On the Application of a Hybrid Harmony Search Algorithm to Node Localization in Anchor-based Wireless Sensor Networks

Diana Manjarres*, Javier Del Ser*, Sergio Gil-Lopez*, Massimo Vecchio†, Itziar Landa-Torres*, and Roberto Lopez-Valcarce†

* TELECOM Unit
TECNALIA RESEARCH & INNOVATION
P. Tecnológico, Ed. 202, 48170, Zamudio, Spain
{diana.manjarres,javier.delser,sergio.gil,itziar.landa}@tecnalia.com
† Departamento de Teoría de la Señal y las Comunicaciones
University of Vigo
C/ Maxwell s/n, 36310 Vigo, Spain
{massimo,valcarce}@gts.uvigo.es

Abstract—In many applications based on Wireless Sensor Networks (WSNs) with static sensor nodes, the availability of accurate location information of the network nodes may become essential. The node localization problem is to estimate all the unknown node positions, based on noisy pairwise distance measurements of nodes within range of each other. Maximum Likelihood (ML) estimation results in a non-convex problem, which is further complicated by the fact that sufficient conditions for the solution to be unique are not easily identified, especially when dealing with sparse networks. Thereby, different node configurations can provide equally good fitness results, with only one of them corresponding to the real network geometry. This paper presents a novel soft-computing localization technique based on hybridizing a Harmony Search (HS) algorithm with a local search procedure whose aim is to identify the localizability issues and mitigate its effects during the iterative process. Moreover, certain connectivity-based geometrical constraints are exploited to further reduce the areas where each sensor node can be located. Simulation results show that our approach outperforms a previously proposed meta-heuristic localization scheme based on the Simulated Annealing (SA) algorithm, in terms of both localization error and computational cost.

Keywords—Wireless Sensor Networks, localization, flip ambiguity, Harmony Search.

I. INTRODUCTION

Recently, Wireless Sensor Networks (WSNs) have emerged as an efficient and cost-effective means for monitoring physical or environmental conditions in many different scenarios. For many existent and novel WSN applications, such as vehicle or animal tracking, control of machinery, agriculture, location-aware route selection and environment monitoring, the availability of information about sensor locations is necessary in order for collected data to be meaningful [1]. However, traditional localization techniques based on installing a Global Positioning System (GPS) device on each sensor unit are not always feasible in practice, because of the excessive cost and the high power consumption of such devices. Furthermore, GPS is generally not suitable for indoor deployments. Based on this rationale, localization in WSNs is considered as an important technical challenge by the research community.

In this regard, the majority of efforts so far have been devoted to the anchor-based scenario where only some nodes of the network (referred to as anchor nodes) know in advance their coordinates (either through GPS or set beforehand). The main goal is then to obtain the position of all non-anchor nodes, assuming that each sensor can obtain (noisy) estimates of the distances to their neighboring nodes, possibly by resorting to distance-based techniques, such as the Received Signal Strength Indicator (RSSI) [2]; time-based methods, such as Time of Arrival (ToA) or Time Difference of Arrival (TDoA) [3]; Angle of Arrival (AoA) [4]; or hyperbolic trilateration [5]. It turns out that, for this framework, the Maximum Likelihood (ML) estimation paradigm results in a non-convex optimization problem, for which different approaches have been extensively proposed. These techniques can be broadly classified into centralized and distributed localization algorithms. In centralized localization algorithms, such as those proposed in [13], [6], [7], the nodes send their data to a fusion center, where they are further processed in order to solve the problem, thus reducing the computational load required at each node. On the other hand, in distributed schemes [8], [9] sensor nodes autonomously process the available distance measurements and eventually communicate with neighboring nodes to further improve their estimated positions.

