Localization using Anonymous Measurements

Niklas Wirström†, Arash Behboodi*, Filip Lemic*, Adam Wolisz*, Thiemo Voigt†
† SICS Swedish ICT
* Dept. of Telecommunication Systems, Technische Universität Berlin

***Abstract*—Range-based IEEE 802.15.4 localization systems currently require relatively high anchor density for indoor deployments. It can therefore be beneficial to use external sources of transmission as additional anchors. We present methods for using WiFi beacons to improve localization accuracy of a range-based IEEE 802.15.4 localization system in cases where only two IEEE 802.15.4 anchor nodes are available. We do this by identifying WiFi beacons from RSSI traces that we dynamically sample online, and applying fingerprinting and range-based methods using the RSSI values of the identified beacons. However, because the data of the WiFi traffic is not decodable by the IEEE 802.15.4 devices, these RSSI measurements lack identifiers that can associate them to specific WiFi Access Points (APs). Therefore, novel methods are required for both fingerprinting and range-based approaches to allow for these additional WiFi APs to be used as anchors. We show by using real-world measurements that our beacon identification method gives a false-positive rate of only 3%, and that if the range measurements to the IEEE 802.15.4 anchors are relatively accurate, with a standard deviation of 1 and 3 m, a localization accuracy improvement of 47% and 24% can be gained, respectively.**

I. Introduction

Interference can be harmful to the performance of RF based localization systems through its influence on signal features used for localization [1]. Nevertheless, signals from interfering systems may also contain information that is useful for localization. In this paper we present methods for identifying WiFi beacons on IEEE 802.15.4 devices, and use the corresponding RSSI values for localization purposes.

This is of high relevance for *range-based* IEEE 802.15.4 indoor localization systems. Such systems currently require relatively high anchor density, due to the fact that at least three anchor nodes are required to produce a unique solution, and that IEEE 802.15.4 radios typically have a limited transmission range in indoor environments. If only two anchors are available, two possible solutions exists, as shown in Figure 1. Moreover, extending the range by increasing the transmission power, potentially results in a high ratio of non-Line-of-Sight (nLoS) measurements which typically have lower accuracy compared to Line-of-Sight (LoS) measurements. Therefore, it can be beneficial to use external sources of transmission as possible *additional* anchors.

However, WiFi signals are not decodable by IEEE 802.15.4 devices and hence the extracted RSSI values are *anonymous*. That is, it is unknown to which APs they correspond. Current range-based and fingerprinting approaches rely on this information. Therefore, to enable the use of anonymous RSSI measurements, new types of localization methods are required. It is unlikely that the anonymous information is sufficient to achieve localization performance comparable to traditional localization solutions by itself. This is because multiple combinations of assigning measurements to anchors can result in seemingly valid location solutions. In this paper, we show that the anonymous measurements can be used in combination with other *non-anonymous* localization measurements of the *principal localization system* to improve localization in scenarios where only two range-based anchors of the principal system are available. To this end, we develop a range-based and a fingerprinting method adapted for anonymous measurements, and use them to select the most probable of the two solutions obtained in such scenarios.

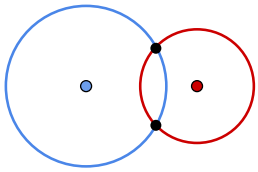


Fig. 1: The two solutions obtained from range-based localization with only two anchors.

Depending on the method used, different types of prior information about the anonymous sources is needed. We refer to this as *environmental awareness* [2]. For range-based methods the locations of the anchors must be known, and some feature of signal must be correlated to the distance to the source. For fingerprinting methods, the only requirement is the existence of a signal feature that is stable over time and that varies in space.

These requirements determine the applicability of the methods with respect to the level of environmental awareness available in a given scenario. However, fingerprinting methods require an extensive calibration phase for learning the fingerprints of different locations, while range-based methods require a much simpler calibration step to estimate different parameters for the transformation of signal features into range estimations. For anonymous localization approaches, however, additional information may be needed. In our case, for the range-based method, we also need to know the channels used by the APs. Moreover, we use a list of relevant beacon-periods to identify the WiFi beacons. This information is provided by the environmental awareness layer in our system, as shown in Figure 2.

To our knowledge, this is the first work that uses anonymous measurements for range-based and fingerprinting methods. Even though the identification of the WiFi beacons, and the extraction of their RSSI values is specific for our approach, the

range-based and fingerprinting methods we develop for combining the anonymous and non-anonymous measurements are generic and work with any anonymous range or fingerprinting measurements.

We evaluate our system using simulated and real-world measurements. In the simulated experiment, we have perfect knowledge of the accuracy of the anonymous measurements, and only two anonymous anchors are used which results in a lower degree of uncertainty than in the real-world experiment.

With real-world anonymous measurements, our beacon identification method gives a false-positive rate of only 3%. For moderate accuracy of the non-anonymous measurements with measurement standard deviation of 3 m, we obtain a localization accuracy improvement of 24% compared to using only the two non-anonymous measurements. For the simulated experiment, the corresponding accuracy improvement exceeds 40%. For high-accuracy non-anonymous measurements with standard deviation of 1 m, the improvement for the real-world and simulated experiments is 47% and 60%, respectively.

The paper continues as follows. In Section II we present related work. In Section III-A, we present our approach for extracting signal features used for localization from WiFi signals that are sampled by IEEE 802.15.4 devices, and Section III-B describes how RSSI measurements can be transformed into range measurements. Sections III-C and III-D describe how measurements from anonymous sources can be used under the two different scenarios outlined above, namely using a fingerprinting approach and a range-based approach. We evaluate the system in Section IV, and finally conclude the paper in Section V.

II. Related Work

The use of anonymous measurements has previously been proposed by Franchi et al. [3] for mutual localization in multi robot systems. In their work they combine the output from a robot's odometer with anonymous bearing and range measurements to other neighboring robots to determine their relative configuration. Cognetti et al. [4] extend this to bearing-only measurements in 3D to enable localization of UAVs. A main difference to our work is that we use range-only measurements and fingerprinting. Moreover, the anonymous RSSI measurement have a much lower level of accuracy than the laser measurements in [3] or the simulated bearings in [4].

The idea that devices of various technologies can be detected by mere processing of the sampled spectrum has already been mentioned in [5] where the authors use hidden structures of signals such as cyclic prefix of OFDM-based technologies to identify different devices. The detected devices and their respective information can constitute a localization system which was implemented in [6]. However a very high sampling rate is needed to reveal these hidden structures. As the transmission period of beacon packets is in the order of milliseconds, e.g. 100 ms, and given the transmission rate, its length is in the order tens of microseconds, we do not need a sampling rate that is as high as the one use by Hong et al. [5].

