HyPAERLoc: Plausible Hybrid Localization for Wireless Sensor Networks

Jakob Salzmann, Ralf Behnke, Philipp Gorski, Dirk Timmermann
Institute of Applied Microelectronics and Computer Engineering
University of Rostock
Rostock, Germany
{jakob.salzmann, ralf.behnke, philipp.gorski, dirk.timmermann}@uni-rostock.de

Abstract— Position estimation is one of the major challenges of sensor nodes in wireless sensor networks. By utilizing the information of messages of some pre-deployed location aware nodes, called beacons, and signal strength based distance estimation, a location unaware node is able to estimate its position. In the recent years, several localization methodologies have been developed. Due to imprecise distance estimations via received signal strength utilization, the localization accuracies of these algorithms differ, depending on the scenario conditions. In the proposed work, we developed an algorithmic approach to improve the localization accuracy of a least squares approach via two strategies. On the one hand, we tackle the imprecise distance measurements with a preceding plausibility check. On the other hand, we evaluate the achieved accuracy and improve the result by weaving the result of adaptive weighted centroid localization into a hybrid localization. The achieved localization accuracy beats the performance of the preceding approaches in all investigated scenarios, particularly for low beacon densities.

Keywords—Wireless Sensor Networks, Localization, Log-normal Fading.

I. INTRODUCTION

The ongoing miniaturization of technical devices allows to combine microcontroller, sensors and radio technology to a single tiny and battery driven device, called sensor node. Due to their radio technology, numbers of the nodes can compose themselves together to a wireless sensor network (WSN), which ranges from few nodes in one-hop distance to each other up to large networks with hundreds of nodes and multi-hop dimension. Such WSNs can be deployed in an area to observe, using their sensor abilities to detect phenomena in various scenarios. Examples for the operation of sensor networks are disaster control, environmental observation, tracking of moving objects [1]. In most of the scenarios, e.g. object tracking, the location of a detected phenomenon is as important as the properties of the phenomenon itself. As result, the detecting nodes have to be aware of their position within a reference system. Due to the WSN operation in inaccessible or indoor scenarios and the large amount of nodes, neither a specific deployment of all nodes nor a global localization system, e.g. GPS, are always feasible for the localization of sensor nodes. A feasible solution is that a fraction of nodes is location aware and used as reference nodes, called beacons [2-6]. These beacon nodes get their position either due to a specific deployment, e.g. deployed by a robot or are additionally equipped with a GPS-receiver. All remaining nodes, called unknowns may be deployed randomly in the monitored area and estimate their position with the help of the beacon nodes. Each beacon broadcasts a message, and each receiving node stores the contained information about the beacons’ position as well as the received signal strength or the derived distance.

The algorithms, which utilize the collected information from all nearby beacon nodes to estimate the position of an unknown node, differ in complexity and achieved estimation accuracy. Interestingly, a high complexity is not mandatory for a competitive accuracy, especially if the distance estimations are inaccurate due to shadowing effects.

The contribution of the paper improves the localization accuracy in two ways. On the one hand, a plausibility check avoids applying impossible distance estimations, on the other hand the uncorrelated localization accuracy of a different algorithm family is utilized. The result is described as an algorithm called HyPAERLoc (Hybrid Plausible Approach for Error Reduced Localization). We compared HyPAERLoc with its preceding localization approaches in different scenarios with a log-normal fading radio channel model. With HyPAERLoc, the localization accuracy increases significantly compared to the preceding algorithms. Additionally, the algorithm avoids outliers.

The remainder of the paper is structured as follows: Section II describes the related work, Section III explains our simulation environment. Section IV analyzes the strengths and weaknesses of the researched localization approaches. In Section V, we describe and evaluate our check for implausible distances, Section VI describes and evaluates the hybrid localization approach. Section VII gives the conclusion and an outlook.

II. RELATED WORK

Beacon based localization algorithms can generally be divided into coarse-grained and fine-grained localization. The actual section describes common representatives of both groups of algorithms.

1. Coarse-Grained Localization

Coarse-grained localization algorithms represent heuristic methodologies to estimate the position of an unknown node. These classes of algorithms are characterized by the disadvantage, that even with exact distance measurements, they are not able to estimate the exact position of the unknown due to their simplification. In the following, three coarse-grained localization algorithms are described and evaluated.
Centroid Localization

By the idea of Centroid Localization (CL), an unknown node is located at the centroid of all received beacons [2]. If \( P_i(x,y) \) represents the position of an unknown node, \( n \) represents the number of received beacons and \( B_j(x,y) \) represents the known position of a beacon node in range, each unknown node \( i \) can perform the algorithm as given in (1).

\[
P_i(x,y) = \frac{1}{n} \sum_{j=1}^{n} B_j(x,y) \tag{1}
\]

It can be seen, that the algorithm is easily to perform and abstain from distance estimations.

