Article

Optimal Data Transmission for WSNs with Data-Location Integration

Shuiyan Wu ¹, Xiaofei Min ² and Jing Li ^{2,*}
¹ College of Mathematics and Information Science, Xianyang Normal University, Xianyang 712099, China; shuiyanwu@xync.edu.cn

² School of Mathematics and Statistics, Xidian University, Xi'an 710126, China; 20071212579@stu.xidian.edu.cn

* Correspondence: lijingmath@xidian.edu.cn; Tel.: +86-136-791-78646

Abstract: Wireless sensor networks (WSNs) have good performance for data transmission, and the data transmission of sensor nodes has the function of symmetry. However, the wireless sensor nodes are facing great pressure in data transmission due to the increasing amount and types of data that easily cause premature energy consumption of some nodes and, thus, affects data transmission. Clustering algorithm is a common method to balance energy consumption, but the existing algorithms fail to balance the network load effectively for big data transmission. Therefore, an optimal data transmission with data-location integration (ODTD-LI) is proposed for WSNs in this paper. For optimal data transmission, we update the network topology once for one round. In the proposed algorithm, we perform calculations of the optimal cluster heads, clustering and data transmission routing through three steps. We first deploy N homogeneous and symmetry nodes in a square area randomly and calculate the optimal number of cluster heads according to the node locations. Then, the optimal number of cluster heads, energy consumption, the distances and degrees of the nodes are taken into consideration during the clustering phase. Direct communication is carried out within a cluster, and the member nodes of the cluster pass the information directly to the cluster head. Lastly, an optimal hybrid routing from each cluster node to Sink is constructed for data transmission after clustering. The simulations verify the good performance of the proposed algorithm in view of the lifetime, average delay, coverage rate (CR) and load balance of the network compared with the existing algorithms. Through the research conducted in this paper, we find that our work has good performance for selecting the hybrid routing in the network with the nodes randomly arranged.

Keywords: wireless sensor networks; optimal cluster head; clustering algorithm; data-location integration; energy efficiency

Citation: Wu, S.; Min, X.; Li, J. Optimal Data Transmission for WSNs with Data-Location Integration. *Symmetry* **2021**, *13*, 1499. <https://doi.org/10.3390/sym13081499>

Academic Editor: Basil Papadopoulos

Received: 20 July 2021

Accepted: 13 August 2021

Published: 16 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wireless sensor networks (WSNs) are a kind of wireless self-organizing network without infrastructure, which can collaborate on monitoring in real time and collect the various environmental or object information within its monitoring area. WSNs deal with the collected data to obtain accurate information, and the WSN finally delivers the information to the user [1]. The network structure diagram is shown in Figure 1. The wireless sensor nodes can be characterized by large quantity, small size, weak communication and computing power, limited and irreparable power supply, heterogeneous energy and so on [2]. Due to the limited attributes such as bandwidth, data storage capacity and node energy, the question of how to satisfy the quality of service (QoS), such as delay and the fault tolerance of the network, while simultaneously maximizing the network lifetime is the biggest challenge for WSNs [3,4]. The main purpose of WSN research includes two aspects: (1) the developments of new networking technologies, which are required to accommodate the highly dynamic environments and self-organizing capabilities; and (2) the developments of networked information processing technologies, which are required

to quickly extract useful, reliable and timely information to users. Therefore, WSNs as one kind of the important technology could have great impact on the 21st century.

Figure 1. The diagram of the network structure.

On the other hand, with the rapid development of wireless communication technology, artificial intelligence (AI), big data and computing technology, WSN-based big data transmission technology has attracted much attention from many countries. The data transmission in WSNs combines sensor technology, embedded technology, distributed information processing technology, communication technology, etc. Moreover, big data transmission technology [5,6] based on multi-layer heterogeneous WSNs plays a crucial role. In general, the sensor nodes have low initial energy but big data transmission always causes a large amount of energy consumption, which could induce conflicts. Furthermore, due to the damage of sensor nodes, loss of hardware, geographical location, signal noise and other reasons, big data transmission often results in inaccurate or delayed data acquisition. However, the repair costs of sensors are relatively high, and it is also hard to avoid the above problems again after repairing. Therefore, maintaining the optimal topology dynamically to satisfy the QoS of the data transmission has important research significance.

