FreeLoc: Calibration-Free Crowdsourced Indoor Localization

Sungwon Yang, Pralav Dessai, Mansi Verma, and Mario Gerla
Department of Computer Science
University of California, Los Angeles
{swyang, pdessai, mansi, gerla}@cs.ucla.edu

***Abstract*—Many indoor localization techniques that rely on RF signals from wireless Access Points have been proposed in the last decade. In recent years, research on crowdsourced (also known as "Organic") Wi-Fi fingerprint positioning systems has been attracting much attention. This participatory approach introduces new challenges that no previously proposed techniques have taken into account. This paper proposes "FreeLoc", an efficient localization method addressing three major technical issues posed in crowdsourcing based systems. Our novel solution facilitates 1) extracting accurate fingerprint values from short RSS measurement times 2) calibration-free positioning across different devices and 3) maintaining a single fingerprint for each location in a radio map, irrespective of any number of uploaded data sets for a given location. Through experiments using four different smartphones, we evaluate our new indoor positioning method. The experimental results confirm that the proposed scheme provides consistent localization accuracy in an environment where the device heterogeneity and the multiple surveyor problems exist.**

I. Introduction

An explosive growth of mobile devices, such as smartphones and tablet computers, is accelerating a demand for more accurate location information. This is imperative for successful realization of mobile Location Based Service (LBS) applications [1]–[3]. Unlike outdoor regions where location of a mobile node is determined accurately by Global Positioning System (GPS), indoor localization still continues to a challenge. A number of techniques [4]–[7] have been proposed in the last decade for robust indoor positioning.

Early work on indoor positioning typically explored specially-designed beaconing hardware installed on walls or ceilings [8]–[10]. Systems using infrared or ultrasound promise fine-grained localization accuracy, however, it is difficult to be deployed on a large scale due to the high deployment cost. As IEEE 802.11 (Wi-Fi) is becoming ubiquitous, research on leveraging Wi-Fi Access Points (APs) is receiving much attention. An approach that estimates the location based on the comparison between observed Received Signal Strength (RSS) values and the "Fingerprint" data in a pre-built radio map which has been obtained through the "Training Phase", has the advantage of avoiding the cost of specialized infrastructure deployment.

The training phase can be classified into two categories depending on the main agent who collects the fingerprint data. The basic "Expert Surveyor" model is first studied [11], [12]. This model provides a robust and precise radio map, but it is costly because all the required jobs, ranging from the initial map building tasks to sporadic maintenance tasks, must be done by trained experts. Due to the cost increasing with the size of the surveying site, researchers recently have started developing systems that normal users can participate in during the training phase [13]–[18]. This crowdsourcing based model can significantly reduce the map-building and maintenance cost, but it also introduces a new set of challenges.

This paper particularly investigates three major technical issues posed in a crowdsourced indoor localization system. First issue is that there are no dedicated surveyors. Unlike expert surveyors who can afford long-enough measurement time for building the fingerprint database, surveyors in a crowdsourced system are all volunteers who should not be forced to sacrifice their time and their device's resources. The radio map should provide robust and accurate fingerprint data even though it is built based on short-duration RSS measurements. Secondly, there is no constraint on type and number of devices. Users carry heterogeneous devices, resulting in a radio map built with RSS values from diverse devices. Since devices may have different chipsets and antenna designs, the RSS measurement data differs across different devices even when they are placed at exactly the same positions. Therefore, a technique that tolerates device diversity is necessary. Finally, it may also be a problem that there are no designated fingerprint collection points. Since the radio map is updated by untrained voluntary users without centralized controls, different users can upload their own fingerprint data that is collected at slightly different locations but has the same location label. Multiple fingerprint data indicating one particular location not only causes slow location estimation but also storage space wastage in a radio map server.

This paper makes the following contributions taking the three issues into account:

- We analyze the characteristics of RSS that is measured for less than one minute. Based on the observations, we present a method that extracts a reliable single fingerprint value per AP from the short-duration RSS measurements.
- We propose "FreeLoc", a novel indoor localization method that requires no calibration among heterogeneous devices.
- We also show that the proposed method resolves the multiple surveyor problem without any calibrations.
- We evaluate the performance of our localization technique in various scenarios through real-world experiments using multiple mobile phones.

978-1-4673-5946-7/13/$31.00 ©2013 IEEE

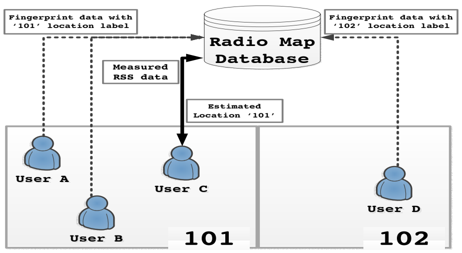


Fig. 1. System Overview

II. System Overview and Challenges

In this section, we briefly describe a high-level usage scenario of our target system that consists of two main entities: The "Contributor" or "Surveyor" records and upload the fingerprint data at a particular location along with the location label. The "User" queries the radio map server for inquiring about its current location. In traditional systems based on Wi-Fi signal fingerprints, contributors and users are typically distinct. In other words, the role of contributors is limited to only a survey. The surveyors, usually trained personnel, collect AP information at each location using specially-designed homogeneous scanners in order to minimize the observational errors. This one-surveyor-multiple-user model is convenient to build an initial radio map in a small area since it requires no calibration for each scanned fingerprint. However, it is not suitable for a large area and it is not easy to update the map frequently.

Instead of the one-surveyor-multiple-user model, we adopt a multiple-surveyor-multiple-user model for our localization system in which anyone can be a contributor, a users, or both. This participatory model has the advantages of fast radio map building and its quick update. Figure 1 depicts the overview of our system. The four users (i.e., *A, B, C, and D*) are untrained normal people who are carrying different mobile phones. In this example scenario, user *A* and *B* play the roles of surveyors for *Location 101*, and upload the measured AP information. The uploaded data is processed on a remote server and converted to fingerprint data for the *Location 101*. When user *C* inquires the current location information, the server returns *Location 101* if user *C*'s RSS measurement matches one from the user *A* and *B*. This crowdsourcing approach has been adopted in many recent researches [13]–[18]. However, none of them addresses the three associated technical issues posed in the following subsections.