Weighted Centroid Localization

An advanced approach is Weighted Centroid Localization (WCL) [3]. In contrast to CL, WCL uses additional information to calculate the centroid. Usually, a sensor node has the ability to measure the received signal strength of a message. The result of this measurement can be utilized to estimate the distance to the beacon node which sent this message. The estimated distance can be utilized to improve the position estimation by weighting the centroid calculation, as done by WCL. A practical approach is to describe the impact of a beacon \( B_j \) with the weight \( w_{ij} \) as reciprocal of the estimated distance \( d_{ij} \) between unknown \( i \) and beacon \( j \), as done in equation (2).

\[
w_{ij} = (d_{ij})^{-1} \tag{2}
\]

With this weight, the calculation of the centroid changes as given in (3).

\[
P_i(x,y) = \frac{\sum_{j=1}^{n} (w_{ij} * B_j(x,y))}{\sum_{j=1}^{n} w_{ij}} \tag{3}
\]

As result, the position of the unknown node is not estimated as real centroid of all surrounding beacons, but nearer beacons pull the position of the unknown in their direction.

Adaptive Weighted Centroid Localization

The goal of Adaptive Weighted Centroid Localization (AWCL) was the improvement of the weighting [4]. If all distances between one unknown and its received beacons are similar to each other, the effect of the weighting becomes relatively low and the result is comparable to CL. To overcome this problem, AWCL adapts the resulting weights by applying the following steps:

1.) Determination of the smallest weight \( w_{i,min} \)
2.) Calculation of the reduction part \( q \). Simulation results in [4] provide to calculate a feasible reduction part as done in (4).

\[
q = w_{i,min} \times 0.55 \tag{4}
\]

3.) Applying the centroid estimation with reduced weighting, as done in (5).

\[
P_i(x,y) = \frac{\sum_{j=1}^{n} (w_{ij} - q) * B_j(x,y)}{\sum_{j=1}^{n} (w_{ij} - q)} \tag{5}
\]

As result, AWCL is able to apply a more influential weighting, which directly impact the localization accuracy in most situations.

2. Fine-Grained Localization

The idea of fine-grained localization approaches is to determine the exact position of an unknown node. The obvious disadvantage of the fine-grained localization approaches is the costly computational effort. In the following, the method of least squares as most common representative and one successor are explained.

Linear Least Squares

The method of linear least squares, also called atomic multilateration in [5], is an approach to approximate the solution of a linear over determined equation system. In least squares, the sum of the squares of the residual of the solved equation system is minimized. Applied to the localization problem, each unknown node \( i \) has to determine two unknown variables, \( x_i \) and \( y_i \). This equation can be transformed as given in (7).

\[
x_i^2 + y_i^2 + x_j^2 + y_j^2 - 2(x_ix_j + y_ix_j) = d_{ij}^2 \tag{7}
\]

Here, \( d_{ij} \) is the known distance between unknown node \( i \) and beacon \( j \). \( x_i \) and \( y_i \) are the unknown coordinates of the unknown node and \( x_j \) and \( y_j \) are the known coordinates of the beacon in a common reference system. For applying linear least squares, the quadratic terms of the unknown have to be removed. This can be done by subtraction of the equation of beacon \( k \), as done in (8).

\[
x_i^2 + y_i^2 - x_k^2 - y_k^2 = 2(x_ix_k + y_ix_k) \tag{8}
\]

This linearization can be done for each couple \( l \) of received beacons. For \( n \) received beacons, the maximum number of different linear equations \( m \) is given by equation (9).

\[
m = 0.5 \left( (n - 1) \times n \right) \tag{9}
\]

After that, the present equation can be transformed into a linear equation \( h_i = g_i x_i + y_i \). Here, \( g_i \) and \( h_i \) are absolute terms and represents single points for the unknown node \( i \) in the
LRG-24413

final equation system, in which a straight line with offset \( y_j \) and slope \( x_i \) can be found.

