Journal of Network Intelligence
Taiwan Ubiquitous Information

Volume **7**, Number **3**, August **2022**

Three-dimensional Localization Algorithm for WSN Nodes Based on Hybrid RSSI and DV-Hop

Jun-Yi He, Xi-Yun Hu, Chao-Chao Hu*, Meng-Xiang Wu, Min Zhang

School of Computer Science and Technology
Henan Polytechnic University
Jiaozuo 454000, China
hejunyi@hpu.edu.cn, huxiyun@home.hpu.edu.cn, 211909020013@home.hpu.edu.cn, zmngd@163.com
*Corresponding Author: Chao-Chao Hu

Received March 24, 2022, revised April 22, 2022, accepted June 10, 2022.

Abstract. *Aiming at the low localization accuracy of wireless sensor network (WSN) nodes with large-scale low-density anchor nodes, a three-dimensional localization algorithm for WSN nodes that mix RSSI and DV-Hop is proposed. Firstly, the genetic tabu numerical optimization hybrid algorithm (GANV) is proposed to solve the problems that the numerical optimization algorithm is easy to fall into local optimum solution, and the genetic tabu algorithm has poor local searchability to obtain the global optimization solution. Secondly, RSSI ranging technology is introduced into the node location part of the DV-Hop algorithm combined with RSSI, and the GANV algorithm solves the initial value. Finally, in the RSSI location algorithm combined with DV-Hop, the location problem of each node is reduced to an unconstrained nonlinear non-convex optimization problem. After the initial value is input, it is solved by the GANV algorithm, and the location estimation is obtained after eliminating outliers by the clustering algorithm. Simulation results show that compared with the traditional node location algorithm in WSN, this algorithm has high location accuracy and is very suitable for large-scale WSN engineering applications with low real-time requirements and high location accuracy requirements.*
Keywords: wireless sensor network; RSSI; DV-Hop; optimization algorithm; three-dimensional localization

1. **Introduction.** With the aggravation of environmental changes and frequent natural disasters, monitoring the geological environment in key areas (such as volcanoes, prone landslides, etc.) is necessary. At the same time, with the promotion of agricultural greenhouse automation, urban informatization and other projects, the demand for WSN is gradually increasing because it can monitor the target area and transmit the data to the observer in time [1, 2]. When the data collected by WSN is closely related to the location, the node location algorithm provides essential support for the practical application of WSN. In the outdoor environment, the anchor node can obtain the three-dimensional localization of the node through the high-precision GPS module and barometer. However, the cost of the GPS module and barometer is high, the power consumption is large, and the density of the anchor node is directly proportional to the deployment cost. According to whether to measure the distance between nodes, the localization algorithm is divided into a ranging localization algorithm and a non-ranging localization algorithm. Typical ranging algorithms include AOA, DTOA and RSSI algorithms [3], in which RSSI ranging does not need synchronization and additional hardware equipment, and the cost is low. The ranging free algorithm mainly relies on the topology of WSN and the connectivity

between nodes to estimate the distance between nodes without additional equipment, but the localization accuracy is low [4].

In the location algorithm based on RSSI, at least four anchor nodes are required to obtain the three-dimensional localization of unknown nodes through the four side location method or maximum likelihood estimation method. The RSSI ranging accuracy is inversely proportional to the distance between nodes. Therefore, many anchor nodes need to be deployed to obtain the high-precision location of nodes, and the cost is high. In this regard, in order to reduce the density of anchor nodes and reduce the cost, for large-scale WSN, the WSN is divided into several sub-regions by using unknown nodes and their communication distance, and the non-convex objective function is constructed by using the ranging data between unknown nodes in the sub-region and the ranging data between unknown nodes and anchor nodes. At this time, in [5], the convex relaxation method is used to approximate the non-convex objective function, and the optimal gradient method is used for convergence, but this method needs the unknown node to be located in the convex hull of the anchor node in order to obtain high localization accuracy. For the problem in [5], the improved Newton method based on Hesse matrix correction is adopted in [6] to converge to the optimal solution quickly. However, it is easy to fall into the local optimum solution when there is a significant error in the initial value and has high requirements for each node's storage and operation performance. The DV-Hop algorithm has low requirements for node performance. It is a widely used non-ranging localization algorithm, but there is a problem of low localization accuracy in WSN with uneven node distribution [7]. In order to improve the localization accuracy of the DV-Hop algorithm and make full use of RSSI ranging data between nodes, researchers have made many improvements based on the DV-Hop algorithm. For example, in [8], RSSI and polynomial approximation are used to estimate the distance between unknown nodes and anchor nodes, and recursive calculation is used to improve the accuracy of location estimation in the process of location. In [9], RSSI ranging technology is combined with the DV-Hop algorithm. After quantifying the number of hops between nodes by RSSI ranging, the average hop distance of anchor nodes and unknown nodes is corrected, and RSSI ranging is used to correct the location estimation of unknown nodes. In [8, 9], the localization accuracy of the improved DV-Hop algorithm combined with RSSI is higher than that of the traditional DV-Hop algorithm, but it still can not meet the requirements of the high-precision location.

In order to further improve the node localization accuracy in WSN with large-scale low-density anchor nodes and reduce the node performance requirements, a three-dimensional WSN node localization algorithm combining RSSI and DV-Hop is proposed by making full use of the RSSI ranging data between nodes, the topology of WSN and the connectivity between nodes without adding additional equipment. Our main contributions are as follows:

(1) In order to obtain the global optimization solution of the unconstrained nonlinear non-convex optimization problem and improve the localization accuracy, the numerical optimization algorithm has robust local searchability. However, it is easy to fall into the local optimum solution, while the genetic taboo algorithm has robust local searchability. It has global searchability, but its local search ability is poor. We combine the advantages of the two algorithms and propose the GANV algorithm;

(2) In order to obtain a higher-precision initial value, a DV-Hop localization algorithm combined with RSSI is proposed. In the unknown node localization estimation part, the RSSI ranging technology is introduced, and the least-squares method is used to obtain a one-step localization estimate, and then the GANV algorithm is used to obtain the higher precision two-step localization estimation;

(3) In order to obtain high-precision localization data, an RSSI localization algorithm combined with DV-Hop is proposed. After reducing the localization problem of each node to an unconstrained nonlinear non-convex optimization problem, the two-step localization estimation is used as the initial value. Use the GANV algorithm to obtain the three-step localization estimation. Because each unknown node has multiple three-step localization estimation results, after using the clustering algorithm to eliminate outliers, the mean value is used as the final localization estimation;

(4) In order to reduce the performance requirements of nodes, a node-server two-stage data processing method is proposed.

