

Determining Node Sequence in a Linear Configuration

M. Onur Ergin, Vlado Handziski, Arash Behboodi, Adam Wolisz
Telecommunication Networks Group
Technische Universität Berlin, Germany
Email: {ergin,handziski,behboodi,wolisz}@tkn.tu-berlin.de

Abstract—Many indoor positioning applications focus on determining the location of a device in a particular area, like a room of a building, rather than its metric coordinates. It is common that most buildings are partitioned in more or less a regular way and discovering the relative positions of the nodes can give us an accurate estimate of the space, e.g. a room, they are located in. In this study, we have developed a probabilistic methodology to detect the sequence of wireless sensor nodes placed in an indoor environment. We assume no preliminary training, preconfiguration or infrastructure other than the identity of one reference node at the beginning of the sequence.

We trigger the nodes under consideration to transmit a series of packets, and use received signal strength information as observed by the neighbors to infer about the sequence of the nodes. In this paper, a centralized processing of the information is used. The proposed methodology is verified both by simulation and by experiments in a building with regular rooms, high shadowing and multipath distortions. Rarely observed classification errors in the experiments have motivated some extension to the basic methodology.

I. INTRODUCTION

Today we are witnessing a rising interest in indoor location based services motivating a huge interest in research for indoor localization. It has been attempted with a large plethora of different technologies, based on usage of radio waves (like narrow- or ultrawide-band signals, acoustic or visual signals). Both combinations of ranging with multilateration and landmark based approaches are being investigated. The individual approach differs strongly in requirements for special infrastructure, energy consumption, cost, and amount of preparation on a given site in which the localization might be utilized.

Depending on the type of application, different level of location accuracy will be desired. One specific class of requirements is namely the space or room level accuracy. In this case, the precise position of the object is not really relevant, but what matters is the fact that within which space(room) the object is. Despite being very natural, let us note that this scenario is only at the first glance simple. To define the room, in which an object is located, from the knowledge of the estimated position and the floor plan, one should achieve the accuracy of the thickness of the wall, especially close to the walls. This is however very demanding. While range-free algorithms, using just “visibility” of some signals/landmarks, might be very useful for this purpose, they mostly require very

dense deployment of signal sources, detectors and landmarks.

In this study, we make an attempt to tackle the indoor position discovery problem by using only common communication radio chips that are available on typical wireless systems like IEEE 802.15.4. Our claim is that we can achieve considerable level of room-level accuracy by processing the available Received Signal Strength (RSS) information provided along with the regular communication of the devices, without leveraging additional infrastructure or intensive calibration and training steps. It is widely agreed within the research community that determining a path loss coefficient and converting RSS values into inter-node distances will not provide reliable information, especially indoors due to the shadowing and multipath effects that disassociate the measured signal strength from the actual distance. This phenomena and the experienced received signal strength measurements versus distance, are further discussed and illustrated in Section IV.

In this paper, we are addressing a specific class of the room level localization problem. We assume that rooms are ordered linearly, and in each room there is an object – such as a container – having a source of radio signals. Such containers will likely be positioned in a similar place in each room. What we would like to know in this case is, in which room a specific one of these containers is located. For this purpose we attempt to make a decision to find out which order is the object of interest in the sequence that maps to the room plan. Therefore, rather than working with a classical localization method, we developed and validated a sequence identification approach.

We propose a probabilistic node sequence discovery (PNSD) solution for the indoor positioning problem for the cases in which precise room-level accuracy is desired. For processing, we take measurements at each channel that the radio chip provides. This diversity in frequency and time provides us sufficient diversity in RSS measurements. We show that approaching the problem probabilistically, the correct classification of the node sequence is possible.

The rest of the paper is structured as follows: Our approach is explained in Section II. In Section III, we describe the experiment design and performance evaluation, followed by Simulation results in Section IV. In Section V we propose a supplementary algorithm to assess the reliability of the results. We provide a short analysis of complexity and overhead. At the end we summarize the related works in Section VII and conclude in Section IX.

II. THE APPROACH

In this study, we focus on discovering positions of a set of nodes, relatively to each other, on a single dimensional setting (i.e. in rooms along a corridor). We denote the set of potential positions (i.e. rooms) with R_1 to R_n . Any adjacent pair of nodes might be separated by a wall or alike. Each position is considered to contain one of n nodes, whose presence is to be discovered and denoted as N_i . First position in the setting contains the *Reference Node*(N_1) and positions of the remaining nodes are not known. We assume that all the positions are occupied or unoccupied positions do not need to be discovered.

The nodes are assumed to be stationary. The system first collects RSS data. Each node transmits a limited number of packets at each channel available to the radio chip at a relatively high power, increasing the chances of correct reception. Each other node saves the RSS information from the correctly received packets. At any time during the measurement process the transmitting node is referred as the Sender while all other nodes are referred as Receivers.

Here we desire diversity in RSS measurements by benefiting from diversity in signal propagation. The sender node transmits t packets in c channels. All other nodes which successfully receive these transmissions keep a record of the sender, channel, transmission number and measured RSS. After all nodes have assumed the sender role and the measurements have completed, the collected data is sent to the processing center. All this data is then used to generate a probability tree of the nodes, root of which is the reference node.

Generating Probability Tree:

The tree is initiated by placing the reference node (N_1) at the root position. The children of this node in the tree will be the receivers of N_1 , with probability weights. The weight is the observed probability of being the closest node. That is, in other words, the ratio of the number of times that the highest RSS value being measured at each transmission to the total number of transmissions across all channels, for all the receivers of a sender. For example, out of 100 transmissions, if node N_2 measures the highest RSS value 60 times and N_3 measures 40 times, then the weight of the edge between node N_1 and N_2 is 0.6 ($P(N_2|N_1) = 0.6$), whereas the weight of the edge between node N_1 and N_3 would be 0.4 ($P(N_3|N_1) = 0.4$). Other receivers will not be placed if they do not measure the highest RSS value for any transmission, therefore $P(N_x|N_y) = 0$. For each added vertex(node) to the tree, this procedure is repeated and the tree is accordingly extended until the depth of the tree is equal to the number of all nodes in the system. The procedure is shown in Algorithm 1.

