Node Localization based on Anchor Placement using Fuzzy C-Means in a Wireless Sensor Network
Sidi Mohammed Hadj Irid, Mourad Hadjila, Mohammed Hicham Hachemi, Sihem Souiki, Reda Mosteghanemi, and Chaima Mostefai

Abstract—Localization is one of the oldest mathematical and technical problems that have been at the forefront of research and development for decades. In a wireless sensor network (WSN), nodes are not able to recognize their position. To solve this problem, studies have been done on algorithms to achieve accurate estimation of nodes in WSNs. In this paper, we present an improvement of a localization algorithm namely Gaussian mixture semi-definite programming (GM-SDP-2). GM-SDP is based on the received signal strength (RSS) to achieve a maximum likelihood location estimator. The improvement lies in the placement of anchors through the Fuzzy C-Means clustering method where the cluster centers represent the anchors’ positions. The simulation of the algorithm is done in Matlab and is based on two evaluation metrics, namely normalized root-mean-squared error (RMSE) and cumulative distribution function (CDF). Simulation results show that our improved algorithm achieves better performance compared to those using a predetermined placement of anchors.

Keywords—WSN; localization algorithm; anchors; GM-SDP-2; WLS; CRLB; Fuzzy C-Means; RMSE; CDF

I. INTRODUCTION

Today, the vision of wireless sensor networks has become a reality with the latest developments in wireless communication and electronics technology, which enable the development of low-cost, low-power, small size, short-range, multi-functional communication sensors. These low-cost smart sensors with wireless networks and mass deployment offer unprecedented opportunities for monitoring and controlling homes, cities and the environment. Networked sensors have a wide range of military, medical, commercial and other applications, generating new capabilities for reconnaissance and surveillance, as well as other tactical applications.

Location estimation capability is essential in most wireless sensor network applications. Wireless sensor network positioning technology is used to estimate the location of sensors with unknown locations in the network, using some prior knowledge of specific sensor locations in the measurement between the network and the sensors, such as distance, time difference of arrival, angle of arrival, and connectivity.

In WSNs, sensor nodes are deployed in the real environment to collect data from the surrounding physical environment. Once the information is collected, it is transmitted from the sensor node to the base station where the information path is displayed. In addition, knowing the geographical location of the node is very important, because if the node does not know its geographical location, any information is useless. Note that if a node judges its location incorrectly, this estimation error propagates around the world through the network and other nodes, leading to the propagation of incorrect location information to other nodes. In order to determine the location of a node, the sensor relies primarily on the distance between a node that knows the coordinates (known location) and a node that does not know the coordinates (unknown location). As a typical solution, GPS is the simplest way to locate nodes. If there are a large number of nodes in the network, this will become very expensive. The topic of low-cost localization has attracted some researchers. So far, many algorithms have been proposed in the literature [1], [2]. These positioning algorithms can be divided into two categories, as shown in Fig. 1.

Fig. 1. Localization protocols

Wireless source localization has attracted thoughtful attentions in the past decades. Among the various localization
methods, energy-based localization via received signal strength (RSS) allows simple implementation compared to other conventional technologies such as time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA). Recent advances have made energy-based localization practical in various networks, including WSNs, wireless local area network (WLAN) [3], and vehicular ad hoc networks (VANETs). Nevertheless, RSS-based localization to achieve the ML estimator of target node coordinates leads to a nonlinear and non-convex optimization problem [4]. Several methods have been proposed to address this problem.

In this paper, we will present algorithms such as WLS and GM-SDP-2 to estimate the location of sensor nodes and transmission power and then propose an improvement to decrease the error rate of these algorithms.

The rest of this paper is organized as follow: some related works are presented in Section II. Section III describes the algorithms used. Our proposed solution is discussed in Section IV where simulation results are presented and we conclude the paper in Section V.

II. RELATED WORK

Location-based systems are defined as a key technology for the development and use of wireless sensor networks. In general, sensors are deployed randomly, as they are used in inaccessible terrain, on mobile machines or at the site of a disaster where their number (very large) does not allow reconfiguring the position. For this, a localization system is needed to provide the nodes with their positions. Node localization is one of the most important services for the development and use of WSN. Several works have been devoted to this research axis.