2. **Related Work.** The three-dimensional DV-Hop localization algorithm is implemented based on the two-dimensional DV-Hop algorithm proposed by Niculescu et al. [10]. Based on the distance vector routing principle. Its primary localization process is as follows:

1. Gets the minimum number of hops between nodes. In WSN, anchor nodes propagate location data, hop value and number to the whole WSN through flooding and distance vector exchange protocol. The initial hop value of each anchor node is 0. When the node receiving the information records multiple hop values from the same anchor node, it ignores the more significant hop value, saves the minimum hop value, and then broadcasts the hop value plus 1 to other neighbour nodes until each node in the WSN obtains the minimum hop value between it and each anchor node.

2. Calculate the distance between the unknown node and the anchor node. Each anchor node calculates its average jump distance through Equation (1) according to the minimum jump value and coordinates to other anchor nodes recorded in the first step.

$$\text{Hopsize}_i = \frac{\sum_{i \neq j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}}{\sum_{i \neq j} h_{ij}} \tag{1}$$

Where, h_{ij} is the minimum number of hops between anchor nodes i and j, (x_i, y_i, z_i) and (x_j, y_j, z_j) is the coordinates of anchor nodes i and j.

After obtaining the average hop distance through Equation (1), the anchor node broadcasts its average hop distance to the WSN. At this time, the average hop distance of the unknown node is the average hop distance information of the first anchor node received. According to the minimum jump value recorded by the unknown node in step 1, calculate the jump distance between it and the anchor node through Equation (2).

$$d_{iu} = \text{Hopsize}_i h_{iu} \tag{2}$$

Where, d_{iu} is the skip distance between the anchor node i and the unknown node u; h_{iu} is the minimum number of hops between the anchor node i and the unknown node u; Hopsize_i is the average hop distance of the anchor node i, and indicating that the first average hop distance data received by the unknown node u comes from the anchor node i.

3. Solve unknown node location information. In step 2, the unknown node has obtained the jump distance between it and each anchor node. The unknown node localization is obtained by solving Equation (3) by the least square method, where d_m is the distance between the unknown node and the anchor node m.

$$\begin{cases} (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = d_1^2 \\ (x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 = d_2^2 \\ \vdots (m \geq 4) \\ (x - x_m)^2 + (y - y_m)^2 + (z - z_m)^2 = d_m^2 \end{cases} \tag{3}$$

2.1. **WSN node location method based on RSSI.** The traditional three-dimensional node localization methods based on RSSI, such as the quadrilateral localization method and weighted centroid method, generally need at least four anchor nodes within the communication range of unknown nodes to assist in localization. Because RSSI ranging accuracy is greatly affected by the distance between nodes, when deploying large-scale WSN, many anchor nodes need to be deployed to ensure the localization accuracy of nodes, resulting in a sharp rise in equipment and later maintenance costs. As shown in Figure 1, A, B, C, D, E and F are anchor nodes. Moreover, M1, M2, and M3 are unknown nodes. The dotted line circle is the effective range of unknown nodes, and the solid line circle is the communication range of unknown nodes. The three unknown nodes in the figure are affected by the number of anchor nodes within the ranging range and cannot obtain three-dimensional localization estimation. At this time, the distance data between unknown nodes within the communication range of unknown nodes and the ranging data between anchor nodes and unknown nodes can be fully used to construct the constraint relationship, and the optimization method can be used to obtain the three-dimensional localization estimation of unknown nodes that minimizes the overall ranging error.

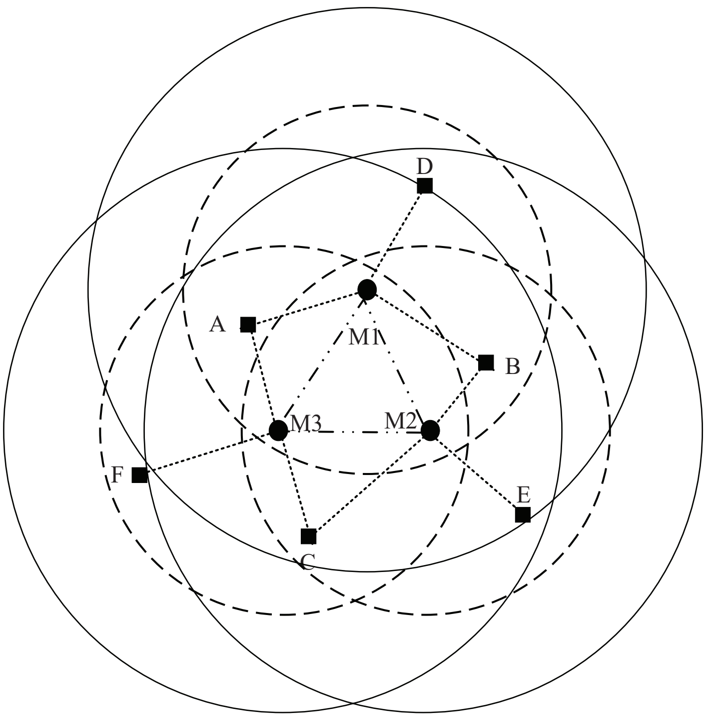

FIGURE 1. Schematic diagram of WSN node location based on RSSI

3. Method.

3.1. **The architecture of the algorithm.** In practical application, the effective communication distance of node d_c is greater than the effective ranging distance d_v. As the distance between nodes increases, the ranging error increases gradually. The range will be seriously distorted when the distance reaches a certain degree. For this, in order to ensure a sure ranging accuracy, set d_v. When the set communication distance d is less than d_v, the ranging range is d. When $d_v < d < d_c$communication distance is d, ranging range is d_v. The range of communication distance d in this paper is $d_v < d < d_c$.

Considering the complexity of the algorithm, to reduce the requirements for node operation performance, this paper adopts the idea of two-stage data processing. In the first stage, all data acquisition and processing are completed by each node. Firstly, the RSSI and distance measurement values of each unknown node and each node within its range are obtained according to the range scheme in Section 3.2. Secondly, for each unknown node, the modified average hop distance and the minimum quantitative hop number between anchor nodes relatively close to it are obtained according to the [9], and the hop distance between each anchor node and each unknown node is calculated. Finally, the processed data is uploaded to the server through the communication node. In the second stage, the data processing is completed in the server. Firstly, the two-step localization estimation of each unknown node is obtained through the DV-Hop algorithm combined with RSSI in Section 3.4. Then, the final location estimation of each unknown node is obtained through the RSSI location algorithm combined with DV-Hop in Section 3.5. At this time, the overall architecture of the algorithm is shown in Figure 2.

