Device Signal Strength Self-Calibration using Histograms

Christos Laoudias*, Robert Piché†, Christos G. Panayiotou*
*KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus
Email: laoudias@ucy.ac.cy, christosp@ucy.ac.cy
†Tampere University of Technology, Tampere, Finland
Email: robert.piche@tut.fi

***Abstract*—Traditionally, positioning with WLAN Received Signal Strength (RSS) fingerprints involves the laborious task of collecting a radiomap with a reference mobile device. Best accuracy can be guaranteed only in case the user carries the same device, while positioning with different devices requires a calibration step to make the new devices' RSS values compatible with the existing radiomap. We propose a novel device self-calibration method that uses histograms of RSS values. First, we use the existing radiomap to create the RSS histogram of the reference device. Subsequently, when the user enters a building and starts positioning, the observed RSS values are recorded simultaneously in the background to create and update the histogram of the user device. We use these RSS histograms to fit a linear mapping between the user and reference device.**

Our calibration method is running concurrently with positioning and does not require any user intervention. Experimental results with five smartphones in a real indoor environment indicate that soon after positioning is initiated the device is self-calibrated, thus improving the accuracy to be comparable with the case of using a radiomap created with the same device.

I. Introduction

Location-oriented services and applications are becoming increasingly popular and are expected to be widely used in the following years, especially in large indoor environments, such as shopping malls, exhibition centers, airports, hospitals, etc. This is mainly due to the fact that people tend to spend much of their time indoors and they are making increasing use of their smart mobile devices that feature high processing power and wireless connectivity.

Positioning based on WLAN technology is a popular solution and has attracted research interest, even though its accuracy is reported to be limited to a few meters. This popularity is because of the ubiquitous WLAN infrastructure, i.e. Access Points (AP), and the availability of WLAN cards on the vast majority of portable electronic devices, which provide access to different types of measurements such as Received Signal Strength (RSS). The RSS values are location-dependent, thus enabling location determination, while they are already monitored as part of the standard WLAN functionality for network operation reasons and can be easily recorded on the device. In this context, several positioning methods use a number of RSS fingerprints collected a priori with a device at some predefined reference locations to create the radiomap of the area. During positioning, location can be estimated by finding the best match between the current fingerprint measured with the same device and the fingerprints in the radiomap [1]–[3].

Using a different device is feasible, but the RSS values are not usually compatible with the radiomap, leading to accuracy degradation and limiting the applicability of the RSS fingerprinting approach. Different mobile devices do not report the same RSS values from the APs in the vicinity, even if they are placed in the same location. This is mainly due to the WLAN (IEEE 802.11) standard specification that defines an 1 byte integer, known as the RSS Indicator (RSSI), for measuring the RSS values in the range between 0 and 255. To make things worse, the actual implementation of each vendor is limited between 0 and a specific maximum RSSI value, so the RSSI levels may differ significantly among various chipsets [4]. The RSSI values are used internally by the device driver to report the actual power in dBm (e.g., for determining the signal quality), however this mapping is proprietary information. The result is that each vendor has its own measurement accuracy, granularity and dynamic range of RSS values, not to mention the receiver sensitivity that makes two WLAN cards detect a variable number of APs at the same location. Even the same chipsets may not report the same RSS values due to different antennas or different device packaging material [5].

Device diversity is one of the reasons that hinders the proliferation of RSS-based positioning systems. This is because best accuracy can be guaranteed only in case the user carries the same device during positioning, while supporting different types of devices requires a *calibration* step to make the new device compatible with the existing radiomap. To this end, several calibration solutions have been studied in the literature; these fall under two main categories. Methods in the first category try to remove the device-dependent component in the RSS values, e.g. due to device-specific antenna characteristics, so that the resulting fingerprints from heterogeneous devices are compatible with each other [6]–[9]. Methods in the second category rely on RSS data fitting to create a mapping between the RSS values collected with different devices, e.g. using standard least squares fitting. However, these methods require the collection of a considerable volume of data at several *known* (denoted as manual calibration) or *unknown* locations prior to positioning [10]–[13]

In this context, our contribution is twofold. First, we study the manual calibration approach and investigate the amount of RSS data that need to be collected at *known* locations with different devices, so that adequate positioning accuracy is achieved for these devices. Second, and more important, we propose a new method based on RSS histograms that runs concurrently with positioning and enables a mobile device to be self-calibrated in a short time, thus improving the positioning accuracy on-the-fly. Moreover, no user involvement in the calibration phase is required (e.g., visiting several locations and pressing a calibration button on the device) and the tedious data collection is avoided.

The rest of this paper is structured as follows. We discuss the related work in more detail in Section II. In Section III, we provide some preliminaries on signal strength fingerprinting and present our measurement setup. Section IV describes the manual calibration approach and investigates the amount of data it requires. We provide the details of the proposed self-calibration method in Section V and we assess its performance in Section VI using experimental data collected with five commercial mobile devices inside a real office environment. Finally, we conclude this work and discuss our future plans in Section VII.