The "Training Phase" in which the surveying-users collect RSS data and build the fingerprint map database is as important as the "Online Localization Phase" because the location estimation is solely based on the radio map information. However, most previous studies seem to overlook it. They rather focus on the online phase, and aim to reduce the localization error in a statistical or probabilistic manner. However, there is a limit to improve accuracy by calibrating only one side. It is obvious that the localization accuracy can improve when the fingerprint data of measured RSS values resemble the trained radio map. The distinct contribution of this paper is to propose an integrated solution covering issues that cause inaccurate radio map construction.

A. RSS Measurement for Short Duration

It is well known that multi-path fading in an indoor environment causes RSS to fluctuate over time even if the receiver is absolutely fixed [19]. One simple method to reduce RSS variance is to record the RSS data for a long time. As the number of RSS samples increases it is easier to identify one single fingerprint value which ensures construction of a more robust and accurate radio map. However, it is almost impossible to have normal users collect RSS data for a long time in a crowdsourced system. We believe that the RSS survey time at each location should not exceed one minute assuming all the surveyors are normal mobile phone users. Therefore, a technique that extracts accurate fingerprint values from short-time measurement is necessary.

B. Device Diversity

In the participatory system we propose, diverse devices inevitably get involved in building the radio map database. Since there are no expert surveyors who use specially-designed hardware, it is highly likely that the RSS data gathered from each user varies even though it is collected at the exact same position. Different Wi-Fi chipsets and antenna designs among devices cause varied RSS recordings per device and make it difficult to calibrate them. This device heterogeneity is the key problem for crowdsourced localization. This paper addresses the problem and proposes an efficient solution.

C. Multiple Measurements for One Location

Another problem of the crowdsourced system is that more than one surveyor can upload one's own fingerprint data which has the same location label, but is obtained at different measurement points. This happens because most of the surveyors are untrained normal users and there are no designated RSS recording points. In Figure 1, both user *A* and *B* provide fingerprint data for *Location 101*. However, the position at which they record the RSS data is different even though both of them are in the same room. This results in multiple fingerprint data sets for one location. Having more than one fingerprint for a particular location is not efficient for the radio map building phase and the location estimation phase. More storage space would be required and more time would be taken to estimate the current location. This, combined with the device diversity problem, is also a principal cause of localization accuracy degradation. However, none of the previous studies take this into account. Our novel method proposed in this paper solves this multiple measurement points issue without any calibrations.

III. Fingerprint Value Extraction

We conducted preliminary experiments using different mobile phones in order to investigate the characteristics of Wi-Fi signal reception from arbitrary APs. This empirical data is used to develop our practical and resilient methods for both radio map building and localization phases. This section presents how we extract each fingerprint value at a particular location from RSS observations and the experimental results with their implications to localization accuracy.

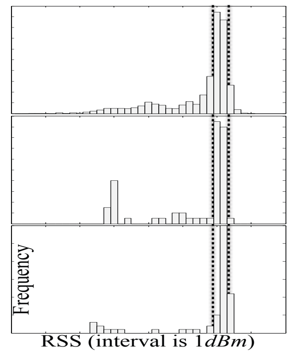
A. AP Response Rate

While we analyzed results from the preliminary experiments, we found that some APs were not recorded in some fractions of the entire Wi-Fi scanning duration. The response rate is different for different APs, which implies that it can be used as independent information for the fingerprint data. Ledlie *et. al.* used this response rate as fingerprint information in combination with RSS in [18] because they found the the correlation between the rate and RSS is rather weak. However, we believe that the correlation is strong enough to conclude that the response rate is redundant information in our method. Our preliminary results show that APs with RSS values of greater than -70*dBm* provide over 90% of response rate and APs with RSS between -70*dBm* and -85*dBm* provide over 50% of response rate. Only APs with RSS of less than -90*dBm* present very poor response rate. Since our localization method naturally gives lower weights to weak RSS values, we discount the AP response rate for fingerprint information.

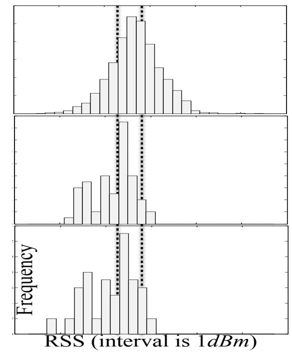
B. RSS Variance over Time

Received signal strength varies over time due to multipath fading and the level of fluctuation gets higher in an indoor environment. This instability degrades the localization accuracy. To overcome this problem, previous work measures RSS for long duration (i.e., more than one hour to a month) and uses the average values [20]. However, this is not appropriate in our system model. Some studies that have adopted the crowdsourced model use probabilistic methods to obtain fingerprint values from a small amount of measurement data [15], [18], [21]. The probabilistic method based on the Gaussian distribution works well in an ideal environment. However, it frequently gives quite a big inaccuracy because of the inherent instability of real world.

Figure 2 illustrates two examples where the probabilistic method provides inaccurate fingerprint values. The top figures are histograms of RSS for one-hour measurements. The middle and the bottom figures are for one-minute measurements. For the one-minute data, each measurement was conducted at intervals of 30 minutes. RSS was recorded at the fastest interval for all the cases in order to get as many RSS values as possible in a limited measurement time (0.5- 1*Hz* depending on the devices). Figure 2(a) shows that the histograms are strongly left-skewed which makes it difficult to model and fit to normal distributions. The left-skewed distribution of RSS is a common phenomenon observed in real life where people move around [22]. Due to the complexity of the radio propagation, Ladd *et. al.* [23] suggested to use only the mean value. Using mean values for the fingerprint works well in the case of Figure 2(a). The mean is -47.5*dBm* for one-hour measurements and -48.5*dBm* and -46.7*dBm* for each one-minute measurement. The variation between long and short measurement times is within 1*dBm* and this may be acceptable. However our experiments show that the mean value is not the best method for representing a fingerprint value. Figure 2(b) presents a counter example showing that the mean value does not work well. The mean values are -76.6, -79.2, and -79.5*dBm*, respectively for the long and short measurement times. The RSS variation is almost 3*dBm* which possibly degrades the localization accuracy.