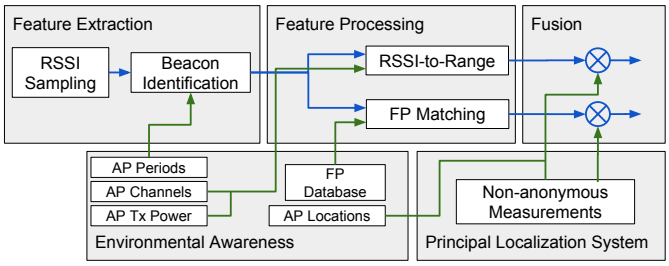


Fig. 2: System diagram.

A major part of our work focuses on the identification of WiFi beacons from sampled IEEE 802.15.4 RSSI traces. This has previously been done by both Hermans et al. [7], and Iyer et al. [8] for the purpose of classifying interference. Hermans et al. use patterns in corrupted IEEE 802.15.4 packets to distinguish between four different types of interference. However, two things make this approach unsuitable for our purposes. Firstly, the crucial information of the RSSI from the interfering WiFi signal is unknown. Secondly, the system does not distinguish between beacons and other WiFi traffic.

On the other hand, the system presented by Iyer et al., which is the inspiring work for this paper, fulfills these two requirements, but is, for other reasons unsuitable for our purposes. We need, for example, higher resolution for the durations of RSSI bursts.

We use *Spray* [9] to fuse the anonymous and non-anonymous measurements. Spray is a multimodal, particle filter based localization system that separates different modalities into different components that are isolated from each other. Particle filters [10] is a widely used approach for highly non-linear localization problems. We implement a new component in Spray for dealing with anonymous range measurements.

III. System Description

The system consists of three phases as shown in Figure 2: A *feature extraction* phase, a *feature processing* phase, and a *fusion* phase. The feature extraction phase is responsible for sampling the wireless medium, and filter out RSSI bursts that correspond to WiFi beacons. The output is the mean RSSI value for each group of beacons believed to correspond to the same AP.

The mean RSSI values are then processed in the feature processing phase. The output depends on the target localization method. In case of fingerprinting, the RSSI values are matched against a database, and the output is a set of location candidates. If the target is the range-based method, the RSSI values are transformed into range estimations.

Finally, the output from the feature processing phase is fused with the non-anonymous measurements from the principal localization system. For this purpose, we use *Spray* [9], a multimodal, particle filter based localization system we have developed earlier. Spray separates different modalities into separate components that are isolated from each other. Spray works in two phases. First a particle generation phase in which each component is allowed to generate particles in areas suggested by their measurements, and then, a particle

evaluation phase in which each component weights all generated particles based on how well they fit their measurements. We implement a new Spray-component for the anonymous range measurements, that we present in Section III-C. In Section III-E we describe the fusion phase.

The first phase is specific to our approach of using WiFi beacons for localization. The second phase is applicable to any RSSI measurements, regardless of the methods of acquisition, and the last phase is applicable to any type of range measurements or set of location candidates.

Depending on the used method, different types of prior information is also needed about the anonymous sources, as described in Section I. These are represented in the *environmental awareness* block. The different types of information are described throughout the succeeding subsections.

A. Feature Extraction

WiFi beacons have the following properties that make them useful for localization: 1) WiFi APs are typically stationary with some exceptions like mobile hotspots. 2) A given AP typically uses a fixed transmission power. 3) Beacons usually have a fixed length for a given SSID (each AP can have multiple SSIDs), and 4) beacons are transmitted periodically. The periodicity is a multiple of 1.024 μs and is typically set to a value around 100 ms. However, periodicity may be broken due to the back-off mechanism of CSMA. Properties 1 and 2 are necessary for the localization data to be meaningful, and properties 2, 3 and 4 can be used to identify the WiFi beacons even if the packets are not decodable.

1) RSSI Sampling: We use an IEEE 802.15.4 based sensor node to sample RSSI traces. We use a threshold of -90 dBm and treat any signal below the threshold as silence. Timestamps are stored for each time the RSSI transits from silence to non-silence and vice versa. The maximum RSSI of each burst, i.e. non-silent period, is also stored.

2) Beacon Identification: Once the sampling phase is finished, the sampled data is off-loaded to a central computer. The data can be compressed before transmission to minimize traffic load and energy consumption. For example, applying a simple dictionary based compression scheme to the RSSI samplings used in Section IV-D, we achieve an average compression ratio of 1.49 in under 100 ms on a sensor node. This is, however, currently not implemented as part of the system.

At the central computer, the relevant information is extracted using a probabilistic approach in which each burst is compared with all other bursts and weighted according to how similar they are in signal strength, duration, and how well the time interval between the bursts fits one of a given set of relevant beacon periods. This information is represented by the *AP Periods* box in Figure 2. In the evaluation of Section IV, the relevant beacon periods are 0.1024 or 0.104448 seconds.

Equations 1, 2, and 3 show how the weights representing the similarity between bursts i and j are computed for duration, signal strength, and periodicity, respectively. $f_{\mathcal{N}}(\cdot)$ is the normal distribution's PDF function, and σ_{dur}, σ_{rssi}, and σ_{per} are standard deviations for the different features. d_i, s_i, and t_i denote burst i's duration, signal strength, and start time, respectively. Q is the set of expected periods. In our case $Q = \{0.1024, 0.104448\}$. In Equation 3, the square brackets denote the *round* function.

$$w_{i,j}^{dur} = f_{\mathcal{N}}(d_i - d_j, \sigma_{dur}) \tag{1}$$

$$w_{i,j}^{rssi} = f_{\mathcal{N}}(s_i - s_j, \sigma_{rssi}) \tag{2}$$

$$w_{i,j}^{per} = f_{\mathcal{N}}\left(\min_{\forall q \in Q}\left(\left|t_i - t_j - \left[\frac{t_i - t_j}{q}\right] q\right|\right), \sigma_{per}\right) \tag{3}$$

We then compute a total weight $w_{i,j}^{tot}$ as the product of the individual weights: $w_{i,j}^{tot} = w_{i,j}^{dur} w_{i,j}^{rssi} w_{i,j}^{per}$, and define a cluster C_i as all the bursts j that with respect to burst i have a total weight higher than a threshold h, i.e, $C_i = \left\{j | w_{i,j}^{tot} > h, \forall i, j\right\}$. For the evaluation in Section IV, we use $h = 0.9$ We chose 0.9 since in our preliminary experiments it gave a good trade-off between false-positive and false-negatives. We remove any duplicated clusters.

