Journal of Network and Computer Applications 44 (2014) 220–229

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

A sensor based indoor localization through fingerprinting

Israat Tanzeena Haque

Computing Science, University of Alberta, Alberta, Canada

ARTICLE INFO

Article history:
Received 4 December 2013
Received in revised form
28 May 2014
Accepted 3 June 2014
Available online 9 July 2014

Keywords:
Indoor localization
RF-based localization
Received signal strength
Fingerprint based localization
Localization error
Robust localization

ABSTRACT

Fingerprint based localization mostly considers to exploit existing infrastructure (APs or FM broadcast) to avoid hardware requirement and deployment cost. On the other hand these solutions confine them with such fixed infrastructures without having any control to reorganize or include further hardware if required. Controlling the infrastructure may help localization, especially indoors where the presence of multipath is high. A dense population of infrastructure nodes placing in all required indoors may capture a precise view of the surveyed area while generating a radio map of the fingerprinting or profiling based localization. The larger the number of infrastructure nodes the higher the cost. Indeed dense node population may introduce high interference. Can we then use low-cost low-power infrastructure nodes to achieve the (1) high density, (2) low cost, (3) minimal interference in addition to have the deployment flexibility? We were curious to know the answer and designed LEMON, an indoor localization system. Extensive experiments show that in addition to have the above characteristics, LEMON also ensures good accuracy. Thus it can be a solution to locate a person (e.g., a security guard in a warehouse) or an object (e.g., equipments) indoors.

1. Introduction

Location-based service is an interesting but challenging task under the roof, where traditional GPS is not a suitable candidate because of (1) cost, (2) form factor, (3) accuracy, and (4) unavailability. However the location-based service has the potentiality to leverage the ease of indoor navigation such as shopping mall and airports. Indeed we spot another application where an accurate location measurement is crucial for on time response; we indicate a home alone elder resident, who may need immediate assistance due to unexpected health condition. Though accuracy is our first concern, we consider cost as the next issue while building an indoor localization system. We thus choose the generic RF-based technology because of its simplicity, ubiquity, low cost, and unobtrusiveness.

We have choice of measuring RSS (received signal strength) (Bahl and Padmanabhan, 2000; Youssef and Agrawala, 2005), TOA (time of arrival) (Humphrey and Hedley, 2008), or AOA (angle of arrival) (Niculescu and Nath, 2004; Amundson et al., 2011) of the signal between a transmitter and a receiver under RF-based technology. The first of these categories of techniques is most attractive from the practical point of view, as it poses minimalistic requirements on the RF technology of the requisite modules, which translates into low cost and off-the-shelf availability. We further consider RSS profiling (Bahl and Padmanabhan, 2000; Youssef and Agrawala, 2005) as the method of localization because of its potentiality of offering high accuracy (Bahl and Padmanabhan, 2000; Youssef and Agrawala, 2005).

In profiling based scheme, RSS is considered as a quantity that depends on the distance between a transmitter and a receiver as well as the indoor environment. Thus we may expect that the RSS readings from similar environment may behave similarly. This hope lies behind the construction of a *radio map* of the monitored area by gathering the RSS readings from known locations. The RSS is captured through a set of infrastructure nodes (*peg*). To estimate the location of a query (*tag*), q, based on a given set of RSS readings Φ, this map is explored to search for a set of nearest neighbors of Φ. In the radio map the locations of those chosen neighbors are also stored that are used to predict the location of q.

We may distinguish the profiling based schemes according to the technologies used to gather the signal strength. In this list WiFi APs come first because of their indoors availability. Almost all wireless devices are equipped with a RSS receiver thus localization with APs does not require any hardware and infrastructure setup. A plethora of indoor localization is done using this technology, and some of the front line solutions are Bahl and Padmanabhan (2000), Youssef and Agrawala (2005), Liu et al. (2012), Wang et al. (2012), Sen et al. (2012), Shen et al. (2013), Yang et al. (2012) and Swangmuang and Krishnamurthy (2008). RFID tags and readers can also be used for indoor localization (Wang et al., 2013; Ni et al., 2004). FM broadcasting is another choice of gathering signal strength signatures (Yoon et al., 2013; Chen et al., 2012). This option does not require any extra hardware or infrastructure as FM

E-mail address: israat@cs.ualberta.ca

http://dx.doi.org/10.1016/j.jnca.2014.06.001
1084-8045/

broadcast is everywhere. Other options include magnetism (Chung et al., 2011), GSM (Otsason, 2005), or electric power-line (Patel et al., 2006).

Specific hardware based approaches (Bahl and Padmanabhan, 2000; Ni et al., 2004) may suffer from large error because of the constraints on the number of costly infrastructure nodes. This problem can be tackled using APs as they are already crowded indoors. Unfortunately localization performance can be affected by the constrains on the number and the locations of these APs (no control over rearranging existing ones or adding more APs). Indeed recent investigation (Yoon et al., 2013; Chen et al., 2012) reveals that the WiFi AP based signals are highly susceptible to indoor environmental changes (like the presence of moving humans, obstacles, etc.). Furthermore WiFi signals have high special and temporal impact that may lead to reconstruct the radio map often, makes WiFi based approaches nonfeasible indoors solution. In Yoon et al. (2013) and Chen et al. (2012) FM signal is used to overcome the shortcomings of WiFi signals. FM signals are less susceptible to the environmental changes and can cover longer distances compared to the WiFi signal. Recent smart phones and other mobile devices are all equipped with FM radio receivers with the existence of plenty of FM broadcasting, makes this technology viable. We still have no control over the number and location of the FM broadcasting, and this limitation may be seen in the localization performance (similar to WiFi, Chen et al., 2012).

We thus introduce a flexible infrastructure based localization without having any constraints on the number and the locations of the infrastructure nodes. In particular we use low-cost low-power small sensor devices (from Texas Instruments) as pegs and tags to design an indoor localization system dubbed *LEMON* (*Location Estimation by Mining Oversampled Neighborhoods*) (presented in Haque et al., 2009). LEMON can effort a large number of pegs (because of their low cost) that can be placed in any locations of indoors (behind and under the furniture) to precisely capture the signal propagation in the constructed radio map. This immediately helps to achieve a good accuracy without applying any further trick to the localization procedure.

A series of experiments in various indoors of the University of Alberta campus show that LEMON can offer an accuracy of 1 m. Note that the surveyed area is divided into a grid with a cell size of $1\ \text{m} \times 1\ \text{m}$. We have analyzed the RSS discrepancy measuring approach, the averaging technique of the coordinates of the nearest neighbors, the number of nearest neighbors and fingerprint samples, the impact of the temporal and special variation on the RSS, and the impact of the types of obstacles to evaluate LEMON. Surprisingly LEMON was not affected by the temporal variations, which makes the radio map construction phase simple. As long as indoors setup remains unchanged we need no more site surveying. However, our deeper analysis reveals that the captured RSS is contaminated by the multipaths and introduced error in the estimation. Thus we further investigated the robustness of LEMON in the presence of noisy RSS and faulty devices and proposed solutions to handle noisy RSS.

