_\rho, \dots, \beta_n - \beta_\rho]$, $M = I - \mathbf{e}^T \mathbf{u}$, $\mathbf{e} = [1, \dots, 1]$, $\mathbf{u} = [0, \dots, 1, \dots, 0]$ and $\boldsymbol{\epsilon} = \sum_{j=1}^n \epsilon_j$.

Following the same process with the MDF approach, we can show that the random variable $R = \boldsymbol{\eta} M \mathbf{s} + \boldsymbol{\epsilon}$ is normally distributed, i.e. $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$ with mean and variance

$$(\mu_R, \sigma_R^2) = (\boldsymbol{\eta} M \mathbf{r}_1^{(1)} + \boldsymbol{\epsilon}, \boldsymbol{\eta} M \Sigma M^T \boldsymbol{\eta}^T). \quad (21)$$

Finally, the probabilities that the NN localization method returns the correct location when it compares the observed SSD fingerprint with just two SSD location fingerprints and several SSD location fingerprints respectively are given by

$$Pr\{R \leq 0\} = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{-\mu_R}{\sqrt{2\sigma_R^2}}\right) \text{ and } \mathcal{P}_R \approx \prod_{i=2}^l Pr\{R_i \leq 0\}. \quad (22)$$

B. Performance Evaluation

We adopt the simple localization setup depicted in Fig. 1a [8], [11]. There are $l = 9$ reference locations (marked with circles) that are uniformly spread over a square grid. The WiFi APs (marked with triangles) are deployed in the perimeter and we start out with $n = 4$ APs that are placed at the four corners of the area, while the maximum number is $n = 16$ APs.

We assume that $D^{(1)}$ is the first crowdsourcing device and the RSS values $r_{ij}^{(1)}$ at the reference locations are given by the propagation model of (3). We use typical values

$A = -22.7$ dBm and $\gamma = 3.3$ for the model parameters. We also assume that $r_{ij}^{(1)}$ is deterministic because it has been averaged over a sufficiently large number of samples to filter out the noise. These RSS values, generated at all 9 locations, constitute the device-specific radiomap $\mathbf{R}^{(1)}$. In addition, a number of crowdsourcing devices $D^{(m)}$, $m = 2, \dots, M$ also cover the whole area and we assume a linear relation between the RSS values reported by any device $D^{(m)}$ and the corresponding RSS values of device $D^{(1)}$ such that

$$r_{ij}^{(m)} = \alpha_{1m} r_{ij}^{(1)} + \beta_{1m}, \quad m = 2, \dots, M, \quad (23)$$

where $(\alpha_{1m}, \beta_{1m})$ are the linear fitting parameters between devices $D^{(m)}$ and $D^{(1)}$ [8]. All M devices contribute their respective radiomaps $\mathbf{R}^{(m)}$ and the crowdsourced RSS radiomap $\mathbf{R}$ is created by aggregating the RSS values according to (1).

We now employ the analytical models in (14), (18), and (22) for RSS, MDF, and SSD to analyse the performance of crowdsourced localization using differential fingerprints. We start out with two devices for crowdsourcing, i.e., $D^{(1)}$ and $D^{(2)}$ with $(\alpha_{12}, \beta_{12}) = (0.95, 10)$, and we assume that a user carries device $D^{(1)}$ during localization while he/she resides at the central location $\ell_1 = (2, 2)$. We generate the RSS fingerprint $\mathbf{s}$ by taking the RSS location fingerprint $\mathbf{r}_i^{(1)}$ and disturbing it with Gaussian noise. Then, we compute the probability of returning the correct location ℓ_1 .

First, we investigate the effect of varying number of APs. In Fig. 1b we observe that the performance of the traditional RSS fingerprints is poor when few APs are considered; however, for more APs (i.e., $n > 8$) they perform better or equally well to SSD and MDF, respectively. Moreover, peaks occur for certain subsets of APs, i.e., $n \in \{4, 8, 12, 16\}$. In these cases the uniform distribution of APs creates symmetries that mask the effect of device diversity³. Similar behaviour was reported with simulations in [8]. Nevertheless, in large-scale real-life setups, the APs are neither expected to be densely deployed nor evenly distributed. In contrast, device diversity does not have any impact on MDF and its performance improves. SSD is inferior due to the lower dimension of the SSD fingerprints, while the error bars indicate the standard deviation around the mean value when using different APs as anchors.

