MDIS: A Node Localization Algorithm Based on Multi-region Division and Similarity Matching

Hao Wang[1], Haixiang Chen[1], Ping Yuan[2], Feng Yin[3], Qian Luo[4], and Liangyin Chen[1(✉)]

[1] School of Computer Science, Sichuan University, Chengdu, China
124324651@qq.com, 1309924894@qq.com, chenliangyin@scu.edu.cn
[2] School of Mathematics and Information Engineering,
Chongqing University of Education, Chongqing, China
pypingyuan@163.com
[3] School of Computer Science and Technology,
Southwest University for Nationalities, Chengdu, China
yf_eagle@aliyun.com
[4] The Second Research Institute, General Administration
of Civil Aviation of China, Chengdu, China
caacsri_luoqian@163.com

Abstract. This paper proposes an improved localization algorithm called MDIS (Multi-Region Division In Shadow). The algorithm divides the traditional overlapping communication regions constructed by neighbor anchor nodes (nodes whose locations are known) into many subregions, and coordinates of centroid points in these subregions could be calculated. Furthermore, the algorithm establishes relation arrays of the unknown node to anchor nodes and centroid points to anchor nodes. Then the location of unknown nodes in subregions is determined by calculating the correlation coefficient of above relation arrays. And simulations demonstrate that MDIS algorithm increases the accuracy and stability compared to other range-free algorithms.

Keywords: Multi-Region Division · MDIS · Relation array
Correlation coefficient

1 Introduction

Sensor nodes are usually randomly deployed and perform a lot of monitoring tasks in different environment in Wireless Sensor Networks (WSNs) [1]. And location information is important for sensor nodes which monitor environmental objectives [2]. In the field of node localization, sensor nodes can be divided into two categories, one is the anchor node (location is known) whose coordinates are usually obtained by GPS or manual deployment. The other is the unknown node (location is unknown), which needs to estimate its coordinates with the

information provided by surrounding anchor nodes. At present, there are two kinds of localization algorithms. One is range-based algorithm that need measure distance among nodes. The other is range-free location algorithm which only completes locating through the angle or hop information without distance measurement. Recent range-based localization algorithms [3] are mostly influenced by the environment, and it leads to large locating error and high algorithm complexity. However, one problem that cannot be ignored in the range-free localization algorithm [4] is also the large locating error. So it is proved that there is still a lot of research space for the localization method, which is worthy of further excavation and exploration. In order to reduce the locating error without raising costs, this paper proposes a range-free localization algorithm by studying the current localization algorithms. The main innovations and contributions are as follows.

(1) This paper proposes MDIS model to implement a novel range-free localization algorithm that determines the unknown node in which subregion through correlation coefficient calculated by relation arrays of the unlocated node to its neighbor anchor nodes and centroid points to these anchor nodes.
(2) The program is designed to verify the rationality of the algorithm by changing Anchor Node Rate, Number of Nodes, GPS Error and Communication Radius of the Anchor Node in simulation scenarios.

2   Related Work

In general, RSSI, TOA, TDOA and AOA are commonly used in range-based localization algorithms. And they calculate location of nodes by using the method of triangulation and maximum likelihood estimation. Some common range-free localization algorithms are introduced as follows. Convex algorithm regards the communication connection between nodes as the geometric constraint problem, and it can build a node location constraint model to locate nodes by using the connection information between nodes. Bounding Box algorithm [5] is to determine the one hop overlapping communication area, which is constituted by neighbor anchor nodes of the unknown node, then the location of centroid point in the region is considered as the position of the unknown node. DV-hop is based on distance vector routing [6]. APIT algorithm [7] is assumed that an unknown node can receive N signals from all anchor nodes. And it finds out all the combinations of three anchor nodes from N anchor nodes. Determine which triangle the unknown nodes are inside and mark such triangles. Centroid points are connected to form a polygon and then centroid point's coordinates of the polygon is regarded as location of the unknown node. And Centroid algorithm regards the centroid point of the polygon constructed by the unknown node's neighbor anchor nodes as the location [8].

