Research Article

Data Aggregation in Heterogeneous Wireless Sensor Networks by Using Local Tree Reconstruction Algorithm

Zhaohui Zhang , **Jing Li** , and **Xiaoli Yang**

School of Mathematics and Statistics, Xidian University, Xi'an 710071, China

Correspondence should be addressed to Jing Li; xidianjing@126.com

Received 13 December 2019; Revised 8 March 2020; Accepted 30 March 2020; Published 5 May 2020

Academic Editor: Roberto Natella

Copyright © 2020 Zhaohui Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the transmission of heterogeneous data in heterogeneous networks, a topology optimization algorithm of heterogeneous wireless sensor networks based on local tree reconstruction is proposed, which can achieve better data transmission in the heterogeneous networks. First, the algorithm divides the nodes of the network into different layers by their hops and chooses different numbers of nodes as relay nodes in different layers. Second, the nodes are set with different initial energies in different layers. Because the packets of different nodes have different sizes, we adopt the corresponding data aggregation coefficients according to the actual data requirements of the network in data transmission. Finally, the lifetime of the network is prolonged by real-time updating of the topology of the tree during the data transmission. The simulations indicate that after the aforementioned three steps, the proposed algorithm prolongs the lifetime of the heterogeneous networks and improves the nodes utilization effectively.

1. Introduction

In wireless sensor networks (WSNs), the transmission data collected by the sensor nodes will be classified as different types. The data are homogeneous if they have the same properties; otherwise, they are heterogeneous data with different properties. In practical applications, the requirements for the data are always heterogeneous such as image, video, and multidimensional big data. The requirements for the heterogeneous data-transmitting function are often for special purposes, because the heterogeneous data are multidimensional data [1]. In the process of data transmission, aggregating the data from the subnodes is called the data aggregation technology. The nodes in networks are heterogeneous for the network extensive applicability: the nodes have different energies, functions, communication abilities, and so on. Some nodes in the network are endowed with stronger ability to enhance the network communication ability, which are called supernodes.

Researchers often adopt the method of abstracting the WSNs into graphs and then build a mathematical model and

solve the model in network topology research. Most of the sources of their ideas are referred to the research on the graph theory. The building of the network topology must satisfy the requirements of the design of network topology optimization and destruction resistance of the network. The graph is an undirected graph because the communication between the nodes is mutual. Therefore, the study of WSNs can be modelled as mathematical problems, such as connectivity, sparse, symmetry, node weights, and duality theory.

The core idea of topology control in the WSNs is to ensure the safety and reliable transmission of the collected data by the sensor nodes. The main method of maintaining the constructed network topology is to ensure the integrity of the maximum connected graph. In recent years, the requirements of the application fields of WSNs are continuously increasing. The topology structure is formed by self-organization network. Not only the complete communication but also the scalability and universality of the network are required in many cases. In other words, a topology is required to satisfy the requirements of various types of users

[2]. This topology greatly expands network operation. The network demands both theoretical operability and practical expansion operation, which put forward higher requirements for the constructed topology [3, 4].

The topology optimization is often designed for a specific network among the study of the current mainstream network topology in WSNs. For example, the goal of the MST algorithm is to maximize the network lifetime, and the lifetime index systems are the main research target for many researchers. The MST algorithm considers multiple target optimization. However, it only optimizes the energy consumption and robustness of the network. The applications of heterogeneous WSNs are becoming more and more widely. The topology optimization technology needs to keep pace with the practical applications.

The heterogeneous data-transmitting network constructed in this paper is a tree topology network, which shows good performance in data transmission [5–12]. The tree topology can transmit the collected data better than other network topologies and also has a strong destruction resistance. The advantages of the tree topology are the high efficiency in data transmission and data aggregation by nonleaf nodes in the tree. In references [7–9], the requirements of energy and network delay are considered comprehensively when constructing the minimum tree topology. Node energy consumption and throughput of the WSNs are the main criteria to measure the advantages and disadvantages of the established network model [10–12]. There are many network topologies constructed for good transmission of the collected data, such as cluster-based topology [13–17], tree-based topology [18–20], and other topologies [21]. In references [22–24], the authors give comprehensive introduction to the performance of the abovementioned topologies in the WSNs, which indicate that the quality of topology has a great influence on data transmission, especially in some practical networks [23, 25, 26]. Some types of computer software commonly used to simulate the algorithm performance are proposed. The simulation results for different network sizes are often different, and the gap is big in some situations. The real time of the transmitted data is strictly required [27], and data security and energy efficiency are also considered in some network topologies [2, 18, 19, 28, 29]. Haseeb et al. [30] provide an energy-efficient and secure routing protocol for intrusion avoidance in IoT-based WSN, which has a better performance in a physical security network. Liu et al. [31] use a disjoint routing for adaptive big data transmission and the data security is verified in simulations. Therefore, the construction of the network topology should be based on specific network requirements. The constructed model needs to satisfy a variety of node types as many as possible and also the network transmission performance.

In view of the different heterogeneous WSNs, we put forward to solve a more practical algorithm, which conforms to the requirements of modern scientific research issue and focuses on future research. The relationship between the graph and its property is studied, and the corresponding conclusion is obtained. We establish the reasonable

transmitting models and aim at practical applications through the network topology model after the above steps.