Since our main goal is to extract reliable information, and not the identification of beacons per se, it is preferable to have false-negatives over false-positives. Therefore, we use a strict approach that aims to only select the two clusters that are most "useful" and most likely to correspond to beacons. We do this by yet another weighting procedure in which high signal strength, number of bursts in the cluster, and a high mean weight of the bursts of the cluster is rewarded. Clusters with high signal strength are more useful since they are more likely to correctly correspond to the estimated distance because there is a higher probability that the corresponding source is in LoS.

B. RSSI-Based Ranging

We now describe the *RSSI-to-Range* step shown in Figure 2. For the range based approach, once the RSSI values are extracted from the RSSI samplings, they are transformed into range estimations. We use the free space propagation model [11] in Equation 4 to transform an RSSI measurement to a range estimation d.

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{4}$$

In the equation, P_r and P_t are the received and transmitted power, respectively. G_r and G_t are the receiver's and the transmitter's antenna gains, λ is the wavelength, and L is called the system loss factor. Instead of determining these constants individually, we combine them into one single constant K as in Equation 5. The value of K is determined through calibration by collecting a number of RSSI measurements and compute an average value for K.

$$P_r = K \frac{1}{d^2} \tag{5}$$

Because the constant K depends on the transmission power, different values of K must be used if APs use different transmission powers. This is represented by the *AP Tx Power* box in Figure 2. Therefore, each RSSI value is transformed into multiple range estimations, each one corresponding to a

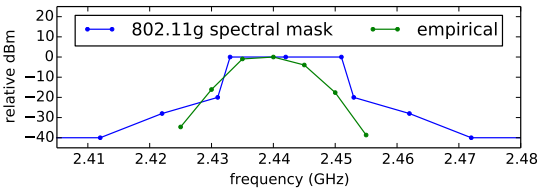


Fig. 3: The measured RSSI depends on distance between the receiving and transmitting channels.

single AP. This does not imply that one needs to know from which AP the signal originates. But the calibration constant K must be known for each AP, and is provided by the environmental awareness layer.

Moreover, the environmental awareness layer must also provide information about the channel used by each AP. This is represented by the *AP Channel* box in Figure 2. This is because the RSSI sample traces contain bursts from all WiFi channels that overlap the 802.15.4 channel used during sampling. The RSSI values observed from an overlapping WiFi channel decrease with the distance between the two channel center frequencies. This constitutes a problem if the channels used by the different APs of interest are relatively close to each other, because a low power signal on a channel close to the sampling channel, and a high power signal on a channel further away can result in similar observed RSSI values on the sampling channel.

However, if the transmission channels of the AP of interest are known, we can compensate the observed RSSI values based on the distance between the sampling channel and the AP channel when constructing the range measurements for a specific AP. That is, if s is the observed RSSI, we can compute the RSSI value, s_i, it corresponds to if the signal originated from AP i with center frequency $f_{c,i}$, as $s_i = s - g(f_{c,sample} - f_{c,i})$, where $g(\cdot)$ is the empirical spectral mask for the used WiFi technology. Figure 3 shows the spectral mask specified by the IEEE 802.11g standard, and relative RSSI values measured using a IEEE 802.15.4 node on different channels for WiFi transmissions on channel 7.

C. Using Anonymous Range Measurements

Once the RSSI values are transformed to range measurements, we need to solve the following problem: Given a set of anonymous measurements and a set of possible anchors to which the measurements correspond, select the most probable of the two localization solutions provided by the principal localization system.

Figure 4(a) gives an intuition of the problem. The blue dots and circles represent non-anonymous anchor nodes that are part of the principal localization system, and their corresponding measurements, respectively. Yellow dots indicate the location of anonymous anchors. The green and red dots represent the correct and incorrect solutions provided by the principal localization system. Finally, the green and red circles around the anonymous sources represent the two anonymous measurements. Green is used for the measurement that corresponds to the source at the center of that circle. As the example

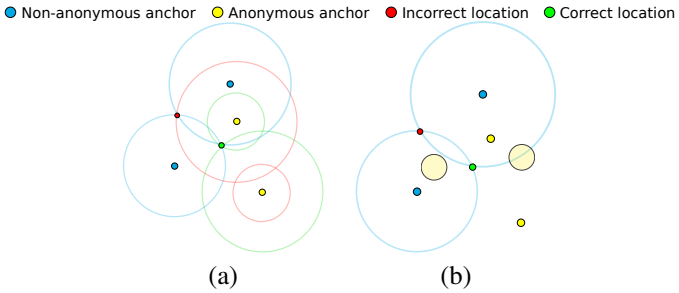


Fig. 4: A simple example of using anonymous ranging (a) and anonymous fingerprinting (b) to resolve a case where too few non-anonymous anchors are available.

in the figure illustrates, if measurements are perfect, the correct location is likely to fit with a higher number of measurements.

We implement a new *anonymous-range* component in Spray which we now describe. The component is used in the fusion process described in Section III-E.

As previously mentioned, Spray works in two phases: a particle generation phase, and a particle evaluation phase. The particle generation phase for the anonymous-range component is straight-forward: For each anonymous measurement, the new component generates particles in a circular cloud with radius corresponding to the measurement, around each of the anonymous anchors.

The goal in the particle evaluation phase is to assign weights to each particle in a way such that the particles near the most likely location obtains the highest weights. This phase is more complex and can be done in different ways.

Solving this optimally requires running the particle filter once for each possible combination of measurement - anchor assignments. If there are n anchors and m range measurements, there are $\frac{n!}{(n-m)!}$ different such combinations. It can be seen as the number of ways to select m from n without repetition, which is $\frac{n!}{m!(n-m)!}$, and for each such combination find all $m!$ permutations. Already with modest settings such as $n = 5$ and $m = 3$, there are 60 such combinations. Hence, this rapidly becomes prohibitively computationally expensive.

Instead, each particle is assigned a weight from the measurement-anchor combination that agrees most with the particle. That is:

$$w_p = \max_{\forall i, \forall j}(f(r_j - d_{i,p}), \tag{6}$$

where r_j is range measurement j, $d_{i,p}$ is the distance between particle p and the anonymous anchor i, and $f(\cdot)$ is the weighting function. This is a greedy solution, and the result can be inconsistent in the sense that anchors can be assigned multiple measurements, and multiple anchors can be assigned the same measurement. This approach is of the order $\mathcal{O}(mn)$.