II. Related Work

Device calibration has recently attracted the research interest, due to the requirement for the provision of accurate and reliable location estimates when heterogeneous mobile devices are considered. To address this issue several works try to remove the device-dependent component in the RSS values, given by the log-distance radio propagation model

$$RSS[dBm] = A - 10\gamma \log_{10} d + X. \quad (1)$$

In this model, d denotes the distance between the transmitter (e.g., a WLAN AP) and the receiver (e.g., a mobile device), while the intercept term A provides the RSS value when $d = 1\,\text{m}$ and encapsulates device specific characteristics, such as the antenna gain. The coefficient γ depends on the propagation environment, while $X \sim \mathcal{N}(0, \sigma_n^2)$ is the Gaussian noise disturbing the RSS values. In this context, authors in [7] use RSS differences, instead of absolute RSS values, to form the fingerprints either by taking the difference between all possible AP pairs or by subtracting the RSS value of an anchor AP (e.g., the AP that provides the strongest RSS value) from the other RSS values. This effectively removes A in (1) and makes the RSS difference fingerprints from diverse devices compatible with each other. However, this may degrade the positioning accuracy because RSS differences exhibit higher noise variance, as explained in Appendix A. A similar approach uses normalized log ratios of the RSS power from different APs to remove the device-dependent component [8]. Other methods rely on ranked, rather that absolute, RSS values (i.e., RSS values from a set of APs are ranked from high to low) because the ranking is not affected by device-specific hardware features [6], [9]. However, ranked-based fingerprints are expected to perform worse, compared to standard RSS fingerprints, because the fine-grain information of the RSS levels is lost when ranks are used.

On a different line, other methods rely on RSS data fitting to create a mapping between different devices. Such methods are motivated by the linear relation between the RSS values reported by heterogeneous devices, which has been observed experimentally in several studies. One approach referred to as manual calibration [10], is to collect a series of RSS measurements at several *known* locations with a pair of devices and subsequently the linear parameters are estimated through least squares fitting [10], [12], [13]. Instead of least squares fitting, authors in [11] create the empirical cdfs for several devices using the RSS values collected at *known* locations and then use the inverse cdf function to build a database of device models that map the RSS values of a user device to a reference device. However, this method is applicable to a limited number of device pairs, while the selection of the appropriate model during positioning relies on the existence of an easily distinguishable location (e.g., building entrance or exit) that may never be visited by the user.

Device calibration with RSS data recorded by the user at *unknown* locations is feasible, but computationally expensive methods are required to obtain the linear fitting parameters. For instance, the parameters can be estimated by maximizing the confidence value produced by Markov localization [10] or with a weighted least squares method [12]; we call these quasi-automatic, as opposed to manual, calibration methods.

In automatic calibration, RSS data collected at *unknown* locations are used as in the quasi-automatic case, however the objective is to minimize the user intervention and ideally perform positioning and device calibration simultaneously, while the user walks freely inside the area of interest. To this end, an expectation maximization (EM) algorithm is proposed in [10], while authors in [12] detect when the user is stationary during positioning in order to divide the data into parts which come from the same *unknown* location and then use these data with their quasi-automatic calibration approach. An unsupervised learning method is proposed in [5] that uses EM and neural networks algorithms to obtain the fitting parameters from the Pearson product-moment correlation coefficient. Authors in [14] make an early, rather sketchy, mention of the possibility of using histograms for automatic device RSS calibration.

III. Signal Strength Fingerprinting

A. Preliminaries

Several solutions to the location determination problem using RSS fingerprints have been studied in the literature. These approaches differ in the underlying positioning algorithm, however they all rely on a RSS radiomap that covers the entire area of interest; see [15] and references therein. Fingerprint-based positioning consists of two phases, namely the offline (training) and the online (positioning) phases.

Offline phase: We use a set of predefined reference locations $\{L : \ell_i = (x_i, y_i),\ i = 1, \ldots, l\}$ to collect RSS measurements from n APs using a reference device D_0. A reference fingerprint $r_i = [r_{i1}, \ldots, r_{in}]^T$ associated with location ℓ_i

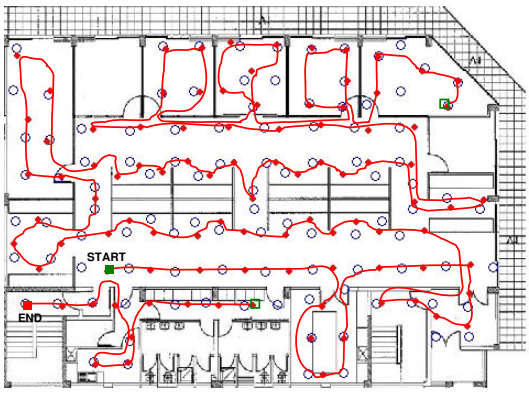


Fig. 1. Experimental setup at KIOS Research Center. Reference locations used during the offline training phase are shown with blue circles; the positioning phase route is shown with a red line.

is a vector of RSS samples and r_{ij} denotes the RSS value related to the j-th AP. Usually, r_i is averaged over the multiple fingerprints collected at ℓ_i so that only one fingerprint, i.e. $\overline{r}_i = \frac{1}{M}\sum_{m=1}^{M} r_i(m)$, is stored in the RSS radiomap followed by the physical coordinates (x_i, y_i). With this preprocessing we reduce the effect of noise in RSS measurements and outlier values, while the radiomap is compressed leading to a significant decrease in the location estimation time.