Here, each vertex of the tree represents a node and each level of the tree represents a potential position in the sequence. If an edge between two vertices V_a (representing node N_a) and V_b (representing node N_b) exists and $level(V_a) < level(V_b) - 1$ then the weight of $Edge(V_a, V_b)$ is the probability of node N_b being at the next position after node N_a , $P(N_b|N_a)$, at

Algorithm 1 Probability Tree Building Algorithm

- Step 0:** Initialize tree T
 - Step 1:** For each node N_i in the sequence
 - Step 1.1:** Measure RSS from multiple frequencies on demand or during normal communication
 - Step 2:** Collect measurements from each receiver
 - Step 3:** Place reference node at the root (T_0) of the tree
 - Step 4:** For each leaf (T_x) of the tree
 - Step 4.1:** Add all receivers of T_x except if they are already in the path $T_0 \rightarrow T_x$
 - Step 4.2:** Assign probability weights to new added edges
 - Step 5:** Repeat Steps 4, 4.1, 4.2 until depth of T reaches the total number of nodes
 - Step 6:** Find the path (\vec{S}^*) with maximum joint probability from T_0 to any of the leaves T_z
-

that level. So the joint probability of a path S_i , $P(S_i)$, is the product of all the weights of edges from T_0 to T_z . Our verdict is the most probable path (sequence) S , which is the path S_i that has the maximum joint probability.

$$\begin{aligned}
 & i \in 1 \rightarrow z \\
 & \vec{S}^* = \{S_1, S_2, S_3, \dots, S_z\}; P(S_1) > P(S_2) > \dots > P(S_z) \\
 & S = \text{argmax}(P(\vec{S}^*)) \\
 \Rightarrow S &= \text{argmax}\left(\prod_{y=T_0}^{T_z} P(V_y|V_{y+1})\right)
 \end{aligned}$$

III. EVALUATION

We have tested the presented probabilistic sequence determination method on the measurements obtained from TWIST testbed [8]. The nodes were exposed to different types of wireless interference, mostly IEEE 802.11g/n and bluetooth traffic, and the experimentation environment has had relatively high density of metal structures.

As measurement hardware, TmoteSky platform is used, which is equipped with Chipcon CC2420 2.4 GHz radio transceivers [3] with 80 MHz bandwidth spread over 16 channels. We selected ten nodes, uniformly positioned roughly along a line, on the fourth floor of the TWIST building. The nodes are 3 meters apart and there is one wall partition in between each adjacent node pair. The setting is illustrated in Figure 1 and we call this set of nodes ‘‘Set-A’’.

Each measurement cycle was composed by all ten nodes, in turn, transmitting $t = 40$ beacon packets with 40ms intervals on each channel $C \in \{11 \text{ to } 26\}$, while the remaining nodes were logging the RSS information from the received packets. All nodes have synchronously switched channels after every round of 40 transmissions, which takes approximately 2 seconds. The full measurement cycle, involving sending 40 packets from each of 10 nodes in the sequence, across all 16 channels, takes approximately 4.3 minutes. First, we have

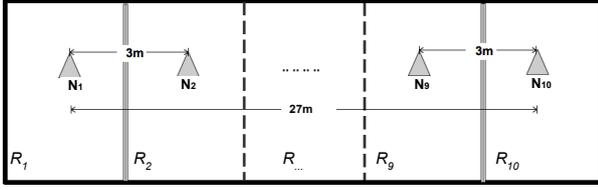


Fig. 1. Schematic representation of the experimental setup, Set-A

performed 2000 repetitions of the measurement cycle over the span of roughly 6.5 days.

All of the measured RSS measurements has been collected using the wired backchannel that is provided by TWIST testbed. This backchannel has also been used to synchronize the channel switching during the experiments. Use of backchannel for synchronizing, however, is not required for running this algorithm. Although not included in this paper, in our untethered tests, we flood a control message among the nodes when it is time to change channel. Total time to receive this packet, change channel and relay this control packet is about 10ms. To reflect this, our wired backchannel experiments also use a 10ms time window to allow all the nodes in the network to change their channels synchronously.

The ten nodes (N_1, N_2, \dots, N_{10}) that we have used in our experiments have been static. When applying our algorithms, we have assumed that N_1 as the reference node, whose position was known to be the first of the sequence. We also repeated our computations a second time, using the reverse order, assuming that N_{10} being the reference node with known position and the rest being at unknown places. Note that, when we're looking for the closeness relationship among nodes, we consider the transmissions of one-to-many, in other words from one sender to multiple receivers. Hence, reversing the computation direction means that the order of senders and receivers are different. So it is a different data set and the results will not be related. We refer these different directions of the same sequence of nodes as Left-to-Right and Right-to-Left in this paper. All these RSS readings are used for generating a weighted probability tree. For improving the computational performance, we have pruned the tree during the runtime of the algorithm, which we elaborate in section VI.

For evaluating, we have compared our algorithm to RSS based ranging and Multidimensional Scaling (MDS). We have used the smacof implementation of MDS [4], since it gave better results than the standard implementation of MDS in R-project. As input to MDS, we converted the RSS values to distance using path loss attenuation equation with path-loss coefficient $\alpha = 3$.

For RSS ranging, we used the greedy approach: starting from the reference node, we added the next closest node to the sequence and iterated this for each last added node until all nodes are placed in a "greedy" manner. As input, we have taken the average of the values collected in all channels (Greedy Avg) and on single channels 11, 18 and 26 (Greedy

AvgCh{11,18,26}) separately. This method is explained in [6] in more detail.

We have initially tested our algorithm on the set of 2000 measurements. We have repeated the computations for both directions, namely Right-to-Left and Left-to-Right. In one direction our methodology (PNSD) has achieved 100% success, while in the other direction it has achieved around 93.6% success, which was influenced by external parameters to a small extend. We discuss this later in this paper. In this scope, "success" means **perfectly recognizing the correct sequence**. In Figure 2 you can see how PNSD outperforms alternative methods.