As weighting function, we use the PDF, $f_{\mathcal{N}}(\mu, \sigma_a)$, of the normal distributions with $\mu = r_j - d_{i,p}$. The value of σ_a determines the level of importance given to the anonymous measurements, and should, on one hand, reflect the measurement standard deviation (i.e. the expected accuracy of the measurements), but on the other hand also take into account

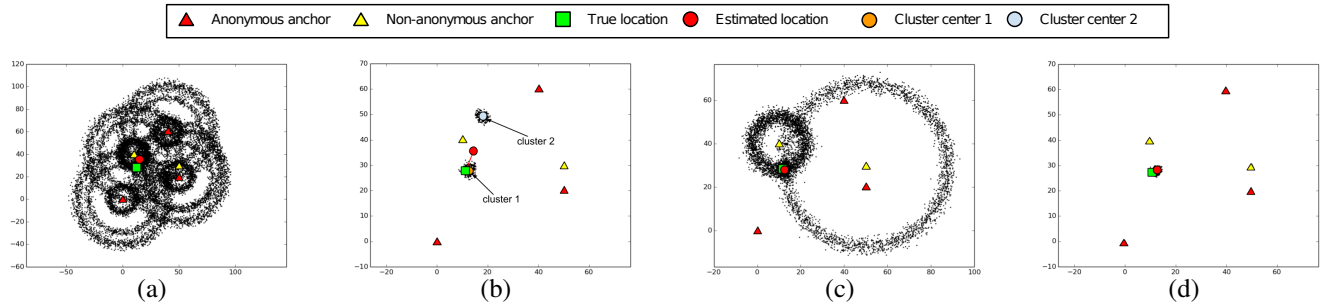


Fig. 5: An illustration of a successful case for fusing two anonymous range measurement with anonymous information to select one of the two possible solutions: (a) Particle generation for anonymous and non-anonymous measurements. (b) Particle evaluation. Two clusters are formed, but cluster 1 (orange) is heavier because of the anonymous measurements (c) Generation for non-anonymous measurements and a proximity component with center at cluster 1. (d) Particle evaluation. The alternative solution (cluster 2) has been completely abandoned.

the uncertainty resulting from the fact that the measurements are anonymous. The measurement standard deviation can easily be estimated for a given environment, by performing and analyzing a series of measurements. The uncertainty resulting from anonymity is, however, harder to predict. This makes it difficult to determine an optimal value for σ_a. In Section IV, we use $\sigma_a = 6$ which was found to give good results in our preliminary experiments.

D. Matching Anonymous Fingerprints

We now describe the step labeled *FP Matching* in Figure 2. We use the RSSI values from the feature extraction phase as fingerprints. In traditional WiFi fingerprinting techniques, the learning phase consists of collecting RSSI measurements and label them with the IDs of their corresponding APs. At run-time, more labeled RSSI values are collected, and compared to those from the learning phase using a matching function. Finally, the locations associated with the best match is reported [12].

For anonymous measurements, however, the identifying labels are not available, and each learned fingerprint consist of a set of unlabeled RSSI values associated with a specific location. Similarly, the run-time fingerprint for a specific location is also a set of unlabeled RSSI values. We use a simple method for comparison between the learned and the run-time fingerprints. The RSSI values of both the run-time fingerprint and the learned fingerprints are sorted and compared so that the highest run-time RSSI value is compared to the highest learned value, and so on. The distance d_i between the run-time fingerprint r and the fingerprint l_i, learned at the ith location is computed as the squared Euclidean distance given by Equation 7, where n is the number of RSSI values output by the beacon extraction phase.

$$d_i = \sum_{k=0}^{n} (l_{i,k} - r_k)^2 \tag{7}$$

Then we select the two locations with the smallest Euclidean distance d_i, and create *proximity components* in Spray with center points according to these locations, as illustrated by the yellow discs in Figure 4(b). The proximity component generates particles in a disc shaped cloud around its center point, and weights particles based on how close they are to this center. If the resulting locations are accurate, the proximity component will cause the particles at the true location to have higher weights. These proximity components are then used in the fusion process described in the next section.

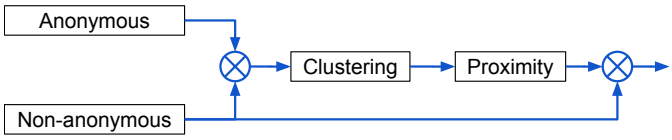


Fig. 6: The localization process. Anonymous and non-anonymous measurements are fused, the heaviest cluster is selected, and finally the non-anonymous measurements are fused and higher weight is given to the particles close to the selected cluster.

The fingerprinting comparison approach in Equation 7 has a drawback, because there is no guarantee that the learning and runtime phases will detect beacons from the same APs. However, as explained in Section III-A, the beacon identification phase is biased towards selecting clusters with high RSSI, which increases the probability of selecting clusters corresponding to the same APs compared to if two clusters were to be selected at random. We show in Section IV that even with this simple matching approach we can improve localization accuracy.

In this approach we have not exploited the fact that different APs in the same environment often are configured to use different channels to minimize congestion. In such cases, the channel can be used to de-anonymize the beacons to some extent.

E. The Fusion Process

Here we describe the fusion phase shown in Figure 2, and illustrated in further detail in Figure 6. This phase is identical for the anonymous ranging and fingerprinting methods, with the difference that the anonymous range component described in Section III-C is used for the ranging method, while the proximity components described in Section III-D are used for the fingerprinting method. The box labeled *Anonymous* in Figure 6 represents either one of these components, while *Non-anonymous* represents Spray's *range* component that is used for the non-anonymous measurements. Spray is run once with the anonymous and non-anonymous components. This generates a number of particles that correspond to both types of measurements. Figure 5(a) shows an example for

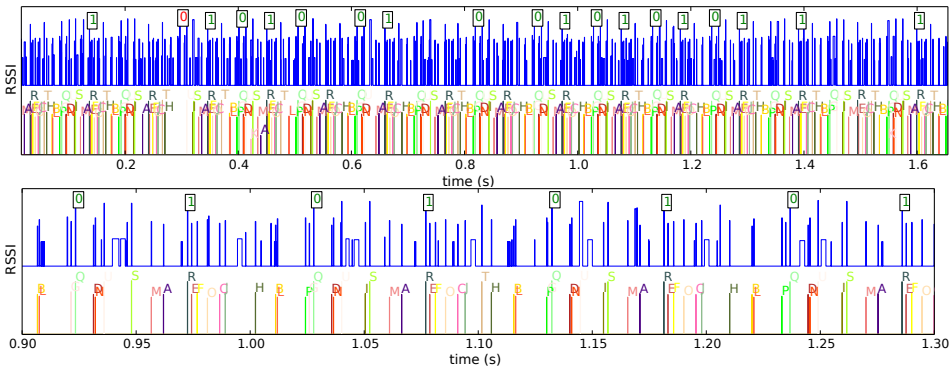


Fig. 7: Bursts classified to different clusters compared to ground truth. The bottom figure is a zoomed version of the top figure. Blue bars are sampled RSSI values. Colored bars represent WiFi beacons from sources indicated by letters. The red zero represents a burst incorrectly classified as belonging to cluster 0. Green numbers represent bursts correctly classified to the cluster corresponding to the given number.

the anonymous range method. Then each particle is weighted with respect to all components, and the particles with the highest weights are selected. If there are only two non-anonymous anchors available, this typically results in two clusters of particles, which both agree equally well with the non-anonymous measurements. This is shown in Figure 5(b). However, the cluster that agrees more with the anonymous measurements will have a slightly higher weight sum than the other. We use a clustering algorithm to identify the two clusters, and we compute the weight sum for each of them. The cluster with the highest weight sum is selected and a proximity component with same center as the cluster is generated.