Online phase: During positioning, we exploit the reference data to obtain a location estimate $\widehat{\ell}$, given a new fingerprint $s = [s_1, \ldots, s_n]^T$ measured at the unknown location ℓ by some device $D_i,\ i = 0, \ldots, N_d$. The positioning algorithm tries to find the best match between the currently observed fingerprint s and the reference fingerprints $\overline{r}_i$ in the RSS radiomap. Various positioning algorithms have been presented, such as deterministic and probabilistic approaches [1]–[3], assuming that the offline and online phases are performed with the same device D_0. In this work we allow the online device to be *any* device $D_i,\ i = 0, \ldots, N_d$ and focus on the improvement achieved solely by our device self-calibration method, rather than the fingerprint-based positioning algorithm itself. Thus, our results are obtained using the well known Nearest Neighbor (NN) method [1] that estimates location by minimizing the Euclidean distance d_i, between the observed fingerprint s and the reference fingerprints $\bar{r}_i$

$$\widehat{\ell}(s) = \arg\min_{\ell_i} d_i, \quad d_i = \sqrt{\sum_{j=1}^{n} \left(\bar{r}_{ij} - s_j\right)^2}. \tag{2}$$

All reference locations are ordered according to d_i and location ℓ_i with the shortest distance between $\bar{r}_i$ and s in the n-dimensional RSS space is returned as the location estimate. Note that the proposed self-calibration approach is independent of the underlying positioning algorithm and using a more sophisticated approach, such as the wide kernel probabilistic method of [13], is expected to further improve accuracy.

B. Measurement Setup

We performed our measurement campaign for collecting RSS data at KIOS Research Center. This is a $560\,\mathrm{m}^2$ typical office environment that consists of several open cubicle-style and private offices, labs, a conference room and corridors (Fig. 1). We have installed 9 APs that use the IEEE 802.11b/g standard and provide full coverage throughout the floor. We used 5 different mobile devices for our data collection, namely a HP iPAQ hw6915 PDA with Windows Mobile, an Asus eeePC T101MT laptop running Windows 9, an HTC Flyer Android tablet and two other Android smartphones (HTC Desire, Samsung Nexus S).

For our training data we recorded fingerprints, which contain RSS measurements from all 9 APs, at 105 distinct reference locations by carrying all 5 devices at the same time. A total of 2100 training fingerprints, corresponding to 20 fingerprints per reference location, were collected at the rate of 1 sample/sec with each device. These data are used to build device-specific radiomaps by calculating the mean value RSS fingerprint that corresponds to each reference location. We point out that the device-specific radiomaps are needed *only* for comparison purposes.

We collected additional test data two weeks later by walking along a predefined route. The route has two segments and consists of 96 locations most of which do not coincide with the reference locations; see Fig. 1. One fingerprint was recorded at each test location. We followed the same route 10 times and collected our test data simultaneously with all devices.

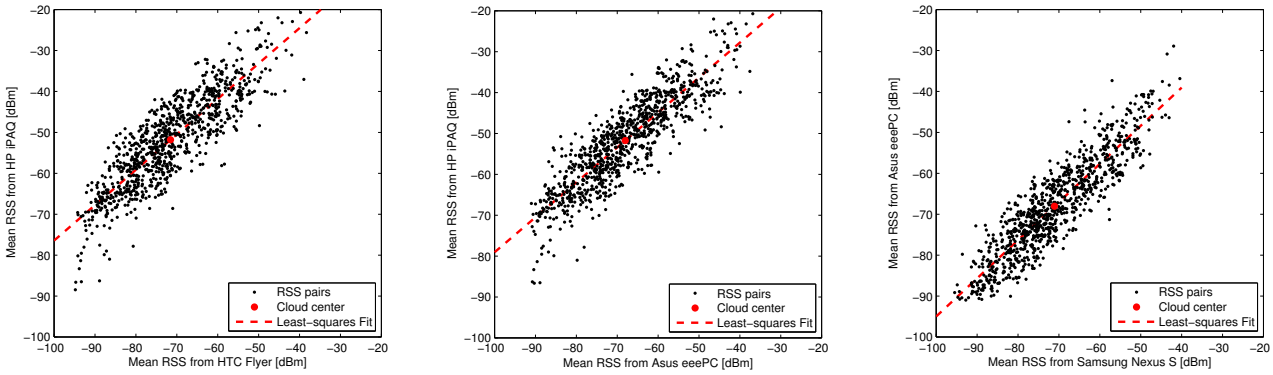
IV. Manual Device Calibration

Several experimental studies have reported a linear relation between the RSS values reported by heterogeneous devices [10], [12], [13]. Thus, if a sufficient number of colocated RSS pairs (i.e., collected at the same location and time with two different devices) is available, then the linear parameters can be estimated through standard least squares fitting. To put it formally, for two devices D_1 and D_2 we use the RSS data in the respective radiomaps to compute the parameters by:

$$\bar{r}_{ij}^{(2)} = \alpha_{12}\bar{r}_{ij}^{(1)} + \beta_{12} \tag{3}$$

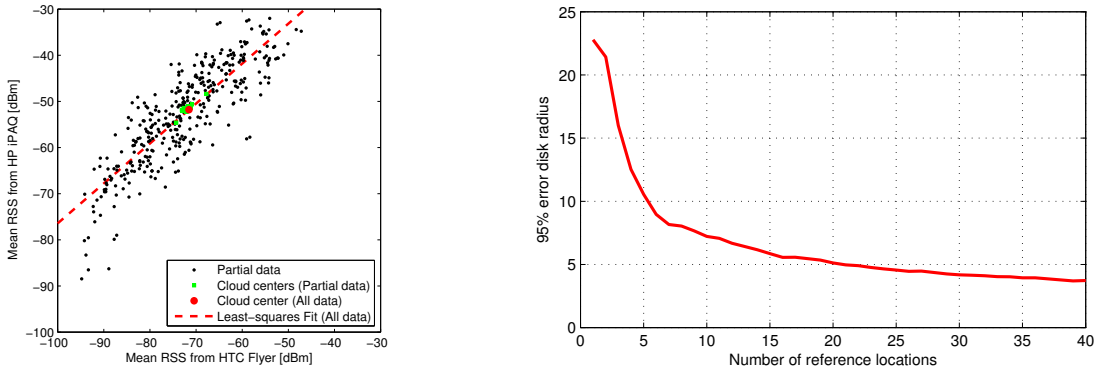
where $\bar{r}_{ij}^{(1)}$ and $\bar{r}_{ij}^{(2)}$ denote the mean RSS value at location ℓ_i from the j-th AP for D_1 and D_2 respectively, while α_{12} and β_{12} are the linear parameters for mapping the RSS values from D_1 to D_2. Some indicative correlation plots for various pairs of devices are shown in Fig. 2. These plots confirm the linear relation between the RSS values reported by heterogeneous devices and justify the effectiveness of first order polynomials for device calibration.