Fig. 1c plots the probability of correct location estimation for increasing noise standard deviation σ with $n = 6$ APs. It is evident that even under low noise conditions RSS fingerprints perform poorly. In contrast, the differential approaches are more robust to noise and their performance degrades smoothly. Although MDF and SSD achieve similar performance when $\sigma = 1$ dBm, beyond that point the MDF approach outperforms SSD by around 10%–20% when $\sigma \leq 6$ dBm.

The effect of growing number of crowdsourcing devices is depicted in Fig. 1d. For $M = 1$, only device $D^{(1)}$ is used for building the radiomap, while the same device is used for localization. Therefore, it is not surprising that RSS performs best. However, when more devices are considered,

²The detailed derivation of (15) and the proof that Q is a normally distributed random variable is omitted due to space limitation.

³This can be easily verified numerically using (23) for two heterogeneous devices and four WiFi APs on the corners of the square area in Fig. 1a.

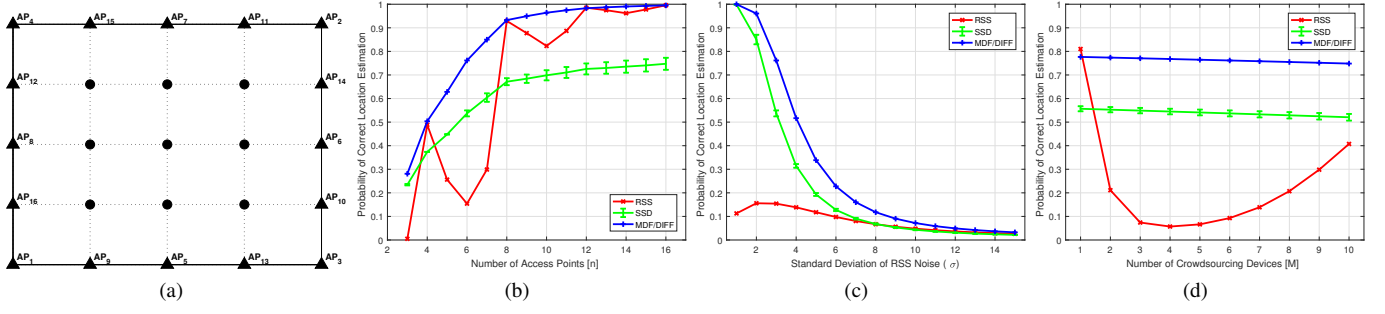


Fig. 1. (a) Localization setup for the performance analysis. Probability of correct location estimation for device $D^{(1)}$ for varying: (b) number of APs [$M = 2$ devices, $\sigma = 3$ dBm]; (c) noise standard deviation [$M = 2$ devices, $n = 6$ APs]; (d) number of crowdsourcing devices [$\sigma = 3$ dBm, $n = 6$ APs].

the performance of RSS drops rapidly and then starts to gradually improve when $M \geq 6$. Still the performance of RSS is well below the differential methods. Both MDF and SSD have a slight linear degradation as more devices contribute to crowdsourcing, while MDF outperforms SSD by around 20%.

IV. EXPERIMENTAL RESULTS

A. Indoor Localization Datasets

KIOS dataset: This small-scale dataset contains WiFi RSS values collected with 5 diverse devices, including a PDA, a laptop, an Android tablet, and two different Android smartphones. The size of the office area is 560 m² and is fully covered by 9 APs. There are 105 reference locations and 20 fingerprints have been collected in every location (i.e., 2,100 training fingerprints in total for each device). There are 96 test locations that do not coincide with the reference locations and 10 test fingerprints are available in every test location (i.e., 960 test fingerprints in total for each device).

UJIIndoorLoc dataset: This large-scale dataset contains WiFi RSS values from tens of APs collected by more than 20 users and 25 Android devices at three buildings of Universitat Jaume I with 4 or more floors and almost 110,000 m² [12]. We report the results for those floors in various buildings where around 100 or more test fingerprints are available.

B. Evaluation Process and Performance Metric

We use both real-life datasets to compare the traditional RSS fingerprints against the differential fingerprinting methods SSD [7] and MDF [9], as well as other approaches such as HLF [5] and RBF that is similar to the rank-based method used in *FreeLoc* [4]. For SSD, the AP that exhibits the least average deviation of RSS values over the whole localization area is set as anchor. We followed the same approach for the HLF method to avoid the computational overhead of the original method that uses all pair-wise AP combinations (e.g., in the UJIIndoorLoc dataset the original HLF fingerprints would contain more than 3,570 values, thus incurring 40x more computations in NN localization). The DIFF method [6] is omitted as it provides exactly the same accuracy with MDF at a considerably higher computational cost.