The proximity component is then combined with the non-anonymous component, and Spray is run a second time, as shown in Figure 5(c). Finally, the outcome is, ideally, a single cluster of particles, as shown in Figure 5(d). The two-step approach is used to increase the probability that the final estimated location is one of the two possible solutions given by the non-anonymous measurements. Using only the first step can easily result in that the final estimated location is different from these two solutions, in cases where multiple anonymous sources are available.

The level of importance given to the proximity component in the second run is determined by a constant σ_p that, due to the same reasons as for σ_a, also is difficult to determine optimally. In Section IV-D, we use $\sigma_p = 5$ which was found from performing preliminary experiments.

IV. Evaluation

In Section IV-A, we present results regarding the beacon identification phase. In Section IV-B, we evaluate the accuracy of the range measurements formed from the RSSI values obtained from the beacon identification phase.

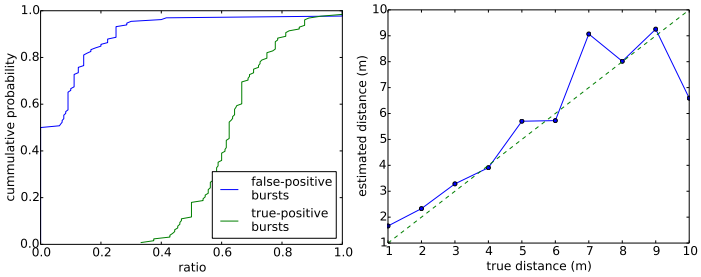
In sections IV-C and IV-D, we evaluate the benefit of using anonymous measurements to find a unique positioning solution in the case when only two non-anonymous anchors are available. In Section IV-C, we perform a simulated experiment to find a baseline of the improvement that can be obtained by using anonymous ranging measurements. In Section IV-D we perform a semi-real-world experiment in an office environment. In this experiment, all anonymous measurements are real, but the two non-anonymous measurements are simulated. This approach is chosen because we want to decouple the results from any specific deployment or technology of the non-anonymous anchors. It turns out that the outcome depends considerably on the location of these anchor relative to the target device, and on the accuracy of the non-anonymous range measurements. Using a fixed real setup would give results for only a single instantiation of the problem.

A. WiFi Beacon Identification

We record over 100 different 802.15.4 RSSI traces in an office environment, and use the approach described in Section III-A to extract WiFi beacons. To get the ground truth we sample the WiFi medium using two WiFi devices on neighboring channels simultaneously to the 802.15.4 sampling. We then align the two logs to get a mapping between WiFi packets and bursts sampled by the 802.15.4 radio. Figure 7 shows an example of this. The bottom figure is a zoomed version of the top figure. The blue lines correspond to the bursts sampled by the sensor node and the multi-colored bars below correspond to WiFi beacons sampled using the two WiFi devices. Beacons of the same color (and letter) correspond to the same AP. The numbers 0 and 1 above the blue bursts indicate the cluster they belong to. Cluster 0 corresponds to AP *Q*, and cluster 1 corresponds to *R*. The red "0" indicates that the corresponding burst has been miss-classified to belong to cluster 0.

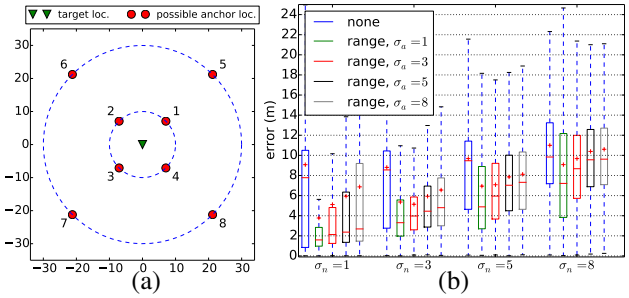
As mentioned in Section III-A, we select the two best clusters with respect to the weighting mechanism, to use for localization. There are two different types of false-positives that can occur. One type is *false-positive clusters*, which are clusters in which the bursts not corresponding to any beacons is the largest group. The other type is *false-positive bursts* which are incorrectly clustered bursts. We consider a burst to be correctly clustered if it corresponds to a WiFi beacon from the AP that has the greatest support in that cluster. That is, the AP to which the largest group of bursts in that cluster corresponds. Similarly, a burst is incorrectly clustered if it corresponds to a beacon from a different AP or to no beacon at all.

False-positive clusters are more severe than false-positive bursts because they will result in range measurements that



(a) (b)

Fig. 8: (a) CDFs for false-positive and true positive bursts in the beacon identification step. (b) Range estimation based on sampled WiFi beacons in LoS.



(a) (b)

Fig. 9: (a) The evaluated anchor locations relative to the target location, for the simulated experiment. (b) Simulation results for different standard deviations of the anonymous (σ_a) and non-anonymous (σ_n) measurement errors. Red "+" indicate mean values.

potentially do not correspond to any AP at all. For the false-positive bursts, however, because of the clustering mechanism used, bursts belonging to the same cluster are guaranteed to be very similar in signal strength which is the feature used for localization. A false-positive burst will therefore not have any greater impact on the final localization accuracy.

Figure 8(a) shows the CDF for the false-positive burst ratio of a given cluster. That is, the ratio of the number of incorrectly classified bursts in a specific cluster to the total number of bursts in that cluster. Approximately 50% of the clusters have no false-positive bursts, and in 3% all bursts are false-positives. The latter occurs when the bursts not corresponding to any beacons is the largest group in a cluster, that is, for a false-positive cluster. Hence, we have a 3% false-positive cluster rate.

The CDF for the ratio of true-positive bursts is also shown in the figure. This is the ratio of the correctly classified bursts in a cluster, to the total number of bursts that correspond to beacons from the AP with the greatest support in that cluster. More than 60% of all the bursts in a sampling are found in 50% of the cases. In the worst case, 30% are found, and 100% are found in 2% of the cases.

B. RSSI Based Ranging

We perform a micro benchmark in a LoS environment to investigate the accuracy for using the sampled bursts RSSI values as range measurements as described in Section III-B. We collect measurements every meter up to 10 m away from a single AP. We calibrate using a leave-one-out approach, that is, to estimate the distance to a specific measurement location, all other measurements are used for calibration, except the ones corresponding to that location. Figure 8(b) shows the result. The accuracy is higher for shorter distances. The mean error is 0.8 m, and the standard deviation is 1.03 m. The average value of K as given in Equation 4, is 0.00055.