3   Model Design

Node localization algorithms still exist many problems, because it is easily influenced by environment. And the cost of range-based location algorithms is still

high and location efficiency is low. So MDIS algorithm introduced in this chapter is to improve the traditional range-free algorithm, unknown nodes are located more accurately without increasing the locating cost and eliminating the signal influence indirectly.

3.1   Background

Before introducing the algorithm, it is necessary to explain the concepts of the algorithm. And two concepts called relation array and correlation coefficient in this paper are introduced.

Definition 1. *Relation Array is a one-dimensional array which represents the distance array of one node to its neighbor anchor nodes.*
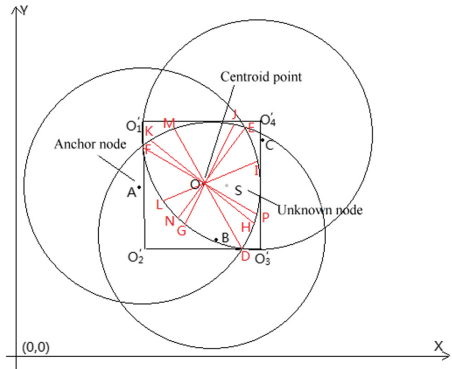
Definition 2. *Correlation Coefficient denotes the degree of distance between nodes, and the greater correlation coefficient, the closer location between nodes. The interval of the correlation coefficient of nodes is* $[-1, 1]$.

$$\cos\theta = \frac{\sum_{i=1}^{n}(A_i \times B_i)}{\sqrt{\sum_{i=1}^{n}A_i{}^2} \times \sqrt{\sum_{i=1}^{n}B_i{}^2}} \qquad (1)$$

$A = (A_1, A_2, \cdots, A_n), B = (B_1, B_2, \cdots, B_n).$
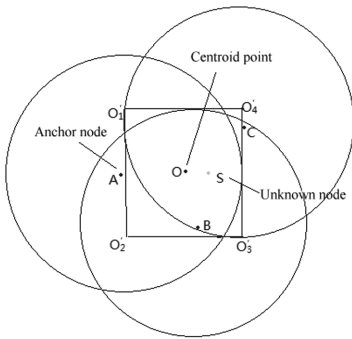
Fig. 1. Convex model

Fig. 2. MDIS model

And there are two theoretical foundations of this paper. One of the theoretical bases is from the convex algorithm, the convex algorithm regards the communication connection between nodes as the geometric constraint problem and it can build a node location constraint model to locate nodes through using the neighbor information among sensor nodes. In Fig. 1, A, B and C are three anchor nodes, S is the unknown node and O_1', O_2', O_3', O_4' are the four vertices of a restricted rectangle with overlapping communication areas of three anchor

nodes. At this time, the centroid point O of $O_1'O_2'O_3'O_4'$ approximately regarded as the location of the unknown node S. Another theoretical basis is from cosine similarity formula. For two vectors, their similarity can be measured by the cosine of the angle.

In Fig. 1, there are two unknown nodes R_1, R_2 and three anchor nodes A, B, C. Meanwhile, the distances of R_1 away from A, B and C are L_{1A}, L_{1B}, L_{1C} respectively. The distances of R_2 to A, B and C are L_{2A}, L_{2B}, L_{2C} respectively. Then the relation array of R_1 is $relation - array_1(L_{1A}, L_{1B}, L_{1C})$ and the relation array of R_2 is $relation - array_2(L_{2A}, L_{2B}, L_{2C})$. So the correlation coefficient of R_1 and R_2 can be denoted by $\frac{L_{1A} \times L_{2A} + L_{1B} \times L_{2B} + L_{1C} \times L_{2C}}{\sqrt{L_{1A}^2 + L_{1B}^2 + L_{1C}^2} \times \sqrt{L_{2A}^2 + L_{2B}^2 + L_{2C}^2}}$.