For each couple of beacons \( l \), values for \( g \) and \( h \) are calculated as given in (10) and (11) for the example beacons \( j \) and \( k \).

\[
    h_{l,i} = \frac{(d_{ij}^2 - d_{ik}^2 - x_j^2 - y_j^2 + x_k^2 + y_k^2)}{2 \cdot (y_k - y_j)} \tag{10}
\]

\[
    g_{l,i} = \frac{(x_k - x_j)}{(y_k - y_j)} \tag{11}
\]

After transforming the equations for each couple of beacons into a system of linear equations, the values for \( x_i \) and \( y_i \) can be computed as given in equation (12) and (13), whereby \( \bar{g}_i \) and \( \bar{h}_i \) are the average values of all available terms of \( g_i \) and \( h_i \).

\[
    x_i = \frac{\sum_{l=1}^{m}(g_{l,i} - \bar{g}_i)(h_{l,i} - \bar{h}_i)}{\sum_{l=1}^{m}(g_{l,i} - \bar{g}_i)^2} \tag{12}
\]

\[
    y_i = \bar{h}_i - x_i \cdot \bar{g}_i \tag{13}
\]

As result, linear least squares is able to calculate the exact position of an unknown node, if the distances between the unknown node and the beacons are correctly estimated. In this case, usually only three beacon nodes and hence two independent couples of beacons are sufficient to determine the exact position.

**Scalable Distributed Least Squares**

Linear least squares have two major drawbacks. The first one is the cost intensive computation of the final position, the second one is the inaccuracy, if the measured distances are inexact. Both drawbacks were tackled with Scalable Distributed Least squares (SDLS) in [6]. While the splitting of the calculation and the involved challenges are not in the focus of this paper, the improvement of the accuracy is significant. Instead of choosing all available or a subset of couples of beacons for calculating \( g_{l,i} \) and \( h_{l,i} \), an unknown node \( i \) in SDLS selects the nearest beacon node as linearizer and subtract the equation of the linearizer from each other node. Hence, the number \( m \) of resulting beacon couples for \( n \) beacons is given by \( m=n-1 \), and each couple contains a part of the linearizer. The reason for selecting the closest beacon as linearizer is given by the fact, that smaller distances could be measured more exactly due to the greater absolute difference in the signal strength.

To get an impression of the work of all localization algorithms with inaccurate distance estimations, example localization is given in Figure 1. LS and SDLS estimate the same position of the unknown node due to the limited number of beacons. Additionally it is recognizable that all coarse-grained algorithms localize the node in a triangle with the beacons as corners, which is always given by the heuristic of the algorithms.

### III. Simulation Environment

The described localization algorithms share one major drawback: Until now, they were only partly analyzed and not optimized to deal with realistic fading situations. Due to the node deployment near to the ground, their motionlessness and the possibly heterogeneous environment conditions, the log-normal shadow fading offers a well performing model for the real-world behavior of communicating sensor nodes [7,8]. Major part of the model is transmission equation of Friis, as given in equation (14).

\[
    P_t = P_r \cdot G_t \cdot G_r \cdot \left( \frac{c}{4\pi f} \right)^2 \cdot \left( \frac{1}{d} \right)^2 \tag{14}
\]

Here, \( P_r \) is the received Power, \( P_t \) the transmitted Power, \( G_t \) and \( G_r \) the antenna gains, \( c \) the speed of light, \( f \) the transmission frequency and \( d \) the distance between transmitter and receiver. To adapt the transmission equation to log-normal shadow fading, an environment depending path loss exponent \( N \) and a Gaussian distributed random variable \( X_\sigma \) with mean value \( \theta \) and standard deviation \( \sigma \) is included into this equation, as given in (15), transformed to dB.

\[
    P_t[dB] = P_t[dB] + 20 \log_{10} \left( \frac{c}{4\pi f} \right) - N \cdot \log_{10}(d) + 10 \log_{10}(G_t \cdot G_r) + X_\sigma \tag{15}
\]

To analyze how the in Section II proposed localization algorithms perform in different environments, realistic values for the path loss exponent \( N \) and the standard deviation \( \sigma \) are required. In [8], the authors performed a widespread analysis of the log-normal shadow fading with a for sensor network feasible frequency band of 900 MHz. For our analysis, we selected two scenarios, from this paper.