For the heterogeneous data transmission, data aggregation by using the local tree reconstruction algorithm (DA-LTRA) in the heterogeneous WSNs is proposed in this paper. The network divides the nodes into different layers. Then, each layer selects a certain number of nodes to be relay nodes and each node has different initial heterogeneous energies in different layers. The network updates the tree topology during data transmission in real time to prolong the nodes' survival time. The simulation results show that the lifetime of the proposed network is prolonged effectively after the proposed adjustment of heterogeneous networks. It indicates that the effective utilization rate of the nodes is improved effectively. Therefore, the heterogeneous network will achieve better performance in data transmission.

Heterogeneous data transmission has a very high requirement for network topology in the context of big data, especially for the network in a specific application environment. The research on real-time data transmission and network energy efficiency is a hot issue at present. Motivated by the aforementioned observations, we explore the data aggregation in the heterogeneous WSNs by using local tree reconstruction algorithm in this paper. The main contributions are summarized as follows:

- (1) The properties of the whole graph can be obtained by extending local optimization to global optimization for the network, which reflects the performance of the maximum connected subgraph of the studied graph. Our research provides a theoretical basis and a new idea for future research on data transmission in distributed multilayer networks.
- (2) The optimal heterogeneous data transmission network topology is constructed in the context of big data. The routing topology of the whole network is realized through the maintenance and reconstruction of local network topology, which is more conducive to the big data transmission technology in some application backgrounds.
- (3) Flexible data aggregation technology is proposed to remove the redundant data and to ensure the authenticity and accuracy of the data in this paper. The constructed routing model guarantees the low latency of data transmission and the low energy consumption of the nodes in the network at the same time and prolongs the lifetime of the network. The operation of the real network is abstracted as a mathematical problem by establishing and solving the mathematic model to get the results. Our model for data transmission has a strong innovation and provides a new way for mathematical and engineering problems of cross research.

The rest of the paper is arranged as follows. The second part introduces the network model and communication model. The third part is the theme algorithm of our scheme, which models and analyzes the local tree reconstruction

technology. The fourth part is the simulation part. The fifth part summarizes this paper.

2. Preliminaries

2.1. Network Model. In order to better describe the proposed algorithm, we define the proposed model as follows. The topology model in WSNs can generally be represented by the set of points, edges, and communication radii. This set can be named as a graph $G = (V, E, r)$, where V represents the set of sensor nodes, E represents the set of edges formed by two nodes that can communicate in the network, and r represents the radius of the sensor nodes. The network deploys N nodes in a circular region randomly, and the nodes in the network are divided into two states: working time and sleeping time. The location of Sink is on the edge of the circular area, and all nodes transmit data to Sink through the relay node in the network. We assume the following:

- (1) All nodes are not moved after deploying in the designated area.
- (2) All nodes are heterogeneous. We set different initial energies for the nodes which are expressed as E_{initial} .
- (3) Each node has the same signal intensity value, and the power of the node signal is adjustable. The communication radius of the nodes is an equal value r .
- (4) The nodes can detect the information of the neighbor nodes.
- (5) There are no attack nodes in the network.
- (6) Sink is located at the edge of the network.

2.2. Energy Model. The energy model is similar as the energy model given in reference [32]. The energy consumption of transmitting data and receiving data according to the wireless communication model is formulated as follows, respectively,

$$E_{\text{tx}}(l, d) = lE_{\text{elec}} + l\epsilon_{\text{fs}}d^2, \quad (1)$$

$$E_{\text{rx}}(l) = lE_{\text{elec}}, \quad (2)$$

where $E_{\text{tx}}(l, d)$ represents the energy consumption of transmitting l -bit data; $E_{\text{rx}}(l)$ represents the energy consumption of receiving l -bit data; E_{elec} is the energy consumption of the wireless transceiver circuit in transmitting or receiving 1-bit data; and ϵ_{fs} represents the magnification of the signal amplifier power in the free-space model.

The node energy consumption while aggregating m -bit data is

$$E_{\text{da}} = mE_{\text{elec}}. \quad (3)$$

2.3. Data Aggregation Model. In the PEDAP algorithm [33], the node i aggregates its own data packets and its children data packets into a single data packet and then transmits the data packet to the next relay node. Sink has a certain amount

of original data in many situations. However, there are very few data remaining when they arrive to Sink based on the above aggregation method. It cannot guarantee the accuracy of the information. We hope that the data information received by Sink is as much as possible as that comes from the source nodes. Assume that the parent of the source node i is the node j and the parent of the node j is the node k . The node k is without the communication radius of the node i , which means the nodes i and k have low correlation. We hope that a few amount of source data of the node i is aggregated by the node j , so the node k can retain more original data.

For example, if we follow the above first data aggregation method, there just remains $1 * 0.5 * 0.5 = 0.25$ unit data packets of the node k which comes from the node i . The original information will be very few if the node k continues to transmit the data to Sink through multihops. It can be seen that the above data aggregation algorithm has a serious shortcoming and will lead to a loss of the data information. Therefore, the node aggregation policy needs to be redefined.