An important question is how much data should be collected with a new device $D_i,\ i \neq 0$ to achieve a good mapping to the reference device D_0 and guarantee acceptable accuracy when this new device is used for positioning. For instance, if a considerable volume of RSS data that spans the whole area of interest needs to be collected with another device, then this new dataset may as well be used as a second radiomap that will be employed whenever this specific device is carried by a user. Thus, manual calibration is only justified if a small number of RSS fingerprints collected at a few calibration locations suffice to obtain an adequate mapping. Authors in [10] report that in



(a) HP iPAQ – HTC Flyer pair. (b) HP iPAQ – Asus eeePC pair. (c) Asus eeePC – Samsung Nexus S pair.

Fig. 2. Correlation plots using colocated pairs of mean RSS values collected at all 105 *known* locations with different devices.



(a) Centers of the RSS cloud. (b) Convergence of the 95% error disk radius.

Fig. 3. Performance of the manual device calibration using a small amount of RSS data.

their setup, where around 15 APs are detected at each location, manual calibration is very effective when RSS fingerprints collected at three to five locations are used for fitting. Our experimental results are in agreement and in the following we provide a justification to support this observation.

Suppose that the colocated RSS pairs from two different devices $y = [\bar{r}_{ij}^{(1)}, \bar{r}_{ij}^{(2)}]^T$ follow a normal distribution, i.e. $y[k]|x \overset{\text{iid}}{\sim} \mathcal{N}(x, R)$, where x is a bivariate random vector and R is the covariance matrix. Roughly speaking, x and R represent the center and the shape of the cloud of RSS pairs in this 2-D space, respectively; see Fig. 2. In the linear fitting (3), the parameter β_{12} is directly related to the cloud center x, while the parameter α_{12} is related to the principal axis of the cloud shape R. For brevity, we assume that R is known, i.e. α_{12} value is fixed, and in the following we study the behavior of the cloud center x for increasing number of RSS pairs that provides insight into the convergence of the β_{12} parameter[1].

Assuming prior distribution $x \sim \mathcal{N}(m[0], P[0])$, the posterior distribution given a series of colocated RSS pairs is $x|y[1{:}k] \sim \mathcal{N}(m[k], P[k])$ with

$$P[k] = P[k-1] - P[k-1](P[k-1]+R)^{-1}P[k-1] \quad (4)$$

$$m[k] = m[k-1] + P[k-1](P[k-1]+R)^{-1}(y[k]-m[k-1]) \quad (5)$$

These equations [17] are used for recursively updating $m[k]$ and $P[k]$ and improve the estimate of x by sequentially processing the RSS observations from pairs of devices. By Chebyshev's inequality, the disk with center $m[k]$ and radius $r = \sqrt{\text{trace}(P[k])/0.05}$ contains x with probability at least 95% [17]. It can be shown that $\text{trace}(P[k]) \leq \text{trace}(R)/k$ and consequently the 95% disk's radius is proportional to $1/\sqrt{k}$. As a rule of thumb, to get for example 10 times better accuracy, 100 times more data are required. We may use the sample covariance estimator to calculate R from a series of colocated RSS pairs $y[1{:}k]$ and study the 95% error disk's radius by plotting $\sqrt{\text{trace}(R)/(0.05k)}$. The center of the RSS cloud (shown in green) for increasing number of locations that contribute their mean RSS pairs to the fitting is illustrated in Fig. 3a. Note that these locations are uniformly distributed and each location contributes 9 RSS pairs. It is observed that the center x, when data from few locations are used, converges quickly to the center obtained when data from all 105 locations (i.e., around 945 RSS pairs) are considered. Moreover, the 95%

[1]The simultaneous estimation of x and R is also possible by treating them both as unknown and assuming a flat prior for x and a Wishart prior for R and then using recursive formulas for the posterior means of x and R given $y[1{:}k]$, as discussed in [16].

TABLE I
POSITIONING ERROR [M] USING VARIABLE NUMBER OF LOCATIONS.

	Uncalibrated	All	20	5
Mean	7.6	2.7	2.7±0.0	2.9±0.2
Median	7.3	2.3	2.3±0.0	2.4±0.1
67%cdf	9.4	3.0	3.0±0.0	3.1±0.3
95%cdf	15.7	6.2	6.3±0.3	7.0±0.8
Max	19.1	16.2	15.0±1.3	14.6±1.1

error disk's radius decreases as the number of locations grows from 1 to 40, indicating that colocated RSS pairs from 15 to 20 locations seem to suffice; see Fig. 3b.