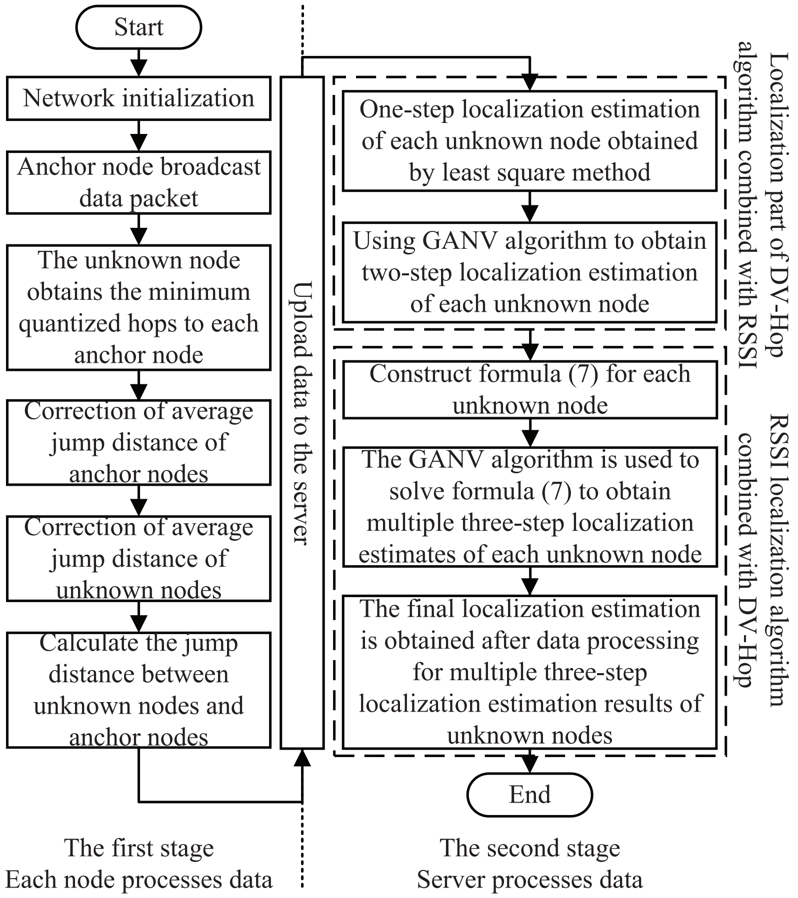

Figure 2. Schematic diagram of the overall architecture of the algorithm

3.2. **RSSI ranging scheme.** In RSSI location, the location accuracy of unknown nodes is greatly affected by the quality of RSSI observation data. In order to improve the ranging accuracy, RSSI data is collected 100 times between the two nodes, which is inconsistent with the reality because there will be fluctuations and coarse outliers in the RSSI data. In order to make the data conform to the objective authenticity, the RSSI data processing method in [11] is adopted: firstly, the Schoville method is used to eliminate outliers; Then, the moving average smoothing filter is used to process the data to suppress the data fluctuation. After obtaining the RSSI value between nodes, the distance between nodes can be obtained according to the ranging model, dramatically impacting the ranging accuracy. Because the Bessel function ranging model proposed in [11] has higher ranging accuracy than the shading model and segmented ranging model, the Bessel function ranging model is adopted.

3.3. **Genetic tabu Numerical optimization hybrid algorithm (GANV).** In order to improve the localization accuracy, an intelligent optimization algorithm is designed to obtain the global optimization solution of the nonlinear non-convex least-squares problem. The genetic algorithm searches for the optimal solution by simulating the natural evolution process. It has high global searchability, but its local search ability is poor, which is easy to produce premature convergence problems. Tabu algorithm is a search method for jumping off local optimization, which has fast searchability. Based on the idea of tabu search, the genetic tabu algorithm [12] proposes a tabu crossover operator and tabu mutation operator with memory function, which makes the algorithm have a more robust ability to jump out of the local optimum solution and improve the probability of obtaining the global optimization solution. However, its local search ability is still poor, the search results are easy to swing near the optimal solution, and the convergence speed is slow. At present, numerical optimization algorithms mainly include traditional nonlinear optimization algorithms (steepest descent method, Newton method, etc.), Gauss-Newton method, Levenberg Marquardt method, etc.; they all have a solid ability to obtain local optimum solutions. However, due to the influence of initial value error, the probability of obtaining global optimization solutions is low. In this section, using the characteristics of the tabu crossover operator and tabu mutation operator in the genetic tabu algorithm to obtain the global optimization solution with high probability in the whole variable space and the characteristics of the improved Newton method based on unit step size in [6] to converge to the local optimum solution quickly, GANV algorithm is proposed. The specific implementation process is as follows:

Step1. Floating-point coding and initial population generation

Floating-point coding is usually used in multi-parameter optimization problems because it has higher speed and accuracy than binary coding. For the unknown node location problem with n unknown nodes, a vector $X = [x_1, y_1, z_1, ..., x_n, y_n, z_n]$ with length $3n$ is used to represent the location estimation scheme of unknown nodes in the location problem, and each variable type in the vector$X$ is a floating-point. In the multi-parameter optimization problem, if each variable takes the 3-axis value range of the location area as the constraint range, the speed and accuracy of the algorithm will be significantly reduced in a large-scale WSN location with a prominent location area. In this regard, each unknown node's initial value vector of localization estimation is passed in. The length is $3n$, the maximum error e and average error $\bar{e}$ of the initial localization estimation value. At this time, the value range of X is set to $[p - e, p + e]$.

The initial population consists of N individuals, of which $N-1$ individuals are generated by random method, and one individual is the initial value vector p of incoming localization estimation.

Step2. Fitness function

When obtaining the location estimation of unknown nodes, the objective function is f generally the minimization function, for which the fitness function is $F = 1/f$.