(a) Top: μ: -47.5 σ: 5.2
Middle: μ: -48.5 σ: 6.3
Bottom: μ: -46.7 σ: 6.2



(b) Top: μ: -76.6 σ: 2.7
Middle: μ: -79.2 σ: 2.3
Bottom: μ: -79.5 σ: 2.7

Fig. 2. Examples of RSS Variance

C. Extraction Method

Neither the normal distribution model nor the mean value provide accurate RSS fingerprint information in crowdsourcing based localization systems. We tried to figure out the best indicator of RSS fingerprint and we found one unique characteristic while we analyzed the recorded RSS data sets. We observed that the most-recorded RSS in the case of the short-duration measurements is very close to the most-recorded RSS in the long-duration measurement case. In Figure 2, for example, the highest frequencies of RSS are found within the small range marked with dotted-lines for all the cases. This implies that the exact same RSS fingerprint value can be obtained regardless of the duration of measurement time if the RSS values between the range are averaged.

Based on our observation, we came up with a simple yet effective method that extracts a single fingerprint value for a particular AP with high tolerance to the RSS variation over short-duration measurements. Equation (1) depicts our method.

$$fpValue = \frac{\sum_{n=1}^{w_{LT}}(RSS_{peak}-n)+RSS_{peak}+\sum_{m=1}^{w_{RT}}(RSS_{peak}+m)}{w_{LT}+w_{RT}+1} \tag{1}$$

where $fpValue$ is the fingerprint value for an AP and RSS_{peak} is the RSS value of highest frequency during the measurement. The width of the range being averaged is set by w_{LT} and w_{RT}. The RSS distribution tends to be left-skewed as discovered in [22], [23]. This implies that more meaningful RSS values for the fingerprint are located at the right side of the distribution (i.e., stronger RSS value is more important). Therefore, we select a stronger RSS value as the $fpValue$ if more than one RSS value has the same frequency in a histogram.

The proposed method extracts fingerprint RSS values by ignoring RSS records that are far away from the most-frequent RSS values and giving the maximum weight to the ones that have their peak value in a RSS histogram. $fpValue$ will always be the same if the most-recorded RSS values fall into the

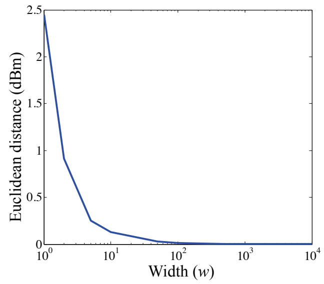


Fig. 3. Impact of w on the fingerprint value extraction

range set by w_{LT} and w_{RT} for every measurement. However, in reality, it is difficult to adjust those two width values to output the same $fpValue$. This is because the peak point moves slightly either to the left or to the right at each measurement depending on the environmental factors involved during the measurement. The difference between the extracted value from the short-time case and the long-time case was up to 3*dBm*, resulting in no advantage over the mean method. This problem that occurs in practice led us to modify the algorithm to the one given below:

$$fpValue = \frac{\sum_{n=1}^{w}(RSS_{peak}-n)+RSS_{peak}+\sum_{n=1}^{w}(RSS_{peak}+n)}{2w+1} \quad (2)$$

In (2), instead of adjusting w_{LT} and w_{RT} independently, we use one width value w and set it to large enough number. By maximizing the w value, we can minimize the differences of $fpValue$ obtained at each measurement. Figure 3 presents Euclidean distances between the $fpValue$ from the one-hour measurement and the $fpValue$ from the one-minute measurement with respect to the w value in a logarithmic scale. The results were based on average values from 50 measurements, and more than 10 APs were recorded for each measurement.

IV. Localization Algorithm

This section introduces our localization scheme in detail taking into account the issues pointed out in Section II.

A. Relative RSS Comparison

Today's widely and densely deployed Wi-Fi APs enable our new indoor localization technique. Previous fingerprint-based localization techniques build the radio map and also estimate the positions using absolute RSS values. They may work well in systems in which users carry homogeneous devices and use fingerprint data sets from trained surveyors. However, they are ineffective when they are applied to a crowdsourced system. Instead of trying to modify existing techniques to make them work in a heterogeneous device environment, we devise a novel technique designed for our target indoor usage scenarios. Thanks to the proliferation of Wi-Fi technology, the number of visible APs anywhere inside buildings usually exceeds 10. Some buildings such as university buildings have private APs as well as densely-designed enterprise wireless networks, providing even more than 30 visible APs. In addition to the number of APs, we can also take advantage of the wide distribution of RSS values. RSS values usually range from very strong one (e.g., -40*dBm*) to very weak one (e.g., -100*dBm*).

Such a large number of visible APs and wide RSS distribution in an indoor environment enables us to use the relationship information between RSS values for the fingerprint data at a particular location. Figure 4 illustrates an example of our localization algorithm. Let us start with the basic scenario, in which one user uploads fingerprint data for one particular location and other users request their positions. Suppose there are four users in *Location 101* and *102* in which six APs are visible to all the users as shown in Figure 4(a). In this example, user *A* and *D* are the surveyors and user *B* and *C* want to know their positions.

First, user *A* and *D* measure RSS values for about one minute and upload fingerprint values with location labels (i.e., location *101* and *102*) to a radio map server. The fingerprint value for each AP is presented in Figure 4(b). Then the server creates radio map data for location *101* and *102* based on the strength relationship among the fingerprint values.