3.2   MDIS Model

The core of MDIS model is to divide the overlapping communication area of anchor nodes into multiple parts. Draw parallel lines of AB, AC, BC over point O and intersect the area DEF with G, H, I, J, K, L, then connect the D, E and F to O and intersect the area DEF with M, N, P. Through the above method, the overlapping area DEF can be divided into twelve subregions in Fig. 2. In this way, the relatively complex location problems are transformed into the problem of finding the best location subregion. According to the circle equations, the intersection coordinates of three circles can be calculated. The coordinates of D, E and F can be denoted by $(x_d, y_d), (x_e, y_e)$ and (x_f, y_f). According to the feature of the circular tangent, the coordinates of O_1, O_2, O_3, O_4 are denoted by $(x_f, y_b + r), (x_f, y_d), (x_a + r, y_d)$ and $(x_a + r, y_b + r)$. Then the coordinates of O can be calculated as $(x_f + \frac{x_a + r - x_f}{2}, y_d + \frac{y_b + r - y_d}{2})$ according to characteristics of the quadrilateral centroid point. Because HK, LI, GJ are parallel to AB, AC, BC respectively, so the slope of HK, LI and GJ can be expressed as $k_{HK} = k_{AB} = \frac{y_b - y_a}{x_b - x_a}, k_{LI} = k_{AC} = \frac{y_c - y_a}{x_c - x_a}, k_{GJ} = k_{BC} = \frac{y_c - y_b}{x_c - x_b}$. The slope of DM, EN and FP can be expressed as $K_{DM} = \frac{y_d - (y_d + \frac{y_b + r - y_d}{2})}{x_d - (x_f + \frac{x_a + r - x_f}{2})}, K_{EN} = \frac{y_e - (y_d + \frac{y_b + r - y_d}{2})}{x_e - (x_f + \frac{x_a + r - x_f}{2})}, K_{FP} = \frac{y_f - (y_d + \frac{y_b + r - y_d}{2})}{x_f - (x_f + \frac{x_a + r - x_f}{2})}$. In summary, equations of HK, LI, GJ, DM, EN, FP can be expressed. Associated with the circular equation, the coordinates of all intersections of lines and the overlapping area DEF can be calculated.

Each subregion is a graphic with a curved edge, in order to find the centroid coordinates of each subregion, the following conversion is needed. Because the subregion is small in reality, we can approximate the curved side of the subregion into a line so that the subregion is converted into a triangular region, as shown in Fig. 3. Then the problem of solving the centroid coordinate of the subregion is transformed into the centroid coordinate problem of solving the triangle. If the centroid point of the triangle OGD is O_1, the coordinates of the centroid point of the triangle OG can be expressed as $(\frac{1}{3} \times (x_O + x_G + x_D), \frac{1}{3} \times (y_O + y_G + y_D))$. So centroid points of all subregions can be calculated, and relation arrays can be denoted as $(d_{ia}, d_{ib}, d_{ic})_{12 \times 3}, d_{ij} = \sqrt{(x_{o_i} - x_j)^2 + (y_{o_i} - y_j)^2}, i \in [1, 12], j \in \{a, b, c\}$. For the unknown node S, although its location is unknown, node S can
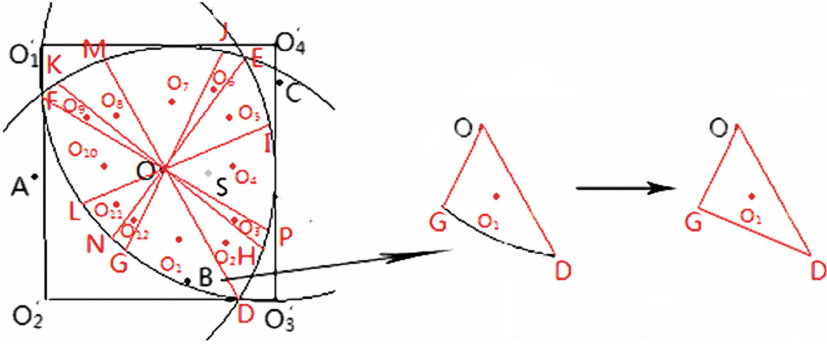
Fig. 3. Subregion approximation into triangle