Step3. Selection operator

In order to ensure that the next generation can inherit individuals with higher fitness than the average fitness, and some individuals with lower fitness can be randomly inherited to the next generation to improve population diversity, the non-playback remainder random selection strategy is adopted to select the operator. The specific process is as follows: 1) After calculating the survival expectation number $n_i = N \cdot F_i / \sum_{j=1}^{N} F_i$ of each individual, the integer part of n_i is the survival number of the corresponding individual in the next generation group, and a total of $\sum_{i=1}^{N} \lfloor n_i \rfloor$ individuals in the next generation group can be determined. 2) After each individual's fitness is equal, the uncertain $N - \sum_{i=1}^{N} \lfloor n_i \rfloor$ individuals in the next generation group are randomly determined by the roulette method.

Step4. Cross taboo

The process of replacing and reorganizing part of the structure of the parent individual to generate a new individual is the crossover, which can improve the search efficiency of the genetic algorithm. First, the adaptive crossover rate P_1 is obtained according to Equation (4):

$$P_1 = \begin{cases} \frac{k_1(F_{\max} - F)}{F_{\max} - F_{\text{avg}}}, F_{\text{avg}} \leq F \\ k_2, F_{\text{avg}} > F \end{cases} \tag{4}$$

Where, k_1 and k_2 are constants between $[0, 1]$, respectively, $F_{\max}$ representing the maximum fitness of individuals in the population, F is the maximum fitness of two crossed individuals, F_{avg} representing the population's average fitness. If the crossover probability P_1 of each individual is less than the random number of each individual, a new random value can be generated between each individual by the crossover probability of each individual. Then, according to the idea of tabu search, after establishing a tabu table with a length L of to record the individual fitness value and setting the desired level of the value as F_{avg}, calculate the new individual fitness. The new individual whose fitness is greater than or equal to the desired level and less than the desired level but does not belong to taboo enters the next generation. For the new individuals who are less than the desired level but belong to taboo, the individuals with the most excellent fitness in the current population are selected to enter the next generation. Finally, the taboo table is updated with the accepted new individual or the individual with the most incredible fitness among the parents.

Step5. Variation taboo

The mutation of individuals can effectively prevent the algorithm from falling into local convergence in the later stage. Firstly, obtain the adaptive mutation rate P_2 according to Equation (5):

$$P_2 = \begin{cases} \frac{k_3(F_{\max} - F(X))}{F_{\max} - F_{\text{avg}}}, F_{\text{avg}} \leq F(X) \\ k_4, F_{\text{avg}} > F(X) \end{cases} \tag{5}$$

Where, k_3 and k_4 are constants between $[0, 1]$ and $F(X)$ is individual fitness of X. Secondly, a random number is generated from $[0, 1]$ for each individual. If the random number is less than P_2, a new individual is generated for its variation. The variation process is as follows: The multi-point mutation method is adopted, and n mutation points are randomly selected because the constraint range of each variable is the maximum error of the initial value of the localization estimation of each unknown node. Generally, the error in each direction in the three-dimensional coordinate is mainly near the average error $\bar{e}$ of the initial value of the localization estimation of the unknown node. In this regard, the Gaussian mutation operator is used for the ith mutation point to generate a random number with mean $p_i (0 < i \leq 3n)$, variance $\bar{e}^2$, and within the constraint range to replace the original variable value. Then, calculate the fitness value of the new individual. The new individual whose fitness value is greater than or equal to the desired level and less than the desired level but does not belong to taboo enters the next generation. For the new individual who is less than the desired level but belongs to taboo, select the individual with the most prominent fitness in the current population to enter the next generation. Finally, the taboo table is updated with the accepted new individual or the individual with the most excellent fitness among the parents.

Step6. Numerical optimization operator

A random number from $[0, 1]$ is generated for each individual in the population. Suppose the corresponding random number of the individual is less than the probability P_3, the individual is taken as the iterative initial value of the function f, and the numerical optimization algorithm is used to search for the local optimum solution (new individual). Due to the significant initial value error of iteration, in order to quickly converge to the local optimum solution, according to the [6], after calculating the gradient vector and Hesse matrix of the objective function and correcting the Hesse matrix to positive definite, the improved Newton method based on unit step size is used to obtain the local optimum solution. After obtaining the new individual, because the numerical optimization algorithm is an excellent local extremum algorithm, the new individual better inherits the excellent quality of the parent individual, makes the new individual directly replace the parent individual in the next generation, and updates the tabu table with the accepted new individual.

Step7. 3D localization optimization

After the localization estimation of each unknown node is obtained by the genetic tabu numerical optimization hybrid algorithm, it is used as the iterative initial value of the function f, and the lewenberg Marquardt algorithm is used for an accurate iteration to obtain a more accurate localization estimation. The algorithm flow chart is shown in Figure 3.

3.4. **DV-Hop algorithm combined with RSSI.** In order to improve the localization accuracy, a DV-Hop algorithm combined with RSSI is proposed to obtain the initial values required for the location estimation using the genetic taboo-numerical optimization hybrid algorithm. In WSN, the nodes are generally unevenly distributed, and the coverage structure is irregular, which leads to errors in the number of hops and average hop distance between nodes. The main reason for this error is: If two nodes are adjacent, the hop value is one regardless of the distance; The jump distance of anchor nodes calculated by Equation (1) is a mean value, but due to the uneven distribution of nodes, the distance between adjacent nodes is generally different. There is an error in estimating the jump distance between nodes using the mean value and the minimum jump value; The unknown node uses the first received average hop distance from the anchor node as its average hop distance, but a single anchor node cannot reflect the whole network. In this regard, the methods of RSSI based hop quantization, anchor node average hop distance correction and unknown node average hop distance correction in [9] are used to obtain the modified average hop distance of each unknown node and the minimum quantitative hop number between each unknown node and each anchor node. Since the number of hops quantization based on RSSI in [9] requires that the communication distance is equal to the ranging distance, the communication distance of the algorithm in this section is equal to d_v.