Shuang et al. [5] propose a selective anchor node localization algorithm based on DV-hop for wireless sensor networks. DV-hop is a classical Range-Free localization algorithm, which allow to unknown nodes to get anchors’ information by estimating distances from themselves, but these latter’s may incur large error and will jeopardize the localization precision. To address this problem, the authors makes unknown node choose three anchors which have the highest precision to localize from all the anchors it received.

In [6], authors proposed a localization technique for grid environment where sensor nodes are deployed in a grid pattern and localization is achieved using a single location aware or anchor node by identifying three type of nodes: anchor nodes, unknown nodes and special nodes. Two metrics are used to evaluate the proposed approach, which are localization time and localization error. Proposed scheme has lower localization error and lower localization time in comparison with Multidimensional scaling (MDS) algorithm.

HAN et al. [7] present a comprehensive review of the recent breakthroughs in the field of mobile anchor node assisted localization algorithms (MANAL) in WSNs. They classify MANAL algorithms into two categories: localization based on mobility model and localization based on path planning scheme, and gave a comprehensive survey for the most interesting and successful advances in them.

To enable a trade-off between location accuracy and energy consumption, authors in [8] proposed a path-planning algorithm combining a Localization algorithm with a Mobile Anchor node based on Trilateration (LMAT) and SCAN algorithm (SLMAT) which ensures that each unknown node is covered by a regular triangle formed by beacons.

SINGH et al. [9] proposed a novel idea of localizing target nodes with moving single anchor node that follow Hilbert trajectory using Computational Intelligence based application of Particle Swarm Optimization (PSO) and H-Best Particle Swarm Optimization (HPSO). The accuracy of HPSO algorithm is more than PSO algorithm, also HPSO has fast convergence rate than PSO. The proposed algorithm can be used for the various applications in logistics and military.

Authors in [10] present the state of the art of localization algorithms in mobile wireless sensor networks where they classify the localization algorithms based on different key features like the localization technique, the anchor based/cooperative, the mobility in the network and the information state.

KUMARI et al. [11] give a comprehensive survey of the algorithms designed for localizing the sensor nodes in terrestrial and underwater regions. These algorithms have been classified based on the nature of anchor nodes. The localization methods have been categorized as static anchor node-based or mobile anchor node based and further range-free or range-based depending upon the number of anchor nodes availability and distance estimation technique, respectively. Authors also explain the commonly used distance estimation methods and the basic node localization techniques. Multidimensional scaling (MDS) is a prominent technique among different approaches used for Terrestrial wireless sensor networks (TWSNs) due to its accuracy, but it requires a lot of computations.

In order to improve the localization accuracy, authors in [12] propose an improved DV-Hop algorithm based on dynamic anchor node set using binary particle swarm optimization (BPSO) algorithm. A novel binary particle-coding scheme and fitness function are designed to select appropriate anchor nodes. Simulation results show that proposed algorithm has excellent localization accuracy compared with the original DV-Hop and other DV-Hop based improved algorithms.

III. ALGORITHMS DESCRIPTION

Let us denote the unknown coordinates of the target node \( j \)-th as \( \varphi_j = [\varphi_{j1}, \varphi_{j2}]^T \) \( (\varphi_j \in \mathbb{R}^2, j = 1, \ldots, M) \), and the known coordinates of the anchor node \( i \)-th as \( \alpha_i = [\alpha_{i1}, \alpha_{i2}]^T \) \( (\alpha_i \in \mathbb{R}^2, i = 1, \ldots, N) \), where \( M \) and \( N \) are the total number of targets and anchors, respectively.