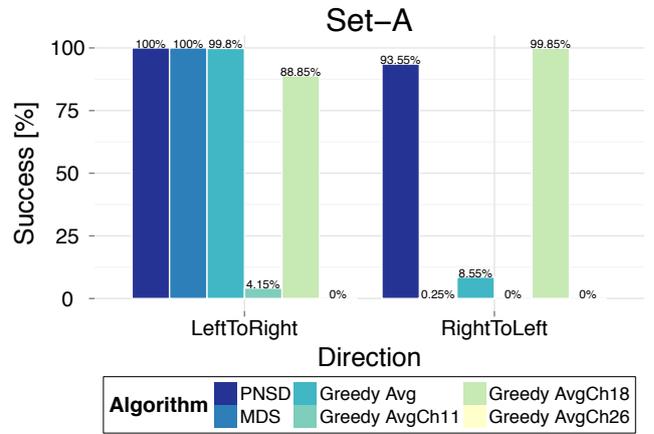


Fig. 2. Success on Set-A

Although being a very good position estimator for range-based localization, MDS could not perform good enough for discovering sequences with RSS information. In one direction it has given 100% success but in the reversed direction it almost never succeeded to generate the correct node sequence. This shows that MDS is too sensitive to the asymmetry of the wireless channel. Although it has been used widely [12], averaging RSS measurements on a single channel does not give us satisfactory results. On channel 11 (Greedy AvgCh11), 4.15% and 0.25% of the time correct sequence were detected using the Greedy approach and on channel 26 (Greedy AvgCh26) the success has been 0% in either of directions. On the other hand, channel 18 (Greedy AvgCh18) has lead to 88.85% and 99.85% correct sequence detection. This shows that the single-channel approaches can sporadically lead to good results, but they have limited practical value because the optimal channel changes with time. The Greedy approach with channel diversity (Greedy Avg) has been quite successful (99.8%) on direction Left-to-Right while it has shown very little success (8.55%) on direction Right-to-Left, which shows us that even channel diversity is not enough by itself for highly successful sequence discovery.

A. Analysis of imperfect results

All the results of PNSD that did not end up in perfectly correct sequence had single position error, which means two nodes are swapped in their estimated positions. The analysis of the results shows that these errors were localized at all cases, namely sixth and seventh nodes falsely appeared in each other's places. We have then examined the exact times at which the RSS measurements were taken for the non-perfect experiments. The measurements were taken around christmas time, which is an official holiday in where TWIST is located. We can see that the measurements which were taken outside of the official holiday time, and were more likely to be interfered, have a localized swap. The other part of the measurements, which fall within the christmas break time have been determined with perfect accuracy. We have plotted the indices of failed experiments in Figure 3.

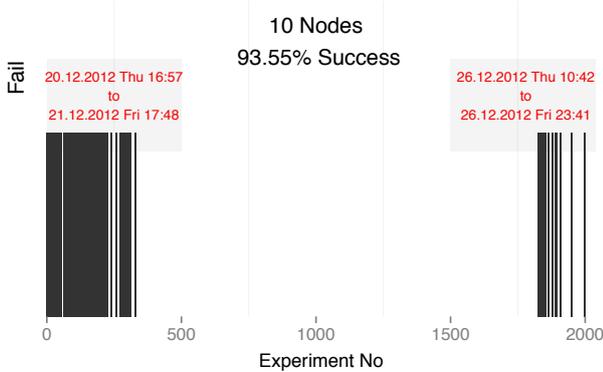


Fig. 3. Failed Experiments

The building we have used in our experiments has a rectangular structure and it has offices on its north and south sides. The above given measurements were taken from the south side, which faces the rest of the campus area. We have also taken measurements from the north side, which faces an open space, with a set of 6 nodes, Set-B. Due to the configuration of the environment, we expect reduced interference effects on this side of the building, as opposed to the campus facing side. Since the building has very similar features for both node sets, the measurements from Set-B hint us that in the absence of the severity of interference that we are facing in our facilities, the above proposed method works flawlessly. We have done 692 (time-limited) experiments and had 100% success, which is shown in Figure 4.

North side has minimal interference in comparison to south side of the facility. This is also improving other methodologies, such as MDS has 100% success in both directions. Greedy Average was about 87% successful in Left-To-Right direction and 100% successful in the other direction. Greedy sequence building on single channels again vary drastically, eg. on channel 18 correct detection was 99.71% and on channel 26 success ratio was as little as 3.61%.

All these experiments suggest that being in a high-interference medium, in comparison with the other evaluated

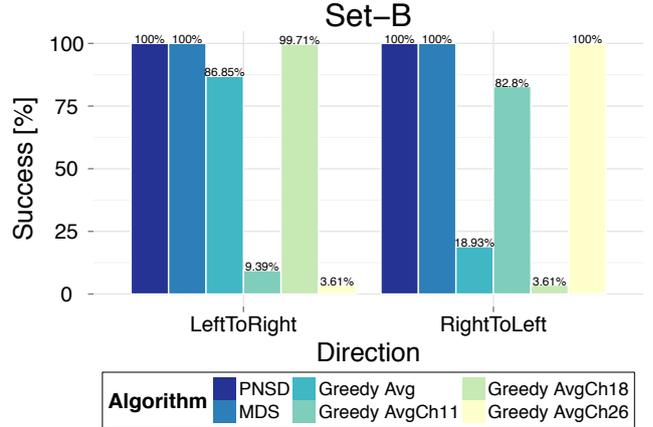


Fig. 4. Success on north side, Set-B

methods, PNSD shows better stability. We are analyzing this further in Section IV.

As we stated before, the misclassifications are not random, but rather localized into a small part of the measurement environment. That is, in vast majority of the cases with Set-A, just two of ten nodes are imprecisely estimated to be in swapped rooms with each other instead of theirs. We have repeated the experiments with the same settings but using a different set of nodes (Set-C) at slightly different positions in the campus-side rooms. The 10 nodes in the experiments again had a linear configuration. The major difference was that this time we performed the experiments during the week, which means a much bigger possibility of higher density of wireless transceivers around the testbed. The ratio of finding the correct sequence was lower as expected. The results are plotted in Figure 5. Although over all success was decreased, our proposed methodology has been much better, compared to alternative methods. It is worth noting that success rate of Greedy Average at channel 11 results are better than others in one direction. However, the results from the other direction, as well as Figure 7, confirms that getting good results without frequency diversity is solely coincidental.

By looking at the results we could notice that there has been an anomaly in some place or time of the testbed-based experiments. Despite of a lot of effort, we have not been able to pinpoint the reason for those anomalies. We have tested and excluded the option of “faulty nodes”. We have intentionally created and monitored WiFi interference, without being able to recreate same type of anomaly. Nevertheless, we have added an additional function to our algorithms for “classification of result reliability”, which is explained in Section V, in an effort to detect how likely it is that there was an anomaly at the time of the measurements.

IV. SIMULATION

The results we present in the previous section has been taken only in one specific building and one specific testbed.