In the remaining paper LEMON is presented in Section 3 and next comes the description of the hardware and the logistics. Experimental results are presented in Section 5 and the robustness study comes in the following section. We wrap up the paper with concluding remarks.

2. Related work

In this section we briefly outline the fingerprint based localization schemes related to our work.

RADAR (Bahl and Padmanabhan, 2000) can be considered the pioneer of the fingerprint based localization indoors. During the radio map construction, RSS samples are gathered from four different directions for the same location to overcome the orientation effect. For localization, a collected sample is compared to the stored set and the coordinates of the closest fingerprint from the signal space is reported as the estimated location. Choosing more nearest neighbors and averaging their locations tends to improve the estimation. However, RADAR still suffers from large errors due to the limited number of infrastructure nodes: three long range APs covering the entire monitored area. One attempt to fix the problem involved a signal propagation model taking into account the presence of walls between transmitters and receivers. That attempt was not effective as the observed performance was even worse, which should be taken as a strong hint that, generally, propagation models cannot compensate for inadequate coverage with infrastructure nodes.

Several popular localization approaches rely on RFID technology. Such a system usually consists of a set of RFID readers, comprising the infrastructure, and trackable RFID tags. An RFID reader is able to detect the signal from a tag, if it gets sufficiently close. For a passive RFID tag, this will happen when the distance to the reader is so small that the scheme becomes range-free: detection by a reader is a sufficient estimation of the tag's location. A localization system like this may not provide a full coverage of the monitored area and be only concerned about detecting the presence of tags in certain “critical” places or regions. With active tags, on the other hand, which act like cheap low-range transmitters, the readers may be able to meaningfully assess the received signal strength and use it as a representation of the tag's distance, e.g., quantized into a few coarse discrete levels.

One RFID-based representative of the fingerprint based schemes is LANDMARC (Ni et al., 2004). The network consists of a set of RFID readers as the infrastructure nodes and RFID tags as the sending (tracked) devices. LANDMARC suffers from the technological limitation of RFID readers (the lack of a direct measurement of RSS by the reader). Also, the large diversity of hardware versions of tags impacts the performance.

In Patel et al. (2006) the existing residential powerline network is used for localization purposes, with the infrastructure nodes being attached to the powerline around the perimeter of the household. The system, called PLP (Power Line Positioning) targets residential applications. The signal transmitted by the infrastructure nodes is received by the tracked tag. Thus, with this approach, tags collect signal samples from the infrastructure nodes, not the other way around, as in RADAR, LANDMARC, and also LEMON. During fingerprinting, signatures of signals from known locations are stored in a database. The estimation stage proceeds in two phases: first the room where the tag appears to be present is identified, and using a respectively trimmed down population of samples, the more exact assessment of the tag's location within that room is carried out. However our experimental results show that the two-phased approach to location estimation in PLP may not be an effective approach. The task of accurately inferring whether a tag is in a particular room is often difficult (especially when the tag is positioned close to the wall), and once that decision is made incorrectly, its subsequent refinement is not useful.

With respect to WiFi-based solutions, Swangmuang and Krishnamurthy (2008) investigate the performance impact of a localization scheme for various AP density. The fingerprints are gathered at every grid point, where the grid consists of $1\ \text{m} \times 1\ \text{m}$ cells. The closest neighbor from the signal space, like in RADAR, is then reported. The authors propose a model for ranking the collected RSS samples with respect to their contamination level and selecting less contaminated samples for location estimation.

The practical performance results reported in Swangmuang and Krishnamurthy (2008) are rather disappointing: in our framework, they would translate into several meters of average error distance, even with five APs. One advantage of LEMON over WiFi-based schemes, worth stressing in this context, is the highly reduced power of RF signals resulting in smaller coverage of a single "AP". As our pegs do not compete in any manner and are cheap, we can easily afford a dense deployment translating into a reduced coverage of a single peg. This automatically implies better resolution and accuracy, even without any further tricks.

Another WiFi-based localization scheme is presented in Hossain et al. (2013) where the authors propose to use pairwise RSS differences from the APs as the fingerprints. This increases the dimensionality of the sample space and, to some extent, may compensate for the limited number of APs. For example, with four APs, the number of samples from a single collection is six rather than four. Our experiments indicate that having a larger number of actual pegs is much more beneficial than such tricks.

In Martin et al. (2010), authors also use WiFi APs as infrastructure nodes and their fingerprint based localization scheme is implemented using smart phones. The accuracy is around 1.5 m where captured RSS is averaged over time and radios to achieve this accuracy. Depending on APs reduces the hardware requirement but that also makes the localization scheme restricted by the number and the locations of those APs, which has impact on the localization accuracy (Martin et al., 2010). LEMON is not depended on such existing hardware, rather it uses low-power low-cost sensors that makes it flexible in terms of the infrastructure deployment and in the analysis of the results section we will see that such flexibility is crucial to obtain good accuracy.

3. The localization algorithm

In general, the KNN based localization explores the radio map to choose K closest neighbors in terms of minimizing the RSS discrepancy between the query RSS (Φ) and the stored fingerprints. The average of the locations of these K neighbors generates the predicted location. LEMON is based on a set of low-cost low-power wireless sensors, where a large number of sensors are used to generate an accurate radio map that immediately helps obtaining good accuracy. The localization steps of LEMON is defined in Algorithm 1.

Algorithm 1. LEMON (Φ, K).

1: Find peg p_h in Φ that received highest signal strength r_h from the tag.
2: Find all the profiled RSS samples that perceived signal strength from p_h.
3: Sort these samples in terms of the RSS discrepancy between these samples and Φ.
4: Choose top K profiled samples (nearest neighbors).
5: $x_e \leftarrow \sum_{i=1}^{K} w_i x_i$
6: $y_e \leftarrow \sum_{i=1}^{K} w_i y_i$

The localization steps of LEMON can be described as follows:

Having received a location estimation request, the server will search through the database of the stored fingerprints in order to choose a small number of best matching ones. The input to the estimation process is a list of readings $<p, r>$, denoted by Φ, representing the set of pegs that have received the current packet sent by the tag (at the unknown location) along with their RSS readings. One simple heuristic used to narrow down the subsequent search consists of selecting from Φ the pair $\langle p_h, r_h \rangle$, such that r_h is the highest among all pairs in Φ. Then, the procedure will only consider those fingerprints from the database whose association lists include p_h as one of the pegs.

After selecting a subset of relevant fingerprints from the database, the server will sort them according to their discrepancy from the query sample Φ. Then, it will select K neighbors in the signal space that are "closest" to the query. Suppose that $\Omega = \{\omega_1, \dots, \omega_m\}$ and $\Psi = \{\psi_1, \dots, \psi_m\}$ are two association lists. The discrepancy between these lists can be defined as

$$D(\Omega, \Psi) = \sqrt{\sum_{j=1}^{m} (R_\Omega(j) - R_\Psi(j))^2} \tag{1}$$

where m is the total number of pegs in the network and $R_\Omega(j)$ is defined as r_j, if the pair $<p_j, r_j>$ occurs in Ω, and 0 otherwise. The same is true for $R_\Psi(j)$.