The above analysis justifies the effectiveness of the manual calibration using a small amount of data and in practice we observed that around 5 known locations uniformly selected inside the area of interest can provide good device calibration. The statistics of the positioning error pertaining to the whole test set are listed in Table I using the HP iPAQ and the HTC Flyer as reference and new device, respectively. These results indicate that the positioning accuracy can be significantly improved when manual calibration is applied. For instance, the mean error decreases to 2.7 m when we use colocated RSS pairs collected with the HTC Flyer at all 105 reference locations, compared to 7.6 m when no device calibration is applied. Interestingly, we observe that using 20 or only 5 locations (i.e., 180 or 45 colocated RSS pairs) for manual calibration has marginal effect on the positioning error. Note that in the cases where few reference locations are considered for manual calibration, the results pertain to 10 experiments assuming different subsets of randomly selected locations.

These results confirm that manual device calibration with colocated RSS pairs collected at *known* locations is a very effective approach. More importantly, when the area of interest is covered by several APs then only a few locations need to be visited with an uncalibrated device, thus reducing the time and labour overhead for calibrating a new device. However, this approach has limited applicability in real-life applications where a user enters an indoor environment, such as shopping malls, airports, etc., carrying an uncalibrated device because he or she has to be guided to specific *known* locations for collecting RSS data. This implies that the user is already familiar with the area of interest which is usually not the case and a considerable data collection effort is still required by the user prior to positioning.

V. DEVICE CALIBRATION

Our objective is to develop a fully automatic approach that does not require any user intervention and we do so by exploiting histograms of RSS values collected with the reference and user device.

A. RSS Histograms

The RSS histograms of three different devices are shown in Fig. 4. These histograms correspond to the mean RSS values of all 9 APs collected at all 105 *known* reference locations. The first observation is that two histograms may differ significantly with respect to the range of RSS values, as well as the probability of each RSS value, e.g. the HP iPAQ and HTC Flyer device pair. On the other hand, the respective histograms for some device pairs can be quite similar, as in the case of Asus eeePC and HTC Flyer. This is also evident from the equivalent empirical cumulative distribution function (cdf) of each device, as shown in Fig. 4d. We have observed that the empirical cdf of the raw RSS values, recorded while walking around with a particular device for a few seconds, resembles the respective empirical cdf of the mean RSS values collected with the same device at several uniformly distributed *known* locations. This implies that we may exploit these empirical cdfs to perform device calibration during positioning.

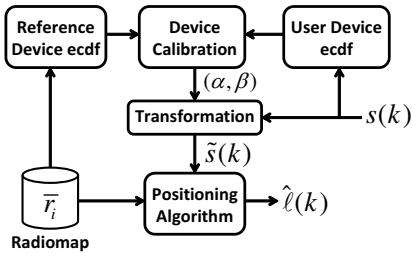


Fig. 5. Block diagram of the proposed self-calibration method.

B. Self-calibration method

The main idea in the proposed self-calibration method is the use of RSS histograms for obtaining a mapping between the reference and various user devices. The block diagram of our method is shown in Fig. 5. First, we use the existing radiomap that contains the mean value fingerprints $\bar{r}_i$ to obtain the RSS histogram of the reference device. Subsequently, when the user enters a building and starts positioning, the RSS values in the currently observed fingerprint $s(k)$ are recorded simultaneously in the background in order to create and update the histogram of raw RSS values for the user device. Then, we use the RSS values that correspond to specific percentiles of the empirical cdf to fit a linear mapping of the form (3) between the user and reference devices. Subsequently, the parameters (α, β) are used to transform the RSS values observed with the user device and obtain the new fingerprint $\tilde{s}(k)$, which is compatible with the radiomap, and then estimate the unknown location $\hat{\ell}(k)$ with any fingerprint-based algorithm.

Let $F_r(x)$ and $F_u(x)$ denote the empirical cdfs of the reference and user devices, respectively. In general the cdf $F(x)$ gives the probability of observing an RSS value that is less than x, while the inverse cdf $F^{-1}(y)$ returns the RSS value that corresponds to the y-th cdf percentile. We use the RSS values that correspond to the 10-th, 20-th, ..., 90-th percentiles of the empirical cdf to fit a least squares linear mapping between the user and reference devices and estimate the parameters (α, β) according to

$$F_r^{-1}(y) = \alpha F_u^{-1}(y) + \beta, \; y \in \{0.1, 0.2, \ldots, 0.9\}. \quad (6)$$

A formal proof on the validity of the least squares mapping (6) that uses the inverse cdf percentile values to reveal the

underlying functional relationship between the RSS values collected with diverse devices is provided in Appendix B. Subsequently, the observed fingerprint $s(k)$ is transformed using $\tilde{s}_j(k) = \alpha s_j(k) + \beta,\ j = 1, \ldots, n$.

While the user is walking the current fingerprint $s(k)$ contributes only a few RSS values and $F_u(x)$ does not change significantly between two consecutive samples. Thus, it is not necessary to update $F_u(x)$ every time a new fingerprint $s(k)$ is available, but rather one can buffer the RSS values contained in a number of successive fingerprints and then update $F_u(x)$ before performing the linear fitting. We have experimentally selected the buffer size $b = 10$ that works well in our setup, i.e. (α, β) are recalculated every 10 seconds. At the beginning, we initialize the parameters to $(\alpha, \beta) = (1, 0)$, i.e. no transformation is performed, to handle the positioning requests until the buffer is full and the parameters are estimated for the first time. Using a lower value for b does not seem to improve the positioning accuracy, while it introduces unnecessary computational overhead. On the other hand, increasing b means that the parameters are not updated frequently enough and the performance is degraded, especially at the beginning until (α, β) are estimated for the first time.