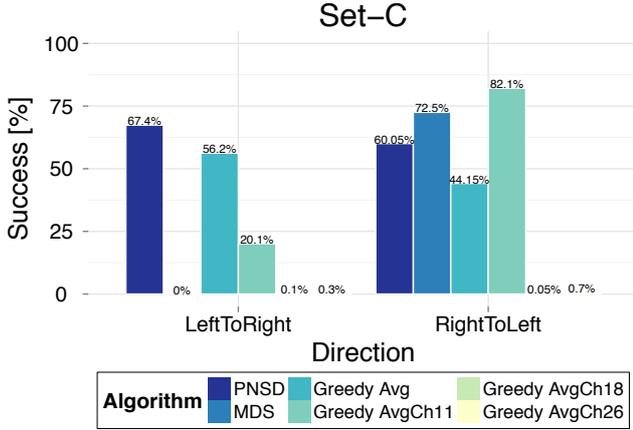


Fig. 5. Success on second set of 10 nodes, Set-C

Although we have repeated tests and measurements in different areas of same testbed, these limited number of measurements will never be enough to guarantee that the current results are not testbed specific. It is also not practical and time-wise feasible to repeat the experiments in many different buildings.

To investigate the testbed dependency of our results, we used a simulation model that represents an indoor environment. The received signal strength (RSS) is assumed to be affected by various propagation effects. Large scale variation of RSS is assumed to be caused by change in distance. We do not consider shadow fading in this model. Small scale variation of RSS is due to the fact that the wireless signals arrive at the receiver from multiple paths and therefore suffer from multipath fading. As each channel bandwidth is smaller than coherence bandwidth of usual office environments with around 100 ns delay spread, we assume a frequency flat fading for each channel but the fading gain is expected to change from one channel to the other channel which provides the frequency diversity of the system. We assume that the environment is static and changes slowly in time which means that the coherence time is very large. RSS is also affected by the thermal noise of receiver.

The propagation effects are modeled using statistical channel models with path loss models [11], [7]. The signal strength values are produced based on the previous assumptions. The transmission power is constant and the path loss exponent α is assumed to be static in the environment. We have used Rayleigh fading model to model small scale fading where the channel fading gains are from a Rayleigh distributed random variable. The fading gains are assumed to be constant in a channel due to slowly fading assumption however it is assumed that the fading gain is different in each channel. The relation between received power P_R and transmit power P_T , both in *milliwatts*, is presented in the following equation:

$$P_R = |h|^2 \times P_T \times \left(\frac{\lambda}{4 \times \pi \times d} \right)^\alpha + P_n \quad (1)$$

where h is the channel gain, which is a Rayleigh distributed random variable. λ is the wavelength at 2.4GHz, α is the path loss exponent and d is the distance between transmitter and receiver antennas in meters. P_n is the power of additive white Gaussian noise in the environment. We have chosen the noise variance as -45dBm and the fading gain variance h as 10. Both random variables have zero mean. Path loss exponent α was taken as 3, which is typical for indoor office environments.

In Figure 6, we show a set of generated RSS values using above formula and parameters. Each produced RSS value corresponds to the reading at different transmitter-receiver distance. The graphs show that the signal strength value is a function of the distance, however it is not always monotonically decreasing with distance due to multipath fading.

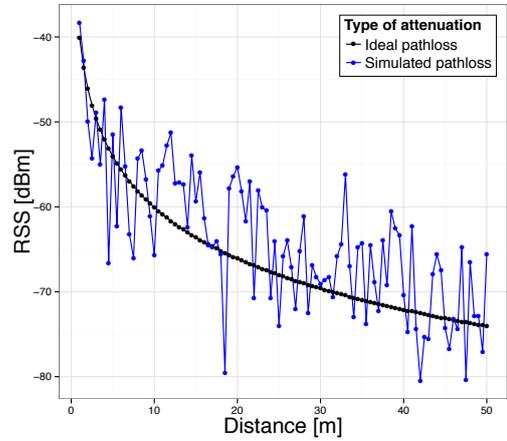


Fig. 6. Produced RSS values

Later, we have produced 16×40 values at 9 different receiver distances of 3m apart and compared them to one of the measurement sets containing the same amount of RSS data. The data plotted side by side in Figure 7 show that the values we produced with above model are not far from what we would expect from our real measurement environment. From the produced values, we have dropped the values that are below the sensitivity level of the radio chips.

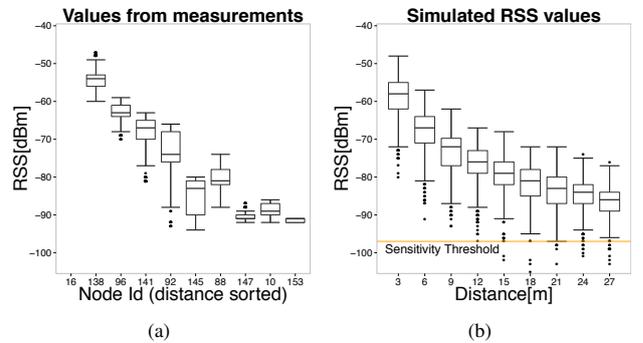


Fig. 7. Real measurements Vs Simulated RSS values

Finally, we have produced same number of simulated results

as we have in our measurements: 2000 repetitions of 10-node setups with nodes 3m apart from each other. On each generated set of data, we have applied the methodology that we have explained in Section II. The result has been 100% perfect identification of the correct sequence for either directions.

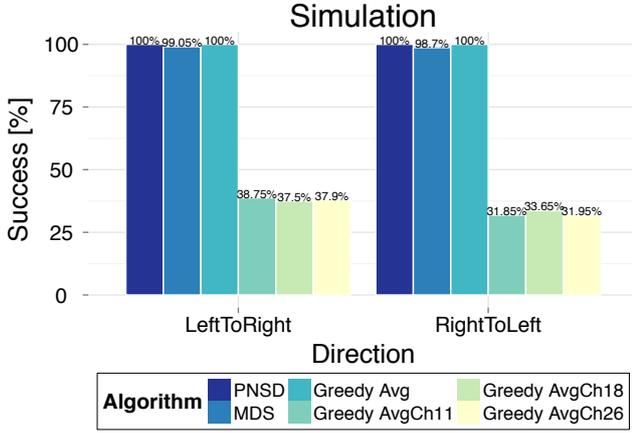


Fig. 8. Success with simulated measurements

Figure 8 is clearly showing us, that in a propagation medium that is fairly affected by fading and multipath, our methodology remains robust, while MDS failed about 1% of the time. We must notice that Left-to-Right and Right-to-Left computations are pretty much symmetrical, since time has not been a parameter in the RSS generation. For each experiment, we have used same standard deviation and mean values for Rayleigh distributed channel gain and Gaussian distributed white noise. Therefore, according to the Law of Large Numbers, the average of the random variables are expected to approach to their mean values. When we average a big number of values, produced with Formula 1, we come closer to what Free-space path loss equation would give us under ideal conditions. Hence, 100% success with averaging all values from all channels is expected in simulations, although it is often not the case in real-world scenarios.