Because the RSSI ranging accuracy is higher than the DV-Hop distance estimation accuracy in the short-range, if there is an anchor node $i (i = 1, 2, ..., n_1)$ in the ranging range of the unknown node s with a coordinate (x_s, y_s, z_s), the RSSI ranging distance d_{si}^{rssi} is used as the distance estimation between the unknown node s and the anchor node

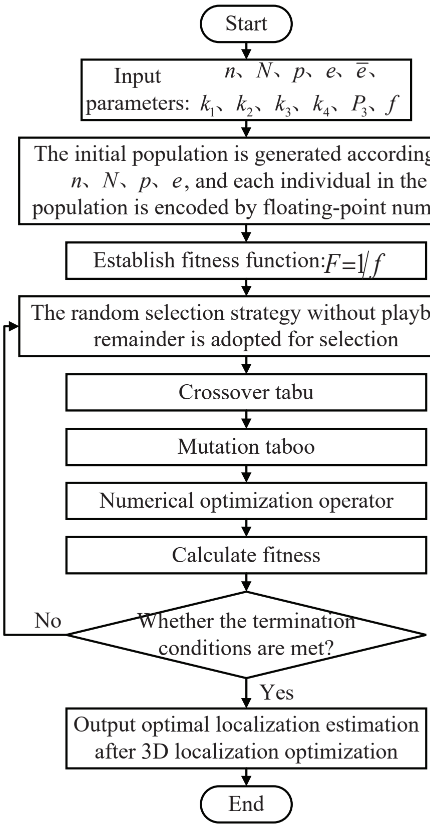

FIGURE 3. Flow chart of genetic tabu numerical optimization hybrid algorithm

i. If $n_1 \geq 4$, then, only use the unknown node s, the anchor node i and d_{si}^{rssi} construction Equation (3) and set $n_2 = 0$ in Equation (6), the corresponding objective function is obtained. If $n_1 < 4$, the average hop distance is corrected by an unknown node s and the minimum number of quantization hops between it and the anchor node $j(j = 1, 2, ..., n_2)$ except the anchor node i, to estimate the hop distance d_{sj}^{dv-hop} between the unknown node s and the anchor node j. At this time, the unknown node s, anchor node i, d_{si}^{rssi}, anchor node j and d_{sj}^{dv-hop} are used to build Equation (3), and Equation (6) is the objective function at this time.

$$f(x_s, x_s, x_s) = \min\left(\sum_{i=1}^{n_1} \left| \sqrt{\begin{array}{c}(x_s - x_i)^2 + (y_s - y_i)^2 + \\ (z_s - z_i)^2 - (d_{si}^{rssi})^2\end{array}} \right| + \sum_{j=1}^{n_2} \left| \sqrt{\begin{array}{c}(x_s - x_j)^2 + (y_s - y_j)^2 + \\ (z_s - z_j)^2 - (d_{sj}^{dv-hop})^2\end{array}} \right| \right) \tag{6}$$

After the one-step localization estimation of the unknown node s is obtained by solving Equation (3) by the least square method, taking the one-step localization estimation as the initial value, the genetic tabu numerical optimization hybrid algorithm is used to solve Equation (6) to obtain a more accurate two-step localization estimation of the unknown node s.

3.5. **RSSI location algorithm combined with DV-Hop.** In order to reduce the amount of calculation and improve the localization accuracy, when obtaining the location of the unknown node s, select the unknown node within its communication distance d, each unknown node within the communication distance (including the unknown node s)

and all anchor nodes within the ranging distance to construct the three-dimensional localization model of the unknown node. For the convenience of representation, the positions of all unknown nodes (including unknown node s) within the communication range of unknown node s are $p_1, p_2, ..., p_i, ..., p_m$, respectively, where $p_i = [x_i, y_i, z_i]^{\mathrm{T}}$. The coordinates of all anchor nodes within the ranging distance of each unknown node (including unknown node s) within the communication distance are $a_1, a_2, ..., a_o, ..., a_n$, respectively, where $a_o = [x_o, y_o, z_o]^{\mathrm{T}}$. RSSI ranging distance with the error between unknown nodes is d_{ij}. d_{io} represents the RSSI ranging distance with the error between the unknown node and the anchor node. $(i, j) \in A$ represents the set of measurable distances between unknown nodes. $(i, o) \in B$ represents the set of measurable distances between unknown nodes and anchor nodes. At this time, the RSSI ranging distance with error is used to transform the unknown node location problem into an unconstrained, highly nonlinear and non-convex optimization problem:

$$\min_{p_1,p_2,\ldots,p_m} \Big(\sum_{(i,j)\in A} \omega_{ij}^2 \left| \|p_i - p_j\|_2^2 - (d_{ij})^2 \right|^2 + \sum_{(i,o)\in B} \omega_{io}^2 \left| \|p_i - a_o\|_2^2 - (d_{io})^2 \right|^2 \Big) \tag{7}$$

Where, ω_{ij} and ω_{io} are the normalized weights based on the inverse distance ratio. Since the reliability of ranging between nodes decreases with the increase of the distance between nodes, the smaller the weight should be. At this time, the weights are:

$$\omega_{ij} = \frac{d_{ij}^{-1}}{\sum\limits_{(i,j)\in A} d_{ij}^{-1} + \sum\limits_{(i,o)\in B} d_{io}^{-1}} \tag{8}$$

$$\omega_{io} = \frac{d_{io}^{-1}}{\sum\limits_{(i,j)\in A} d_{ij}^{-1} + \sum\limits_{(i,o)\in B} d_{io}^{-1}} \tag{9}$$

Using $p = [p_1^{\mathrm{T}}, p_2^{\mathrm{T}}, ..., p_m^{\mathrm{T}}]^{\mathrm{T}}$, it represents the column vector composed of the coordinates of all unknown nodes (including node s) within the communication distance of the unknown node s, and $a = [a_1^{\mathrm{T}}, a_2^{\mathrm{T}}, ..., a_n^{\mathrm{T}}]^{\mathrm{T}}$ is the column vector composed of the coordinates of all anchor nodes. The $2i - 1$, $2i$ and $2i + 1$ columns of the $3m \times 3m$ identity matrix are respectively expressed as e_{i1}, e_{i2}, e_{i3}. The $2o - 1$, $2o$ and $2o + 1$ columns of the $3n \times 3n$ identity matrix are respectively expressed as e_{o1}, e_{o2}, e_{o3}. At this time, $x_i = e_{i1}^{\mathrm{T}} p$, $y_i = e_{i2}^{\mathrm{T}} p$, $z_i = e_{i3}^{\mathrm{T}} p$, $x_o = e_{o1}^{\mathrm{T}} a$, $y_o = e_{o2}^{\mathrm{T}} a$, $z_o = e_{o3}^{\mathrm{T}} a$. For this, Equation (7) can be written as:

$$\begin{aligned} f(p) = \min_{p_1,p_2,\ldots,p_m} \Big(&\sum_{(i,j)\in A} (p^{\mathrm{T}} A p - (d_{ij})^2)^2 \\ &+ \sum_{(i,o)\in B} (p^{\mathrm{T}} B p - p^{\mathrm{T}} C a - a^{\mathrm{T}} D p + a^{\mathrm{T}} E a - (d_{io})^2)^2 \Big) \end{aligned} \tag{10}$$