C. Simulation Experiment

In this experiment, we use simulation to evaluate the improvement of using anonymous range measurements for selecting the most plausible of the two solutions obtained when only two non-anonymous anchor nodes are available. As opposed to the semi-real-world experiment in Section IV-D in which the anonymous APs have fixed locations, here we also want to evaluate how the result varies depending on the locations of these APs. In addition to the anonymous AP locations, we also evaluate for different configurations for non-anonymous anchor node locations, and different measurement variances for both the anonymous and non-anonymous measurements.

Only two anonymous APs are used in the simulations. Typically, in a real situation, as in the experiment presented in Section IV-D, more anonymous APs are available which leads to higher uncertainty of which measurement corresponds to which AP. Therefore, the results from this simulated experiment should be seen as upper limits for the improvement of using the proposed anonymous ranging approach for the evaluated configurations.

For each simulation run, we locate the two anchor nodes at one of 8 different positions indicated by the red dots in Figure 9(a). These are four different locations, $\frac{\pi}{2}$ radians apart, at two different distances from the target location: 10 and 30 m. We use the same procedure to select the anonymous AP locations. For the non-anonymous anchors, we do not evaluate the case when both anchors have the same angular configuration, to avoid the degenerate case in which both anchors have the same position. We use four different values (1,3, 5, and 8 m) for the standard deviations σ_n and σ_a used to simulate non-anonymous and anonymous range measurements, respectively.

Figure 9(b) shows a standard box plot for all different configuration of σ_n, and σ_a, with a red "+" indicating the mean error for each setting. For high-accuracy anonymous and non-anonymous measurements with $\sigma_n = \sigma_a = 1$ m, the benefit, with respect to the mean errors, is approximately 60% for combining non-anonymous measurements with anonymous ranging, compared to using only the non-anonymous measurements. In the figure, these are labeled *range* and *none*, respectively. For $\sigma_n = \sigma_a = 3$ m and $\sigma_n = \sigma_a = 5$ m, the benefit is approximately 40% and 20%, respectively. For low-accuracy measurements with $\sigma_n = \sigma_a = 8$ m, the benefit is negligible. That the benefit decreases when σ_n increases is intuitive, because if the error is high for both solutions given by the non-anonymous measurements, it matters less which one we chose. This is also illustrated in Figure 13(b), in which e_n is the mean error from using only non-anonymous measurements, and e_b, and e_w are the errors if we always select the best and the worst solutions, respectively. As can be seen in the figure, the difference between e_n and e_b decreases

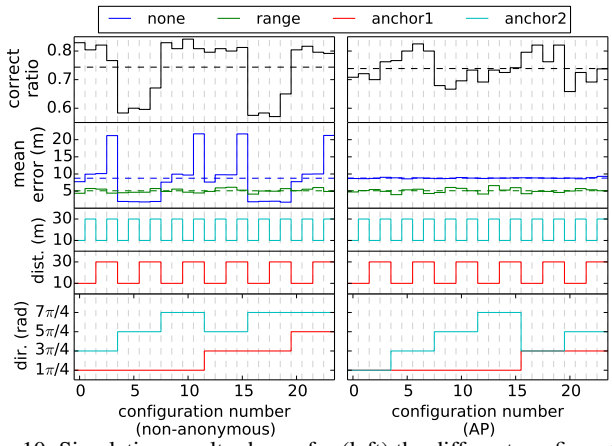


Fig. 10: Simulation results shown for (left) the different configuration of the non-anonymous anchors, averaged over all configurations for the anonymous anchors, and (right) the different configuration of the anonymous anchors, averaged over all configurations for the non-anonymous anchors.

with increasing σ_n. The results in Figure 13(b) is obtained by analytically computing the two possible solutions for a high number of random positions for both anonymous and non-anonymous anchors.

The accuracy of the non-anonymous measurements (σ_n) has a higher impact on the estimation error than the accuracy of the anonymous measurements (σ_a). The reason for this is that the final position estimation is, in general, one of the two solutions given by the non-anonymous measurements.

Figure 10 show the results for a measurement standard deviation of $\sigma_n = \sigma_a = 3$ m, sliced in two different ways. The left graph shows the results for each evaluated location configuration of the two non-anonymous anchors, averaged over all AP configurations, while graph to the right shows the results for each anonymous AP configuration, averaged over all non-anonymous anchor configurations.

In the left graph, the anchor location configuration is shown in the three bottom graphs of the figure as the angle and distance from the target location. For example, for configuration number 10, anchor 1 is located 30 m away in direction $\frac{\pi}{4}$ from the target position, and anchor 2 is located 10 m away in direction $7\frac{\pi}{4}$. This corresponds to anchor locations 5 and 4 in Figure 9(a). In the right graph, the three bottom graphs shows the AP configuration in the same way.

The top graphs in the two figures show the average probability of selecting the correct cluster for each configuration. The dashed lines indicates the average overall ratio. We define the correct cluster to be the one that is closest to the true target position. In the left graph of Figure 10, there are two low-probability regions in which the probability decreases to between 55 and 65%. This corresponds to the cases when the difference between the two anchor nodes' angular configurations is π radians. In these cases the target node lies on a straight line between the anchors, and a unique solution is actually available. The resulting particles are, however, still divided into two clusters by the localization algorithm. Because the two clusters are relatively close to each other, it is difficult to select the correct one. However, because the clusters are close to each other, there is only a limited impact on the localization error, if the wrong one is selected. This can be seen in the second graph from the top in which shows the estimation errors for running the simulation with and without the use of anonymous measurements, labeled *range* and *none*, respectively. When anonymous range measurements are used, the estimation error has low variance for all anchor configurations, also in the low-probability regions. In this region, however, not using the anonymous measurements results in a smaller error. Therefore, detecting the cases when the two anchor nodes and the target are positioned on a straight line, could further improve the results. This is possible by computing the sum of the two non-anonymous range measurements. If the sum is close, or equal to the distance between the two anchors, then the target must be on a straight line between the two. This has, however, not been implemented in our system.