In summary, many achievements have been made in the WSNs research and applications in recent years [3–36]. By constructing the optimal data traffic transmission path between the cluster heads and Sink, these clustering algorithms optimize the network data transmission routing and have been used in many practical applications, such as environmental monitoring, internet of things, intelligent city and vehicle-to-everything communications. However, there are still many new problems that need to be solved urgently because of people's increasing demands for big data transmission. The energy of sensor nodes is likely to be drained unnecessarily and quickly, which results in the death of some sensor nodes in the network. Therefore, the question of how to propose an optimal multi-layer heterogeneous network topology with fast convergence and low energy consumption for big data transmission in WSNs has been an important topic. The innovative achievement of this research lies in the fact that data and location information of the sensor nodes are combined together to obtain the global optimization of the network data transmission topology.

Motivated by the aforementioned observation, we explore an optimal data transmission for WSNs with data-location integration. The proposed algorithm uses a mathematical modeling method to study the distributed multi-layer network data transmission algorithm and finally builds an optimal hybrid multi-layer heterogeneous network topology to meet the QoS requirements of big data transmission and strong scalability and universality.

The main contributions are listed below.

(1) The optimal hybrid topology can provide multi-QoSs (the lifetime of network, delay, CR and load balance) support for data transmission in WSNs.

(2) The designed big data transmission strategy can meet the target of low-power consumption and real-time data transmission under the condition of limited network resources.

(3) Frontier crossover research driven by practical problems is the feature of this paper. Research problems are highly applicable, and the research results provide strong theoretical support for many fields. For example, the efficient parameter self-adaptation in the proposed algorithm provides a theoretical basis for the automatic network.

Here, we emphasize the necessity of our research again. In a WSN with limited resources, the constructed optimal data transmission mathematical model is always a NP-hard problem when there are multi-constraint conditions. The optimization target includes the maximum network CR, lowest energy consumption, highest data accuracy, etc. In order to solve this problem, we propose a new hybrid optimal multi-layer heterogeneous network topology with data-location integration for data transmission.

The remainder of the paper is organized as follows. Section 2 provides the related work of the problem. Section 3 introduces the optimal cluster head selection algorithm based on geographical location. Section 4 presents the system model and the ODTD-LI algorithm in detail. The simulation results and performance analysis are presented in Section 5. Section 6 provides the conclusion of this paper.

2. Related Work

Different network topologies result in different performance for data transmission. In 2000, Heinzelman, Chandrakasan and Balakrishnan proposed the classical clustering LEACH algorithm for data transmission in homogeneous WSNs for the first time [12]. The LEACH algorithm ensures low energy consumption on account of the rotation mechanism of cluster head, and the network topology is shown in Figure 2.

Figure 2. The network topology of the LEACH algorithm.

The LEACH algorithm has been recognized and improved by a large number of researchers, which also ignites researchers' enthusiasm for the research of data transmission based on the multi-layer network. In 2002, Heinzelman improved their LEACH algorithm and then proposed the C-LEACH algorithm [13]. The C-LEACH algorithm renders energy consumption more stable by reducing the probability of the nodes being selected as cluster heads again, for which its residual energy would be lower. Younis and Fahmy propose another classical HEED clustering algorithm [14]. The HEED algorithm considers the

maximum residual energy of the nodes in cluster head selection phase, and load balancing is further improved. In the I-SEP algorithm [15], the super sensor nodes are deployed in the monitoring area for data transmission, which renders the network performance stronger than homogeneous WSNs.