In this paper we will focus on a centralized localization technique which relies on a stochastically-driven, yet intelligent, search procedure able to efficiently explore the solution space of the underlying optimization problem. This technique, despite its suboptimality, allows a computationally
efficient method for locating the nodes of a WSN with full coverage of their constituent anchor nodes. Unfortunately, as the node connectivity is reduced, the underlying network may become not uniquely localizable. In such situations the optimization problem may fail to obtain accurate solutions, due to the so-called "flip ambiguity" phenomenon, by which multiple configurations are compatible with the inter-node distance measurements [13]. The effects of the flip ambiguity can be catastrophic, from a localization point of view, when the localization algorithm relies on some previously estimated node locations that may have been flipped [12].

In this context, a recently proposed algorithm has tackled the localization problem with flip ambiguity risks, by using a two-stage Simulated Annealing (SA) optimization framework [13]. There, in the first stage SA is applied to obtain an initial estimate of the node locations, while in the second stage a refinement phase exploits the neighborhood information to identify and relocate those nodes that had been flipped in the first stage. In this paper we take a step forward in this direction by proposing an hybridization of a Harmony Search (HS) algorithm with a novel local search procedure acting at flipped node locations and exploiting certain geometrical constraints. In particular, the localization problem is formulated as the combination of two terms into a fitness function to be minimized. The first term represents the squared error between the estimated and the measured inter-node distances, whereas the second establishes a penalty for all the neighborhood violations in the estimated network topology. Some preliminary simulation results will show that the proposed approach outperforms the SA scheme, in terms of Normalized Localization Error (NLE) and computational complexity.

The remaining of the paper is structured as follows: Section II formally defines the node localization problem, whereas Section III presents the proposed hybrid HS algorithm. Next, Section IV shows a simulation-based comparison between our hybrid HS approach and the SA algorithm in [13]. Finally, in Section V we draw some concluding remarks.

II. PROBLEM STATEMENT

The node localization problem can be formulated as follows: given $n$ nodes uniformly deployed in $T = [0, 1] \times [0, 1] \subset \mathbb{R}^2$, from which the first $m$ nodes (with $m < n$) represent the anchor nodes and whose coordinates $p_i = (x_i, y_i) \in \mathbb{R}^2$ ($i = 1, \ldots, m$) are fixed and known a priori, we are interested in estimating the positions of the remaining $n-m$ non-anchor nodes of the network as $p_i = (\hat{x}_i, \hat{y}_i)$, $\forall i \in \{m+1, \ldots, n\}$. We define a binary connectivity matrix $C$ such that each entry $c_{ij} = 1$ iff sensor nodes $i$ and $j$ are within the connectivity range of each other (i.e., $r_{ij} \leq R$, where $r_{ij}$ denotes the real distance between nodes $i$ and $j$ and $R$ is the circular transmission range common to any sensor node). Without loss of generality, we assume that $C$ is available as a priori information for the localization algorithm, as each node can determine which other nodes it can communicate with. The measured inter-node distances $d_{ij}$, which can be obtained by resorting to any of the measurement techniques mentioned in Section I, can be modeled as

$$d_{ij} = \begin{cases} r_{ij} & \text{if } (i, j) \in \{1, \ldots, m\} \times \{1, \ldots, m\}, \\ r_{ij} + e_{ij} & \text{otherwise}, \end{cases}$$

(1)

where $r_{ij} \triangleq ||p_i - p_j||$, and $e_{ij}$ denotes the measurement error. Let us define the set of nodes which are neighbors of node $i$ as

$$N_i \triangleq \{ j \in \{1, \ldots, n\}, j \neq i : r_{ij} \leq R \},$$

(2)

and the corresponding complementary set $\overline{N}_i$ containing the nodes located outside the connectivity range of node $i$. It is worth to observe that the positions of the anchor nodes, together with the connectivity range and the connectivity matrix $C$, can be exploited to define regions of the network where some non-anchor nodes can (or cannot) be located. Indeed, those nodes inside the coverage area of the generic anchor node $i \in \{1, \ldots, m\}$ must lie in the circular area centered in $p_i = (x_i, y_i)$ and with radius $R$, whereas all nodes with no connection with any anchor node must be located outside the union of the circular areas of radius $R$ and centered in all anchor nodes. This information is exploited by the proposed algorithm during both the generation of the initial population and the iterative refining process.