obtain the signals sent by A, B, C and can calculate the RSSI. According to the RSSI distance decay model formula $d = \frac{|RSSI|-A}{10 \times n}$, where d denotes the distance between receiving sensor node and sending sensor node, and $RSSI$ is received signal strength. A is signal strength when the distance between the sender and the receiver is $1m$, and n is the environmental attenuation factor. The relation array of unknown node S can be denoted as $(10^{\frac{|RSSI_a|-A}{10 \times n}}, 10^{\frac{|RSSI_b|-A}{10 \times n}}, 10^{\frac{|RSSI_c|-A}{10 \times n}})$, where $RSSI_a, RSSI_b, RSSI_c$ represent the signal strength of S to receive the anchor nodes A, B and C respectively. Thus, formula 1 is applied to compute the correlation coefficient of the unknown node S and each centroid point of twelve subregions. According to the experimental situation, the above mentioned location area DEF is relatively small area in the actual situation, and the interference is same for any point in area DEF. So it is reasonable to denote the distance of two nodes by using RSSI signal attenuation model. According to formula 1 and the relation array, the correlation coefficient matrix between the unknown point S and each centroid point of all subregions can be denoted as,

$$r_c = \left(\frac{10^{\frac{|RSSI_a|-A}{10 \times n}} \times d_{ia} + 10^{\frac{|RSSI_b|-A}{10 \times n}} \times d_{ib} + 10^{\frac{|RSSI_c|-A}{10 \times n}} \times d_{ic}}{\sqrt{d_{ia}^2 + d_{ib}^2 + d_{ic}^2} \times \sqrt{(10^{\frac{|RSSI_a|-A}{10 \times n}})^2 + (10^{\frac{|RSSI_b|-A}{10 \times n}})^2 + (10^{\frac{|RSSI_c|-A}{10 \times n}})^2}} \right)_{12 \times 1} \quad (2)$$

The larger the correlation coefficient between the unknown node S and the centroid point of a subregion, the closer their location is. Then the location of the centroid point with the largest correlation coefficient can be regarded as the location of S. When the number of anchor nodes whose signals could be received by node S is 2, its location can be estimated as the coordinates of the centroid point of the bounding box for communication circles of these two anchor nodes. When S could only receive signals from one anchor node, its location can be estimated as the location of this anchor node. When there is no neighbor anchor node around node S, S cannot be located. In summary, the location coordinates of the node can be computed by using formula 3,

$$L_S = \begin{cases} unsure, & \text{n} = 0 \\ (x_a, y_a), & \text{n} = 1 \\ (\frac{x_a+x_b}{2}, \frac{y_a+y_b}{2}), & \text{n} = 2 \\ (x_{o_i}, y_{o_i}), & r_{so_i} = ||r_c||_\infty, \text{n} > 2 \end{cases} \qquad (3)$$

where n is the number of neighbor anchor nodes for S, and r_{SO_i} denotes correlation coefficient of S and O_i.

3.3 MDIS Algorithm

(1) Initialize the entire wireless sensor network, the maximum value of communication radius is r. For node k, its neighbor discovery process can use neighbor discovery algorithms [9]. When it finds its neighbor node, the set of neighbor anchor nodes is marked as $\{A(k)\}$, where $\{A(k)\} = (\cup_{i=1}^{a_k} A_i)$, and the number of neighbor anchor nodes is a_k.
(2) For a node k, when the number of neighbor anchor nodes is less than 3, the coordinates of node k can be calculated according to formula 3. And when the number of its neighbor anchors is equal or greater than 3, the coordinates of the centroid point in each subregion can be estimated by selecting 3 nodes with strongest signal strength of k neighbor anchor nodes to construct MDIS model.
(3) Calculate relation arrays of the unknown node and centroid points of all subregions, and calculate correlation coefficient matrix of the relation array of the unknown node and the relation array of the centroid point of each subregion. At last, the location of the centroid point with the largest correlation coefficient is estimated as the location denoted as $L_k(x_k, y_k)$.
(4) Take turns to locate other nodes.