In the last step, the coordinates of the K selected fingerprints are averaged to produce the predicted location as

$$(x_e, y_e) = \frac{\sum_{i=1}^{K} (x_i, y_i) \times (d_{max} - d_i)}{K \times d_{max} - D} \tag{2}$$

where (x_i, y_i) are the coordinates of the fingerprints i. d_{max} is the maximum discrepancy from Φ among the best K selected fingerprints and $D = \sum_i^K d_i$ is the sum of all those discrepancies.

4. System design

To evaluate and compare the performance of LEMON and other profiling based localization, we conducted a series of experiments at various indoors of the University of Alberta campus. The experimental setup and hardware are presented in the following subsections.

4.1. The hardware

In our experiments, the wireless device used for both pegs and tags alike is the EMSPCC11[1] from Olsonet Communications shown in Fig. 1. EMSPCC11 is a low-cost low-power mote for wireless sensor networking programmable in PicOS (Gburzynski and Olesinski, 2008). The mote employs the MSP430F1611 microcontroller and the CC1100 RF module, both from Texas Instruments. The RF module operates within the 916 MHz band. The transmission power is settable from −30 dBm to 10 dBm (in 8 discrete steps), the bit rate options are 5 kbps, 10 kbps, 38 kbps, and 200 kbps, and there are 256 different channels (numbered 0–255) with 200 kHz spacing. All combinations of options are possible and, in principle, sensible. The experiments reported in the rest of this paper were carried out at the second lowest power setting (to reduce interference and save energy) with 5 kbps transmission rate using channel 0.

4.2. Monitored area setup

In general we generate a grid in the monitored area with each grid cell size 1 m × 1 m. The RSS readings for the fingerprints are gathered from every grid point, whereas the test data is collected from the center of each grid cell (see Fig. 2). Note that the above-mentioned cell size is chosen based on an initial experiment.

4.3. Monitored area description

In early single-room experiments, the size of the surveyed area was a 8 × 8 grid with cell size 1 m × 1 m. The rooms were

[1] See http://www.olsonet.com/Documents/emspcc11.pdf for details.

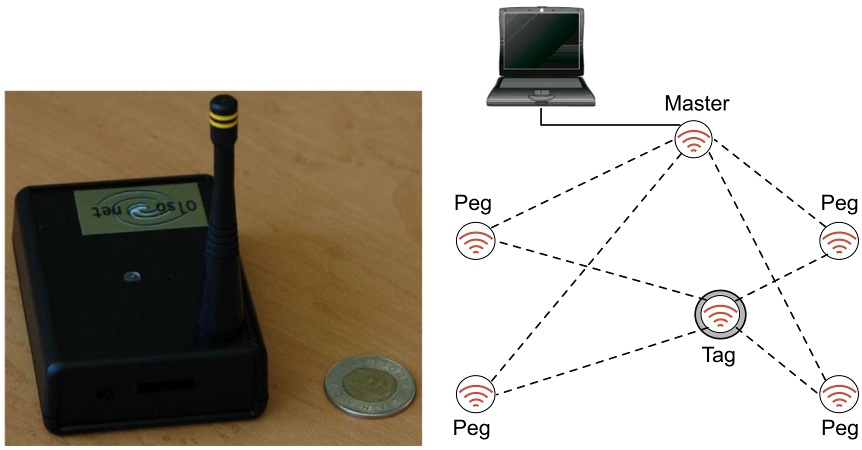

Fig. 1. The EMSPCC11 mote and the outline of experimental setup.

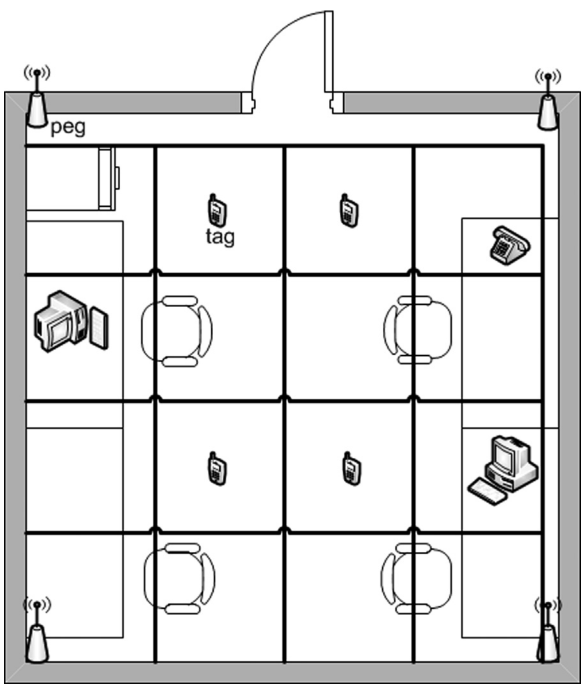

Fig. 2. A sample distribution of nodes in a room.

populated with obstacles, i.e., furniture and equipment (mostly tables, cabinets, chairs, computers, and related items), to a varying degree, from sparse (a few chairs moved to the walls), to dense (a typical packed office inhabited by graduate students). Subsequent experiments were carried out in a three-room scenario: three adjacent graduate student offices (each $3\ \text{m} \times 5\ \text{m}$), with each office outfitted with four wooden tables, four chairs, several wooden shelves mounted on the wall, two PC computers, and a steel file cabinet. Also, people were allowed to move around during the data collection.

4.4. Data collection

Our experiment starts by deploying a number of nodes within the monitored area. The networked program run in the nodes (the *praxis* according to PicOS terminology Gburzynski and Olesinski, 2008) allowed us to obtain RSS readings between all pairs of directly reachable nodes and for any selected setting of the transmitter (output power, bit rate, channel number). The nodes then exchange a massive number of packets, conveying to the central node (dubbed the *master*) their parameters (sender/receiver ID, serial number, transmitter parameters, RSS). The master node was connected to a laptop, where all the data collected by the network were deposited. Figure 1 depicts schematically the data collection phase of our experiment, where the master node initiates the exchange of packets and also collects the reported signal readings from the pegs sending them to the server.

4.5. Impact of time

We initially tested the impact of time on the gathered RSS. In particular we gathered the profiled data over two (in case of single room) and three (in case of multiple rooms) days, then started gathering the test data over more than a day. We have noticed that gathered RSS has almost the same behavior as long as the indoor environment (setup) remains same. However, if we rearrange the setup of furniture, a change in the RSS is observed.

4.6. RSS scaling

Initial RSS analysis reveals that the RSS from a distant peg may contribute to error rather than information if proper care is not taken. Thus to give less relevance to distant pegs, we define *RSS scaling*. It is a process of assigning different priorities to the gathered RSS, r, to control the impact of RSS noise (Haque et al., 2009). For example, we can use following scaling formula:

$$R_s = \left(\frac{r - MIN}{MAX - MIN}\right)^{\gamma},$$

MIN and MAX represent the minimum and maximum obtainable readings across all sets of measurements, respectively, and $\gamma > 1$ is the amplification factor.

4.7. Performance evaluation

Once we finish the localization, we measure the *error distance* (error in meter) between the actual and the estimated locations as the Euclidean distance between them. All our results report *average error distance* measured in meter. All the average error results that are presented in this work consider 95% confidence interval.