Our goal is to accurately estimate the non-anchor positions by minimizing a global metric, which is the sum of two objective functions. The first function, namely Cost Function (CF), refers to the squared error between the estimated and the measured inter-node distances between nodes within range of each other, and is defined as

$$\text{CF} \triangleq \sum_{i=m+1}^{n} \left( \sum_{j \in N_i} (d_{ij} - \hat{d}_{ij})^2 \right),$$

(3)

where $d_{ij}$ and $\hat{d}_{ij} \triangleq \sqrt{(\hat{x}_i - \hat{x}_j)^2 + (\hat{y}_i - \hat{y}_j)^2}$ represent the measured and estimated distance between node $i$ and $j$, respectively. The second objective function, namely Soft Constraint Violation (SCV), takes into accounts the connectivity neighborhood violations of the non-anchor nodes in each candidate topology. In words, if a node $j$ has been placed in the neighborhood of node $i$, while $j \in \overline{N}_i$ or, equivalently, if a node $j$ has been placed in a position such that $\hat{d}_{ij} > R$, while $j \in N_i$, then it is likely to have a constraint violation and a constant error term $(\hat{d}_{ij} - R)^2$ is added to SCV. Formally,

$$\text{SCV} \triangleq \sum_{i=1}^{n} \left( \sum_{j \in N_i, \hat{d}_{ij} > R} (\hat{d}_{ij} - R)^2 + \sum_{j \in \overline{N}_i, \hat{d}_{ij} < R} (\hat{d}_{ij} - R)^2 \right).$$

(4)
It is important to remark that as argued in [13], the term \((\hat{d}_{ij} - R)^2\) represents the minimum error due to a localization flip. Finally, in order to assess the quality of the final estimate, the Normalized Localization Error (NLE) is calculated as

\[
\text{NLE} = \frac{1}{R} \left( \frac{1}{(n-m)} \sum_{i=m+1}^{n} ||\mathbf{p}_i - \hat{\mathbf{p}}_i||^2 \times 100, \quad [%]\right)
\]

where clearly, a value of NLE close to 0% would correspond to the asymptotic case of error-free estimation.

III. PROPOSED HARMONY SEARCH ALGORITHM

The Harmony Search algorithm, originally formulated in [14], is a meta-heuristic population-based optimization procedure which mimics the improvisation process of musicians when seeking the best harmony. In essence, this technique iterates on a set of \(K\) candidate solutions \(\{(\hat{\mathbf{p}}_i^k)_{i=1}^m\}_{k=1}^K\) (referred to as harmonies or melodies) to the optimization problem at hand, to whose compounding elements \((\hat{x}_i^k, \hat{y}_i^k)\) (correspondingly, notes) several combinatorial operators are applied so as to progressively refine their associated fitness. This refinement procedure is repeated until a maximum number of iterations \(\mathcal{I}\) is reached. The HS algorithm has been so far utilized for a wide variety of communication-related optimization problems, e.g. multicast routing [15], engineering design [16], multiuser detection [17], [18], or radio resource allocation in OFDMA networks [19]. To the authors’ knowledge, this work embodies the first attempt to apply HS to node localization.

The steps of the proposed HS-based localization algorithm can be summarized as follows:

1) The initialization process is only considered at the first iteration. At this step, the first \(m\) notes \(\{(\hat{x}_i^k, \hat{y}_i^k)\}_{i=1}^m\) corresponding to the anchor nodes are filled, \(\forall k = 1, \ldots, K\), with the actual node positions \((x_i, y_i)\) of such anchor nodes. The remaining \(n - m\) notes are chosen at random, again \(\forall k = 1, \ldots, K\), from the areas of the network resulting from the constraints imposed by their connectivity to the \(m\) anchor nodes.