The proposed method is simple, yet very effective and does not require any user intervention, unlike the existing RSS data fitting approaches. Moreover, our self-calibration method is running concurrently with positioning, while the user walks freely inside the area of interest, and there is no need to dwell at various locations for collecting several RSS samples.

VI. Experimental Results

We evaluate the proposed self-calibration method using experimental data collected with five devices in a real office environment, as described in Section III-B. We compare its performance for all device pairs against the manual calibration approach [10], [12] that uses the mean RSS values collected with the new device at all 105 locations visited with the reference device. We also report the positioning accuracy in the two extreme cases of no calibration and using a device-specific radiomap collected with each device that provide the upper and lower bound on the performance, respectively.

We investigate the mean positioning error $\bar{\epsilon}$ per route, which is averaged over the 96 distinct test locations. The statistics for $\bar{\epsilon}$ pertaining to all 10 routes are depicted as boxplots in Fig. 6a and Fig. 6b for the iPAQ – Flyer and the eeePC – Flyer pairs, respectively. The first observation is that $\bar{\epsilon}$ does not vary significantly among different test routes. For the iPAQ – Flyer pair, positioning without calibrating the Flyer device leads to significant performance degradation; the median of the mean error $\bar{\epsilon}$ is 7.5 m compared to 1.9 m in case we use a radiomap collected with the Flyer, rather than the iPAQ, device. Our self-calibration method is very effective and achieves performance similar to the manual calibration approach (i.e., the median of $\bar{\epsilon}$ is 3 m compared to 2.6 m as shown in Fig. 6a), but with considerably less effort. On the other hand, the Asus eeePC seems to be more appropriate for positioning the Flyer device, as the median of $\bar{\epsilon}$ is increased by 1 m when no calibration is performed compared to the case of using a device-specific radiomap. Yet, our method is beneficial (the median of $\bar{\epsilon}$ is 2.3 m) and interestingly it performs slightly better compared to manual calibration; see Fig. 6b.

We demonstrate the efficiency of our method on a single route using the iPAQ radiomap, while the user carries the Flyer device. The performance of our method is illustrated in Fig. 6c, where we observe that in the first 10 seconds the accuracy is not adequate, because the device is still uncalibrated. While the user is walking the raw RSS values are collected in order to build the RSS histogram that will be used for the calibration. It is obvious that beyond that point, the user device has been automatically calibrated and the positioning system delivers accuracy as good as in the case of using a radiomap that is created from data collected with the Flyer device.

The results for five devices are summarized in Table II. We examine all device pairs and each row indicates the device used as reference, while each column indicates the test device. For every device pair the median of the mean positioning error $\bar{\epsilon}$ using our device self-calibration method is reported and the corresponding positioning error when no calibration is used is shown in parentheses. The entries in the diagonal cells represent the best case scenario where the test and reference devices are the same, i.e. a device-specific radiomap is used for positioning. Looking at these results it is evident that the proposed calibration method improves the accuracy achieved by the NN positioning algorithm for all device pairs. The improvement is bigger when the HP iPAQ or the Asus eeePC are involved as either reference or test devices. For the pairs involving only Android devices the improvement is marginal, which is probably due to the fact that they report the RSS values in a very similar way, thus device calibration cannot offer significant improvement. We highlight that in all cases, assuming a specific test device (e.g., the Asus eeePC in the second column), our calibration method reduces the positioning error compared to the no calibration case, while the accuracy is very close to the accuracy achieved using device-specific radiomaps.

As a last remark, in our self-calibration method we observed that the successive α values, estimated while the user is walking, are always around 1. This implies that we can actually fit a unit slope linear mapping, i.e. fix $\alpha = 1$ and only estimate the parameter β using the following median estimator

$$\beta = \mathrm{med}(F_r^{-1}(y) - F_u^{-1}(y)),\ y \in \{0.1, 0.2, \ldots, 0.9\}. \quad (7)$$

The advantage is that only one parameter, instead of two parameters, needs to be estimated and the computational overhead of the least-squares fitting in (6) is reduced. Our experimental results with various devices indicate that the positioning accuracy when (7) is employed in our self-calibration method is very close to the accuracy reported in Table II.

VII. Conclusions

Addressing the device diversity is important for the wide acceptance of RSS-based indoor positioning systems. In this paper, we presented a low-complexity, yet effective method

TABLE II
MEDIAN OF THE MEAN ERROR $\bar{\epsilon}$ [M], WITH AND WITHOUT CALIBRATION.

	iPAQ	**eeePC**	**Flyer**	**Desire**	**Nexus S**
iPAQ	2.7	2.8 (6.6)	3.0 (7.5)	2.9 (8.4)	2.6 (7.7)
eeePC	2.8 (4.4)	2.3	2.3 (2.8)	2.6 (3.5)	2.5 (2.9)
Flyer	3.2 (5.9)	2.6 (3.0)	1.9	2.1 (2.3)	2.6 (2.7)
Desire	3.4 (6.1)	2.8 (3.2)	2.5 (2.5)	2.4	2.5 (2.6)
Nexus S	3.0 (6.2)	2.6 (2.8)	2.7 (2.7)	2.4 (2.5)	2.3

that allows any mobile device to be self-calibrated, shortly after the user has started positioning. We use the existing RSS radiomap and the RSS values observed while the user is moving normally in order to create the RSS histograms of the reference and new device, respectively. Self-calibration is then achieved by a fitting a linear mapping between the histograms of these heterogeneous devices, so that the observed RSS values are compatible with the available radiomap. Experimental results with several mobile devices in a real office environment demonstrate the effectiveness of the proposed approach.