V. RELIABILITY THROUGH REVERSE VALIDATION

With presented PNSD methodology we have achieved a significant improvement over alternative methods in identifying node sequence correctly with relatively small inter-node distances. We have shown that we can achieve up to 100% correct identification of the node sequence where other approaches are found to be statistically unstable. We have also noticed that occasionally, due to some anomalies, two adjacent nodes can be imprecisely identified in swapped spots. In this study we are motivated by using only low power radio chips to be used in regular communication of the nodes. So we do not find it feasible with such low cost hardware to spend energy to detect if there are anomalies in the medium during measurements. Even though one position mistake might be acceptable for most needs, we still want to be sure that the

verdict is dependable. With this aim, we have developed a post-processing methodology to assess the reliability of the computed sequence and change the verdict if a better one exists. So along with the candidate sequence S_1 which has the highest probability, we give chance to the other sequence S_2 with less probability.

We have noticed that the channel does not stay symmetrical even though there is very little time between measurements from node A to B and from B to A. In this case, assuming a 4-node scenario (A-B-C-D), node A wants to decide between adjacent nodes B and C, and it might imprecisely decide that C to be the closer one instead of node B. However, the odds are very little for D to choose B as the closest node too. Such a case would require same type of measurement distortion between different nodes and different places. But the correct decision is much more likely to be in agreement with the reverse direction. With this justification in mind, we construct another sequence of the nodes for validating the result, assuming that the last node in the forward constructed sequence as the reference node.

Our probabilistic method has also shown us that imperfect results are not just random. When we sort the sequences with their probabilities (S^*), the second best sequence (S_2) has been in most cases the correct one if it is not the first one (S_1). The control(reverse) sequence (S'_i) of an imperfect sequence (S_i) also has the probability of being the correct (reverse) sequence, or even the second best control sequence (S''_i) being the correct sequence has a probability greater than zero.

For reliability assessment, we compare the most probable sequence (S_1) with most probable reverse sequence (S'_1) for S_1 and second most probable reverse sequence (S''_1) for S_1 . Then we compare second most probable sequence (S_2) with most probable reverse sequence (S'_2) for S_2 and second most probable reverse sequence (S''_2) for S_2 .

Our claim is, the sequence (S_i) which matches one of its reverse sequences (S'_i or S''_i) is more likely to be the correct sequence (hence the verdict will be S_i) and by looking at the rank (i) of that sequence (among other competing sequences of the same dataset) we will decide how reliable the final decision is. If there are multiple matches, we choose the one with highest probability ($S = \text{argmax}(P(S_i))$), as long as the probabilities are distinctly different from each other.

We have 4 comparisons:

$$\begin{aligned} S_1 = S'_1 &\Rightarrow \text{match_rank} = 1 \\ S_1 = S''_1 &\Rightarrow \text{match_rank} = 1 \\ S_2 = S'_2 &\Rightarrow \text{match_rank} = 2 \\ S_2 = S''_2 &\Rightarrow \text{match_rank} = 2 \end{aligned}$$

The number of comparisons (s) where we have a match:

$$t(S_i) = \begin{cases} 1, & \text{if } S_i = S'_i \\ 1, & \text{if } S_i = S''_i \\ 0, & \text{otherwise} \end{cases}$$

$$s = t(S_1) + t(S_2)$$

As the final verdict, we choose between S_1 and S_2 , depending on which one has a match with a reverse sequence. If both of them have a matching sequence, then we choose the one with the higher probability.

$$finalVerdict = \begin{cases} S_n, & \text{if } S_n = S'_n \text{ AND } s = 1 \\ S_m, & \text{if } P(S_m) = \max(P(S^*)) \text{ AND } s \neq 1 \end{cases}$$

We define 3 levels of reliability:

High:

if $s = 1$ and $match_rank = 1$

Medium:

if $s > 1$ or $match_rank = 2$

Low:

if $s = 0$

or

if S'_1 or S'_2 cannot be produced in the same length as S_1 using the same dataset. This happens in case of too many packet losses.

or

if $\frac{|P(S_1) - P(S_2)|}{\min(P(S_1), P(S_2))} < 0.1$ and $s \neq 1$

If an experiment has one or two matching sequences, but any of these matching sequences does not have a probability difference of at least 0.1 from the competing sequence (i.e. $P(S_1)$ vs $P(S_2)$), then we put this experiment into the “Low” reliability class.

We have applied the described reliability verification methodology to our experiments that we presented in Section III. The results of classification is shown below. In the experiment, in which we had 100% success (Figure 7), 9 experiments have ended up with wrong verdicts in oppose to 1991 correct verdicts. Those wrong verdicts however were all classified as “Medium” reliable. The other direction for the nodes in the test, where we had 93.55% success, the success ratio is even increased to 96.95% success, changing 68 wrong verdicts with the correct ones. In both cases, 100% of the sequences that were classified as “High” reliable, have been the correct verdicts. Results are plotted in Figure 9.

In the second set of experiments from the south part of the building, in which we had lower success ratio (Figure 5) the verified success has increased in both directions. More importantly, we have associated most of the wrong verdicts with “Low” reliable verdicts. Again 100% of the “High” classified verdicts have been correct. See Figure 10.

In the Set-C, the success ratio is increased from 60.05% to 86.7% in Right-to-Left direction and from 67.4% to 89.5% in the other direction. This is due to the nature of our “reverse validation” algorithm which chooses the more consistent result as the final verdict. So by applying reverse validation, we do not only assess the reliability of a verdict, we also increase the chance that it is correct.