We compute the localization error defined as the Euclidean distance between the actual user location and his/her location

estimated with the NN approach. We report statistics for the cumulative error distribution and the central mark in the box plots indicates the median error, the box edges correspond to the 25th and 75th percentiles, and the whiskers extend to the 5th and 95th percentiles, respectively. The bar charts present the mean localization error together with the 95% confidence interval given by $\pm 1.96\sigma_e/\sqrt{|T|}$, where σ_e is the standard deviation of the error and $|T|$ is the number of test samples. The 95% confidence interval indicates that the mean error falls within the interval with a high degree of certainty.

The results for the KIOS dataset are illustrated in Fig. 2. First, we focus on one device for localization (i.e., HTC Desire smartphone) and consider all five devices for crowdsourcing in Fig. 2a. We observe that the MDF and SSD approaches outperform the HLF and RBF methods. MDF is slightly better than SSD with the median error being well below 2 m. Even though the performance of RSS fingerprints is improved with more crowdsourcing devices, still they fail to deliver the same level of accuracy as MDF. This observation is confirmed in Fig. 2b where we study the error of localizing the Samsung Nexus S device for increasing number of crowdsourcing devices M . When $M = 1$, only the Nexus S is used to build the radiomap and the RSS fingerprints achieve the lowest localization error. However, in case $M = 2$ their performance degrades significantly, as the second device reports RSS very differently from the Nexus S. This effect is gradually smoothed out as more devices contribute data. The behavior of RSS is in agreement with our analytical findings in Fig. 1d.

Figure 3 depicts the localization error using data from Building 'x' and Floor 'y', denoted Bx_{Fy} for brevity, of the UJIIndoorLoc dataset. The MDF method outperforms significantly all other methods at B1F1 as shown in Fig. 3a. This is explained by the number of crowdsourcing devices (i.e., 5 devices) and the difficulty in selecting the anchor AP in such large setups for the SSD and HLF methods. Regarding B0F1, where two devices are used for crowdsourcing, the MDF method still provides lower error than SSD and HLF by 1 m and RBF by 2.5 m, as shown in Table I. The RSS method performs equally well with MDF because the devices in both cases are Android devices that seem to report RSS similarly. Figure 3b plots the localization error at B1F2 for

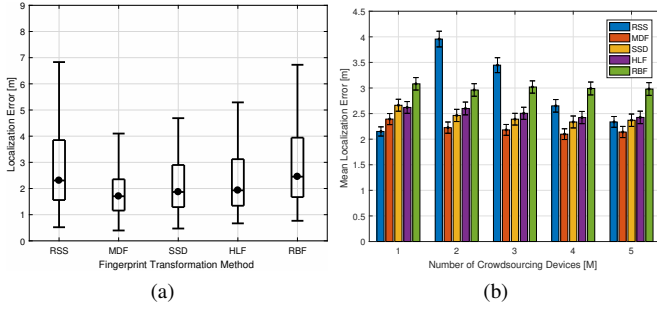


Fig. 2. KIOS dataset. (a) localizing the HTC Desire smartphone with crowdsourced data for the radiomap from all five devices; (b) localization of the Nexus S smartphone for increasing number of crowdsourcing devices.

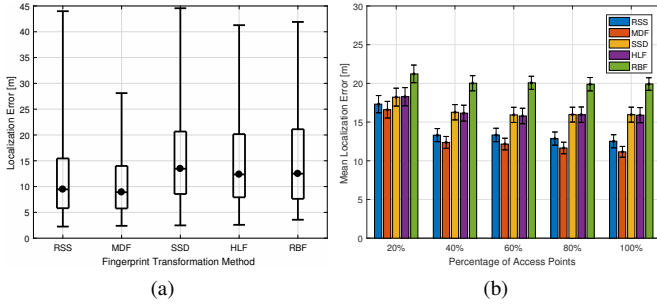


Fig. 3. Results with the UJIIndoorLoc dataset in Building ‘x’ and Floor ‘y’. (a) B1F1; (b) error at B1F2 for varying percentage of APs.

varying number of APs (shown as % of all APs in that floor). The results are averaged for ten scenarios with randomly selected subsets of APs. The RSS fingerprints do not have the oscillations predicted analytically (see Fig. 1b) due to the irregular deployment of the WiFi APs. Even though the performance of RSS may seem to be close to MDF, the median error in Table I suggests that in some cases, e.g., B1F1 or B1F2, the RSS fingerprints may lead to lower accuracy.