Define the node aggregation coefficient as

$$\rho_i = \frac{\text{hop}(i) - 1}{\text{hop}_{\text{max}}}, \quad (4)$$

where hop_{max} is the maximum number of the node hops in the network and $\text{hop}(i)$ is the hop numbers from the node i to Sink.

Therefore, there is a relationship between the data packets before data aggregation and after data aggregation for the node i :

$$\text{packet}(i)_{\text{num}_{\text{back}}} = \text{packet}(i)_{\text{num}_{\text{front}}} \times (1 - \rho_i). \quad (5)$$

3. Data Aggregation in Heterogeneous WSNs by Using the Local Tree Reconstruction Algorithm

The tree network topology structure in the WSNs is more suitable for the fusion of the global information network. For instance, in a nuclear power plant, we can establish a data aggregation tree to learn the useful information of each workshop without the nuclear radiation. Figure 1 is the schematic diagram of the data-transmitting topology structure model in this research.

The tree structure has a good scalability. The nodes will cause fast energy consumption because they need to receive and transmit a lot of data when they are near Sink. Therefore, the first step of the tree reconstructing method is to reselect the relay node at the time of tree updating. As a result, the main purpose in the network topology is that the lifetimes of bottleneck nodes are prolonged.

As shown in Figure 1, the nodes a–o are the parent nodes. The parent nodes need to collect and receive the data from their children nodes and then transmit the data back to their parent nodes after aggregating all data. If the nodes have the same initial energy in the network, the bottleneck nodes will cause lots of energy consumption and will also lead to the nonuniformity of the node energy consumption.

FIGURE 1: Tree data-transmitting topology structure model.

Therefore, the end result is that the lifetime of the network is shortened and the network topology fails.

With a view to the information similarity property of the data collected by the nodes whose geographic positions are close, we let some parent nodes only transmit the data based on the proportion considered above. The amount of data will be reduced; meanwhile, the premature occurrence of the bottleneck nodes will be restrained after this treatment.

3.1. Selecting Method of Relay Nodes. In this paper, the first step is to improve the energy balance of the whole network. The method we adopted is to choose the nodes as relay nodes whose positions are near Sink.

As shown in Figure 2, assume that the node data aggregation rate is 1 and the packet size generated by each node is 1 in the whole network. In Figure 2(a), the network leading to the bottleneck node appears too early because the nodes *a*, *c*, *d*, and *g* have massive data packet sizes. We find that the transmit data packet size will be reduced to nearly half proportion of nodes *a*, *c*, *d*, and *g* after we select other nodes as relay nodes in a certain proportion as shown in Figure 2(b). It means that the network load is balanced.

However, the network needs to collect the accurate data while we prolong the lifetime of the network. The key here is that we do not allow more nodes to be relay nodes. This ensures that more raw data information is collected and transmitted.

In order to ensure that the more nodes a layer has, the more nodes are chosen as relay nodes, and the fewer nodes a layer has, the fewer nodes are chosen as relay nodes, the proportion of relay nodes is selected as

$$\rho_{i-\text{hop}} = \frac{n_{i-\text{hop}}}{N}, \quad (6)$$

according to the number of nodes in different layers, where $\rho_{i-\text{hop}}$ is the proportion of relay nodes in the layer *i*, $n_{i-\text{hop}}$ is the number of nodes in the layer *i*, and *N* is the number of total nodes in the network.

Therefore, the number of relay nodes in the layer *i* is

$$n_{\text{relay},i-\text{hop}} = \left\lceil \frac{n_{i-\text{hop}}^2}{N} \right\rceil + 1. \quad (7)$$

In this way, the nodes can save energy effectively.

3.2. Method of Initial Energy Setting for Heterogeneous Nodes.

In heterogeneous networks, the task of the nodes is to monitor the network in real time. The method of data transmission is introduced in Section 2.3, so that each node will transmit different size packets to the relay node. Obviously, the nodes that have lower hop will transmit and receive more data and will consume a lot of energy and die early either.

As shown in Figure 3(a), assume that the initial energy of the whole nodes is 40. The remaining energy of the node *d* is 4 after 3 rounds. It is not enough to sustain the next round of data transmission as shown in Figure 3(b).

In view of the above deficiencies, the method we adopt is to set different initial energies for the heterogeneous nodes in different layers. We set higher energy for the nodes whose layers are lower. It means that the nodes near Sink have higher initial energy. Assume that the node energy with the maximum hop is E_{initial} and the influence of aggregation proportion is taken into account in this research. The energy of the nodes in different hop layers is defined as

$$E_{i-\text{hop}} = \frac{E_{\text{initial}} * (n - i + 1)}{\rho_i}, \quad (8)$$

FIGURE 2: Selecting method of relay nodes.

FIGURE 3: Energy consumption diagram.

where $E_{i-\text{hop}}$ is the node energy in the layer i .