Where:

$$A = (e_{i1} - e_{j1})(e_{i1} - e_{j1})^{\mathrm{T}} + (e_{i2} - e_{j2})(e_{i2} - e_{j2})^{\mathrm{T}} + (e_{i3} - e_{j3})(e_{i3} - e_{j3})^{\mathrm{T}} \tag{11}$$

$$B = e_{i1} e_{i1}^{\mathrm{T}} + e_{i2} e_{i2}^{\mathrm{T}} + e_{i3} e_{i3}^{\mathrm{T}} \tag{12}$$

$$C = e_{i1} e_{o1}^{\mathrm{T}} + e_{i2} e_{o2}^{\mathrm{T}} + e_{i3} e_{o3}^{\mathrm{T}} \tag{13}$$

$$D = e_{o1} e_{i1}^{\mathrm{T}} + e_{o2} e_{i2}^{\mathrm{T}} + e_{o3} e_{i3}^{\mathrm{T}} \tag{14}$$

$$E = e_{o1} e_{o1}^{\mathrm{T}} + e_{o2} e_{o2}^{\mathrm{T}} + e_{o3} e_{o3}^{\mathrm{T}} \tag{15}$$

At this time, the initial localization of each unknown node can be obtained by the quadrilateral localization method, and the lewenberg Marquardt algorithm is used to solve Equation (10). However, in practical application, when the density of anchor nodes is small or unevenly distributed, there are no more than three anchor nodes or unknown nodes with known positions near some unknown nodes. In this regard, about [6], when there are $1 \sim 4$ anchor nodes near the unknown node, the nearest anchor node localization of the unknown node is taken as the initial position; When there is no anchor node near the unknown node, the centre of the localization area is taken as the initial position. However, the initial localization error obtained in this way is significant, and Equation (10) is a highly nonlinear and non-convex optimization problem, which leads to the decrease in solution speed and makes it easy to fall into the local optimum value. In this regard, after obtaining the two-step localization estimation of each unknown node through Section 3.4, take it as the initial localization of each unknown node, and use the genetic tabu numerical optimization hybrid algorithm to solve Equation (10) to obtain the three-step localization estimation of the unknown node and the unknown node within its communication distance.

Because each unknown node obtains a three-step localization estimation of the unknown node within its communication distance, each unknown node has multiple three-step localization estimation results. According to the centralized location algorithm experiment based on RSSI in [13], the location estimation error of the unknown node at the edge of the location area is significant, and the location estimation error of the unknown node at the centre of the location area is small. When using this algorithm, there may be some unknown nodes in the edge area within the communication range of the unknown node. Therefore, there is a significant error in each unknown node's multiple three-step localization estimation results. In this paper, the density-based clustering algorithm (DBSCAN) [14] is used to eliminate the outliers in multiple three-step localization estimation results of unknown nodes. After eliminating outliers, the average of the three-step localization estimation results is used as the final localization estimation of unknown nodes.

4. **Simulation results and analysis.** In order to verify the localization performance of this algorithm, the DV-Hop algorithm, Jiang et al.'s algorithm [6], Ren and Pan's algorithm [9] algorithm and this algorithm are simulated and compared by MATLAB. Data processing environment: Xeon e5-2678v3 processor and 16g ram. Simulation environment: sensor nodes are randomly distributed in the region; The value range of node quantity is; The value range of anchor node proportion is; The effective ranging distance is 20m; The value range of communication distance is 20m $\sim$ 40m; In the actual environment, the wireless signal is affected by the environment and has problems such as reflection, multipath propagation and background interference, resulting in different ranging errors. The ranging scheme in the [11] can only suppress some errors. In order to simulate the authenticity, Gaussian noise with the mean value of 0 and variance of 1 is added to the Bessel function ranging model in [11], as shown in Equation (16):

$$\mathrm{RSSI}(dis) + X_\sigma = k_1 \mathrm{J}_0(k_2 \cdot dis) + k_3 \mathrm{J}_1(k_4 \cdot dis) \tag{16}$$

Where, dis is the distance of signal propagation; $\mathrm{RSSI}(d)$ is the RSSI value; J_1 and J_0 are the 0-order and 1-order function expressions of Bessel functions of the first kind. The unknown parameters k_1, k_2, k_3, k_4 of the ranging model are set as -32.3, 0.095, 150.6 and 0.042, respectively, in the experiment.

In order to show the robustness of the algorithm, five experiments were carried out randomly every time the experimental parameters were changed. In addition, taking the average localization error e and normalized average localization error $\bar{e}$ as the localization performance index, its definition is shown in Equation (17):

$$\begin{cases} e = \frac{1}{KN} \sum\limits_{j=1}^{N} \sum\limits_{i=1}^{K} |q - \hat{q}| \\ \bar{e} = \frac{e}{d} \end{cases} \tag{17}$$

Where, q and $\hat{q}$ represent the real localization and estimated localization of the unknown node respectively; K is the number of unknown nodes; N is the simulation times, taken as N=5; d is the communication radius of the node.

Firstly, the normalized average localization error change with the proportion δ of anchor nodes in all nodes is analyzed. Take $r = 30$ and $K = 500$, the change range of anchor node proportion is 10 % $\sim$ 30 % , and the results are shown in Figure 4.

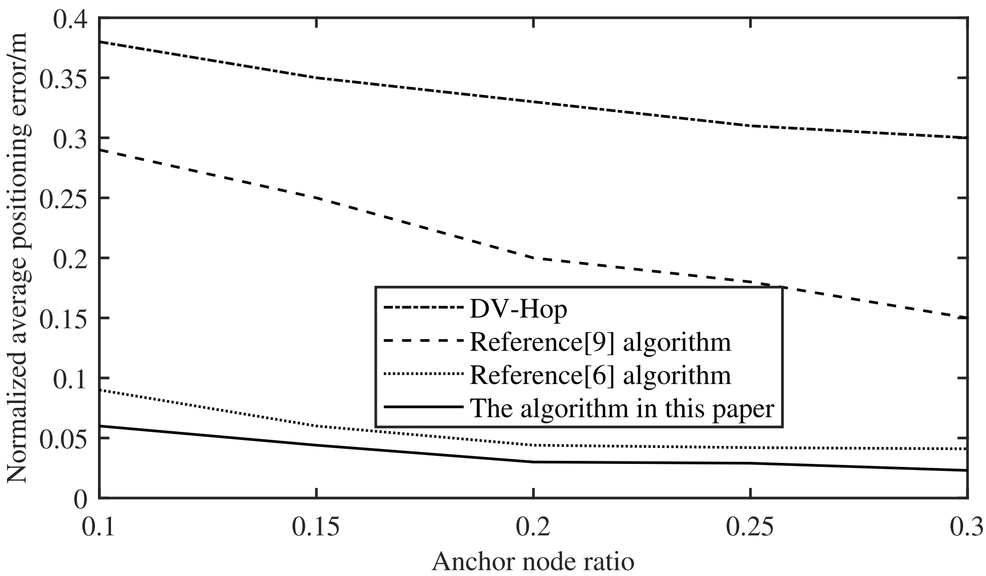

FIGURE 4. Variation curve of normalized average localization error with the proportion of anchor nodes