The radio map database consists of the following data sets:

$$\begin{aligned} fp_{l_x} = \{ & l_x, (KEY_0, VALUE_{0_\delta}, KEY_1, VALUE_{1_\delta}, ... \\ & , KEY_n, VALUE_{n_\delta})\} \end{aligned}$$

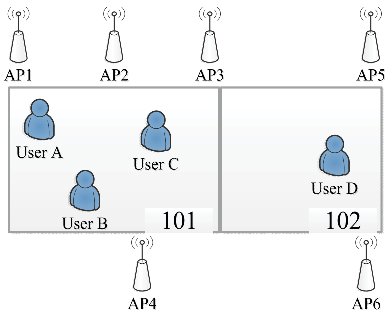
where $VALUE_{i_\delta}$ is a vector containing BSSIDs of which RSS is δ weaker than KEY_i. The *delta* (δ) is the core parameter, enabling our method to work well with heterogeneous devices with no calibration efforts. The proposed *Key-Value* mechanism and the δ value allow the fingerprint data to be meaningful in itself, not jut a set of RSS values.

fp_{l_x} represents the fingerprint data that is taken at location l_x. l_x contains the information of the location x (e.g., room number), KEY_i is the BSSID that has the i^{th} strongest $fpValue$. Figure 4(c) presents the fingerprint data created based on each user's fingerprint values and the δ value. Now the fingerprint data from the User *A* and *D* is stored in the radio map.

When user *B* and *C* request their positions, their observed fingerprint values are sent to the server, and then the data is compared with the radio map. The detailed localization algorithm using the fingerprint data is presented in Algorithm 1.

Based on our localization algorithm, the fingerprint data from user *B* scores 8 for location *101* and 1 for *102*. For the user *C*, the scores are 9 and 2 for location *101* and *102*, respectively. Therefore we conclude that both user *B* and *C* are in the location *101*.

Since the location estimation time is proportional to the number of fingerprint data to be compared, it is necessary to select optimal number of possible fingerprint data samples in the database in order to achieve fast result retrieval (line 4 in Algorithm 1). To this end, we utilize a "Importance Flag", an one-bit flag given to the *KEY* in fingerprint data. The flag is set if the *KEY* ranks high. When the localization server receives a location request, it selects fingerprint data that has at least one flagged *KEY* that matches the flagged *KEY* in the request data. For example, in the same abovementioned scenario, the flagged *KEYs* are {*AP1* and *AP2*} for the location *101* and {*AP6* and *AP5*} for the location *102* if we set the criterion for the flag



(a) User and AP locations

User A		User B	
AP	fpValue	AP	fpValue
AP1	-50	AP4	-52
AP2	-55	AP1	-65
AP3	-67	AP2	-68
AP4	-72	AP3	-74
AP5	-88	AP6	-87
AP6	-90	AP5	-91
User C		**User D**	
AP	fpValue	AP	fpValue
AP2	-51	AP6	-54
AP3	-53	AP5	-58
AP1	-65	AP3	-67
AP4	-66	AP4	-75
AP5	-82	AP2	-80
AP6	-85	AP1	-92

(b) RSS measurements for each user (ordered by strength)

User A		User B	
KEY	VALUE	KEY	VALUE
AP1*	{AP3,AP4, AP5, AP6}	AP4*	{AP1,AP2, AP3,AP5, AP6}
AP2*	{AP3,AP4, AP5, AP6}	AP1*	{AP5, AP6}
AP3	{AP5, AP6}	AP2	{AP5, AP6}
AP4	{AP5, AP6}	AP3	{AP5, AP6}
AP5	{}	AP6	{}
AP6	{}	AP5	{}
User C		**User D**	
KEY	VALUE	KEY	VALUE
AP2*	{AP1,AP4, AP5, AP6}	AP6*	{AP1,AP2, AP3, AP4}
AP3*	{AP1,AP4, AP5, AP6}	AP5*	{AP1,AP2, AP4}
AP1	{AP5, AP6}	AP3	{AP1, AP2}
AP4	{AP5, AP6}	AP4	{AP1}
AP5	{}	AP2	{AP1}
AP6	{}	AP1	{}

(c) Fingerprint data for each user (δ is $10dBm$)

Fig. 4. Example of Localization Algorithm

Algorithm 1 Estimate the location

Input: fingerprint data $fp_{unknown}$
Output: estimated location l

$score \leftarrow 0$, $score_{MAX} \leftarrow 0$
$value_{MAP} \leftarrow \{\}$

for each possible fp_{l_x} in the radio map **do**
  for each $KEY_{unknown}$ in $fp_{unknown}$ **do**
    if $KEY_{unknown}$ is found in fp_{l_x} **then**
      $value_{MAP} \leftarrow VALUE$ vector where its $KEY = KEY_{unknown}$
      for each $BSSID$ in $KEY_{unknown}$ **do**
        if $BSSID$ is found in $value_{MAP}$ **then**
          $score \leftarrow score + 1$
        end if
      end for
    end if
  end for
  if $score > score_{MAX}$ **then**
    $score_{MAX} \leftarrow score$
    $l \leftarrow l_x$
  end if
end for
return l

as the top two. The flagged *KEYs* are marked with asterisk in Figure 4(c). When user *B* and *C* request their positions, at least one flagged *KEY* in their fingerprint data is found in the fingerprint data of the location *101* (i.e., *AP1* for user *B* and *AP2* for user *C*) while no matching flagged *KEY* is found in the fingerprint data of location *102*. Therefore, only location *101* is considered for the possible location in this example. The reason we are bothered about high rank of *KEYs* is because adjacent locations have high probability of seeing the same APs that have strong RSS values. If no matches are found, we regard the position where the measurement is taken as an unlabeled location.

B. Heterogeneous Devices

Our radio map building and localization methods also work well in the case in which heterogeneous devices are involved because we do not use absolute RSS values but utilize only the relative relationship information among RSS values. The different Wi-Fi chipsets and antenna designs by devices cause different RSS for the same measurement location. Previous studies like [21] suggest to use a linear transformation to calibrate the RSS variation across different devices. However, as discovered in [16] we also find that the linear transformation method is not sufficient for an accurate localization system that adopts cross-device participation, and our method provides more robustness and accuracy.