VI. COMPLEXITY AND OVERHEAD REDUCTION

A. Computation with pruning

Constructing the probability tree and trying to achieve the correct sequence is highly effective as shown in Section II, the

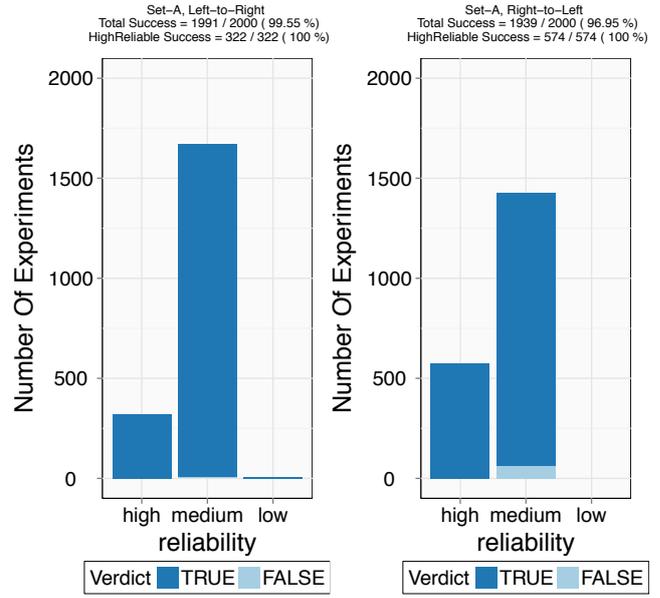


Fig. 9. PNSD with verdict reliability classification, Set-A

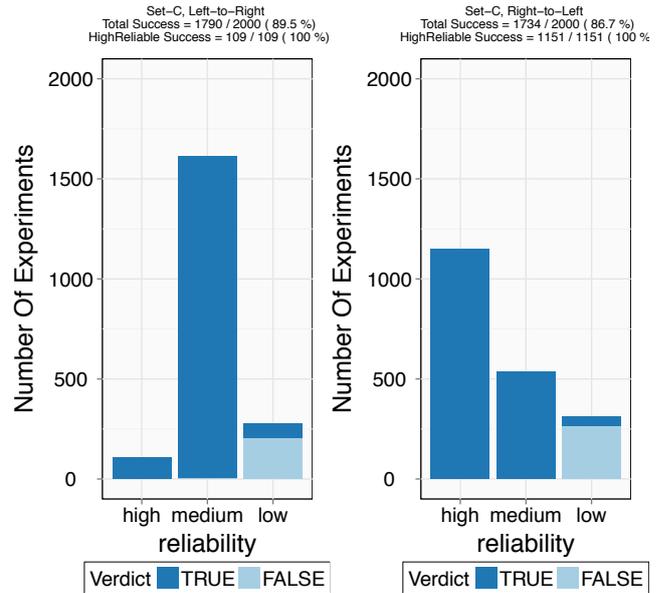


Fig. 10. Sequence Discovery with verdict reliability classification, Set-C

tree construction might be too costly in terms of computational complexity. Theoretically, in the worst case scenario, the tree would result in $N!$ leaves for N nodes and the computational complexity would be $O(N!)$, which is not feasible. Though, the nature of the wireless communication plays in our favor and allows only the nodes in each other’s communication range cause the growth of the tree. If the transmission range of

the nodes is r meters and the nodes are d meters apart from each other, then the computational complexity of the worst case scenario is $O(\frac{r}{d}!)$.

This might still end up being too costly in very dense environments. Therefore we reduce the computational time by pruning the probability tree by half of its root-to-leaf paths, when the weight (joint probability) of a path becomes too small as the sequence grows into a less likely sequence. We prune the tree at every m^{th} level (in other words, after each m^{th} node in a candidate sequence). This reduces the complexity to $O(N^m)$, assuming that all of the receivers of a sender will measure one maximum RSS at least once through out a complete transmission campaign of one sender. This worst-case scenario, however, almost never occurs in practice, because in a linear configuration competition with RSS occurs only among few closest nodes. Very far away nodes do not have a real chance of measuring as high RSS as very close nodes due to distance related attenuation, despite multipath and RF noise. Thus the real computation time never reaches $O(N^m)$.

Our pruning criterion at m^{th} level is “half of the maximum root-to-leaf joint probability(P^{max})”. Any root-to-leaf path that has a smaller probability than $0.5 \times P^{max}$ is pruned out at level m and the tree is iterated for the remaining leaves.

We applied this pruning on the same dataset (Set-A) that we showed initially. The computation time was drastically improved while maintaining the success ratio the same. Figure 11 shows that success ratio does not decrease when we prune the tree of 10 nodes at levels $m = 5$ and $m = 1$. However pruning affects the number of leaves in the resulting tree (number of candidate sequences).

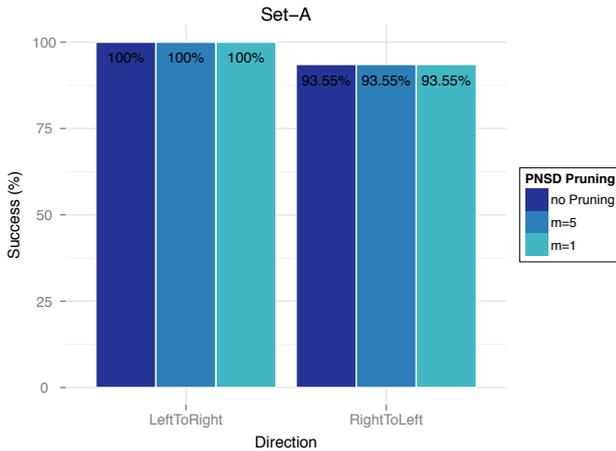


Fig. 11. Success with pruning, Set-A

B. Impact of Transmissions per Channel

Throughout this paper we have demonstrated all our results from the experiments that contained 40 transmissions per channel. The computation was centralized, meaning that all the measurements needed to be collected at one node.

Although we have a distributed implementation approach for the proposed algorithm, we do not elaborate on it here, since it is beyond the scope of this paper.

Regardless of whether the implementation is distributed or centralized, one would desire to do as few transmissions as possible to save energy and time. Therefore, we downsized our measurements to fewer number of transmissions per channel while keeping the number of channels same. We have repeated all our computations by filtering the input measurement set down to only 1, 8 and 16 transmissions per channel. In Figure 12, we observe that the success ratio was not affected noticeably by the changing number of transmissions, which is also discussed in more detail in [6]. This proves that the success we get with channel diversity is not because of the added number of transmissions into the computations, but the changing of the multipath conditions.