![Figure 1. Example Localization (A) Example setup with erroneous distance estimations, (B) Visualization and results. All values in meters.](image-url)
The first one is the “Sandy Flat Beach” scenario. This scenario represents a sensor network which is deployed on a beach, a desert or another sandy and relatively flat area, for example to detect vehicles. The second selected scenario is the “Dry Tall Underbrush” scenario, which emulates a sensor network deployed in a forest, e.g. to detect or prevent forest fires. To compare the algorithms with each other in the appropriate scenarios, we set up a simulation environment as given in Table 1 in Prowler [9] with minimized edge effects by creating an internal area for unknown nodes and beacons and a surrounding are only for beacon nodes. As precondition for the distance estimation, each node has knowledge about the path loss exponent $N$ of its environment and is able to compute a distance to a beacon on the basis of the received signal strength by assuming an undisturbed channel and applying equation (16).

$$d[m] = 10^{\left(\frac{P_i[dB]-P_r[dB]+20\log_{10}\left(\frac{\lambda}{4\pi R}\right)+10\log_{10}(G_tG_r)}{10^N}\right)}$$

(16)

It is assumed, that the path loss exponent is estimated by all communicating beacons and provided in the network.

To get a statement about the number of received beacons versus the achieved accuracy, we varied the beacon density. For each selected density, we created 200 beacon arrangements and in each arrangement 50 unknown nodes localized their position with all described algorithms, if possible. The increased beacon density correlates with a certain number of beacons in range, as shown in Figure 2, and for low density, a fraction of nodes did not received enough beacons to localize themselves, as shown in Figure 3. After applying the simulation with all beacon densities, the overall number of nodes with certain beacons in range differed extremely, as shown in Figure 4.

For statistical significance, we decided that a beacon number had to be used at least 1000 times for localization. Hence, we were able to analyze the accuracy with up to 15 received beacon nodes in the Sandy Flat Beach scenario and up to 44 received beacon nodes in the Dry Tall Underbrush scenario.

IV. PERFORMANCE ANALYSIS

To estimate the performance of the algorithms, we noted the localization error as distance of the origin position of a node and computed position of each algorithm. The simulation for the Sandy Flat Beach scenario is shown in Figure 5 and for the Dry Tall Underbrush scenario in Figure 6 as Boxplot-diagram. We selected this kind of presentation for a better visualization of the dispersion of the achieved results. The result shows that each algorithm performs better if more messages of different beacons are received. This is not surprising and covers our expectations.

The simulations allow comparing the coarse-grained and later the fine-grained algorithms among each other. In both

![Figure 2. Beacons in range versus beacon density](image)

![Figure 3. Fraction of localizable nodes versus beacon density. Coarse-grained algorithms require at least 1 received beacon, fine-grained algorithms at least 3 received beacons](image)

![Figure 4. Overall number of simulated unknowns with certain beacons in range](image)

![Figure 5. Beacons in range](image)

![Figure 6. Boxplot-diagram](image)
scenarios, AWCL outperforms CL and WCL by a reduced median and arithmetic mean error. The reason is given by the improved weighting of the different distances to the beacons. The outliers of all three algorithms are similar and there is no predication about the best performance possible.

In the fine-grained algorithms, the arithmetic mean and the median of SDLS outperform the linear least squares algorithm with random choice of beacon couples. The reason is the careful choice of the linearizer in SDLS, which is always one of the closest beacons. Due to the channel model, the average distance estimation error is reduced if the distance is shorter. Hence, the impact of the more accurate linearizer distance allows a performance increase compared to randomly selected beacon couples.

Furthermore, the simulations allow a comparison between the fine-grained and the coarse-grained algorithms. It is recognizable that the fine-grained algorithms perform only marginal better or even worse than the coarse-grained algorithms. Additionally, both fine-grained algorithms are characterized by a number of extreme outliers, which achieve errors in the range of hundreds up to several thousand meters. There are two reasons for such outliers.

The first one is an inauspicious beacon arrangement where the received beacons do not surround the unknown node, but are (nearly) arranged in a line. The second one is erroneous distance measurements to the beacon nodes. Due to the algorithm, which does not optimize until the smallest distance error is found, but optimize the terms as given in (10) and (11), the final positions are not forced to reflect real possible positions. In contrast, the localization results of the coarse-grained algorithms localize the unknown node always anywhere between the beacons. On the one hand, this limits the maximum possible error, on the other hand, this limits the maximum achievable localization accuracy.

As last part of our analysis, we investigated the correlation between errors in coarse-grained and fine-grained localization by selecting randomly 100 localized nodes with each received 3 beacon messages, and compared the localization error AWCL with WCL and SDLS of each localized node in a scatter plot, as shown in Figure 7. It is recognizable that there is a strong correlation between the two coarse-grained localization algorithms, while there is nearly no correlation between AWCL and SDLS. This different behavior is later utilized by HyPAERLoc.