As shown in Figure 4, the node d and the other nodes near Sink have higher initial energy after the initial node energy resetting. Assume that 4 is the highest hop of the nodes in the network and the initial energy of the node d is $40 \times 1.5 = 60$. Therefore, we improve the usability of the node d effectively because the node d can run 5 rounds and the network run time is increased by 2 rounds. So the lifetime span of the nodes in the heterogeneous network is prolonged obviously after the second step.

3.3. Load-Balancing Method of the Nodes. The network collects, aggregates, and transmits data according to the topology (as shown in Figure 5(a)). The energy consumption of the node d is 12 in each round, and the node d contains 12 units of energy after running 4 rounds. The node d will die after the next round if the tree still runs in the same way. Therefore, the node lifetime in the network is influenced greatly. We find the node e which is near the node d has higher remaining energy in the same layer, and then, the node d transfers its children to the node e . Subsequently, the

FIGURE 4: Setting of initial energy.

FIGURE 5: Continued.

FIGURE 5: Load-balancing method of the nodes in heterogeneous networks.

node d is transformed into the leaf node and is no longer responsible for aggregating and transmitting data. At the same time, the node e is transformed into the relay node, and it is responsible for aggregating and transmitting data. Finally, the network is rebuilt according to the above method.

As shown in Figure 5(b), the remaining energy of the node f is 0 after running 4 rounds if we continue to run the network. But it still prolongs the network lifetime to 4 rounds (as shown in Figure 5(c)) when we use the above load balance method similarly. Therefore, the lifetime of the network is prolonged if we follow the same node topology adjustment method.

We extend the overall network lifetime after using the above method repeatedly, and the energy consumption of the nodes in the same layer is balanced after the above operation.

3.4. Flow Chart of the Algorithm. To better describe the proposed algorithm in this research, a flow chart of the algorithm is shown in Figure 6.

First, the greedy algorithm is used to build a minimum routing tree network topology from Sink to the leaf node. The relay nodes are selected according to the first step. Then, the different initial node energies are set up according to disparate tasks in the different layers in the second step. Finally, the network performs the dynamic local routing tree evolution. The number of surviving nodes is used to determine the state of the network in practical applications. The network dies and outputs the results when the nodes die by a certain proportion. Otherwise, the surviving nodes in the network are reconstructed. We repeat the network

optimization operations until the given node death proportion is reached.

Our research on max-flow is a maximum connected subgraph optimization problem actually. The desired results are obtained by studying the maximum connected subgraph.

4. Theoretical Analysis on the Effectiveness of Energy Consumption of DA-LTRA

In this section, we analyse the effectiveness of the energy consumption in the proposed DA-LTRA. Through the analysis on the proposed model, we come to the conclusion that the proposed energy consumption model is mainly embodied in the following two aspects: (i) for each node in local network, node energy and load of packet have been greatly reduced owing to the participation of relay nodes and heterogeneous energy; (ii) for the whole nodes in the global network, in the process of relay transmission, the node selects the relay node in the upper layer for data transmission when the node's energy is more than others. To sum up, the average transmission distance, packet size, and energy consumption of the network are much smaller than those of network which are not dynamically adjusted and not divided into multilayers. We compare the proposed DA-LTRA with DADAT [32] in the context of the same initial energy.

In DADAT, all the nodes can be chosen as relay nodes. The parent of the node i is the node j . Assume there are n nodes in the network, and the relay node i receives l_i -bit data and produces m_i -bit data. The energy consumption of the node i has three aspects: data reception, data aggregation,

FIGURE 6: Flow chart of the algorithm.

and data transmission, which is shown as follows based on equations (1)–(3):

$$E_{\text{tran},i} = (l_i + m_i) * (1 - \rho_i) * E_{\text{elec}} + (l_i + m_i) * (1 - \rho_i) * \varepsilon_{\text{fs}} * d_{i-j}^2, \quad (9)$$

$$E_{\text{rec},i} = l_i * E_{\text{elec}},$$

$$E_{\text{da},i} = (l_i + m_i) * \rho_i * E_{\text{elec}}.$$

The energy for the node i is

$$\begin{aligned} E_{\text{con},i} &= E_{\text{tran},i} + E_{\text{rec},i} + E_{\text{da},i} \\ &= (2l_i + m_i) * E_{\text{elec}} + (1 + m_i) * (1 - \rho_i) * \varepsilon_{\text{fs}} * d_{i-j}^2. \end{aligned} \quad (10)$$

Therefore, the energy consumption of all the nodes in each round is

$$E_{\text{con}} = \sum_{i=1}^n E_{\text{con},i}. \quad (11)$$

To ensure the smooth and efficient operation of the network, we simplify equation (11) to find minimum energy consumption. The result is shown in the following equation:

$$E_{\text{con}} \geq \sum_{i=1}^n \left\{ (2n-1) * E_{\text{elec}} * (l_i + m_i) + \frac{(l_i + m_i) * (1 - \rho_i) * \varepsilon_{fs} * d_{i-j}^2}{n} \right\}. \quad (12)$$

Assume that the average data produced by each node are m and the average distance between the node and its relay node is d_{C-S} . We can obtain the minimum value of the energy consumption according to equation (12), which is shown as follows:

$$E_{\text{con}}^{\min} = 2mn^3 * E_{\text{elec}} + 2mn^2 * d_{C-S}^2 * \varepsilon_{fs} * \sum_{i=1}^n (1 - \rho_i). \quad (13)$$

In a similar way, for our design of the DA-LTRA, assume that there are k leaf nodes and $n - k$ relay nodes in the network. In the layer i , the number of nodes is $n_{i-\text{hop}}$. Therefore, the relay nodes' number in layer i is $n_{i-\text{hop}} * \rho_{i-\text{hop}} = [n_{i-\text{hop}}^2/n]$.