A. Calculation of the received power

From [13] and [14], the received power at target \( j \)-th of anchor \( i \)-th (or vice versa) is generally modeled as [4]:

\[
P_{r,i} = P_0 - 10\beta\log_{10}\left(\frac{d(\varphi_j, \alpha_i)}{d_0}\right) + n_{r,i} \tag{1}
\]

Where, \( P_0 \) is the transmitted power at distance \( d_0 \); \( \beta \) is the path loss exponent with the common value between 2 and 6; \( d(\varphi_j, \alpha_i) \) is the Euclidean distance between target \( j \)-th of
anchor node $i$-th: $d(\varphi_j, \alpha_i) = \| \varphi_j - \alpha_i \|_2^2$; $d_0$ is the reference distance of the receiver; and $n_{i,j}$ is the additive noise following Gaussian distribution represents the log-normal observation effect in multipath environments.

### B. GM-SDP-2 algorithm

GM-SDP-2 is an improved RSS-based node location algorithm, called Gaussian Mixture Semi-Definite Programming (GM-SDP) estimator, created to achieve ML estimation of node positions in WSNs [4].

1) **Position estimation**: The design goal of the GM-SDP-2 localization algorithm is to obtain the ML estimate of the target node $\varphi^*_j$ by finding the parameter $\tau_{i,s}$. To avoid the problem of convexity and linearity we use the semi-definite relaxation for C6 (inequality constraint $\psi \geq \varphi_j \varphi_j^T$). Jensen’s inequality for C8 ($y_i = \sum_{s=1}^{S} \tau_{i,s} \xi_{i,s} \geq \sum_{s=1}^{S} \tau_{i,s} \xi_{s,i}$) and Schur’s complement for C5 and C6 [4]. The ML estimator can be formulated as [4]:

$$\min_{\varphi_j, \tau_{i,s}} \| x \|_1 \| y \|_1$$

(2)

- $C1: \sum_{s=1}^{S} \tau_{i,s} = 1, \forall i$
- $C2: 0 \leq \tau_{i,s} \leq 1, \forall s, i$
- $C3: d(\varphi_j, \alpha_i) \neq 0, \forall i$
- $C4: Tr(\psi) - \varphi_j^T \alpha_i + \| \alpha_i \|_2 \leq \gamma^2_s \sigma_s \xi_{s,i}$
- $C5: \left[ Tr(\psi) - \varphi_j^T \alpha_i + \| \alpha_i \|_2 \right] \leq \gamma^2_s \sigma_s \xi_{s,i} \geq 0, \forall i, s$
- $C6: \left[ \frac{\psi^T \varphi_j}{\varphi_j^T} \right] \geq 0, \psi \in S^2$
- $C7: x_i \geq Tr(\tau_i \eta T)$
- $C8: y_i \geq \sum_{s=1}^{S} \xi_{s,i}, \forall i, s$

Where,

$$\gamma^2_s = d_0^2 \frac{P_0 + \mu_s - P_{i,j}}{5\beta}$$

$$\tau_i = [\tau_{i,1}, ..., \tau_{i,S}]^T$$

$$\eta = \left[ \ln \sqrt{2\pi\sigma_1}, ..., \ln \sqrt{2\pi\sigma_S} \right]^T$$

$$\xi_i = [\xi_{i,1}, ..., \xi_{i,S}]^T$$

$$\max \left[ \frac{\sigma_s \gamma^2_s}{\sigma_s^2 \gamma^2_s}, \frac{\sigma_s \gamma^2_s}{\sigma_s^2 \gamma^2_s} \right] \xi_{s,i} \triangleq (\text{Chebyshev norm})$$

Noting that: $\| \varphi_j - \alpha_i \|_2^2 = Tr(\psi) - 2\varphi_j^T \alpha_i + \| \alpha_i \|_2^2$

Now the convex optimization problem can be solved by existing numerical tools [15] to obtain the globally optimal solution $\varphi^*_j$. In MATLAB, CVX [16] is used to simulate this estimation.