As previously mentioned, stronger network topology needs to be designed to satisfy the modern big data transmission characteristics. The distributed EEUC algorithm [16] considers the node energy consumption and the distance between the source node and Sink in the cluster head selection phase. The EEUC algorithm marks all the nodes and equalizes network energy consumption. In the CUCA algorithm [17], the selection of the cluster head considers not only the distance factor but also the factors of network coverage and node residual energy. In ERA algorithm [18], the cluster head establishes multi-path data transmission according to the residual energy, but it is easy to cause premature emergence of the energy hole. Nabajyoti et al. [19] proposed a distributed fuzzy logic algorithm based on the energy perception and CR in 2017, where the selected cluster heads were non-uniform. However, the non-uniformity of cluster heads cannot meet the requirements of minimum energy consumption and maximum CR of the network. In 2018, a distributed multi-layer DFCR clustering algorithm [20] was proposed with better big data transmission performance. The DFCR algorithm takes into account the degree, centrality, residual energy, distance and other factors of the nodes and then constructs a multi-objective optimization algorithm to meet the uneven consumption of the nodes' energy. Zhang et al. [21] proposed an adaptive consensus-based distributed target tracking with dynamic cluster in sensor networks. The paper is concerned with the target tracking problem over a filtering network with dynamic cluster and data fusion. In recent two years, the clustering algorithm has been improved for some practical applications and achieved good performances [22,23].

Some tree-based hybrid topologies also have good performance for big data transmission in some practical multi-layer heterogeneous WSNs. Zhang et al. proposed an effective hybrid link-adding strategy in order to enhance the network traffic for scale-free networks, and then they proposed a related routing link-adding strategy for large-scale heterogeneous WSNs in 2019 [24]. Chowdhury et al. [25] proposed a hybrid heterogeneous topological structure in order to meet the real-time performance for big data transmission. Yao et al. [26] have proposed an energy-adaptive and bottleneck-aware tree-based communication scheduling for battery-free WSNs, and the simulations show that the algorithm has high performance in terms of communication latency and energy usage ratio. Zhang et al. [27] proposed an optimal tree-based topology in order to meet the requirements of big data transmission by improving local optimization relative to global optimization.

Nowadays, WSNs are typically used with dynamic conditions of task-related environments for monitoring and gathering raw sensor data. It is hard to use the traditional techniques to solve these dynamic data transmission tasks. As a good solution to these problems, machine learning techniques were able to successfully handle these dynamic situations [28]. For optimization of microgrid-connected WSNs with load balancing and intrusion detection, some authors [29,30] use the data-driven deep learning approach to solve the above-mentioned data transmission tasks with dynamic conditions. We find that the proposed method [29,30] can outperform the state-of-the-art solutions in terms of recognition accuracy. Zhou et al. [31] also trained deep neural network in WSNs for image target recognition when data, energy and computation resources are limited.

Wang et al. [32] proposed an efficient big data transmission EEICS technology. The main contribution of EEICS is the safe data transmission mechanism within the cluster and energy-efficient model. In EEICS, the network is divided into a number of different layers based on the distance from node to Sink, and the optimal node is selected as the relay node in each layer. Thus, EEICS maintains the network overhead effectively. There are also some other hybrid topologies with good performance for big data transmission [33–36]. Much of the research achievements have been widely used in the military, Internet of Things (IoT), artificial intelligence (AI), environmental monitoring (EM), modern intelligent

transportation, convenient express, modern medical science, industrial application, space exploration and other fields. These applications reflect the importance of this issue.

3. Optimal Clustering Method Based on the Node Locations

In order to better describe the proposed algorithm, we introduce the following assumptions of the network.

- (1) Deploy N homogeneous and symmetry nodes in a square area randomly, and the side length of the square is L . The nodes cannot be moved after they are deployed.
- (2) Each node has an identical transmission radius r .
- (3) Each node knows the information of its neighbor node.
- (4) The initial energy of each node is identical.
- (5) The node has the function of data aggregation.
- (6) The position of Sink can be adjusted.
- (7) The network is updated once for one round.
- (8) There is no attack node in the network.