Motorola Bionic	
AP	RSS
A	-61
B	-63
C	-64
D	-73
E	-75
F	-76
G	-76
H	-76
I	-79
J	-82
K	-83
L	-84
M	-87
N	-87
O	-87
P	-88
Q	-90
R	-90

(a) fingerprint

Samsung Galaxy II	
AP	RSS
C	-63
A	-64
B	-68
D	-70
E	-72
F	-72
H	-72
J	-75
L	-79
I	-80
K	-81
G	-83
Galaxy1	-85
R	-85
T	-86
Galaxy2	-86
Galaxy3	-87
P	-87

(b) Device 1

HTC NexusOne	
AP	RSS
A	-66
C	-67
B	-68
D	-70
E	-72
F	-72
H	-72
K	-74
G	-79
L	-80
N	-84
R	-86
J	-86
One1	-87
P	-87
I	-87
One2	-88
One3	-88

(c) Device 2

Samsung NexusS	
AP	RSS
A	-61
B	-63
C	-64
H	-73
F	-75
E	-76
D	-76
K	-76
G	-79
I	-82
L	-83
J	-84
R	-87
P	-87
U	-87
T	-88
Nexus1	-90
Q	-90

(d) Device 3

Fig. 5. Example of Device Heterogeneity

Figure 5 is an example that shows how our method provides consistent localization performance, regardless of device heterogeneity and without any calibrations, while linear transformation do not. Suppose that RSS values in Figure 5(a) are used to create fingerprint data for a particular location in which three different users are requesting their current locations using three different devices. The RSS values measured by each user's device are presented in Figure 5(b) (c) (d). For the sake of simplicity, all the recorded BSSIDs are replaced with single letters and the ones with weak RSS values are excluded. Device name with numbers indicate BSSIDs that are visible to the device but not found in the fingerprint data.

None of the three devices shows the same AP list order as compared to the one used for the fingerprint data. Not only list of out-of-order APs but also different APs that are visible to a particular device, hinder application of linear transformation. Uniform application of the transformation to devices may significantly degrade the localization accuracy. In contrast, our method has tolerance towards the discrepancies

Location 101 (User A+B)	
KEY	*VALUE*
AP1	{AP3,AP4, AP5, AP6}
AP2	{AP3,AP4, AP5, AP6}
AP3	{AP5, AP6}
AP4	{AP1,AP2,AP3,AP5, AP6}
AP5	{}
AP6	{}

Fig. 6. New Fingerprint Data for Location 101

in RSS values and visible AP lists among the heterogeneous devices. The δ value relieves the degradation of localization accuracy caused by inconsistent sequence of few adjacent APs among the heterogeneous devices. Since the proposed method focuses on the overall relationship between RSS from APs rather than individual RSS values, it is immune to the device diversity.

C. Multiple Surveyors

In addition to the device diversity, another unique problem of crowdsourced localization systems is that more than one user can upload their own fingerprint data with the same location labels. This results in more than one fingerprint data for one location in a radio map in previous systems, which leads to slow location estimation due to the increase in the number of data sets to be compared. Unlike the previous techniques, we maintain only one fingerprint data for one location. Unlike any other localization methods, our fingerprint data building mechanism supports this unique feature with no calibrations. Updation of fingerprint map for new data is possible because we utilize the information on the overall RSS relationship among APs. The merging procedure for multiple fingerprint data is extremely straightforward and simple. The proposed *Key-Value* fingerprint data structure and the δ value increase the similarity among multiple fingerprint data although they are measured at slightly different locations with different devices, which allows us to simply merge the *Value* vectors for the same *Key* value. For example, the new fingerprint data will be Figure 6 if user *A* and *B* upload their RSS measurements with the same location label, location *101* in Figure 4. This merging scheme can be also applied to the case when users upload their fingerprint data with different devices at the same measurement positions.

V. Evaluation

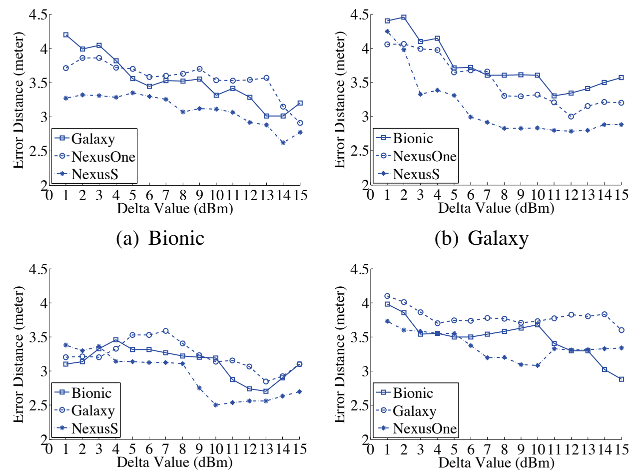
This section presents the methodologies, results, and analyses of our experiments.

A. Experimental Setup

The Wi-Fi fingerprint data was collected at approximately 70 different locations at one of the engineering buildings in our university. Every fingerprint comprises of the following information: timestamp, BSSID (MAC address of the AP), and received signal strength indicator. The fingerprints were collected using four different devices (Motorola Bionic, HTC Nexus One, Samsung GalaxyS and GalaxyS2). For evaluations, we selected two sites which had different environments. The first experiment site was our laboratory that has many small rooms which contain different types of furniture (Figure 7(a)). Another experiment site was the corridor of third floor of the