Concluding, both algorithm families perform similar in the terms of accuracy in average, but with different accuracy in single beacon arrangements and distance estimations. Due to the higher accuracy potential of the fine-grained algorithms, our developed algorithm HyPAERLoc is based on the idea of SDLS, which accuracy is improved in two steps.

V. DETECTION OF IMPLAUSIBLE DISTANCE MEASUREMENTS

The major idea to improve the distance measurement is given by a plausibility check, which benefits from the known positions of each couple of beacons $j$ and $k$, whose distances to the unknown $i$ are used to create a common linear equation. With the knowledge of the beacons’ positions and the estimated distances to and between them, an unknown node is able to recognize implausible gaps as result of erroneous distance measurements. A node can
apply the plausibility check by solving equations (17), (18) and (19).

If one of the resulting gaps are greater than 0, at least one distance estimation is erroneous, even if the unknown and the beacons are deployed in a line, as shown in Figure 8.

\[
Gap_1 > 0 \quad d'_{ij} = d_{ij} \left( 1 + \frac{Gap_1}{d_{ij} + d_{ik}} \right) \\
Gap_2 > 0 \quad d'_{ik} = d_{ik} \left( 1 - \frac{Gap_2}{d_{ij} + d_{ik}} \right) \\
Gap_3 > 0 \quad d'_{ij} = d_{ij} \left( 1 - \frac{Gap_3}{d_{ij} + d_{ik}} \right) \\
\]

Our solution to deal with this knowledge about the implausible gaps is to adapt the estimated distances between the node and the beacons, until the conditions are not longer fulfilled, as given in Table II. As result, a more realistic distance estimation is performed, which can be applied to any fine-grained localization algorithm. Applied to SDLs, the increased accuracy is shown in Figure 9. Although the dispersion is not reduced, the plausibility check increases the arithmetic mean of the algorithm in both scenarios by up to 20%.

VI. HYBRID LOCALIZATION

Unfortunately, the plausibility check is not able to tackle the outliers due to inauspicious beacon arrangements or unrecognized erroneous distance estimations. For a further improvement of the localization accuracy, we tackle extreme outliers by a comparison of the plausibility of the fine-grained algorithm result with the result of a robust coarse-grained algorithm. This hybrid approach completes our HyPAERLoc algorithm and is performed in 3 steps:

1.) Estimation of the position of an unknown node with SDLs with preceding plausibility check and with AWCL.

2.) Rating of the accuracy of the estimated position of the unknown by comparing the estimated distances to all beacons \( RATE_{SDLS} \) and \( RATE_{AWCL} \) with equation (20).

\[
RATE = \frac{x_{i,h} = x_{i,SDLS}}{RATE_{SDLS} + RATE_{AWCL}} \cdot \left( x_{AWCL} - x_{i,SDLS} \right) + \frac{y_{i,h} = y_{i,SDLS}}{RATE_{SDLS} + RATE_{AWCL}} \cdot \left( y_{AWCL} - y_{i,SDLS} \right) 
\]

Here, \( x_i \) and \( y_i \) are the estimated position of an unknown \( i \), \( x_j \) and \( y_j \) the position of the beacon \( j \), \( d_{ij} \) the estimated distance between beacon \( j \) and node \( i \) and \( n \) the number of beacons in range. As a result, a high rate correlates with bad position estimation, because the resulting position does not correlate with the estimated distances to the beacons.

3.) Computation of the final position for HyPAERLoc \( x_{i,h} \) and \( y_{i,h} \) with equations (21) and (22).

After applying equations (20) and (21), the position of the node is estimated between the position of SDLs and AWCL and the node is located nearer to the position with the lower rate. The result for the position estimation of HyPAERLoc as boxplot diagram is given in Figure 10. One can see that the outliers of SDLs are completely eliminated and also the accuracy of the median is increased. For a further comparison, the resulting mean averages of SDLs, AWCL and HyPAERLoc compared to the beacon density are given in Figure 11 for both scenarios. One can see that HyPAERLoc always outperform its preceding algorithms with round about 50% accuracy increase in average.

![Figure 8. Implausible distance estimations](A) Condition 1, (B) Condition 2. For condition 3, exchange \( j \) and \( k \) in (B)](A)

![Figure 9. Accuracy of SDLS with plausibility check in both investigated scenarios](A)

![Figure 10. Accuracy of HyPAERLoc in both investigated scenarios](A)
and how a less random beacon deployment would impact the algorithms performance and the number of required beacon nodes. A comparison to alternative localization algorithms, e.g. MDS-map [10], is also intended in the near future.

REFERENCES