As shown in Figure 4, the localization accuracy of each algorithm will improve with the increase of the proportion of anchor nodes. However, the average value of the normalized average localization error $\bar{e}$ of the algorithm in this paper under different proportions of anchor nodes is 3.72cm, which is 32.9% lower than 5.54cm of the algorithm in the [6], and is significantly better than DV-Hop and the algorithm in [9]. Therefore, this algorithm can use fewer anchor nodes to achieve a similar localization effect and reduce cost.

Secondly, the average localization error variation with the communication radius is analyzed. Set δ=0.2 and $K = 500$. The results are shown in Figure 5.

Due to different communication radii, the number of nodes within the communication radius of unknown nodes will change, and the localization accuracy will also change. As shown in Figure 5, the localization accuracy of the algorithm in this paper increases slightly with the increase in communication radius, and the average localization error under different communication radius is 0.98M, which is 30% lower than that of 1.4m in [6], and is significantly better than DV-Hop and [9]. Therefore, this algorithm can use a smaller communication radius to achieve a similar localization effect to reduce energy consumption, increase the sensor's service life, and reduce the dimension of the unknown node vector in Section 3.5 to reduce the amount of calculation and running time.

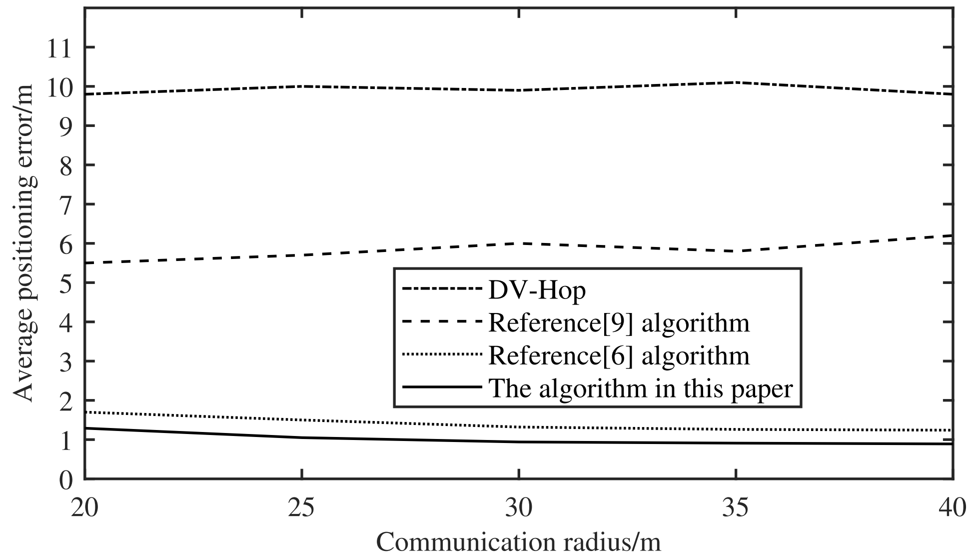

Figure 5. Variation curve of average localization error with communication radius

Then, the normalized average localization error variation with the number of node K is analyzed. Take δ=0.2 and d=30. The variation range of the number of nodes is 400 $\sim$ 900, and the results are shown in Figure 6.

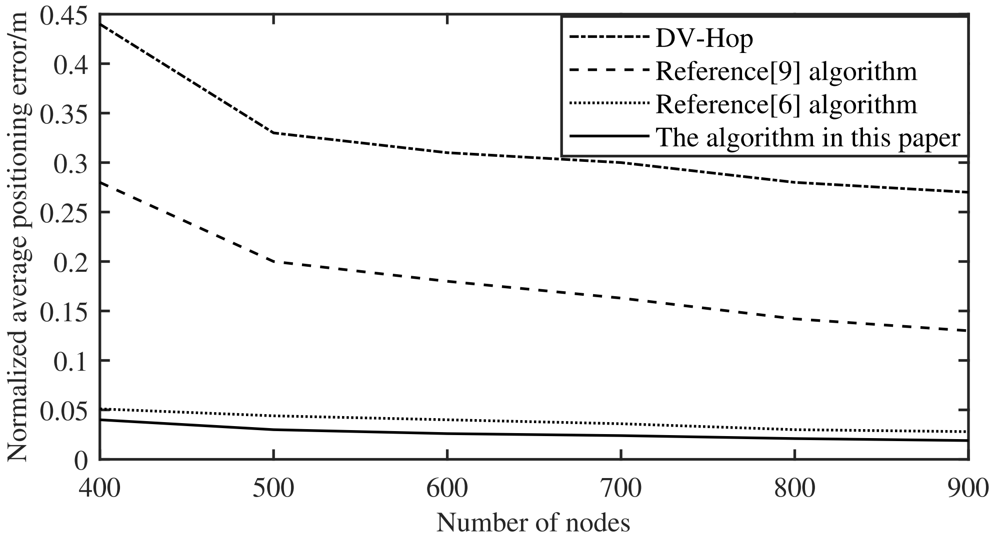

Figure 6. Variation curve of normalized average localization error with the number of nodes

As shown in Figure 6, the localization accuracy of the algorithm in this paper gradually improves with the increase of the number of nodes, and the average value of normalized average localization error under the different numbers of nodes is 2.63cm, which is 31% lower than 3.82M in [6], and is significantly better than DV-Hop algorithm and Ren and Pan's algorithm [9]. For the DV-Hop algorithm and Ren and Pan's algorithm [9], the total number of nodes in the WSN is increased, the node density in the WSN is increased, the network structure is more regular, the network connectivity is ensured, and the positioning accuracy is improved. For the RSSI ranging part of the algorithm in [6], the algorithm in [9] and the algorithm in this paper, with the increase of node density, the distance between adjacent nodes decreases, the ranging accuracy is improved, and then the localization accuracy is improved.