4 Simulation

In this section, each experiment runs for 20 times and then calculates the average value as the result. In order to make the experimental simulation more realistic, the communication radiation model is constructed in the simulation algorithm, and nodes are deployed in $500\,\text{m} \times 500\,\text{m}$ area randomly. The degree of irregularity (DOI) is also considered in the simulation. And the main simulation parameters are Anchor Node Rate, Number of Nodes, GPS Error and Communication radius of the Anchor Node (Table 1).

(1) The impact of anchor node rate (Figs. 4 and 5)
(2) The impact of the number of nodes (Figs. 6 and 7)
(3) The impact of GPS Error (Figs. 8 and 9)
(4) The impact of anchor node communication radius (Figs. 10 and 11)
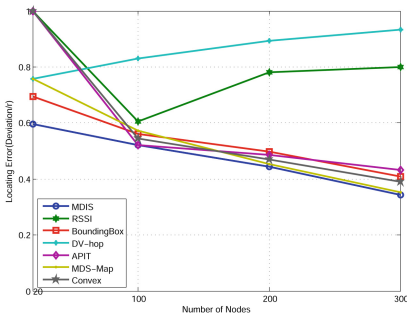
Table 1. Simulation parameters

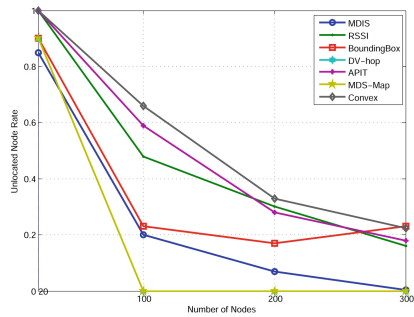| Parameter	Value
Length of area	$500\,\mathrm{m}$
Width of area	$500\,\mathrm{m}$
Number of nodes	0–300
Anchor node rate	0.1–0.2
DOI	0.015
GPS error	10–$40\,\mathrm{m}$
Radius of the unknown node (r)	$r = r_{max}(1 - DOI), r_{max} = 100\,\mathrm{m}$
Radius of the anchor node (r_a)	$r_a = \beta r,\ \beta \in [1, 4]$
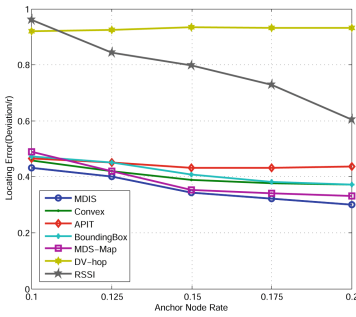
Fig. 4. Impact on locating accuracy
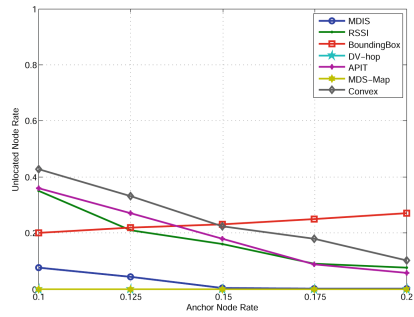
Fig. 5. Impact on unlocated node rate

Fig. 6. Impact on locating accuracy

Fig. 7. Impact on unlocated node rate
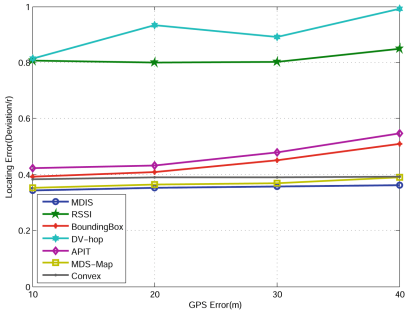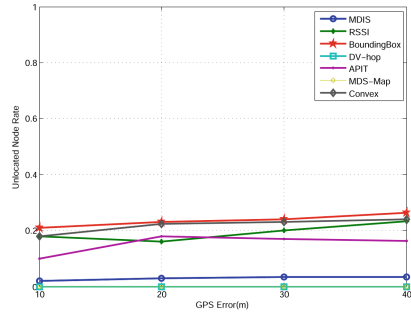
Fig. 8. Impact on locating accuracy