For the leaf node j , the relay node is o and the other assumptions are the same as the algorithm DADAT. The energy consumption is

$$E_{\text{con},j} = E_{\text{tran},j} + E_{\text{da},j} = l_j * E_{\text{elec}} + l_j * \varepsilon_{fs} * d_{j-o}^2 + (l_j + m_j) * \rho_j * E_{\text{elec}}. \quad (14)$$

For the relay node p , the relay node is q and the energy consumption is

$$\begin{aligned} E_{\text{con},p} &= E_{\text{tran},p} + E_{\text{rec},p} + E_{\text{da},p} \\ &= \sum_{i=1}^{n_{i-\text{hop}}} \{ (l_p + m_p) * (1 - \rho_p) * E_{\text{elec}} + (1 + m_p) * (1 - \rho_p) * \varepsilon_{fs} * d_{p-q}^2 + l_p * E_{\text{elec}} + (l_p + m_p) * \rho_p * E_{\text{elec}} \}. \end{aligned} \quad (15)$$

Therefore, the energy consumption of all the nodes in each round for the DA-LTRA is

$$E_{\text{con}} = \sum_{j=1}^m E_{\text{con},j} + \sum_{p=2}^{[n_{p-\text{hop}}^2/n]+1} E_{\text{con},p} * (\text{hop}_{\max} - 1). \quad (16)$$

Similarly, we can obtain the maximum of the energy consumption according to equation (16), which is shown as follows:

$$\begin{aligned} E_{\text{con}}^{\max} &= 2mn^2 * E_{\text{elec}} + 3mn * E_{\text{elec}} \\ &+ 2mn^2 * d_{C-S}^2 * \varepsilon_{fs} * \left[\sum_{i=1}^m (1 - \rho_i) + \sum_{p=1}^{n-m} (1 - \rho_p) \right]. \end{aligned} \quad (17)$$

By analysing the result of equations (13) and (17), we can see that the maximal energy consumption of DA-LTRA is

smaller than that of DADAT obviously. This means that we can get a simple mathematical analysis and the optimization of our algorithm reduces the redundant data, namely, reduces the energy consumption. Furthermore, our model enhances the initial energy of the nodes based on the network layer and thus prolongs the lifetime of the network.

5. Simulation Experiment

5.1. Experimental Environment and Simulation Parameters.

In order to verify the effectiveness of the algorithm proposed in this research, we perform the simulation experiment to compare the proposed algorithm with the similar DADAT and GIT (Greedy Incremental Tree) algorithms. The simulation system environment is Win7, 64 bit, and the software is MATLAB 2012a, the CPU is i7-4720hq, and the memory is 8.00 GB. The proposed and compared algorithms are simulated under the same network environment, and the initial setting of the network is similar.

The simulation parameters in the network are shown in Table 1.

5.2. Simulation of the Optimization on the Network Update Rounds.

Under the optimal parameters, the simulation results can better reflect the network performance. The proposed algorithm is compared with DADAT in algorithm simulation experiment. The DADAT tree is adopted to reestablish the tree every 90 rounds under the same simulation environment, that is, the number of update rounds is 90.

It can be seen from Figure 7 that the lifetime of the network is reduced double when the number of network update rounds is more than 110 or less than 70. It indicates that the energy consumption of the network in data calculation and data transmission has a great impact on the lifetime of the network. The network has highest lifetime while the update rounds are 90. Therefore, we select 90 as the update rounds in this simulation.

If the network is updated too frequently (less than the optimal value 90), the data transmission and computation of the nodes will cause a lot of energy consumption in the network, which will lead to the shortening of the overall network lifetime. On the other side, some nodes will also consume energy rapidly if the network is updated too slowly because the nodes undertake too much data forwarding mission in the network. In the simulation experiment, the optimal tree topology update times are selected and fixed in advance in the research. Therefore, the topology evolution in the optimal environment can maximize the utilization rate of the network.

5.3. Simulation of the Network Lifetime.

The number of surviving nodes can reflect the variation of the maximum connected subgraph in the network. As shown in Figures 8 and 9, the proposed DA-LTRA improves the lifetime of the network effectively. When the first node in the network dies, the network lifetime of the DA-LTRA is 9160 rounds, DADAT is 5945 rounds, and GIT is 4121 rounds

TABLE 1: Simulation parameters.

Parameters	Value
Network scale	300m × 300m
Nodes number	300
Radius (m)	30
Base station coordinates	(0, 0)
E_{fs} (nJ/b)	50
fs (pJ/b/m ²)	10
E_{da} (nJ/b/m ²)	5
Packet size (bits)	1K
Initial energy (J)	2
Aggregation packet per second	200 kbps
Channel width	250 kbps

FIGURE 7: Relationship between the update rounds and the lifetime of the network.