Finally, the variation of algorithm execution time with the number of node K is analyzed. Take δ=0.2 and d=30. The variation range of the number of nodes is 400 $\sim$ 900, and the results are shown in Table 1.

It can be seen from Table 1 that the algorithm in this paper has higher time complexity than other algorithms, but at present, most WSN application scenarios have low requirements for real-time because frequent localization will aggravate the energy loss of nodes. Because this algorithm adopts the node server two-stage data processing scheme, this algorithm adopts multi-threaded concurrent execution in the simulation experiment. In practical application, more thread concurrent or distributed execution algorithms can be used in servers with a solid performance to reduce the algorithm's execution time further.

Table 1. Variation of algorithm execution time with the number of nodes

Number of nodes	Time			
	Ours	DV-Hop	[6]	[9]
400	136.8s	0.8s	102.1s	1.1s
500	154.8s	1.1s	146.3s	1.5s
600	176.4s	1.3s	193.7s	1.8s
700	208.8s	1.5s	272.4s	2.3s
800	262.8s	1.8s	378.2s	2.8s
900	352.8s	2.1s	502.1s	3.2s

5. **Conclusion.** In order to improve the accuracy of node location and reduce the requirements of node operation and storage performance, a three-dimensional WSN node location algorithm combining RSSI and DV-Hop is proposed. Firstly, in the case of low anchor node density and uneven distribution, the initial value error obtained by the DV-Hop algorithm is large, and it is easy to obtain the local optimum solution using a numerical optimization algorithm. The DV-Hop algorithm combined with RSSI is used to obtain the initial value, and a GANV algorithm is proposed to obtain the global optimization solution. DBSCAN algorithm eliminates outliers for multiple location estimates of unknown nodes. Secondly, to improve the accuracy of the initial value, the RSSI ranging technology is introduced into the node localization part of the DV-Hop algorithm combined with RSSI, and the GANV algorithm is used to obtain the initial value. Finally, due to the high complexity of the algorithm, the node server's two-stage data processing method is adopted to reduce the performance requirements of the node. Simulation results show that compared with the existing related algorithms, this algorithm has high localization accuracy and meets the requirements of node location accuracy in large-scale WSN. However, using a single-thread execution algorithm in simulation experiments leads to a waste of time. Multi-thread or distributed data processing should be considered in the application; Because the RSSI ranging accuracy significantly impacts the localization accuracy of the algorithm, the dynamic environment adaptive distance model and extensive NLOS error detection method will be considered in the subsequent research.

Acknowledgement. This work is partially supported by National Natural Science Foundation of China (No.61872126, No.61772159), Key Scientific and Technological Projects in Henan Province, China (No.212102210092), Key Research Fund for University in Henan

Province, China (No:20A520015), Doctoral Fund (No.60907023), School Youth Backbone Teacher Fund(NO.2022XQG03).

REFERENCES

[1] T. Y. Wu, L. Yang, Z. Lee, S. C. Chu, S. Kumari and S. Kumar, “A Provably Secure Three-factor Authentication Protocol for Wireless Sensor Networks,” *Wireless Communications and Mobile Computing*, Vol.2021, 5537018, 2021.

[2] J. N. Chen, Z. J. Huang, Y. P. Zhou, F. M. Zou, C. M. Chen, J M. T. Wu and T. Y. Wu, “Efficient Certificate-Based Aggregate Signature Scheme for Vehicular Ad Hoc Networks,” *IET Networks*, vol.9, no.6, pp.290-297, 2020.

[3] O. Cheikhrouhou, G. M. Bhatti and A. Roobaea, “A hybrid DV-Hop algorithm using RSSI for localization in large-scale wireless sensor networks,” *Sensors*, vol.18, no.5, pp.1469-1482, 2018.

[4] L. Song and D. S. Huang, “An improved DV-Hop location algorithm based on improved wolf colony algorithm,” *Computer Engineering and Science*, vol.43, no.7, pp.1210-1218, 2021.

[5] C. Soares, J. Xavier and J. Gomes, “Simple and fast convex relaxation method for cooperative localization in sensor networks using range measurements,” *IEEE Transactions on Signal Processing*, vol.63, no.17, pp.4532-4543, 2015.

[6] J. Z. Jiang, Y. J. Li, H. B. Zhao and S. Ouyang, “A distributed node localization algorithm for large scale sensor networks,” *Journal of Electronics & Information Technology*, vol.41, no.12, pp.3022-3028, 2019.

[7] Y. Wang, Z. Y. Fang and L. Chen, “A new type of weighted DV-Hop algorithm based on correction factor in WSNs,” *Journal of Communications*, vol.9, no.9, pp.699-705, 2014.

[8] S. Messous, H. Liouane, O. Cheikhrouhou and H. Haman, “Improved recursive DV-Hop localization algorithm with RSSI measurement for wireless sensor networks,” *Sensors*, vol.21, no.12, pp.4152-4168, 2021.

[9] K. Q. Ren and C. M. Pan, “Improved DV-Hop algorithm based on RSSI hop quantization and error correction,” *Chinese Journal of Sensors and Actuators*, vol.33, no.5, pp.718-724, 2020.

[10] D. Niculescu and B. Nath, “DV based positioning in ad hoc networks,” *Telecommunication Systems*, vol.22, no.1, pp.267-280, 2003.

[11] Y. Y. Jiang, Y. Yu, M. L. Shi and Y. B. Liu, “RSSI ranging method of bessel function ranging model,” *Chinese Journal of Sensors and Actuators*, vol.33, no.2, pp.279-285, 2020.

[12] X. W. Yu, L. P. Huang, Y. Liu, H. Yu and P. Li, “Three-dimensional DV-Hop location algorithm based on genetic-tabu search optimization in WSN,” *Journal of Beijing University of Posts and Telecommunications*, vol.44, no.4, pp.75-81, 2021.

[13] “S. S. Wang, J. P. Yin, Z. P. Cai and G. M. Zhang, A RSSI-based self-localization algorithm for wireless sensor networks,” *Journal of Computer Research and Development*, vol.45, no.S1, pp.385-388, 2008.

[14] “Z. W. Feng, S. P. Zhu, Z. H. Zhao, M. L. Sun, M. Dong and D. R. Song, Comparative study on detection methods of wind power abnormal data,” *Advanced Technology of Electrical Engineering and Energy*, vol.40, no.7, pp.55-61, 2021.