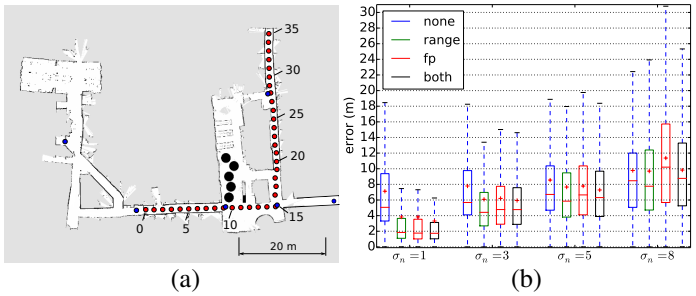
The results also show that, when not using anonymous measurements, the errors depend on the anchors' distances to the target location when the non-anonymous anchor nodes are not π radians apart. For the error peaks of approximately 20 m, both anchors are at a distance of 30 m from the target, and the lowest error for a given angular configuration is found when both anchors are at a distance of 10 m. This is expected because the distance between the two solutions resulting from the non-anonymous measurements increases with the anchor-to-target distances, unless the target is on a straight line between the anchors.

In the right graph of Figure 10, the ratio for correct cluster selection has smaller variance than in the graph to the left. This means that this ratio is less sensitive to the locations of anonymous anchors than to that of the non-anonymous anchors. The highest probabilities of selecting the correct cluster are found when the APs are π radians apart. The lowest errors for anonymous ranging are, as expected, found when the two APs are π radians apart *and* the target-to-AP distances are short. Naturally, if anonymous measurements are not used, the error is not affected by the AP locations.

The small variances of the mean errors for *range* in Figure 10 show that the performance of anonymous ranging is stable with respect to the positions of both anonymous and non-anonymous anchors. This is not the case if anonymous measurements are not used, as in the left graph. However, the errors in the two figures are, as previously mentioned, averages over the anonymous anchor configurations, and the non-anonymous anchor configurations, respectively. The individual errors have a much higher variance, as Figure 9(b) shows.

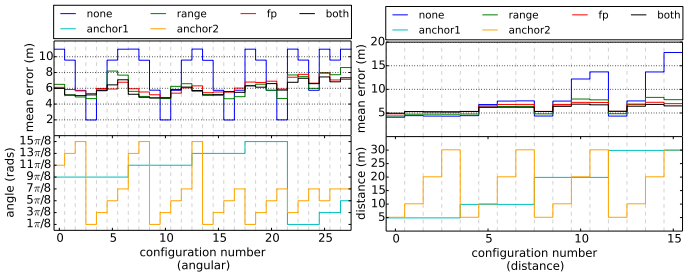
D. Semi-Real World Experiment

We perform a real world experiment in an office environment. The purpose is to evaluate the entire system described in Section III in a real environment. Figure 11(a) shows the evaluation points of the experiment as red dots. The points are 2 m apart. The experiment is carried out by letting a mobile robot [13] visit the evaluation points in order and collect measurements using an STM32W sensor node [14], equipped with an external antenna. The sensor node samples RSSI as



(a) (b)

Fig. 11: (a) Evaluation points (red) and anonymous WiFi APs (blue) for the real-world experiment. (b) Estimation error for all different standard deviations of the non-anonymous measurements (σ_n) for the real-world experiment.



(a) (b)

Fig. 12: (a) Estimation error per angle, averaged over all distances and locations for the semi-real-world experiment, and (b) per distance, averaged over all angles and locations for the semi-real-world experiment.

described in Section III-A. The timestamps are taken using a 24 MHz clock, and the sampling frequency is approximately 1.5 MHz. The sensor node has 16 KB of RAM that is shared with the Contiki OS [15], and can therefore only store a limited number of bursts at the same time. The total sampling time therefore depends on the current traffic load. The sample traces are transferred via USB to a central computer for processing.

The robot uses information from its odometers and a Kinect 3D camera to localize itself. The ground truth is given by the predefined evaluation points, which means that there is a built-in error due to inaccuracies in the robot's own localization mechanism. This error is in our experiments in the order of up to 1 m. At each evaluation point the medium is sampled on three different 802.15.4 channels which overlap with the channels of the APs of interest. The APs are shown as blue dots in Figure 11(a). In addition to these APs, other APs from neighboring floors are present on the sampled channels. These APs have unknown locations and cannot be used for localization. Due to the anonymity, however, it is impossible to differentiate these beacons from the beacons of interest. We also construct a second data set by repeating the procedure at each location in between the evaluation points shown in Figure 11(a). This data set is used as a database for the fingerprinting method.

Rather than deploying *non-anonymous* anchors throughout the evaluation area, we use simulation to generate range measurements to imaginary anchor node locations. This approach is chosen so that we can evaluate multiple settings for both the anchors relative locations to the evaluation points, and the accuracy of the non-anonymous range measurements. As for the simulation experiment in the previous section, the benefit of using the extra information obtained from the anonymous sources largely depends on these parameters.

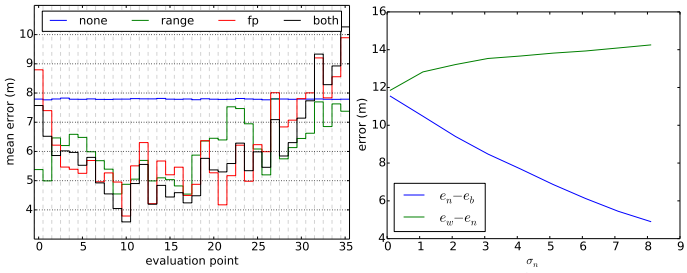
For each evaluation point, we use 448 different anchor-location configurations, and evaluate this for four different values (1, 3, 5, and 8 m) on the measurement standard deviation σ_n that controls the simulated accuracy of the evaluation-point-to-anchor measurements. In each of the 448 different anchor-location configurations, each anchor can be at a distance of 5, 10, 20 or 30 m away from the evaluation point, and in one of the 8 different directions $\frac{\pi}{8}$, $3\frac{\pi}{8}$, $5\frac{\pi}{8}$, ...,$15\frac{\pi}{8}$. As in the simulation experiment, we have excluded the case where both anchors have the same angular configuration.

Figure 11(b) shows the average error for the four different values of σ_n. These are computed over all combinations of evaluation points, distances and angular configurations for the non-anonymous anchors. The curve labeled *none* represents the case where only non-anonymous measurements are used, and the curves labeled *range*, *fp*, and *both* represent the case when non-anonymous measurements are combined with anonymous range, anonymous fingerprinting, and both, respectively. For $\sigma_n = 1$ and $\sigma_n = 3$, the average improvement with respect to the mean errors, for using any combination of anonymous localization, is 47% and 24%, respectively, compared to only using non-anonymous measurements. The benefit is negligible for $\sigma_n = 5$ and $\sigma_n = 8$. For the latter, using anonymous fingerprinting actually decreases performance. This is due to many high errors for the evaluation points 25 through 35 shown in Figure 11(a). This is further elaborated on at the end of this section.