(a) Our Laboratory (b) 3rd Floor of Our Building

Fig. 7. Floorplans



(a) Bionic (b) Galaxy

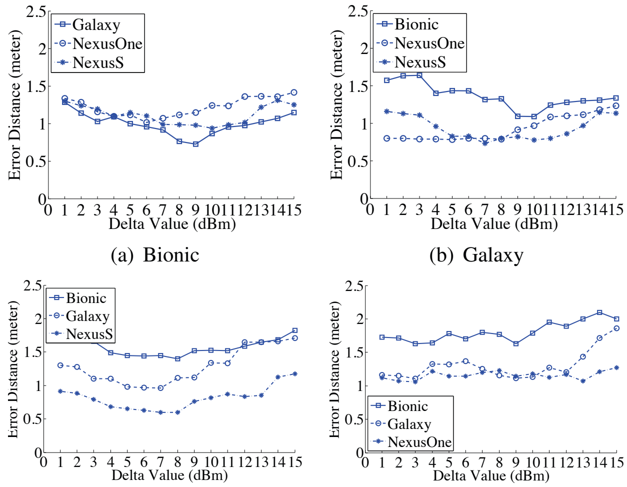
(c) NexusOne (d) NexusS

Fig. 8. Pairwise Evaluation (Laboratory)

building, the width of whose corridor is approximately 2.5*m*, and some APs are visible with line-of-sight (Figure 7(b)). Small dots in Figure 7 indicate the locations where RSS measurements were taken. The data collected at these points was used for the experiments in Section V-B and V-C. The adjacent points are approximately 6*m* and 1.5*m* apart for the corridor and the laboratory cases, respectively. Different sets of reading were taken at different times of the day, and over multiple days, keeping the simulations as close to real-world scenarios as possible.

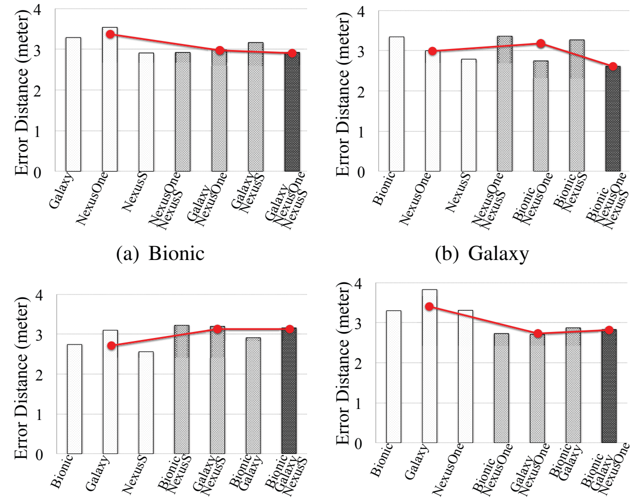
B. Pairwise Device Evaluation

In the first evaluation, the Wi-Fi fingerprinting data for each location was collected from one phone and the data from all other three phones were compared against it for different δ values. There were two goals of this experiment. The first was to find out whether the proposed method of building fingerprint and using it for indoor localization works well with heterogeneous devices. The second goal was to find out the optimal δ value, to be used for subsequent experiments. The δ values were varied from 1 to 15. By repeating the same experiment for all the devices and all the δ values, we found the optimal empirical δ value. The results for all the four devices are shown in Figure 8 for the laboratory (which had lots of objects, walls, furniture etc) and in Figure 9 for the third floor (which had relatively less objects hindering the signal paths).



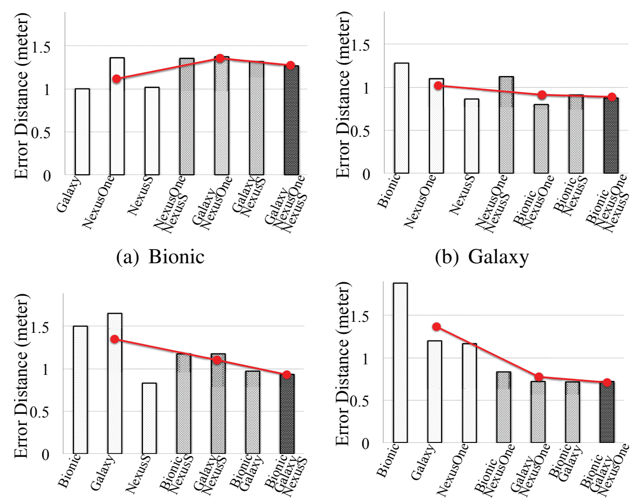
(a) Bionic (b) Galaxy (c) NexusOne (d) NexusS

Fig. 9. Pairwise Evaluation (3rd Floor)



(a) Bionic (b) Galaxy (c) NexusOne (d) NexusS

Fig. 10. Impact of Device Heterogeneity (Laboratory)



(a) Bionic (b) Galaxy (c) NexusOne (d) NexusS

Fig. 11. Impact of Device Heterogeneity (3rd Floor)

We ran the evaluation for data collected over three days. This data was taken at all the locations as shown on the maps at different times. For a majority of cases, we were able to accurately locate the position of a user using our algorithm. For the cases where the predicted location of the user was different from the real location, we calculated the error distances. The error in location detection was lesser in the case of third floor of the building as compared to the laboratory. The reason for this observation of degraded accuracy in case of the laboratory is because the lab structure is more complicated, and the walls in the floor plan are made of thin plywood material, not thick concrete. This hinders each room from having a unique fingerprint. We believe that more unique fingerprints will be formed in buildings with thick concrete walls.

As can be seen from Figure 9, the optimal δ value is around $9dBm$ for the third floor case. If we define the error when we form fingerprints from one device and use another device for indoor location detection as cross device error, then it has been observed that cross-device error does not exceed $2m$ for hallway and $4m$ for laboratory case. This justifies that our method provides consistent performance among heterogeneous devices. We believe that error will decrease considerably as the homogeneity of devices increase in the crowdsourced system. As can be seen from Figure 8, the optimal delta value is around $12dBm$ for the laboratory. Considering overall performance of the whole system, using $12dBm$ for the δ value is better than using the value $9dBm$. We have made $12dBm$ as optimal threshold for the rest of the evaluations.

C. Impact of Device Heterogeneity

In the second evaluation, the Wi-Fi fingerprinting data for each location was taken from multiple devices and data from all other mobile phone devices was compared against it. Having established the fact that our algorithm works well for calibration-free crowdsourced indoor localization, the aim for this evaluation was to find out how device heterogeneity affects the localization performance. Hence for this evaluation, we varied the number of devices which participate in the fingerprint map for each user device. The same device was never used for both generating fingerprinting map and for localization. The results have been plotted in Figure 10 and 11. First three bars are the cases when only one device was used for the fingerprinting. The next three bars are the cases when the fingerprint maps were generated from two different devices and these fingerprint maps were merged to form one fingerprint map. The last bar denotes the case when fingerprint maps have been generated from three different devices and merged. Dots in Figure 10 and 11 indicate average values of the same colored bars.