2) **Positioning technique with GM-SDP-2**: The objective of this algorithm is to calculate the position of the nodes from the received power (power calculation part) and the ML estimation (position estimation part) to estimate the position of the nodes ($\hat{\varphi}$). Its RMSE is defined by:

$$\text{RMSE}(e) = \sqrt{\frac{\sum_{j=1}^{M} \| \varphi_j - \hat{\varphi}_j \|_2^2}{M}}$$

(4)

### C. Cramer-Rao Lower Bound

CRLB is a lower bound on the variance of all unbiased estimators. It is well known that the Cramer-Rao Lower Bound (CRLB) which is the inverse of the Fisher information matrix (FIM), where the element $[J]_{v,r}$ of FIM $J$ is defined by [4]:

$$[J]_{v,r} = E \left[ \frac{\partial \ln(P_j|\varphi_j)}{\partial \varphi_{j,v}} \frac{\partial \ln(P_j|\varphi_j)}{\partial \varphi_{j,r}} \right], v, r \in V$$

(5)

$V$ is the set of dimensions in the coordinate axis. For a specific target node in our two-dimensional scenario ($|V| = 2$), we have:

$$[J]_{v,r} = \left[ \frac{10\beta}{\ln(10)} \right]^2 \sum_{j=1}^{N} \frac{\left( \varphi_{j,v} - \alpha_{i,v} \right) \left( \varphi_{j,r} - \alpha_{i,r} \right)}{|| \varphi_j - \alpha_i ||_2^2}, v, r \in V$$

Where,

$$I_n = E \left[ \left( \frac{\nabla \ln(p(n))}{p(n)} \right)^2 \right] = \int \frac{[\nabla \ln(p(n))^2]}{p(n)}$$

(7)

$I_n$ can be numerically estimated by Monte Carlo integration [17].

The localization estimation error is defined by:

$$e = \| \hat{\varphi} - \varphi \|$$

Its lower RMSE is bounded by:

$$\sqrt{E(e^2)} \geq \frac{1}{\sqrt{Tr(J)^{-1}}} \triangleq \text{CRLB}(\varphi)$$

### D. WLS Algorithm

The weighted least squares algorithm, also known as weighted linear regression [18], [19] is a low-complexity localization technique that owes its high accuracy to the ability to complete and approximate the EDM samples constructed from incomplete and error-perturbed information collected by the sensors [20].

1) **WLS positioning technique**: After calculating the received power, we can estimate the distance $d$ as:

$$d(\varphi_j, \alpha_i) = d_0 \times \frac{10^{-\frac{r_{i,j}}{10\beta}}}{\eta}$$

(11)

We will re-center the origin on the first anchor location to construct the matrices $H$ and $b$:

$$H = 2 \times \begin{bmatrix} \alpha_{2,1} - \alpha_{1,1} & \alpha_{2,2} - \alpha_{1,2} \\ \alpha_{3,1} - \alpha_{1,1} & \alpha_{3,2} - \alpha_{1,2} \\ \vdots & \vdots \\ \alpha_{N,1} - \alpha_{1,1} & \alpha_{N,2} - \alpha_{1,2} \end{bmatrix}$$

(12)
Where,
\[ b = \begin{bmatrix} (\alpha_{2.1} - \alpha_{1.1})^2 + (\alpha_{2.2} - \alpha_{1.2})^2 \\ (\alpha_{3.1} - \alpha_{1.1})^2 + (\alpha_{3.2} - \alpha_{1.2})^2 \\ \vdots \\ (\alpha_{N.1} - \alpha_{1.1})^2 + (\alpha_{N.2} - \alpha_{1.2})^2 \end{bmatrix} - \begin{bmatrix} d_2 - d_1 \\ d_3 - d_1 \\ \vdots \\ d_N - d_1 \end{bmatrix} \] (13)

Where, \( d_i \) is the estimated distance between the target node and the anchor \( i \) from the RSS calculation. From the variance \( V \), we obtain the matrix \( S \):

\[ S = \begin{bmatrix} V_2 + V_1 & V_1 & \cdots & V_1 \\ V_1 & V_3 + V_1 & \cdots & V_1 \\ \vdots & \vdots & \ddots & \vdots \\ V_1 & V_1 & \cdots & V_N + V_1 \end{bmatrix} \] (14)