FIGURE 8: Lifetime comparison under different rounds.

FIGURE 9: Lifetime comparison under different numbers of nodes.

accordingly. DA-LTRA runs 11725 rounds, DADAT runs 8041 rounds, and GIT runs 7746 rounds accordingly when there are 10 nodes dead. The lifetimes of the network are 5604 rounds, 4612 rounds, and 3223 rounds under 200 nodes, and 7294 rounds, 5290 rounds, and 4100 rounds under 250 nodes for DA-LTRA, DADAT, and GIT, respectively. The performance improvement rates of network lifetime can reach up to 122% and 45% compared with GIT and DADAT, respectively. This is because the proposed algorithm is better in the process of the initial node energy setting and the relay node selecting than the DADAT and GIT algorithms in the same heterogeneous condition. Additionally, we use the local tree reconstruction technology to improve the network load balance in the tree maintenance phase.

5.4. Number of Packets of the Network. Figure 10 gives the number of packets of the network per round under 200 nodes and 300 nodes, respectively. We can see from the simulation diagram that there are more packets of our proposed algorithm DA-LTRA per round. The packets number in one round is 1550, 1315, and 1250 under 200 nodes, and 1735, 1437, and 1367 under 300 nodes for DA-LTRA, DADAT, and GIT, respectively. Therefore, the proposed algorithm still maintains high data volume to maintain data accuracy even after data aggregation. In GIT algorithm, the data aggregation ratio for each relay node is 1, which leads to loss of some packets easily. In DADAT algorithm, the aggregation ratio is considered according to node distance and a lot of nodes are selected as relay nodes. The proposed algorithm in this paper improves the selection scheme of relay nodes and aggregation ratio to avoid the above deficiencies. By selecting the high-energy nodes to be relay nodes for data transmission, the accuracy of data can be guaranteed and energy consumptions are reduced.

FIGURE 10: Number of packets of the network per round.

FIGURE 11: Variance of energy consumption under 200 nodes.

5.5. Variance of Node Energy Consumption. In this section, we compare the overheads of the nodes by comparing the energy variances with different algorithms. The variances of node energy consumption are given in Figures 11 and 12 under different numbers of nodes in the network. The

simulation diagrams can reflect that the variance of node energy consumption of the proposed DA-LTRA is smallest, which indicates that the overheads of the nodes are balanced through the local tree reconstruction. When the network runs 5000 rounds, the variance of node energy consumption

FIGURE 12: Variance of energy consumption under 300 nodes.

FIGURE 13: Overheads of energy under different rounds.

is 0.02, 0.083, and 0.138 under 200 nodes and 0.06, 0.12, and 0.172 under 300 nodes for DA-LTRA, DADATA, and GIT, respectively.

The topological stability for data transmission can reach a very high level. Therefore, our model achieves an optimal network topology for data transmission, and the whole network is extended to fast convergence of global optimization through the local optimization.

5.6. Routing Overheads Evaluation. Figure 13 gives the overheads of energy under different rounds. When the network runs 5000 rounds, the energy consumptions for

FIGURE 14: Overheads of delay under numbers of nodes.

FIGURE 15: Packets drop ratio under different rounds.

DA-LTRA, DADATA, and GIT are 0.17 J, 0.22 J, and 0.39 J, respectively. And those are 0.3 J, 0.7 J, and 0.9 J, when the rounds reach 9000. The performance improvement rates of the DA-LTRA algorithm reach 129% under 5000 rounds and 200% under 9000 rounds compared with GIT. Therefore, the performance of routing overheads of energy has been improved a lot with the increase in the number of network running rounds. Due to different energy settings of the nodes in different layers, the performance of the proposed algorithm suggests that the heterogeneous network topology for big data transmission has better properties.

Real-time data transmission is another important measure of network overhead. In our work, the data

transmission delays are compared for different algorithms. From Figure 14, we can see that the proposed algorithm DA-LTRA has low delays compared with GIT and DADAT. The delays are 20.75 ms, 21.42 ms, and 21.85 ms under 200 nodes and 21.60 ms, 22.25 ms, and 22.45 ms under 300 nodes for DA-LTRA, DADAT, and GIT, respectively. The performance improvement ratios almost reach 5% and 2% compared with GIT and DADAT. We can see from Figure 14 that the overall delays of all algorithms grow with the increase in network nodes. This is because the density of nodes leads to the increase of data and subroutes, which results in an increase in data transmission delays from the source node to Sink. However, the nodes do not select nodes in the same layer as relay nodes and only part of nodes in each layer are selected as relay nodes to reduce the subroutes in DA-LTRA. In this way, the delay is reduced effectively.