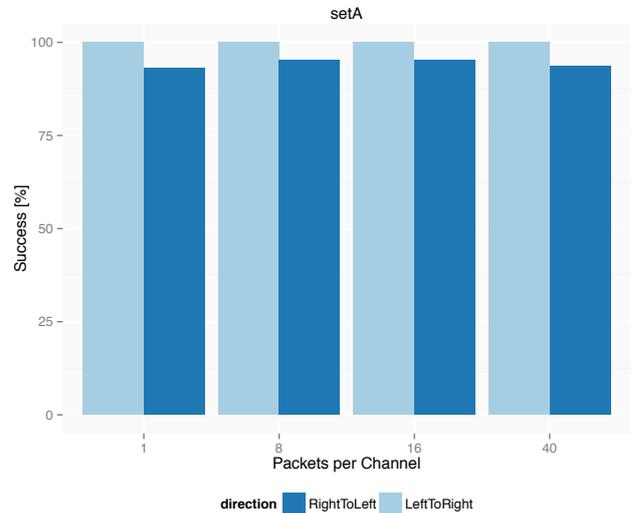


Fig. 12. Impact of number of transmissions per Channel, Set-A

We have also examined the impact of different number of transmissions per channel for MDS in comparison to PNSD. In Set-A (Left-to-Right direction), using fewer number of transmissions did not have a significant impact on MDS either. However, for Set-C (Right-to-Left direction), wherein the success ratio was worse for all of the algorithms in consideration, MDS got affected more significantly than PNSD by the reduced number of transmissions. The comparison is shown in Figure 13.

Although it is not shown in the plots (due to paper size limitations), “Greedy” approach was not affected either by the changed number of transmissions per channel.

And finally when we examined the impact of different number of transmissions per channel on “reverse validation”, we have noticed that using as many as 40 packets is indeed increasing the reliability of the results. This analysis on Set-A Right-to-Left direction is shown in Figure 14.

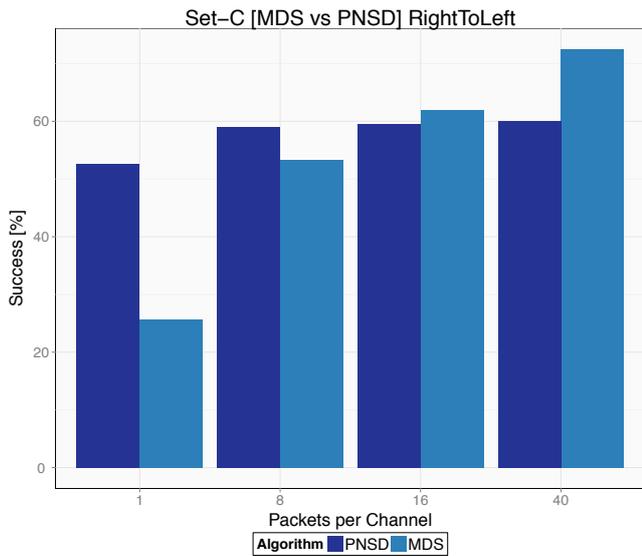


Fig. 13. Impact of number of transmissions per Channel, Set-C Right-To-Left

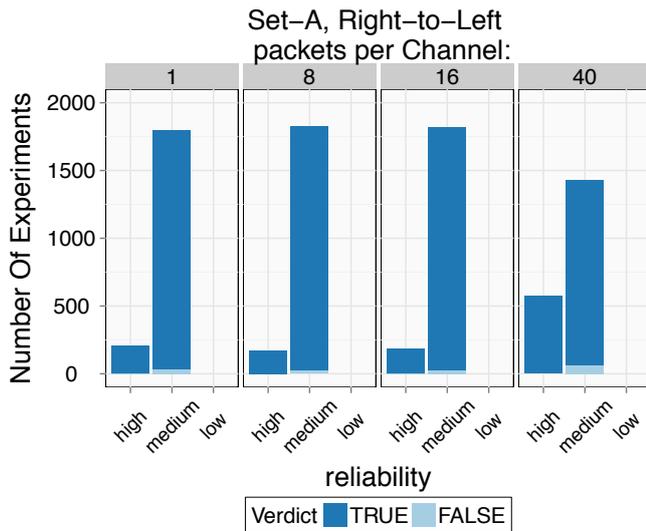


Fig. 14. Impact of number of transmissions per Channel on reliability, Set-A Right-To-Left

To sum up, we have noticed that our suggested approach (PNSD) is considerably robust to varying number of transmissions per channel while MDS is affected in some cases. However, PNSD benefits from bigger number of transmissions per channel resulting in statistically increased levels of reliability. Analyses on other datasets that we have previously mentioned also support our conclusions, but are not presented here due to space limitations.

VII. RELATED WORK

Similarly to classical range-free localization methods like centroid, APIT or DV-Hop [9], our sequence determination approach is not based on direct inversion of the path-loss function for estimating the inter-node ranges. Due to the fading-related distortions discussed in Section I, this can lead to significant range estimation errors [12], [14], [2], leading to erroneous estimation of the node sequence.

Instead of focusing on complex approaches to estimate and compensate the distortion effects of the fading, we leverage the diversity in the frequency domain which allows us to achieve slightly different fading behaviors even for static nodes [16]. Averaging RSS samples across different channels, as means of improving the stability of the mean RSS estimate, has been explored in the work presented in [1], [15]. The use of frequency diversity has proved useful also for non-RSS range-based methods like Time-of-Flight [10].

In [13], the authors introduce the concept of sequence-based localization, a 2D range-free approach where the space is divided into regions that can be uniquely identified using sequences representing the rank of distances from a set of reference nodes. Their procedure for determining the unknown node location sequence (the distance rank table from the unknown node to all reference nodes) faces the same challenges due to the distortion of the RSS signal through shadowing and multipath as in our work. In contrast to our work, they do not attempt to mitigate these effects using frequency diversity and probabilistic sequence estimation methods, and only rely on comparing the corrupted distance rank sequences to the set of all possible rank sequences.

Another 2D range-free localization and tracking scheme, based on relative distance between neighbors, has been presented in [17]. The approach aims at improving the performance of pure connection-based, rang-free methods through the concept of a “Relative Signature Distance (RSD)”, new metric expressing the expected proximity between 1-hop neighbors. Even though the approach has also been evaluated for a long 1D node configuration outdoors, the focus is on estimation of the spatial coordinates of the nodes instead of just the node sequence. The approach also lacks any mechanisms to deal with the challenging indoor conditions that are in the focus of our work. Similar attempt to improve the connection-based range estimates is presented in [5].