5. Discussion on results

In this section we present the analysis on the performance of LEMON and related solutions. Before starting the discussion on the results, we briefly outline the localization procedure of LEMON.

We consider the deployment setup presented in Section 4 to generate a radio map of the monitored area during offline. The radio map consists of the fingerprints to be explored during online localization. After receiving a localization request, the localization algorithm presented in Section 3 is used to choose K nearest neighbors from the radio map to predict the location. Note that the above is a general localization approach usually followed by any localization schemes. The difference of our approach with the existing ones lies in the system design, the localization algorithm, and the RSS noise mitigating techniques.

5.1. The impact of profiled samples and pegs density

As the first step of the evaluation of LEMON, we analyzed the impact of the number of fingerprints and pegs on the estimation accuracy. Initially we considered a single room and a 8×8 grid to gather a varying number of fingerprint samples using 4, 5, and 9 pegs. The outcome is shown in Fig. 3. It is apparent that the higher the density of the samples the better the accuracy. The same statement is also true for the pegs. This is expected as with a higher density we capture the signal strength variation better, which is reflected in the estimated accuracy. However, in case of K, the best result is obtained with $K=6$. Selecting more neighbors deteriorates the performance because of adding distant neighbors that may not ensure good proximity in the signal space due to multipaths. A small number of K is not useful either as we may have lack of proximity information. Thus choosing the right number of nearest neighbors, fingerprint samples, and infrastructure nodes (pegs) all have significant impact on the localization performance and must be tuned as the first step of the fingerprint based localization.

5.2. The effect of RSS discrepancy and averaging bias

As part of the estimation process, we first choose a number of *close* fingerprint samples from the signal space, and then average the coordinates of the respective samples into the estimated coordinates of the tag. The notion of closeness assumes some metric in the space of vectors consisting of RSS readings. The most natural choice for this metric is straightforward Euclidean distance; however, we have also tested other candidates listed in Table 1.

The observed difference in performance is not very pronounced. The average error observed under different sample selection metrics is shown in Fig. 4(a). The best performance is observed for the Euclidean metric.

In the last step of our location estimation procedure, the coordinates of the selected samples are averaged to produce the

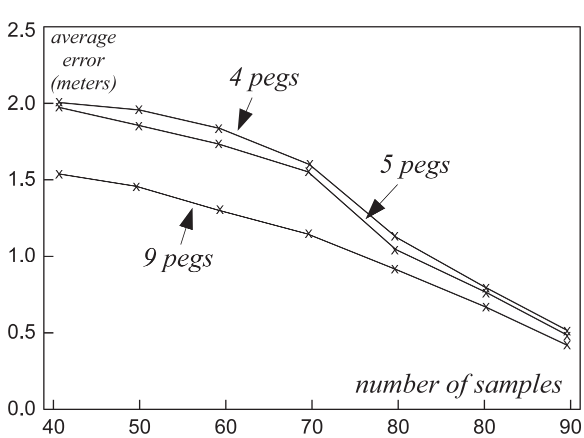

Fig. 3. The impact of profiled samples and pegs density.

Table 1
RSS vector proximity metrics.

Metric	Formula
Euclidean	$\sqrt{\sum_{j=1}^{m}(R_\Omega(j)-R_\Psi(j))^2}$
Manhattan	$\sum_{j=1}^{m}\lvert R_\Omega(j)-R_\Psi(j)\rvert$
Supreme	$\max_{1\le j\le m}\lvert R_\Omega(j)-R_\Psi(j)\rvert$
Lorentzian	$\sum_{j=1}^{m}\ln(1+\lvert R_\Omega(j)-R_\Psi(j)\rvert)$

estimated locations of the tag. It is possible to define various averaging formulas for this purpose (in addition to the one used for LEMON). In the following paragraphs such averaging techniques are outlined:

$$x_e = \frac{\sum_{i=1}^{K} w_i \times \left(\sum^* x_j\right)}{(K-1)}, \quad y_e = \frac{\sum_{i=1}^{K} w_i \times \left(\sum^* y_j\right)}{(K-1)}, \tag{3}$$

where $w_i = d_i/D$ and $(\sum^* x_j)$ is the sum of all K x_j coordinates except x_i. Note the difference between formulas (2) and (3) (we call it LEMON2). Let S be a database sample and (x_i, y_i) be the coordinates. Let d_S be the distance of the RSS vector of S from the RSS vector of the tag's readings. The LEMON formula includes (x_i, y_i) in the averaged estimate with a factor directly equal to the difference between d_S and the maximum d over all fingerprints selected for the averaging (d_{max}). In particular, the sample with $d_S = d_{max}$ will be completely ignored from the final result (its factor will be zero). The sole purpose of that fingerprint is to provide the maximum value of distance, i.e., d_{max} to be used for weighting the contribution of fingerprints from the remaining ones. On the other hand, with formula (3), all the coordinates of all selected fingerprints contribute to the final average. The relative distance of a fingerprint is used as factor weighting the contribution of all other fingerprints, i.e., the further the fingerprint's RSS vector is from the tag's readings, the more all the remaining fingerprints should count in the final estimate.

In the LANDMARC (Ni et al., 2004), the authors suggest averaging formula:

$$(x_e, y_e) = \sum_{i=1}^{K} w_i(x_i, y_i), \quad \text{where } w_i = \frac{1/d_i^2}{\sum_{i=1}^{K} 1/d_i^2} \tag{4}$$

where the distance exponent (equal to 2) is in principle a parameter. This formula amplifies the impact of closer fingerprints with respect to formulas (2) and (3). Of course, one can apply arbitrary functions to the distance, preferably the ones preserving the monotonicity of factors (i.e., closer fingerprints having larger contributions to the average). In particular, this formula proposed in Otsason (2005) (we name it exponential):

$$(x_e, y_e) = \sum_{i=1}^{K} w_i(x_i, y_i), \quad \text{where } w_i = \frac{e^{-bd_i}}{\sum_{i=1}^{K} e^{-bd_i}} \tag{5}$$

where $b > 0$ amplifies the distance exponentially.

We have tested all the above formulas, and the differences in performance turned out to be marginal. The results are compared in Fig. 4(b).