Where,
\[ Var_{j,i} = \begin{bmatrix} d_{i,1}^2 & d_{i,1}^2 & \cdots & d_{i,1}^2 \\ d_{i,2}^2 & d_{i,2}^2 & \cdots & d_{i,2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{i,M}^2 & d_{i,M}^2 & \cdots & d_{i,M}^2 \end{bmatrix} \]

and,
\[ V_i = \frac{\sum_{j=1}^{M} (Var_{j,i} - \frac{\sum_{j=1}^{M} Var_{j,i}}{M})^2}{M} \]

The WLS solution is formulated as follows:

\[ \hat{\varphi} = (H^T \times S^{-1} \times H^{-1}) \times H^T \times S^{-1} \times b \] (15)

Its RMSE is defined by:

\[ RMSE(e) = \sqrt{\frac{\sum_{j=1}^{M} \| \varphi_j - \hat{\varphi}_j \|^2}{M}} \] (16)

IV. PROPOSED ALGORITHM

In this section, we describe our solution to improve the results of the previous simulation part. In the GM-SDP and WLS algorithms [4], the anchors have been placed in a predetermined way. We proposed a slight improvement of the above algorithms by placing the anchors in the WSN with the Fuzzy C-Means clustering method. Once the nodes are deployed in an environment, we proceed to group them in clusters according to the number of anchors to be used and the centers of these clusters constitute the anchor positions.

A. Fuzzy C-Means

Fuzzy C-Means is a data clustering technique in which each data point belongs to a cluster to some degree that is specified by a membership degree. This technique was originally introduced by Jim Bezdek in 1981 [20] as an improvement on previous clustering methods. It provides a method that shows how to group data points that populate a multidimensional space into a specific number of different clusters [21].

The FCM method is based on the minimization of the following objective function [22]:

\[ J_m = \sum_{i=1}^{D} \sum_{j=1}^{N} \mu_{i,j}^{m} \| \mathbf{x}_i - \mathbf{c}_j \|^2 \] (17)

Where, \( D \) and \( N \) represent the number of data points and the number of clusters, respectively. \( m \) is the exponent of the fuzzy partition matrix to control the degree of fuzzy overlap. Fuzzy overlap refers to the degree of fuzziness of the boundaries between clusters, i.e., the number of data points that have significant membership in more than one cluster. \( x_i \) and \( c_j \) represent the \( i \)th data point and the center of the \( j \)th cluster, respectively, and \( \mu_{i,j} \) is the degree of membership of to the \( j \)th cluster. For a given data point \( x_i \), the sum of the membership values of all clusters is equal to one.

B. Simulation results

To evaluate the performance of the studied algorithm improvement, simulations were performed by MATLAB. The objective of this improvement is to place the anchor nodes with the FCM clustering method where the center of each cluster represents the anchor placement. Upper part of Fig. 2 below shows the position anchors in the case of 4 and 8 anchors. In the case of 4 anchors and for the same deployment of the nodes, these anchors are placed respectively at the coordinates (3.75, 3.75), (11.25, 3.75), (3.75, 11.25), (11.25, 11.25) in the case of predetermined positioning while they are placed at coordinates (3.1299, 3.8411), (10.7628, 3.0481), (3.8798, 11.5938), (11.8612, 12.1947) in the case of positioning based on FCM clustering. In the case of 8 anchors and for the same deployment of the nodes, these anchors are placed respectively at the coordinates (3.75, 3.75), (11.25, 3.75), (3.75, 11.25), (11.25, 11.25), (0, 0), (15, 0), (15, 15), (0, 15) in the case of a pre-determined positioning whereas they are placed at the coordinates (2.4063, 3.0166), (12.5989, 3.6800), (2.7264, 12.9875), (10.4629, 11.2448), (2.4063, 3.0166), (15, 0), (13.0046, 13.8124), (2.7264, 12.9875) in the case of FCM clustering-based positioning. This is illustrated in the bottom part of Fig. 2.