V. CONCLUSIONS

We investigate the effectiveness of WiFi RSS data transformation methods against device diversity for crowdsourced localization systems. Our key findings are i) RSS values should be avoided as in most cases they degrade the quality of the crowdsourced radiomap and lead to higher errors; ii) the use of RSS values could be justified in areas covered by a large number of WiFi APs, when the volume of crowdsourcing

devices is high, or if the devices are not highly heterogeneous, e.g., Android devices equipped with similar WiFi adapters; iii) differential methods deliver higher accuracy compared to other transformation methods; and iv) the MDF method achieves the best performance under all crowdsourcing scenarios. These findings can be viewed as design guidelines for real systems depending on the size of the area and number of WiFi APs, the volume and the heterogeneity of the crowdsourcing devices.

Compared to traditional RSS, the MDF method addresses device diversity with a marginal overhead (i.e., to compute the mean and subtract from the RSS values in the original fingerprint), while preserving the fingerprint dimension. Importantly, MDF can be applied as a data pre-processing technique regardless of the localization algorithm, e.g., NN or more advanced fingerprint matching approaches. As part of our future work, we plan to integrate the MDF method into our *Anyplace*⁴ crowdsourced service [13] to further assess its value and measure the power consumption on real mobile devices.

REFERENCES

- [1] “Location based service market report - forecast till 2023,” Market Future Research, Tech. Rep., 2018. [Online]. Available: <https://www.marketresearchfuture.com/reports/location-based-service-market-5439>
- [2] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, and C. Fischione, “A survey of enabling technologies for network localization, tracking, and navigation,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3607–3644, 2018.
- [3] B. Wang, Q. Chen, L. T. Yang, and H.-C. Chao, “Indoor smartphone localization via fingerprint crowdsourcing: Challenges and approaches,” *IEEE Wireless Communications*, vol. 23, no. 3, pp. 82–89, 2016.
- [4] S. Yang, P. Dessai, M. Verma, and M. Gerla, “FreeLoc: Calibration-free crowdsourced indoor localization,” in *IEEE International Conference on Computer Communications INFOCOM*, 2013, pp. 2481–2489.
- [5] M. B. Kjergaard, “Indoor location fingerprinting with heterogeneous clients,” *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 31–43, 2011.
- [6] F. Dong, Y. Chen, J. Liu, Q. Ning, and S. Piao, “A calibration-free localization solution for handling signal strength variance,” in *2nd international conference on Mobile Entity Localization and Tracking in GPS-less environments (MELT)*, 2009, pp. 79–90.
- [7] A. Mahtab Hossain, Y. Jin, W.-S. Soh, and H. N. Van, “SSD: A robust RF location fingerprint addressing mobile devices’ heterogeneity,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 65–77, 2013.
- [8] C. Laoudias, D. Zeinalipour-Yazti, and C. G. Panayiotou, “Crowd-sourced indoor localization for diverse devices through radiomap fusion,” in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2013, pp. 1–7.
- [9] C. Laoudias, P. Kolios, and C. Panayiotou, “Differential signal strength fingerprinting revisited,” in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014, pp. 30–37.
- [10] C. Laoudias, “Localization and tracking in wireless networks for fault tolerance and device diversity,” Ph.D. dissertation, University of Cyprus, Faculty of Engineering, May 2014. [Online]. Available: <http://hdl.handle.net/10797/14156>
- [11] K. Kaemarungsi and P. Krishnamurthy, “Modeling of indoor positioning systems based on location fingerprinting,” in *IEEE International Conference on Computer Communications (INFOCOM)*, vol. 2, 2004, pp. 1012–1022.
- [12] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, “UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems,” in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2014, pp. 261–270.
- [13] D. Zeinalipour-Yazti, C. Laoudias, K. Georgiou, and G. Chatzimilioudis, “Internet-based indoor navigation services,” *IEEE Internet Computing*, vol. 21, no. 4, pp. 54–63, Jul. 2017.

⁴<https://anyplace.cs.ucy.ac.cy/>

TABLE I
MEDIAN LOCALIZATION ERROR [M] IN UJIINDOORLOC DATASET

BxFy*/ Method	B0F1	B0F2	B1F1	B1F2	B2F1
RSS	5.62	4.83	9.46	8.51	5.57
MDF	5.44	4.87	8.91	7.44	5.53
SSD	6.44	6.27	13.45	9.67	7.38
HLF	6.44	5.55	12.35	9.57	6.40
RBF	8.03	8.63	12.50	18.09	7.86

*Building ‘x’ and Floor ‘y’.