The accuracy did not degrade when more than one fingerprint data from different devices are merged. In most cases, the most robust fingerprint has been formed when fingerprint maps from three different devices have been merged. This re-affirms our belief that our method will work well in crowdsourced environment. The reason for the above findings is that our

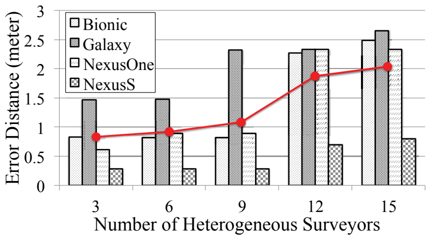


Fig. 12. Impact of Multiple Surveyors (Laboratory)

fingerprint map gets better as more fingerprint data is merged. If certain APs went unnoticed by one device which could have contributed for missing information from the fingerprint maps, then it is likely that other devices would have sensed these APs. There is less probability that a particular AP will get unnoticed by three devices than by one device. Hence, more information helps in forming more accurate, robust and reliable fingerprint map.

D. Impact of Multiple Surveyors

In the third evaluation, our aim was to evaluate how our algorithm worked for the case of multiple surveyors. We constructed the fingerprint map for a particular room using heterogeneous devices placed at different parts (levels) of the room. Each level corresponded to a set of locations that were at a specified distance from the center of the room. The user requesting for location information was assumed to be standing at the center of the room. Every level had three devices, that were different from the user's device. We conducted these evaluations across the five levels and across the eight rooms in our lab. The five levels span the entire room (e.g., level 1 was the center of a room and level 5 was the edges of a room). The results are presented in Figure 12. Dots indicate average values of the four devices' results.

As expected, the error in finding the user's location increases as more users at different locations contribute to the fingerprint map for a particular room. However, our algorithm limits the error in accuracy to less than 3 meters. This is commendable given the fact that as more fingerprinting data is collected at the "edges" and "corners" of different rooms, it is more likely to predict the location of the user as being the adjacent room.

VI. Discussion

In this paper, we proposed a novel calibration-free technique for a crowdsourced indoor localization system, and evaluated its performance through laboratory-level experiments. The results showed that our scheme provided robust and consistent localization accuracy in various experimental scenarios. However, there are still several challenging technical issues that need to be addressed in order to deploy our system in a real large-scale environment. We briefly describe two of these issues that we are currently working on.

Filtering erroneous fingerprint data is essential in crowdsourced systems. Since the entire system is based on participation of untrained normal users, erroneous uploads can happen frequently. Apart from malicious users who intentionally upload fingerprint data with wrong location labels, even well-intentioned users may contribute wrong fingerprint data due to a simple mistake in location labeling. Currently, we consider an approach that utilizes the "importance flag" introduced in Section IV-A. We anticipate that detection of erroneous fingerprint data is possible by comparing properly-selected flagged APs in adjacent areas. We plan to perform more experiments to find the proper flag criterion.

Another issue is how our system would update fingerprint data. Outdated fingerprint data may significantly degrade the localization accuracy. Since Wi-Fi APs can be easily added or removed, a technique that detects the change of APs is imperative. To this end, we are investigating an approach that keeps track of "hit rate" of APs in fingerprint data sets. If a particular AP is never used to estimate locations for a certain amount of time, there is a high chance that this AP has been removed. Similarly, if an AP that does not exist in the fingerprint data is observed frequently, then that AP is probably installed recently. We may be able to automatically update fingerprint data if we make the best use of such information.

VII. Related Work

Many indoor localization techniques have been proposed over the past decade. Amongst these, the techniques that use existing Wi-Fi infrastructure have attracted special attention due to the ubiquity of wireless access points in indoor environments. Our work builds upon the crowdsourcing based (also known as "organic") approaches, and deals with the problems posed by the device heterogeneity and multiple surveyors for a particular location.

Early indoor localization techniques required specialized hardware to determine the devices location. Systems like Active Badge [8] and Active Bat [9] used methods based on tags which transmitted infrared and ultrasound pulses. These were detected by fixed sensors placed in buildings. The Cricket and Cricket Compass [10], [24] used a combination of RF and ultrasound technologies. Techniques using active RFID were also proposed [25], [26]. Later systems used existing infrastructure in buildings and depended on radio frequency signal measurements. RF signal intensity was first used in RADAR [12] for the purpose of indoor localization. More recently many researchers have focused on techniques that use Wi-Fi RSS data. Studies like [22] analyzed the properties of received signal strength values reported by wireless network interface cards.

In general, the techniques for localization using signal strength values are either based on triangulation methods [11] or are based on predicting the location based on stored Wi-Fi fingerprints. Recent systems like [27], [28] use probabilistic techniques to predict the location of a Wi-Fi device. In these systems, a "training" phase is required in which the system uses a set of tagged data to build its internal localization model. Various studies, like [21], have researched on the amount of training data required to sufficiently train the localization system. [21] presents a system which is trained using signal strength measurements, which are of the order of about a minute per location.

The Wi-Fi based systems that use signal strength fingerprints can be classified into two categories. In the first category, the training data set is collected by expert surveyors. In the second category, the training data set is contributed by users, and is referred to as organic localization [15]–[18]. Due to the complexity, effort and cost involved in expert surveyor based system, crowdsourced or organic indoor localization techniques are being rigorously explored. The system proposed in [13], [14] was among the first to explore such an approach. The key idea here is to use a participatory system through which users can contribute Wi-Fi fingerprints at various locations. [14] carried out experiments in indoor environments and have discussed encouraging results. Park *et al* in [15] have designed and deployed an organic location system and achieved position accuracy that is comparable to the accuracy achieved by a survey driven system. A recent crowdsourcing based approach by Ledlie *et. al.* models the world as a tree of hierarchical namespaces, and provides an algorithm that explicitly accounts for temporal variations in signal space [18].