As future work, we plan to investigate the performance of our calibration method in larger scale setups featuring non uniform AP infrastructure layouts. In such cases, weak RSS values from APs located far from the user may prevail, thus increasing the skewness of the RSS histograms.

ACKNOWLEDGEMENT

The work of the first and third authors is partially supported by the Cyprus Research Promotion Foundation and is co-funded by the Republic of Cyprus and the European Regional Development Fund.

APPENDIX A

We show that the RSS difference values do not contain the device-dependent intercept term A, however they suffer from increased noise variance. Assume that a mobile device resides at a location ℓ, which is covered by 2 WLAN APs, namely AP_1 and AP_2. The RSS values recorded by the device are given by

$$RSS_1 = A - 10\gamma \log_{10} d_1 + X_1 \quad (8)$$

$$RSS_2 = A - 10\gamma \log_{10} d_2 + X_2 \quad (9)$$

where $d_i,\ i = 1, 2$ is the distance from the i-th AP, while $X_1, X_2 \sim \mathcal{N}(0, \sigma_n^2)$ are independent Gaussian noise components disturbing the RSS values. Taking the difference of these RSS values, denoted as $RSSD_{12}$, gives

$$RSSD_{12} = RSS_1 - RSS_2 = 10\gamma \log_{10} \frac{d_2}{d_1} + X' \quad (10)$$

where $X' \sim \mathcal{N}(0, 2\sigma_n^2)$ is the linear combination of X_1, X_2.

APPENDIX B

If $\mathbf{u}$ is a continuous random variable and $\mathbf{y} = f(\mathbf{u})$ with monotonically increasing f then $f = F_{\mathbf{y}}^{-1} \circ F_{\mathbf{u}}$. In particular, the inverse cdf ordered pairs

$$\{(u_i, y_i) = (F_{\mathbf{u}}^{-1}(q_i), F_{\mathbf{y}}^{-1}(q_i)) : q_i \in \{0.1, \ldots, 0.9\}\}$$

lie on the curve $y = f(u)$.

Proof: We have

$$\begin{aligned} F_{\mathbf{u}}(u) &= P(\mathbf{u} \le u) = P(f(\mathbf{u}) \le f(u)) = \\ &= P(\mathbf{y} \le f(u)) = F_{\mathbf{y}}(f(u)). \end{aligned}$$

Applying $F_{\mathbf{y}}^{-1}$ to both sides gives the identity $f = F_{\mathbf{y}}^{-1} \circ F_{\mathbf{u}}$. Also, the components of the inverse cdf ordered pairs satisfy

$$y_i = F_{\mathbf{y}}^{-1}(q_i) = F_{\mathbf{y}}^{-1}(F_{\mathbf{u}}(u_i)) = f(u_i).$$

REFERENCES

[1] P. Bahl and V. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *IEEE International Conference on Computer Communications INFOCOM*, vol. 2, 2000, pp. 775–784.

[2] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, Jul. 2002.

[3] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *3rd ACM International Conference on Mobile systems, applications, and services*, 2005, pp. 205–218.

[4] K. Kaemarungsi, "Distribution of WLAN received signal strength indication for indoor location determination," in *1st International Symposium on Wireless Pervasive Computing*, 2006, p. 6.

[5] A. W. Tsui, Y.-H. Chuang, and H.-H. Chu, "Unsupervised learning for solving RSS hardware variance problem in WiFi localization," *Mobile Networks and Applications*, vol. 14, no. 5, pp. 677–691, 2009.

[6] Y.-C. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm, "Accuracy characterization for metropolitan-scale Wi-Fi localization," in *3rd ACM International conference on Mobile systems, applications, and services (MobiSys)*, 2005, pp. 233–245.

[7] A. Hossain and W.-S. Soh, "Cramer-Rao bound analysis of localization using signal strength difference as location fingerprint," in *IEEE INFOCOM*, 2010, pp. 1–9.

[8] M. B. Kjærgaard, "Indoor location fingerprinting with heterogeneous clients," *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 31–43, 2011.

[9] J. Machaj, P. Brida, and R. Piché, "Rank based fingerprinting algorithm for indoor positioning," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2011, pp. 1–6.

[10] 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 *10th international conference on Mobile computing and networking (MobiCom)*, 2004, pp. 70–84.

[11] P. Misikangas and L. Lekman, "Applications of signal quality observations," WO Patent WO/2004/008,796, 2005.

[12] M. Kjaergaard, "Automatic mitigation of sensor variations for signal strength based location systems," pp. 30–47, 2006.

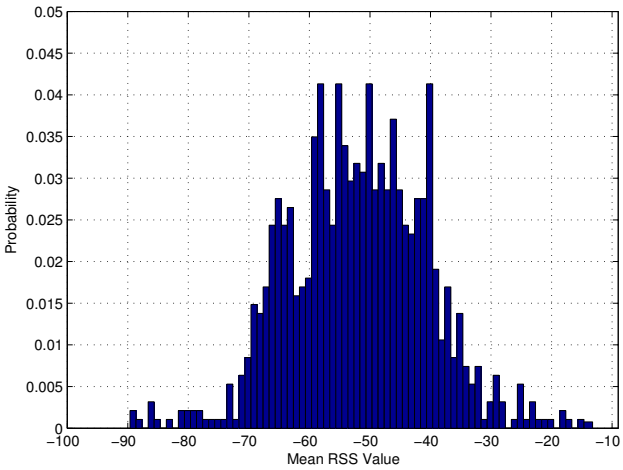
[13] J. Park, D. Curtis, S. Teller, J. Ledlie *et al.*, "Implications of device diversity for organic localization," in *IEEE International Conference on Computer Communications INFOCOM*, 2011, pp. 3182–3190.