5.3. The effect of obstacles

In this series of experiments we studied the impact of obstacles placed in the monitored area. We considered three scenarios deployed in the monitored area:

1. No obstacles.
2. Four stacks of chairs placed symmetrically in the four centers of the grid quadrants.

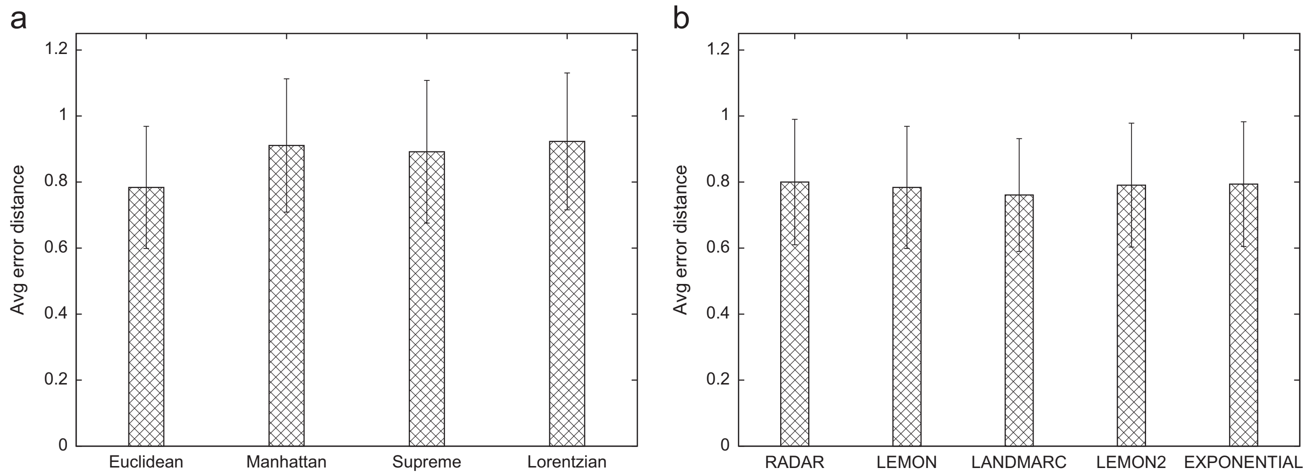

Fig. 4. The effect of distance metrics (a) and averaging factors (b).

3. A large table made of wood and steel placed in the center, with a large metal object placed on top of the table.

The fingerprinting stage was carried out in an empty room. The idea was to see the performance of our localization scheme in a situation when the layout of the monitored area has been changed after the fingerprinting. In particular, one would like to know in what circumstances the fingerprinting data should be deemed obsolete, calling for a new radio map.

Somewhat surprisingly, the disturbance caused by the obstacles turned out to be quite minimal (see Fig. 5). In particular, the piles of chairs cause absolutely no reduction in accuracy.[2] The impact of the table (including a significant amount of metal) is more perceptible and results in a 3.8% increase in the average error. We further conducted experiments in even complex multiroom environment to investigate the impact of obstacles on the accuracy.

After the initial set of experiments presented above, we carried out experiments in three adjacent rooms and gathered 52 fingerprint samples to test 33 tag locations. RSS scaling was used to control the impact of RSS noise, i.e., giving weight to the captured RSS proportional to their proximity. The best results for LEMON in the multiroom case were obtained with $K=6$, scaled RSS, and scaling factor $\gamma=3.5$. To show the error versus structure (impact of obstacles) relationship in the multiroom space, we present the error distribution of LEMON in Fig. 6, where the highest estimation error is observed near the wall separating the rooms. Indeed worst error is observed near a steel file-cabinet beside the wall. This gives us a hint that the performance of RSS based localization is impacted by the indoor structure and furniture.

5.4. The impact of human presence

In this experiment, the sending tag was carried by a person and data was gathered from four directions for both fingerprinting and location measurement. Localization was carried out with and without considering the direction, i.e., in the former case all the fingerprints were considered to estimate the location of a tag in spite of its direction, whereas, in the latter scenario only the fingerprints that align with the direction of the tag were used. The estimation results are shown in Fig. 7. It is apparent from the results that the presence of human body has impact on the localization performance, which could be alleviated by a trick like the orientations of the tag. However, the

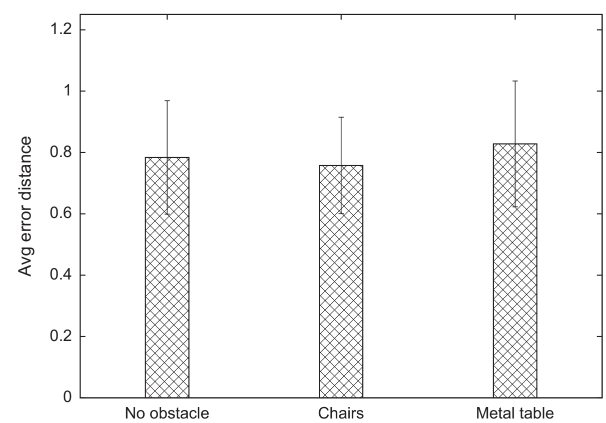

Fig. 5. The effect of obstacles.

orientation improves the accuracy by 12 cm, which might not be worth incorporating into the localization design considering the associated fingerprinting overhead to capture the directions. Indeed, localization may require special hardware to capture smooth orientations (instead of 4 discrete directions). This leads to an extra cost and energy consumption and goes against the low-cost low-power philosophy of LEMON.

5.5. The impact of different localization schemes

In this section we consider the RSS fingerprinting based localization that relies on KNN search. We then compare their accuracy. Note that different localization schemes consider different hardware and setup that make the comparison process tedious. However, we may consider our experimental data to compare the KNN searching approach. For instance, RADAR and precise (Martin et al., 2010) consider a single nearest neighbor as the predicted location, whereas LANDMARC uses weighted average of the chosen K neighbors, where the averaging technique differs from LEMON. In Hossain et al. (2013) RSS preprocessing is used to overcome the constraints on the number and the locations of the APs (4 APs to cover $30\,\text{m}\times 18\,\text{m}$ area). In particular pairwise RSS differences for a given set of APs are used to expand the dimension (from m to $(m/2)$) of the stored RSS vectors of the fingerprints.

Figure 8(a) presents the accuracy comparison of different KNN based schemes. It is apparent that higher accuracy can be obtained with $K>1$, LEMON and LANDMARC are better than RADAR and precise. Indeed the averaging technique of the chosen K neighbors has

[2] We could even see a slight improvement which has to be attributed to a statistical fluctuation.

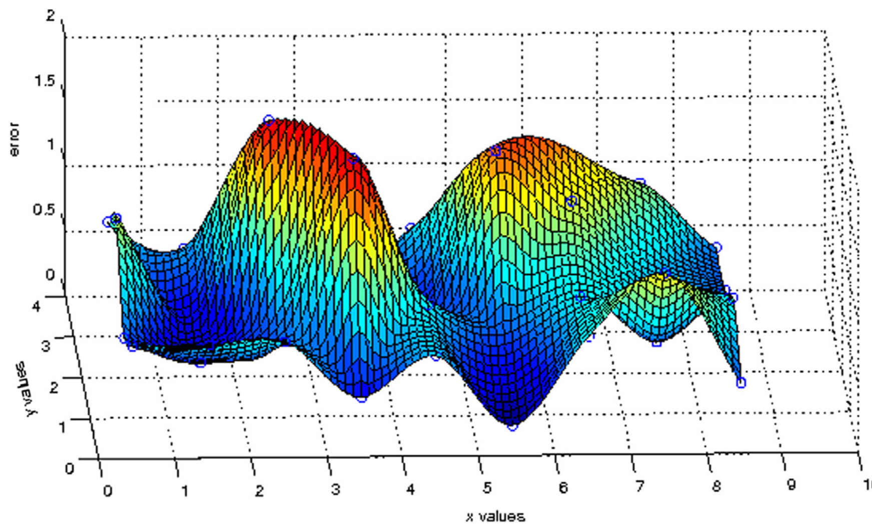

Fig. 6. The error distribution of LEMON.