2) At each iteration of the algorithm, the improvisation process is applied sequentially to each note \(\{(\hat{x}_i^k, \hat{y}_i^k)\}_{i=m+1}^{K}\) of the total set of melodies. This improvisation method is driven by three different probabilistic parameters:

- The Harmony Memory Considering Rate, HMCRR \(\in [0, 1]\), sets the probability that the new value for a certain note \((\hat{x}_i^k, \hat{y}_i^k)\) is drawn uniformly from the values of this same note in all the remainder \(K - 1\) melodies.
- The Pitch Adjusting Rate, PAR \(\in [0, 1]\), refers to the probability that the new value for a given note \((\hat{x}_i^k, \hat{y}_i^k)\) is taken randomly from its coverage area. Geometrical constraints posed by anchor nodes for the sensor at hand are also considered in this operator.
- The Random Selection Rate, RSR \(\in [0, 1]\), establishes the probability to pick the value for the new note \((\hat{x}_i^k, \hat{y}_i^k)\) randomly from the subset \(T_i \subset \mathcal{T} = [0, 1] \times [0, 1]\), which is defined by the intersection of all geometrical constraints established by the coverage region of anchor nodes. Notice that, as opposed to the PAR procedure, the RSR parameter operates network-wide, i.e. an estimated position for a node could be moved beyond its coverage area.

3) An additional local search method aims at improving the fitness of the candidate with potentially lowest metric value\(^1\). This local search procedure is carried out for every node \(i\) 1) lying outside the coverage region of any anchor node; and 2) whose any of its neighbors in the estimated topology violates the connectivity constraints imposed by the \(i\)-th row of \(\mathbf{C}\). In this case, those anchor nodes located within the connectivity range \(R\) of its real neighbors (if any) are first selected. The node at hand is then moved to the intersection of the annuli with inner and outer radii \(R\) and \(2R\) respectively, centered in these selected anchor nodes, under the condition that the number of suspicious\(^2\) neighbors decreases. The rest of real neighbors, which are not connected to any anchor, are randomly set inside the coverage region centered on the new location of this node.

4) The evaluation of the new generated candidate solutions and the update of the harmony memory are performed at each iteration based on the global metric function CF+SCV. To this end, only those harmonies improving the fitness of those from the previous iteration are included in the next harmony memory. Once this has been done, the harmony memory is sorted in ascending order of the fitness values of its compounding melodies.

5) After a fixed number of iterations \(\mathcal{I}\), the algorithm is halted, and the set of estimated positions is given by the first melody \(\{(\hat{x}_i^1, \hat{y}_i^1)\}_{i=1}^n\) in the harmony memory.

IV. SIMULATION RESULTS

In order to assess the performance of the proposed approach, this section presents some preliminary, though encouraging, simulation results and a comparison with the SA-based localization technique proposed in [13]. In all experiments, the inter-sensor distance measurement error is modeled as Gaussian-distributed with zero-mean and variance \(\sigma^2\).

\(^1\)In the first iteration no metric evaluation has been done at this point. Nevertheless, the local search is applied to the first melody in the memory.