![Fig. 2. Sensor placement with and without clustering in the 4 and 8 anchor cases.](Image)
TABLE I

<table>
<thead>
<tr>
<th>Number of anchors</th>
<th>4</th>
<th>8</th>
<th>12</th>
<th>16</th>
<th>20</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (m)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WLS</td>
<td>11,448</td>
<td>6,387</td>
<td>4,498</td>
<td>3,048</td>
<td>3,002</td>
<td>6.7%</td>
</tr>
<tr>
<td>Improved WLS</td>
<td>10,112 ↓</td>
<td>5,389 ↓</td>
<td>3,792 ↓</td>
<td>3,482 ↑</td>
<td>3,344 ↑</td>
<td></td>
</tr>
<tr>
<td>GM-SDP-2</td>
<td>3,475</td>
<td>2,397</td>
<td>2,172</td>
<td>1,716</td>
<td>1,683</td>
<td>21.08%</td>
</tr>
<tr>
<td>Improved GM-SDP-2</td>
<td>3,234 ↓</td>
<td>2,051 ↓</td>
<td>1,473 ↓</td>
<td>1,22 ↓</td>
<td>1,053 ↓</td>
<td></td>
</tr>
<tr>
<td>CRLB</td>
<td>3,132</td>
<td>1,835</td>
<td>1,317</td>
<td>1,048</td>
<td>0,915</td>
<td>12.78%</td>
</tr>
<tr>
<td>Improved CRLB</td>
<td>2,829 ↓</td>
<td>1,69 ↓</td>
<td>1,078 ↓</td>
<td>0,893 ↓</td>
<td>0,743 ↓</td>
<td></td>
</tr>
</tbody>
</table>

nodes with a number of anchors varying from 4 to 20. Nodes are randomly deployed in an area of sides 15m × 15m. Anchors are placed using Fuzzy C-Means clustering method. Since the nodes are deployed randomly in a zone, the simulations were run 5 times for each algorithm and the average value of the RMSE performance criterion was calculated based on the simulation results obtained. The same simulation parameters from [4] were used in order to make a comparison. Besides, the corresponding Cramer-Rao lower bound (CRLB) is derived for performance comparison.

Fig. 3 shows a comparison of RMSE versus anchor number for various estimators between the predetermined anchor placement and that based on the FCM clustering method.

The performance of the GM-SDP-2 algorithm was improved for all cases of anchor numbers. For the WLS algorithm, we observe an efficiency of the anchors clustering on the performance when the number of anchors is equal to 4, 8 and 12. On the other hand, the performance decreases when the number of anchors exceeds 16.

1) Evaluation based on CDF: The performance of the studied algorithms with anchor clustering is further evaluated by the CDF metric measurement.

Fig. 4 shows the CDF of the location estimation errors of various algorithms with up to 120 sensors (including 20 anchors) with fixed anchor placement and placement with clustering. This figure shows that the performance of the GM-SDP-2 estimator was improved with anchor clustering. We notice that with clustering, for example, the GM-SDP-2 reaches 99% of its errors for a range of 2m while it reaches 84% with the fixed placement of the anchors, i.e., it presents an improvement of 15%.

Table II represents a value comparison between GM-SDP-2 and Improved GM-SDP-2 in term of CDF metric.

In contrast, the CDF of estimation errors by the WLS algorithm shows a total decrease in performance such that the error to reach 100% CDF varies from 10 to 20.5 m for WLS and improved WLS respectively.
V. CONCLUSION

In this paper, we have proposed an improvement of a localization algorithm in wireless sensor networks based on the received signal strength named GM-SDP-2. This algorithm has been simulated in MATLAB with the help of the CVX modeling system. In the improved GM-SDP-2 algorithm, we proposed a modification on the anchor placement using the Fuzzy C-Means clustering method. The latter gives a better placement of the anchor nodes. The results obtained from the comparison between these two algorithms show the effectiveness of the proposed improved GM-SDP-2 protocol for WSNs, which can provide better accuracy in estimating the position of nodes in a network with an improvement of 21.08% in terms of the RMSE metric and a better performance for the evaluation based on the CDF metric.

REFERENCES