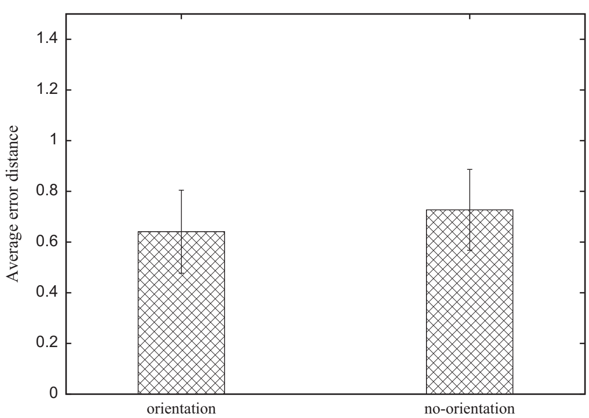

Fig. 7. The impact of human presence.

impact on the accuracy as LEMON is better than LANDMARC for the same K. Finally, we observe that the dimensionality expansion might be helpful with limited number of infrastructure nodes covering a large area (like in Hossain et al., 2013), but deploying a large number of low-cost low-power devices yields good estimation without restoring such technique. We have already shown that the higher peg density improves the accuracy. In Fig. 8(b) we again present the impact of the density of the pegs to show that LEMON ensures high accuracy because of dense population of low-cost low-power pegs.

Thus the flexible deployment of LEMON (using a large set of pegs to render an accurate radio map) is the key reason behind its high accuracy. Furthermore it uses an effective averaging technique on the chosen K neighbors to report the predicted locations. Note that high density of pegs does not introduce interference as the pegs are tuned to the lowest possible transmission power. However, the low-power cheap pegs introduce noise in the perceived RSS, which is even worse indoors due to multipaths. Thus in the next section we discuss the behavior of LEMON in the presence of noisy RSS, i.e., its robustness and introduce solutions to take care of such noise.

6. Study of robustness of LEMON

In this section we consider various ways in which the inherent unreliability of RSS impacts the localization and mitigation techniques that can be applied to deal with such shortcomings.

6.1. The impact of outlier RSS values

We consider a closer inspection of the collected RSS measurements with the purpose of discarding the ones that are likely less useful. Eyeballing the collected RSS values (see Fig. 9) one can identify three “regions,” two at the extreme ends, and a midrange of values that are related to distance. RSS values are ranging from -42 to -110 dBm, where the high values fall around -70 dBm, low ones around -95 dBm, and the rest are midrange RSS values. We observe that the low and the high values can be unreliable compared to the midrange values, i.e., they are unrelated to the corresponding physical distances between sender and receiver. For example the points that fall below -100 dBm in Fig. 9 appear to be outliers (e.g., at distances 4 m and 5 m) and so appears to be the point above -60 dBm at distance 1 m. The low RSS values are not far from the noise floor, while the high RSS values denote a peg that is probably close enough to the transmitting tag and its receiver may be saturated by the impact of nearby transmission. To mitigate the effect on localization of noisy RSS from the “unreliable” extremes of the range of values, we introduce a two-step localization process. First, we compute the cross correlation coefficient c between RSS and distance for each peg:

$$c = \frac{\sum_{i=1}^{n}(r_i-\overline{r})(d_i-\overline{d})}{\sqrt{\sum_{i=1}^{n}(r_i-\overline{r})^2\sum_{i=1}^{n}(d_i-\overline{d})^2}} \tag{6}$$

where n, r, and d are the number of fingerprint samples, the RSS value, and the known distance between a peg and the fingerprint data, respectively. The value of c is between -1 and 1, and the smaller the value the weaker the correlation between the distance and the RSS. We notice that, for each peg, c can be further improved by trimming off the extremes of the RSS values for the pegs. This simply implies that those two extremes of the range introduce noise to the captured RSS values. We therefore isolate the “midrange” values of RSS as the most useful, while replacing the extremes with two representative values (one for the high end and one for the low end values).

Thus we first compute the coefficient c for each peg and then attempt to trim t% of RSS values from both ends of the RSS range for that peg. We eliminate outliers by starting with the furthest RSS extremes and trimming performed in such a way that the value of c improves. The trimmed values are replaced as we mentioned them above. This is done (instead of discarding the trimmed RSS completely) because we would not like to distort the dimensionality of the measurements. Clearly, we do not

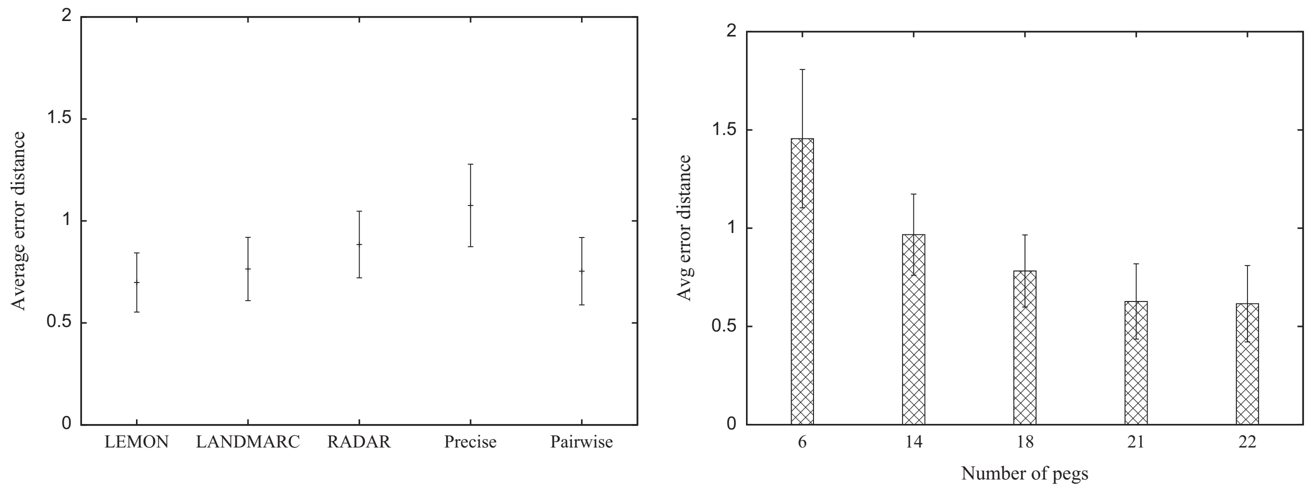

Fig. 8. The accuracy comparison and the impact of the density of the infrastructure nodes.