5.7. Packets Drop Ratio. Packets drop ratio (PDR) is the ratio of the number of dropped data to Sink. The lower value of the PDR results in better performance of the algorithm. Figure 15 shows the packets drop ratio of the proposed algorithm DA-LTRA with the algorithms DADAT and GIT. The proposed system results in a lower packets drop ratio than the compared algorithm. Especially, when the network runs 9000 rounds, the packets drop ratio of the proposed algorithm DA-LTRA is 3.3%, 4.6%, and 5% for DADAT and GIT, respectively. The proposed algorithm still maintains a low packet drop ratio when there are some dead nodes. From the above experimental results, it obviously proves that the proposed algorithm DA-LTRA can perform better than the algorithms DADAT and GIT. Hence, the objective of the DA-LTRA is achieved after three steps of adjustment.

6. Conclusions

The tree topology has a strong damage resistance, and the maximum connected subgraph is the basis of studying topology properties. In this research, the tree topology optimization algorithm is proposed based on the heterogeneous data, which have higher requirements on network topology.

In general, the homogeneous network is difficult to meet the service quality requirements of heterogeneous data transmission. Therefore, the strategy we adopted in this research is to use the heterogeneous nodes to complete the data transmission of these different types. Such a data-transmitting mode can meet the real-time and accuracy requirements of data transmission. In the process of topology optimization, this research adopts three steps to adjust the structure of the network. The network has a strong topology after the three steps of the network topology optimization. The simulation results of the proposed algorithm have higher performance and reduce the energy consumption of the nodes in the network effectively under the same simulation conditions and parameters.

The data transfer in heterogeneous networks contains different types of data, which include two-dimensional data, audio, image, and video. These data extremely demand for good network topology, so it is difficult to realize the

transmission requirements of these data in homogeneous networks. Usually, the size of heterogeneous data packets is very large.

The research work in this paper only involves the optimization technology of data transmission in WSNs. The tree-based sensor network topology proposed in this paper only considers the data transmission between link layers. But there are many problems need to be studied further, for example, the topology power control of sensor nodes between different layers and the node power control strategies at different layers (such as physical layer, link layer, and network layer). Furthermore, for the reason of technical difficulty, in this paper, we have not studied the optimization problem under the data-sending rate and it is necessary to study this problem in the data transmission-based network. In our future work, we will study the related network topology optimization research problems under the data-sending rate scenarios.

Data Availability

The data (including the simulation parameters and simulation program) used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant no. 61673014), the Fundamental Research Funds for the Central Universities (Grant no. JB190702), and the project supported by Shaanxi Natural Science Foundation (Grant no. 2018JM6026).

References

- [1] D. Alrababah, E. Al-shammari, and A. Alsuh, "A survey: authentication protocols for wireless sensor network in the internet of things; keys and attacks," in *Proceedings of the 2017 International Conference on New Trends in Computing Sciences (ICTCS)*, IEEE, Amman, Jordan, pp. 270–276, September 2017.
- [2] A. Wang, J. Shen, P. Vijayakumar, Y. Zhu, and L. Tian, "Secure big data communication for energy efficient intra-cluster in WSNs," *Information Sciences*, vol. 505, pp. 586–599, 2019.
- [3] H. Wang, Z. Xu, H. Fujita, and S. Liu, "Towards felicitous decision making: an overview on challenges and trends of big data," *Information Sciences*, vol. 367–368, pp. 747–765, 2016.
- [4] D. Liu, J. Shen, A. Wang, and C. Wang, "Lightweight and practical node clustering authentication protocol for hierarchical wireless sensor networks," *International Journal of Sensor Networks*, vol. 27, no. 2, pp. 95–102, 2018.
- [5] W. Liang and Y. Liu, "Online data gathering for maximizing network lifetime in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 1, pp. 2–11, 2007.
- [6] Y. Wen, F. Lin, and H. Wang, "A TDMA-based scheduling and routing algorithm for data-centric wireless sensor networks," in *Proceedings of the IEEE Global Telecommunications*