\(^2\)Fake neighbors in the estimated network that violate the connectivity constraints posed by \(\mathbf{C}\).
First, for the sake of completeness, we briefly summarize the SA-based approach used as comparison. SA is a stochastic optimization algorithm inspired to the physical process of temperature annealing in metallurgy. In contrast with gradient-based search methods that aim at performing each optimization step using the idea of steepest descent, SA probabilistically allows uphill perturbations, thus accepting worse candidate solutions to avoid getting stuck in local minima. At the beginning, a control temperature $T_c$ is initialized at a high value to allow an initial random search of the space. Then, as the iteration counter $i$ increases, $T_c$ and the perturbation distance $\Delta D$ are slowly decreased according to $T_c(i) = \alpha \cdot T_c(i-1)$ and $\Delta D(i) = \beta \cdot \Delta D(i-1)$, respectively (with $\alpha < 1$ and $\beta < 1$). At each temperature value, the SA-based algorithm perturbs $(n - m) \cdot P \cdot Q$ random selected non-anchor nodes positions and evaluates the perturbed configuration, where $P$ and $Q$ are operational parameters to be set beforehand. If the perturbed estimation is characterized by a better fitness value, then it is accepted, otherwise its acceptance is subject to the probability $\exp\{-\frac{\Delta(CF)}{T_c(i)}\}$, where $\Delta(CF)$ represents the difference between the current and previous values of the metric function. As the number of iterations increases, the temperature cools and fewer worse candidate estimations are allowed, leading to a progressive refinement of the final solution.

The SA localization approach presented in [13] performs a two-stage optimization procedure. In the first phase, a naïve SA is used to estimate the positions of the non-anchor nodes minimizing the objective function defined in (3). The first phase ends once the value of this function is below a predefined threshold, or an arbitrary final temperature $T_{c,f}$ is reached. Then, the goal of the refinement phase is to identify the non-uniquely localizable nodes that cause flip ambiguities, and elevate to virtual anchor nodes those that keep their correct neighborhood in the estimated network deployment. In this stage, simulated annealing is performed again in order to minimize a new cost function, which is formulated as

$$CF_{SA} \triangleq \sum_{i=m+1}^{n} \left( \sum_{j \in N_i} (\hat{d}_{ij} - d_{ij})^2 + \sum_{j \in N_i, \hat{d}_{ij} \leq R} (\hat{d}_{ij} - R)^2 \right).$$

Table I

<table>
<thead>
<tr>
<th>Parameters used in the simulations.</th>
<th>SA in [13]</th>
<th>Proposed HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{c,i}$</td>
<td>0.1</td>
<td>HMCRC</td>
</tr>
<tr>
<td>$T_{c,f}$</td>
<td>10^{-11}</td>
<td>0.9</td>
</tr>
<tr>
<td>$P$</td>
<td>100</td>
<td>PAR</td>
</tr>
<tr>
<td>$Q$</td>
<td>2</td>
<td>RSR</td>
</tr>
<tr>
<td>$D_0$</td>
<td>0.1</td>
<td>$K$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.80</td>
<td>50</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.94</td>
<td>$I$</td>
</tr>
<tr>
<td>$96$</td>
<td>2000</td>
<td></td>
</tr>
</tbody>
</table>

Table II

<table>
<thead>
<tr>
<th>NLE statistics obtained by the HS-based and SA-based algorithms.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
</tr>
<tr>
<td>0.13</td>
</tr>
<tr>
<td>0.15</td>
</tr>
<tr>
<td>0.17</td>
</tr>
</tbody>
</table>

We assess the performance of the proposed HS algorithm over the three network topologies in terms of NLE defined in (5), and compare the results with the ones achieved by the SA-based algorithm, for the different values of $R$. Table II summarizes the results in the format best/mean/std NLE value (averaged over 20 experiments). The first thing to observe is that the mean and standard deviation of the NLE obtained by the HS-based algorithm are significantly lower with respect to the ones achieved by the two-stage SA-based approach. Furthermore, the respective best NLE values (i.e., the minimum) are quite similar. In light of these preliminary results, one may conclude that the localization estimation obtained by SA is on average much less accurate and stable with respect to the one achieved by HS. This conclusion is further buttressed by a Wilcoxon two-sided rank sum test performed on all the obtained result sets, which verifies that the medians of the NLE distributions corresponding to HS and SA are statistically different with a confidence level of 95%.