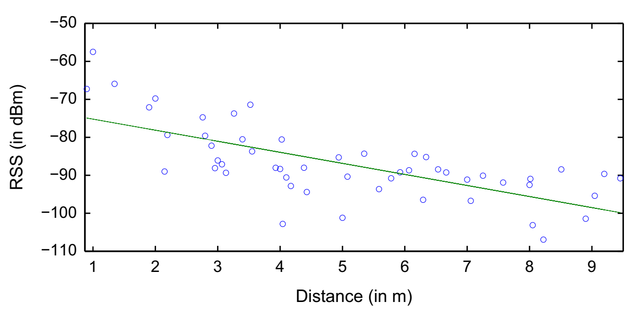

Fig. 9. The RSS vs. distance for peg 11.

perform any trimming of the RSS values for the measurements of the tag that is querying about its location because it would have been dubious at best given that we do not know its distance from the pegs. For each peg it is possible to find its best c by trimming any percentage of RSS values from either end. However, this also means that different pegs may loose different percentage of RSS values and may have different impact on localization. Thus we trim up to a fixed percentage of RSS values such that we are not loosing much but at the same time are able to avoid extreme RSS as well as improve c. Finally, we perform localization using the newly constructed association list of the fingerprints.

The performance of midrange RSS based localization is shown in Fig. 10(a). Note that we used unscaled RSS in midrange as well as in the next section on peg reliability. The reason of doing so is to capture the relationship between RSS and absolute distance; whereas, if we perform scaling, we lose that RSS vs. distance relationship. We start with 5% trimming, then move to 10%, and finally consider 20%. For 5% trimming localization accuracy remained almost the same. However, it improved when trimming was increased to 10%. Further trimming though makes the localization results worse. Thus there is a trade-off between improving c and localization accuracy. If we trim more RSS values c improves but that also decreases the localization accuracy. Trimming very small percentage also does not help as we still may have outlier RSS values in our set of measurements. We found 10% is a good choice to improve the localization accuracy.

We compare the midrange localization performance of LEMON with LANDMARC and RADAR. All three schemes experience performance improvement when consider midrange localization. The approach seems promising in terms of improving localization accuracy, where we need to reconstruct the association list of each profiled sample based on the cross correlation coefficient of each peg and trimming the unwanted RSS values. An open question is that of determining a systematic RSS measurement trimming/rejection strategy that prescribes what needs to be trimmed automatically, and takes the appropriate action.

6.2. Reliability assessment of pegs

Another way to study the RSS data set is to attempt to characterize the reliability of the pegs without resorting to any assumption of a relation between distance and RSS value. RSS measurements, in addition to the contamination by multipath propagation, may also be affected by the hardware unreliability of the infrastructure nodes. We have mentioned before that our infrastructure nodes (pegs) are low-cost and low-power devices, which may introduce unreliability on the measured RSS. Here, we will attempt to introduce metrics to measure the reliability of the pegs. We use the correlation between the measured RSS and the corresponding distance between the transmitter and the pegs based on the fingerprints. However, instead of using the original RSS, we consider midrange values to compute the correlation coefficient, c_m, and use it as a reliability measure. The higher the coefficient the better its reliability. This measured reliability factor can be used to control the influence of a peg in the RSS discrepancy as follows:

$$(RSS_{ti} - RSS_{fi})^2 \times w_i \tag{7}$$

where RSS_{ti} and RSS_{fi} are the RSS values from the association list of a tag and a fingerprint, for peg i, respectively. We consider the following weight w_i for a peg i:

$$w_i = \frac{c_{m_i}}{\sum_{i=1}^{m} c_{m_i}}.$$

The accuracy of localization based on the peg reliability is tested using both original and midrange RSS. We found that the midrange RSS helps to further improve the accuracy if peg reliability is introduced. However, the original RSS along with peg reliability does not help. This may demonstrate that trimming noisy or unreliable RSS is a good idea to improve the localization accuracy. The outcome of our experiments is shown in Fig. 10(b). The performance of LEMON with original RSS is not improved, it rather deteriorates a bit under consideration of peg reliability. However, accuracy experiences improvement if midrange RSS values and peg reliability is used.

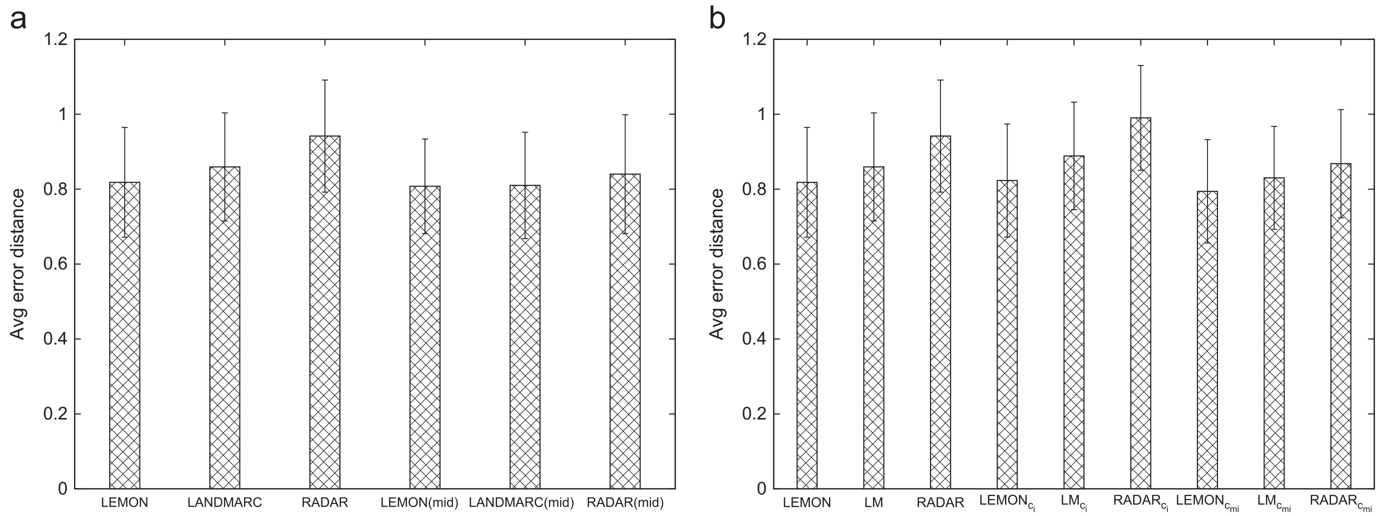

Fig. 10. The midrange RSS based localization (a) and the effect of peg reliability (b).

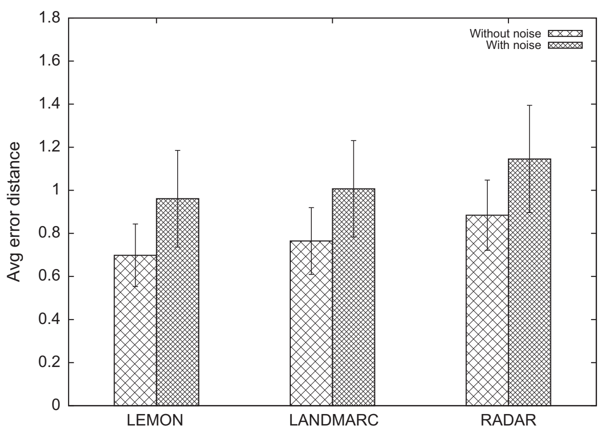

Fig. 11. Effect of noisy RSS on localization.