- Conference, IEEE, Washington, DC, USA, pp. 586–591, November 2007.
- [7] S. Kin, H. Abbasizadeh, B. Rikan, S. Oh, B. Jang, and Y. Park, “A-20 to 30 dBm input power range wireless power system with a MPPT-based reconfigurable 48% efficient RF energy harvester and 82% efficient A4WP wireless power receiver with open-loop delay compensation,” *IEEE Transactions on Industrial Electronics*, vol. 34, no. 7, pp. 6803–6817, 2019.
 - [8] M. Song and M. Zheng, “Energy efficiency optimization for wireless powered sensor networks with nonorthogonal multiple access,” *IEEE Sensors Letters*, vol. 2, no. 1, pp. 1–4, 2018.
 - [9] S. Chowdhury and C. Giri, “EETC: energy efficient tree-clustering in delay constrained wireless sensor network,” *Wireless Personal Communications*, vol. 109, no. 1, pp. 189–210, 2019.
 - [10] Q. Wu, M. Tao, D. W. Kwan Ng, W. Chen, and R. Schober, “Energy-efficient resource allocation for wireless powered communication networks,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2312–2327, 2016.
 - [11] X. Lin, L. Huang, C. Guo, P. Zhang, M. Huang, and J. Zhang, “Energy-efficient resource allocation in TDMS based wireless powered communication networks,” *IEEE Communications Letters*, vol. 21, no. 4, pp. 861–864, 2017.
 - [12] H. Ju and R. Zhang, “Throughput maximization in wireless powered communication networks,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 418–428, 2014.
 - [13] N. Gharai, K. A. Bakar, S. Z. M. Hashim, A. H. Pourasl, M. Siraj, and T. Darwish, “An energy-efficient mobile sink-based unequal clustering mechanism for WSNs,” *Sensors*, vol. 17, no. 8, p. 20, 2017.
 - [14] Y. Hu, Y. Niu, J. Lam, and Z. Shu, “An energy-efficient adaptive overlapping clustering method for dynamic continuous monitoring in WSNs,” *IEEE Sensors Journal*, vol. 17, no. 3, pp. 834–847, 2017.
 - [15] N. T. Hanh, H. T. T. Binh, N. X. Hoai, and M. S. Palaniswami, “An efficient genetic algorithm for maximizing area coverage in wireless sensor networks,” *Information Sciences*, vol. 488, pp. 58–75, 2019.
 - [16] M. Sajwan, D. Gosain, and A. K. Sharma, “Hybrid energy-efficient multi-path routing for wireless sensor networks,” *Computers & Electrical Engineering*, vol. 67, pp. 96–113, 2018.
 - [17] G. Xiao, J. Shi, and S. Ning, “Adaptive freeshape clustering for balanced energy saving in the wireless HART networks,” *Complexity*, vol. 2019, Article ID 2836981, 11 pages, 2019.
 - [18] Y. Liu, Q. Wu, T. Zhao, Y. Tie, F. Bai, and M. Jin, “An improved energy-efficient routing protocol for wireless sensor networks,” *Sensors*, vol. 19, no. 4579, pp. 1–20, 2019.
 - [19] J. Chang and T. Shen, “An efficient tree-based power saving scheme for wireless sensor networks with mobile sink,” *IEEE Sensor Journal*, vol. 16, no. 20, pp. 545–7557, 2016.
 - [20] H. Salarian, K.-W. Chin, and F. Naghdy, “An energy-efficient mobile-sink path selection strategy for wireless sensor networks,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2407–2419, 2014.
 - [21] X. Fu, H. Yao, and Y. Yang, “Sink-convergence cascading model for wireless sensor networks with different load-redistribution schemes,” *Complexity*, vol. 2019, Article ID 7630168, 9 pages, 2019.
 - [22] A. J. Al-Mousawi and H. K. AL-Hassani, “A survey in wireless sensor network for explosives detection,” *Computers & Electrical Engineering*, vol. 72, pp. 682–701, 2018.
 - [23] M. Pule, A. Yahya, and J. Chuma, “Wireless sensor networks: a survey on monitoring water quality,” *Journal of Applied Research and Technology*, vol. 15, no. 6, pp. 562–570, 2017.
 - [24] A. M. Obeid, N. Atitallah, K. Loukil, M. Abid, and M. Bensalah, “A survey on efficient power consumption in adaptive wireless sensor networks,” *Wireless Personal Communications*, vol. 101, no. 1, pp. 101–117, 2018.
 - [25] A. Toor, S. U. Islam, N. Sohail et al., “Energy and performance aware fog computing: a case of DVFS and green renewable energy,” *Future Generation Computer Systems*, vol. 101, pp. 1112–1121, 2019.
 - [26] U. Ullah, A. Khan, M. Zareei, I. Ali, H. A. Khattak, and I. U. Din, “Energy-effective cooperative and reliable delivery routing protocols for underwater wireless sensor networks,” *Energies*, vol. 12, no. 13, p. 22, 2019.
 - [27] C. Lin, J. Zhou, C. Guo et al., “TSCA: a temporal-spatial real-time charging scheduling algorithm for on-demand architecture in wireless rechargeable sensor networks,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 211–224, 2018.
 - [28] K. Xie, X. Ning, X. Wang et al., “An efficient privacy-preserving compressive data gathering scheme in WSNs,” *Informaion Sciences*, vol. 390, pp. 82–94, 2019.
 - [29] D. Vinodha and E. A. Mary Anita, “Secure data aggregation techniques for wireless sensor networks: a review,” *Archives of Computational Methods in Engineering*, vol. 26, no. 4, pp. 1007–1027, 2019.
 - [30] K. Haseeb, A. Almogren, N. Islam, I. U. Din, and Z. Jan, “An energy-efficient and secure routing protocol for intrusion avoidance in IoT-based WSN,” *Energies*, vol. 12, pp. 18–21, 2019.
 - [31] X. Liu, A. Liu, T. Wang et al., “Adaptive data and verified message disjoint security routing for gathering big data in energy harvesting networks,” *Journal of Parallel and Distributed Computing*, vol. 135, pp. 140–155, 2020.
 - [32] X. Qi, Z. Zhang, L. Liu, and M. Xie, “The degree-constrained adaptive algorithm based on the data aggregation tree,” *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 870792, 7 pages, 2014.
 - [33] H. Ö. Tan and I. Körpeoğlu, “Power efficient data gathering and aggregation in wireless sensor networks,” *ACM SIGMOD Record*, vol. 32, no. 4, pp. 66–71, 2003.