As a further performance assessment, we compare the number of fitness evaluations computed by the two approach, in order to characterize them in terms of computational complexity. SA performs $(n - m) \cdot P \cdot Q$ fitness evaluations at each iteration of the first optimization phase, whereas in the refinement phase the number of fitness evaluations can not be determined in advance, as the number of non-anchor nodes elevated to virtual anchor nodes is not predictable. Anyway, we have verified that the average number of fitness case are based on an exhaustive optimization study for a large combination of the parameter set.

To compare the performance of both algorithms, we have built three different network topologies by uniformly placing $n = 200$ nodes in a square region of $1 \times 1$. A fraction (10%, hence $m = 20$) of these nodes corresponds to the anchor nodes set, whose fixed positions are known a priori. Different connectivity ranges $R \in \{0.13, 0.15, 0.17\}$ are considered to model different network sparsity levels. For each scenario, 20 Monte Carlo runs of the algorithms are executed. The noisy distance measurements defined in (1) are assumed to be based on RSSI, which is generally affected by log-normal shadowing [20]. Therefore, the variance $\sigma^2$ of the error $e_{ij}$ is given by $\lambda^2 r_{ij}^2$, where $\lambda$ is set to 0.1 in all simulations. At this point it is also important to note that the proposed HS local search procedure is applied every 100 iterations, rather than at each iteration; we have verified that this complexity-reduction strategy does not degrade the performance of the algorithm.

Anyway, we have verified that the average number of fitness
evaluations needed by SA to run to completion over the 20
trials of each network topology is 710000. On the other
hand, HS employs a fixed number of \( I = 2000 \) iterations,
at each of which the objective function is evaluated for each
of the \( K = 50 \) newly improvised melodies. Therefore, the
number of fitness evaluations equals \( K \cdot I = 100000 \) in all
cases, thus drastically reducing the computational load with
respect to the SA counterpart (in a approximated 7:1 ratio).

Figure 1. Node positioning for HS with \( R = 0.15 \) and for the experiment
corresponding to the minimum value of NLE.

In order to visually show the improvement in the ob-
tained results, we depict the estimated network deployment
obtained for \( R = 0.15 \) by both algorithms in Figs. 1-3.
The following notation is adopted: the anchor nodes are
represented with diamonds (♦) and their coverage areas
with empty circles. The real non-anchor nodes positions
are plotted with filled triangles (▼) and connected to the
estimated positions (represented by empty squares □) by
solid lines. Fig. 1 shows the estimated positions of non-
anchor nodes obtained by the HS-based algorithm for the
experiment characterized by the minimum value of NLE
(10.695%). On the other hand, Figs. 2-3 depict the estimated
node deployment for the highest values of NLE achieved by
SA (112.49%) and HS (37.69%), respectively. The reader
may observe that HS, jointly with the proposed local search
procedure, is able to alleviate the flip ambiguity effects
and to estimate the positions of isolated non-anchor nodes
significantly better than the SA approach.

Figure 2. Node positioning for HS with \( R = 0.15 \) and for the experiment
corresponding to the maximum value of NLE.

Figure 3. Node positioning for SA with \( R = 0.15 \) and for the experiment
corresponding to the maximum value of NLE.

V. CONCLUSION

This paper has presented a novel iterative localization
technique for anchor-based WSN, based on hybridizing a
Harmony Search algorithm with a local search procedure to
mitigate the so-called flip ambiguity phenomenon. Besides,
the algorithm takes into account the geometrical constraints
imposed by the connections between nodes to further limit
the zones of the network where some sensor nodes can be
located. We have shown that the proposed approach outperforms a previously proposed algorithm based on simulated annealing, in different transmission ranges, in terms of the localization error, stability and computational complexity. Current ongoing research is being conducted towards extending this work to localization based on multi-objective criteria.

ACKNOWLEDGMENTS

This work has been supported in part by the Spanish Ministry of Science and Innovation through the CONSOLIDER-INGENIO 2010 (CSD200800010) and the Torres-Quevedo (PTQ-09-01-00740, PTQ-06-01-0159) funding programs.

REFERENCES