6.3. *The impact of malfunctioning pegs*

In the presence of multipath propagation, RSS may correlate with distance in an unpredictable way. While the primary role of profiling is to capture *any* correlation that might exist between RSS and locations in space, the ultimate case of a difficult environment occurs when there is no apparent correlation at all. Clearly, the total lack of any correlation whatsoever renders all attempts to use RSS for location estimation futile; however, it still makes sense to ask this question: how resistant an RSS-based scheme is to the partial loss of correlation. Such a loss may result from a particularly malicious propagation properties in a certain region of the monitored area (e.g., lots of metallic objects), or by a malfunctioning peg. The latter problem may become particularly relevant in systems with a large numbers of cheap nodes, as the ones targeted by LEMON.

To test the sensitivity of our scheme to such properties of the environment, we have performed an experiment whereby a selected peg (one such peg considered at a time) would produce uniformly distributed random RSS readings between *MIN* and *MAX* (minimum and maximum value of RSS). We have tested 33 localization queries over all possible scenarios in our setup involving one peg (a different peg in each scenario) misbehaving in this fashion. As the same problem can affect any RSS-based scheme, we have subjected LANDMARC and RADAR to the same test. Note that the implementation of those schemes was in a sense virtual, and more favorable than of their original versions, as they operated on our hardware (with the same coverage by pegs as LEMON), the only difference being their way of transforming the RSS data into location estimates. Needless to say, all schemes suffered a drop in the accuracy of their estimates (see Fig. 11) but the relative order in terms of performance, with LEMON outperforming LANDMARC and RADAR, was preserved.

7. Conclusion

We have presented a RSS fingerprint based indoor localization system named LEMON. It is an infrastructure based approach where the surveyed area is crowded with a dense collection of low-power low-cost wireless sensors. We performed a series of experiments to analyze the performance of LEMON with varying degree of signatures and node density, impact of averaging techniques and discrimination measuring approach, impact of obstacles and methods of localization. The maximum error distance we obtained is less than a meter. We then investigated the robustness of LEMON in the presence of noisy RSS and faulty devices. We also suggested solutions to take care of such noise and faults.

As part of the future work we may design an infrastructure that encounters stair tread while considering multiroom multifloor. Our experiments were carried out in the presence of people in the monitored area; however, a continuous flow of highly crowded environment might exhibit interesting multipath phenomenon of the RSS. Thus performance evaluation of LEMON under such dynamic environment could be another research direction.

References

Amundson I, Sallai J, Koutsoukos X, Ledeczi A, Maroti M. RF angle of arrival based node localisation. Int J Sens Netw 2011;9(3/4):209–24.

Bahl P, Padmanabhan VN. RADAR: an in-building RF-based user location and tracking system. In: Proceedings of the IEEE international conference on computer communications (INFOCOM 2000). Israel: Tel Aviv; 2000. p. 775–84.

Chen Y, Lymberopoulos D, Liu J, Priyantha B. FM-based indoor localization. In: Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys'12). Lake district, UK; 2012. p. 169–82.

Chung J, Donahoe M, Schmandt C, Kim I, Razavai P, Wiseman M. Indoor location sensing using geo-magnetism. In: Proceedings of the 9th international conference on Mobile systems, applications, and services (MobiSys'11). Maryland, USA; 2011. p. 141–54.

Gburzynski P, Olesinski W. On a practical approach to low-cost ad hoc wireless networking. J Telecommun Inf Technol 2008;2008(1):29–42.

Haque I, Nikolaidis I, Gburzyński P. A scheme for indoor localization through RF profiling. In: Proceedings of the international workshop on synergies in communications and localization (SyCoLo'09). Dresden, Germany; 2009.

Hossain M, Jin Y, Soh W, Van HN. SSD: A robust RF location fingerprint addressing mobile devices' heterogeneity. IEEE Trans Mob Comput 2013;12(1):65–77.

Humphrey D, Hedley M. Super-resolution time of arrival for indoor localization. In: Proceedings of the international conference on communications (ICC' 08). Beijing, China; 2008. p. 3286–290.

Liu H, Gan Y, Yang J, Sidhom S, Wang Y, Chen Y, et al. Push the limit of WiFi based localization for smartphones. In: Proceedings of the 18th annual international conference on mobile computing and networking (Mobicom'12). Istanbul, Turkey; 2012. p. 305–16.

Martin E, Vinyals O, Friedland G, Bajcsy R. Precise indoor localization using smart phones. In: Proceedings of the international conference on multimedia. Firenze, Italy; 2010. p. 787–90.

Niculescu D, Nath B. VOR base stations for indoor 802.11 positioning. In: Proceedings of the 10th international conference on mobile computing and networking (MobiCom 2004). New York, NY, USA; 2004. p. 58–69.

Ni LM, Liu Y, Lau YC, Patil AP. LANDMARC: indoor location sensing using active RFID. Wirel Netw 2004;10(6):701–10.

Otsason V. Accurate indoor localization using wide GSM fingerprinting [Master's thesis]. Department of Computer Science, University of Toronto; June 2005.

Patel SN, Truong KN, Abowd GD. Powerline positioning: a practical sub-room-level indoor location system for domestic use. In: Proceedings of the international conference of ubiquitous computing (UBICOMP 2006). California, USA; 2006. p. 441–58.

Sen S, Radunovic B, Choudhury RR, Minka T. Spot localization using PHY layer information. In: Proceedings of the 10th international conference on Mobile systems, applications, and services (MobiSys'12). Lake district, UK; 2012. p. 183–96.

Shen G, Chen Z, Zhang P, Moscibroda T, Zhang Y. Walkie-Markie: indoor pathway mapping made easy. In: Proceedings of the 10th USENIX conference on networked systems design and implementation (NSDI'13). Lombard, IL, USA; 2013. p. 85–98.

Swangmuang N, Krishnamurthy P. An effective location fingerprint model for wireless indoor localization. Pervasive Mob Comput 2008;4(6):836–50.

Wang H, Sen S, Elgohary A, Farid M, Youssef M, Choudhury RR. No need to war-drive: unsupervised indoor localization. In: Proceedings of the 10th international conference on mobile systems, applications, and services (MobiSys'12). Lake District, UK: Low Wood Bay; 2012. p. 197–10.

Wang J, Katabi D. Dude, where's my card?: RFID positioning that works with multipath and non-line of sight. In: Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM (SIGCOMM'13). Hong Kong, China; 2013. p. 51–62.

Yang Z, Wu C, Liu Y. Locating in fingerprint space: wireless indoor localization with little human intervention. In: Proceedings of the 18th annual international conference on mobile computing and networking (Mobicom'12). Istanbul, Turkey; 2012. p. 269–80.

Yoon S, Lee K, Rhee I. FM-based indoor localization via automatic fingerprint DB construction and matching. In: Proceeding of the 11th annual international conference on mobile systems, applications, and services (MobiSys'13). Taipei, Taiwan; 2013. p. 207–20.

Youssef M, Agrawala A. The Horus WLAN location determination system. In: Proceedings of the 3rd international conference on mobile systems, applications, and services (MobiSys'05). Seattle, Washington; 2005. p. 205–18.