The method for relative localization of wireless sensor nodes in linear typologies, presented in [18] is the most similar to our work, from the point of view of the common focus on determination of the correct node sequence, instead of estimation of spatial coordinates. The proposed method shares resemblance with our single-channel, greedy, average-RSS approach, introduced and more extensively evaluated in our prior work [6]. Similarly to the other related work, however, their method has not been evaluated in indoor environments. As demonstrated by our previous work and the results presented in Section III, this simple approach is not able to deal with the challenging fading disturbances that are typical for indoor

environments. In that paper, accuracy was also commented on, however any indication of accuracy detection does not exist like we provide in our work.

VIII. ACKNOWLEDGEMENTS

This work has been partially funded by the European Commission (FP7-ICT-FIRE) within the project EVARILOS (grant No. 317989).

IX. CONCLUSIONS

In this work we have presented a robust methodology, PNSD, for discovering the node sequence in a linear configuration for wireless sensor networks. We have shown that contrary to popular belief, the problem of sequence determination in challenging indoor environments is solvable with the imprecise RSS measurements from low-power radio chips. In contrast to many RSS-based indoor localization solutions, our methodology does not rely on special hardware, fingerprinting, calibration or preconfiguration phases.

Building on our claims and findings in our previous work [6], we have again shown that frequency diversity is essential for relating RSS measurements to position information. For the sake of fairness, we have incorporated the same frequency diversity to other approaches to which we have compared our algorithm.

We have compared the suggested methodology (PNSD) to greedy RSS-based sequence building algorithm and well accepted Multidimensional Scaling algorithm. The results show that when the measurement medium is not affected heavily by external factors (Set-B) we achieve 100% success as well as MDS. However, when the measurement medium gets challenged by interference or other types of anomalies, MDS and Greedy approaches get affected heavily while ours stay robust (ie. Set-A, Set-C). It is worth mentioning that the presented sequence discovery algorithm has also been tested in many other experiments under varying conditions with different, also very small (down to 0.5 meter), distances and it has demonstrated good performance.

The results show that there are often conditions for which correct construction of the sequence is not possible 100% of the time. Despite hard efforts on determining the reasons for the instabilities in the raw data, leading to false sequences when our and other approaches are applied, we were not able to repeat the conditions that impaired the measurements. There have been periods of time when successful determination was possible and there have been other periods of time when correct determination of the sequence was not possible even after many repetitions. This has motivated us to create an additional quality assessment step.

By generating the reverse sequence from our initial result we are able to assess the reliability of the sequence estimate or even retroactively change the decision of the previous step to a more probable estimate. Extensive tests have shown that all the classification for which we could obtain “High” reliability verdicts have always been correct and the vast majority of those with “Medium” reliability verdicts have been correct,

but rarely wrong. All the remaining wrong verdicts can be attributed to “Low” level of reliability. Using this information, users can accept the sequence with its provided reliability rank, or they might want to repeat the process until a “High” reliable sequence is found.

REFERENCES

- [1] A. Bardella, N. Bui, A. Zanella, and M. Zorzi. An experimental study on ieeec 802.15.4 multichannel transmission to improve rssi-based service performance. In *Proceedings of the 4th international conference on Real-world wireless sensor networks*, REALWSN'10, 2010.
- [2] J. Cavers. *Mobile Channel Characteristics*. Kluwer Academic, 2000.
- [3] 2.4 GHz IEEE 802.15.4/ZigBee-ready RF Transceiver. Online. <http://focus.ti.com/lit/ds/symlink/cc2420.pdf>.
- [4] J. de Leeuw and P. Mair. Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3):1–30, 2009.
- [5] A. El Assaf, S. Zaidi, S. Affes, and N. Kandil. Hop-count based localization algorithm for wireless sensor networks. In *Mediterranean Microwave Symposium (MMS), 2013 13th*, pages 1–6, Sept 2013.
- [6] M. Ergin, V. Handziski, and A. Wolisz. Node sequence discovery in wireless sensor networks. In *Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on*, pages 394–401, May 2013.
- [7] A. Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005.
- [8] V. Handziski, A. Köpke, A. Willig, and A. Wolisz. Twist: a scalable and reconfigurable testbed for wireless indoor experiments with sensor networks. In *Proceedings of the 2nd international workshop on Multi-hop ad hoc networks: from theory to reality*, REALMAN '06, 2006.
- [9] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher. Range-free localization schemes for large scale sensor networks. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, MobiCom '03, 2003.
- [10] P. Pettinato, N. Wirstrom, J. Eriksson, and T. Voigt. Multi-channel two-way time of flight sensor network ranging. In *Proceedings of the 9th European Conference on Wireless Sensor Networks*, EWSN 2012, Feb. 2012.
- [11] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, New York, NY, USA, 2005.
- [12] K. Whitehouse, C. Karlof, and D. Culler. A practical evaluation of radio signal strength for ranging-based localization. *SIGMOBILE Mob. Comput. Commun. Rev.*, 11(1), Jan. 2007.
- [13] K. Yedavalli and B. Krishnamachari. Sequence-based localization in wireless sensor networks. *Mobile Computing, IEEE Transactions on*, 7(1), 2008.
- [14] G. Zanca, F. Zorzi, A. Zanella, and M. Zorzi. Experimental comparison of rssi-based localization algorithms for indoor wireless sensor networks. In *Proceedings of the workshop on Real-world wireless sensor networks*, REALWSN '08, 2008.
- [15] A. Zanella and A. Bardella. Rss-based ranging by multichannel rss averaging. *Wireless Communications Letters, IEEE*, 3(1):10–13, February 2014.
- [16] D. Zhang, Y. Liu, X. Guo, M. Gao, and L. Ni. On distinguishing the multiple radio paths in rssi-based ranging. In *INFOCOM, 2012 Proceedings IEEE*, Mar. 2012.
- [17] Z. Zhong and T. He. Rsd: A metric for achieving range-free localization beyond connectivity. *Parallel and Distributed Systems, IEEE Transactions on*, 22(11):1943–1951, Nov 2011.
- [18] X. Zhu, X. Wu, and G. Chen. Relative localization for wireless sensor networks with linear topology. *Computer Communications*, 36(1516):1581 – 1591, 2013.