A main issue with crowdsourcing based localization systems is the usage of diverse devices, usually a variety of mobile phones, to collect Wi-Fi fingerprints during the training phase. This usually leads to variation in the values of observed signal strength measurements due to the different chipsets present on different devices. Park *et. al.* explore this issue in [16] and compare various methods used to mitigate this problem. The main parameters that are used are signal strength values and access point detection. The Kernel function based estimation, which predicts user location using a naive Bayes classifier, has been found to obtain results that are better than linear transformation based approaches [21]. Arvin *et. al.* proposed an unsupervised learning method that automatically tries to solve hardware variance problem in wifi localization [29].

VIII. Conclusion

FreeLoc is a novel, calibration-free indoor localization scheme that uses existing Wi-Fi infrastructure. The proposed radio map building and localization techniques are based on the overall relationship among RSS by APs. Our techniques provide robust localization accuracy in a crowdsourced environment in which device heterogeneity and multiple untrained surveyors mainly cause performance degradation. The main contribution of our work is a novel approach to fingerprint data management and a localization algorithm that are capable of handling diverse users and their devices without complicated calibrations or transformations. Initial experiments using heterogeneous devices in two sites that have different environments have confirmed that our novel scheme is reliable and feasible.

For future work, we plan to develop schemes covering the issues discussed in this paper. We will then expand the scale of our experiments to cover the entire university and perform long-term usability testing.

References

[1] S. Wang, J. Min, and B. K. Yi, "Location based services for mobiles : Technologies and standards," *Electronics*, p. 7892, 2008.

[2] D. Mohapatra and S. Suma, "Survey of location based wireless services," in *Personal Wireless Communications, 2005. ICPWC 2005.*

[3] M. Mohammadi, E. Molaei, and A. Naserasadi, "A survey on location based services and positioning techniques," *International Journal*, vol. 24, no. 5, pp. 1–5, 2011.

[4] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *Communications Surveys Tutorials, IEEE*, vol. 11, no. 1, pp. 13 –32, quarter 2009.

[5] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *Systems, Man, and Cybernetics, IEEE Transactions on*, vol. 37, no. 6, nov. 2007.

[6] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche, "A comparative survey of wlan location fingerprinting methods," in *Positioning, Navigation and Communication, 2009. WPNC 2009.*

[7] D. Zhang, F. Xia, Z. Yang, L. Yao, and W. Zhao, "Localization technologies for indoor human tracking," in *FutureTech, 2010.*

[8] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.*, vol. 10.

[9] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," in *Proceedings of MobiCom '99.*

[10] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of MobiCom '00.*

[11] I. Smith and et al., "Place lab: Device positioning using radio beacons in the wild," in *In Proceedings of the Third International Conference on Pervasive Computing*. Springer, 2005.

[12] P. Bahl and V. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *INFOCOM 2000.*

[13] P. Bolliger, "Redpin – adaptive, zero-configuration indoor localization through user collaboration," in *MELT*, San Francisco, USA, Sep. 2008.

[14] M. Lee, H. Yang, D. Han, and C. Yu, "Crowdsourced radiomap for room-level place recognition in urban environment," in *(PERCOM Workshops), 2010.*

[15] S. Teller, J. Battat, B. Charrow, R. Ryan, J. Ledlie, J. Hicks, S. Teller, R. Ryan, J. Battat, B. Charrow, J. Ledlie, D. Curtis, and J. Hicks, "Organic indoor location discovery," Tech. Rep., 2008.

[16] J. geun Park, D. Curtis, S. Teller, and J. Ledlie, "Implications of device diversity for organic localization," in *INFOCOM, 2011.*

[17] J.-g. Park, B. Charrow, D. Curtis, J. Battat, E. Minkov, J. Hicks, S. Teller, and J. Ledlie, "Growing an organic indoor location system," in *Proceedings of MobiSys '10.*

[18] J. Ledlie, J. geun Park, D. Curtis, A. Cavalcante, and L. Camara, "Mole: A scalable, user-generated wifi positioning engine," in *IPIN, 2011.*

[19] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks: A Unified Approach*, 1st ed. Prentice Hall PTR, 2001.

[20] A. Smailagic, J. Small, and D. P. Siewiorek, "Determining user location for context aware computing through the use of a wireless lan infrastructure," *Institute for Complex Engineered Systems Carnegie Mellon University Pittsburgh PA*, vol. 15213, 2000.

[21] 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 *Proceedings of MobiCom '04*, 2004.

[22] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for wlan location fingerprinting," in *MOBIQUITOUS 2004.*

[23] A. M. Ladd, K. E. Bekris, A. Rudys, G. Marceau, L. E. Kavraki, and D. S. Wallach, "Robotics-based location sensing using wireless ethernet," in *Proceedings of MobiCom '02*.

[24] N. B. Priyantha, A. K. Miu, H. Balakrishnan, and S. Teller, "The cricket compass for context-aware mobile applications," in *Proceedings of MobiCom '01.*

[25] L. Ni, Y. Liu, Y. C. Lau, and A. Patil, "Landmarc: indoor location sensing using active rfid," in *PerCom 2003.*

[26] G. yao Jin, X. yi Lu, and M.-S. Park, "An indoor localization mechanism using active rfid tag," *Sensor Networks, Ubiquitous, and Trustworthy Computing, International Conference on*, vol. 1, pp. 40–43, 2006.

[27] P. Castro, P. Chiu, T. Kremenek, and R. R. Muntz, "A probabilistic room location service for wireless networked environments," in *Proceedings of UbiComp '01*. London, UK, UK: Springer-Verlag.

[28] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A Probabilistic Approach to WLAN User Location Estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, Jul. 2002.

[29] A. W. Tsui, Y.-H. Chuang, and H.-H. Chu, "Unsupervised learning for solving rss hardware variance problem in wifi localization," *Mob. Netw. Appl.*, vol. 14.