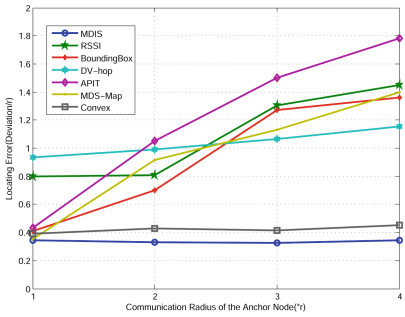**Fig. 9.** Impact on unlocated node rate
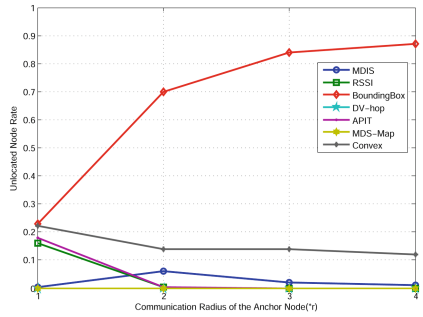

Fig. 10. Impact on locating accuracy

Fig. 11. Impact on unlocated node rate

5   Conclusion

In conclusion, the main idea of MDIS algorithm is to segment the overlapping localization area. And then position of the unknown node can be estimated by constructing relation arrays of nodes and calculating correlation coefficient of nodes. According to (1), (2), (3), (4) of Sect. 4, MDIS algorithm is applied in locating scenarios with different parameters. Compared to other range-free and RSSI based localization algorithms, the locating error of MDIS algorithm is the smallest, and the unlocated node rate is less than 10%. In the future, security of nodes will be considered in this algorithm by using hop count information.

References

1. Chen, L., Xiong, X., Liu, K., et al.: POPA: localising nodes by splitting PLA with NTH-HOP anchor neighbours. Electron. Lett. **50**(13), 964–966 (2014). https://doi.org/10.1049/el.2013.4046

2. Rodriquez, R.Y., Julcapoma, M.R., Jacinto, R.A.: Network monitoring environmental quality in agriculture and pisciculture with low power sensor nodes based on ZigBee and GPRS technology. In: IEEE Xxiii International Congress on Electronics, Electrical Engineering and Computing, pp. 1–6. IEEE (2017). https://doi.org/10.0.4.85/INTERCON.2016.7815578

3. Qiang, R., Wang, W., Wang, H., et al.: 3D maximum likelihood estimation positioning algorithm based on RSSI ranging. In: Advanced Information Technology, Electronic and Automation Control Conference, pp. 1311–1314. IEEE (2017). https://doi.org/10.1109/IAEAC.2017.8054226

4. Shahzad, F., Shaltami, T., Shakshukhi, E.: DV-maxHop: a fast and accurate range-free localization algorithm for anisotropic wireless networks. IEEE Trans. Mob. Comput. **PP**(99), 2494–2505 (2017). https://doi.org/10.1109/TMC.2016.2632715

5. Simic, S.N., Sastry, S.: Distributed localization in wireless ad hoc networks (2001). https://doi.org/10.1145/1653760.1653768

6. Niculescu, D., Nath, B.: DV based positioning in ad hoc networks. Telecommun. Syst. **22**(1–4), 267–280 (2003). https://doi.org/10.1023/A:1023403323460

7. He, T., Huang, C., Blum, B.M., et al.: Range-free localization schemes for large scale sensor networks. In: International Conference on Mobile Computing and Networking, pp. 81–95. ACM (2005)

8. Kim, K.Y., Shin, Y.: A distance boundary with virtual nodes for the weighted centroid localization algorithm. Sensors **18**(4), 1054 (2018). https://doi.org/10.3390/s18041054

9. Wei, L., Zhou, B., Ma, X., et al.: Lightning: a high-efficient neighbor discovery protocol for low duty cycle WSNs. IEEE Commun. Lett. **20**(5), 966–969 (2016). https://doi.org/10.1109/LCOMM.2016.2536018