Figures 12 and 13 show the cases for which $\sigma_n = 3$ m. The top graph of Figure 12(a) shows the estimation errors for a given angular configuration for the two non-anonymous anchor nodes, averaged over all evaluation points and evaluation-point-to-anchor distances. The bottom graph shows the values for the angular configurations. As in the simulated experiment in the previous section, it is only in the special case in which the difference of the directions of the two anchor nodes is exactly π radians, that anonymous information impairs the localization accuracy. This happens for angular configurations 3, 10, 16 and 21 in Figure 12(a).

Similarly, Figure 12(b) shows the estimation errors per evaluation-point-to-anchor distances averaged over all evaluation points and angular configurations. Using anonymous information improves the accuracy when both evaluation-point-to-anchor distances are greater or equal to 10 m. The highest benefit is obtained for long distances. When both anchors are 30 m away from the target, the absolute accuracy improvement is approximately 10 m. This can be contrasted to the accuracy decrease for short distances, which is approximately 0.5 m. Moreover, cases where anchors are close to evaluation points can easily be identified directly from the range measurements, and in such cases, the use of anonymous range information can be avoided.

Figure 13(a) shows the errors per evaluation point, averaged over all distance and angular configurations. Because the sim-



(a) (b)

Fig. 13: (a) Estimation error per location, averaged over all angles and distances for the semi-real-world experiment. (b) The difference of errors when using only non-anonymous measurements (e_n), and always selecting the best solution (e_b), and the worst (e_w), respectively

ulated non-anonymous range measurements are independent of the actual location, the average errors for using only non-anonymous measurements have minimal variance. By comparing the evaluation point indices to those given in the map in Figure 11(a), we see that, for all the anonymous localization methods, the estimation errors are in general lower for evaluation points with higher AP density. This is expected for the range based method, because the accuracy of the anonymous range measurements generally decreases with distance. For the fingerprinting method, the reason can be that the probability increases that the beacon extraction phase outputs two beacon clusters corresponding to the same AP. For example, the anonymous measurements obtained for evaluation points 25 through 35 suffer from this problem. This makes any two locations at similar distance from the closest AP have similar fingerprints. This also explains that the fingerprinting method, when $\sigma_n = 8$, results in worse performance than using only the non-anonymous measurements (Figure 11). Because if the fingerprinting method, more often results in that the "wrong" solution is selected, then the difference between the errors of using only non-anonymous measurements and combined with fingerprinting increases with σ_n. This is illustrated in Figure 13(b) by the line labeled $e_w - e_n$. Here e_w is the error resulting from always selecting the worst of the two solutions given by the non-anonymous measurement, and e_n is the error when using only the non-anonymous measurements.

V. Conclusion

We have proposed a method for extracting RSSI values corresponding to WiFi beacons, from RSSI traces sampled using IEEE 802.15.4 devices. These RSSI values lack identifiers as to which source they originate from. To this end, we have proposed new range-based and fingerprinting methods for using such anonymous measurements to for localization purposes. We have shown that our beacon identification method is able to identify groups of beacons with a false-positive of only 3%, and that the extracted beacon RSSI values can, with relatively high accuracy, be transformed to range measurements for distances up to, at least 10 m. We have evaluated our localization methods under the assumption that two non-anonymous range based anchors are present, and we use the anonymous measurements to distinguish between the two possible solutions resulting from the two non-anonymous measurements. We show with both simulation and real-world experiments, that the benefit of combining anonymous and non-anonymous measurements largely depends on the non-anonymous anchors' relative location to the target position, and on the accuracy of their corresponding range measurements. For relatively accurate non-anonymous range measurements with a standard deviation of 3 m, we obtain a 24% and 40% improvement in the real-world and simulated experiments, respectively.

VI. Acknowledgement

This work has been partly performed within the FP7 EVARILOS project (grant agreement 317989).

References

[1] A. Behboodi *et al.*, "Interference effect on localization solutions: signal feature perspective," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81th*, 2015.

[2] J. M. Rabaey *et al.*, *Connectivity brokerage - enabling seamless cooperation in wireless networks*, A White Paper, OtherTechnicReportsTKN, Unpublished article, 2010. [Online]. Available: http://www.tkn.tu-berlin.de/fileadmin/fg112/Papers/papers_all/rabaey10connectivity_brokerage_enabling.pdf.

[3] A. Franchi, G. Oriolo, and P. Stegagno, "Mutual localization in multi-robot systems using anonymous relative measurements," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1302–1322, 2013.

[4] M. Cognetti *et al.*, "3-d mutual localization with anonymous bearing measurements," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, IEEE, 2012, pp. 791–798.

[5] S. S. Hong and S. R. Katti, "Dof: a local wireless information plane," in *Proceedings of the ACM SIGCOMM 2011 Conference*, ser. SIGCOMM '11, Toronto, Ontario, Canada: ACM, 2011, pp. 230–241, ISBN: 978-1-4503-0797-0. DOI: 10.1145/2018436.2018463. [Online]. Available: http://doi.acm.org/10.1145/2018436.2018463.

[6] K. Joshi, S. Hong, and S. Katti, "Pinpoint: localizing interfering radios," in *Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation*, ser. nsdi'13, Lombard, IL: USENIX Association, 2013, pp. 241–254. [Online]. Available: http://dl.acm.org/citation.cfm?id=2482626.2482651.

[7] F. Hermans *et al.*, "Sonic: classifying interference in 802.15. 4 sensor networks," in *Proceedings of the 12th international conference on Information processing in sensor networks*, ACM, 2013, pp. 55–66.

[8] V. Iyer, F. Hermans, and T. Voigt, "Detecting and avoiding multiple sources of interference in the 2.4 ghz spectrum," in *EWSN*, 2015, pp. 35–51.

[9] N. Wirstrom, P. Misra, and T. Voigt, "Spray: a multi-modal localization system for stationary sensor network deployment," in *Wireless On-demand Network Systems and Services (WONS), 2014 11th Annual Conference on*, IEEE, 2014, pp. 25–32.

[10] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, IET, vol. 140, 1993, pp. 107–113.

[11] J. Xu *et al.*, "Distance measurement model based on rssi in wsn," *Wireless Sensor Network*, vol. 2, no. 08, p. 606, 2010.

[12] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piché, "A comparative survey of wlan location fingerprinting methods," in *Positioning, Navigation and Communication, 2009. WPNC 2009. 6th Workshop on*, IEEE, 2009, pp. 243–251.

[13] *Turtlebot 2*, http://kobuki.yujinrobot.com/home-en/about/reference-platforms/turtlebot-2/, Accessed: 2015-02-23.

[14] *Stm32w108hb stm32w108cc stm32w108cb stm32w108cz datasheet*, STMicroelectronics.

[15] A. Dunkels, B. Gronvall, and T. Voigt, "Contiki-a lightweight and flexible operating system for tiny networked sensors," in *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, IEEE, 2004, pp. 455–462.