[14] L. Koski, T. Perälä, and R. Piché, "Indoor positioning using WLAN coverage area estimates," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010, pp. 1–7.

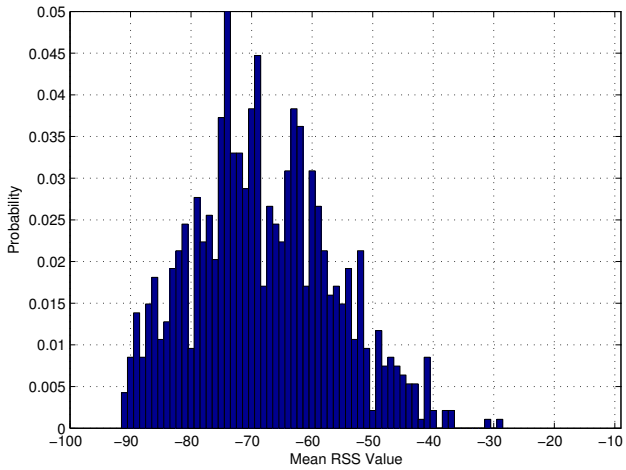
[15] V. Honkavirta, T. Perälä, S. Ali-Löytty, and R. Piché, "A comparative survey of WLAN location fingerprinting methods," in *6th Workshop on Positioning, Navigation and Communication (WPNC)*, 2009, pp. 243–251.

[16] C.-F. Chen, "Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 41, no. 2, pp. 235–248, 1979.

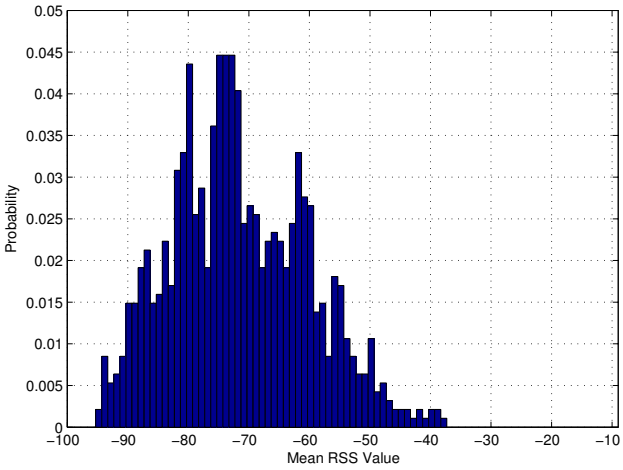
[17] R. Piché, "Stochastic processes," 2010. [Online]. Available: http://urn.fi/URN:NBN:fi:tty-201012021377



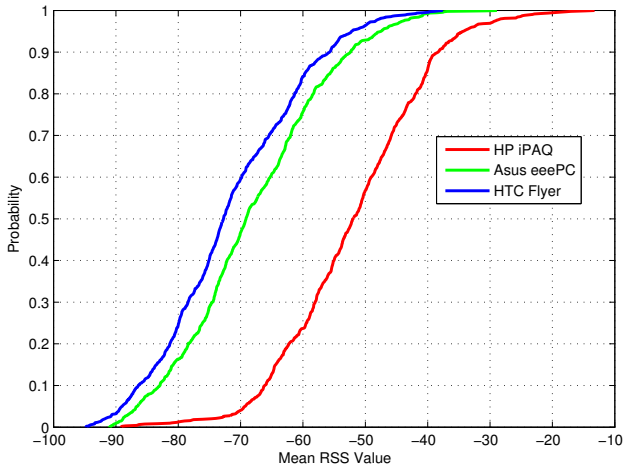
(a) HP iPAQ.



(b) Asus eeePC.

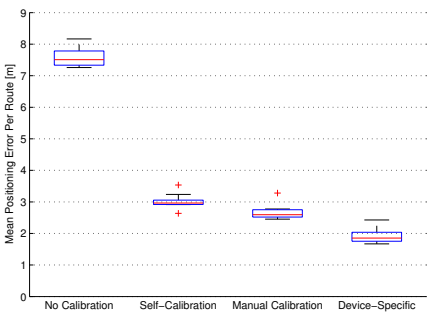


(c) HTC Flyer.

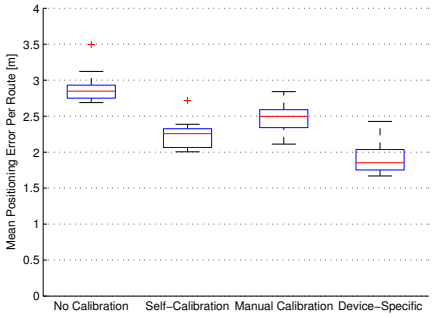


(d) Empirical cumulative distribution function.

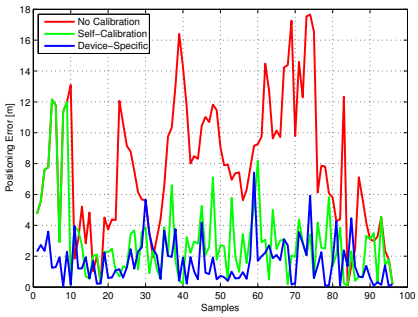
Fig. 4. Histograms of the mean RSS values from available APs at all 105 locations.



(a) HP iPAQ radiomap – HTC Flyer user.



(b) Asus eeePC radiomap – HTC Flyer user.



(c) Positioning error in a single route.

Fig. 6. Performance of the signal strength self-calibration method.