3.1. Energy Model

The energy model in this paper is similar to that of the LEACH algorithm. The energy consumption of transmitting, receiving and aggregating data is formulated as follows, respectively:

$$E_{tr}(l) = \begin{cases} \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d^2 & d \leq d_0, \\ \cdot E_{elec} + l \cdot \varepsilon_{mp} \cdot d^4 & d > d_0, \end{cases} \quad (1)$$

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

$$E_{ag}(l) = lE_{da}, \quad (3)$$

where d is the distance from source node to Sink; $d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$; $E_{tx}(l)$, $E_{rx}(l)$, $E_{ag}(l)$ represent the energy consumption, respectively, when the node transmits, receives and aggregates l bits data; E_{elec} is the energy consumption of wireless transceiver circuit in transmitting and receiving data; E_{da} represents the energy consumption while aggregating 1 bit data; ε_{fs} and ε_{mp} represent the magnifications of the signal amplifier power under the free-space model and the multi-channel attenuation model, respectively.

Obviously, if $d \leq d_0$, the communication model of the network is a free-space model, otherwise a multi-channel attenuation model.

3.2. Optimal Number of the Cluster Heads

The first step in this paper is to construct the network topology according to the LEACH algorithm. Since the network selects the optimal cluster heads to obtain the maximum utilization of node energy, we need to calculate the optimal number of cluster heads based on the node locations. Firstly, we use the integral to compute the expectation of distances between the cluster heads and sink. Then, we obtain the expectation of distances between the cluster heads and the nodes within its cluster. At this time, we may obtain the optimal number of cluster heads by calculating the entire expected value of energy consumption. We select the optimal cluster heads by considering the aspects such as energy consumption of the nodes, degrees and distances of the nodes in the subsequent process.

Assume that there are k clusters and N nodes in the network. The energy consumption of the network under the free-space model is as follows:

$$E_{CH} = \left(\left\lfloor \frac{N}{k} \right\rfloor - 1 \right) \cdot l \cdot E_{elec} + \frac{N}{k} \cdot l \cdot E_{da} + l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d_{CH-Sink}^2, \quad (4)$$

where $d_{CH-Sink}$ represents the expected distance between the cluster head and Sink; the average number of nodes in each cluster is $\left\lfloor \frac{N}{k} \right\rfloor$; $\left(\left\lfloor \frac{N}{k} \right\rfloor - 1 \right) \cdot l \cdot E_{elec}$ represents the energy

consumption for the cluster head to receive $\left(\left\lfloor \frac{N}{k} \right\rfloor - 1\right) \cdot \text{bits data}$; $\frac{N}{k} \cdot l \cdot E_{da}$ is the energy consumption for the cluster head to aggregate the receiving data; and $l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d_{CH-Sink}^2$ represents the energy consumption for the cluster head to transmit l bits data.

In a free-space model, denote the location of the optimal cluster head as (x, y) , and the coordinate of Sink as (a, b) . Therefore, the expectation of squared distance between the cluster head and Sink is the following:

$$\begin{aligned} \mathbf{E}[d_{CH-Sink}^2] &= \iint ((a-x)^2 + (b-y)^2) d_{CH}(x, y) dx dy \\ &= \iint \frac{(a-x)^2 + (b-y)^2}{S} dx dy, \end{aligned} \quad (5)$$

where $d_{CH}(x, y)$ represents the distance from source node to cluster head; S is the square area of node distribution.

Assume that there are l bits data within one cluster to be transmitted. The energy consumption within one cluster in one round is as follows:

$$E_{CH-members} = l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d_{CH-members}^2 \quad (6)$$

where $d_{CH-members}$ is the distance from cluster head to source node within one cluster.

The expectation of squared distance from cluster head to source node within one cluster is as follows.

$$\mathbf{E}[d_{CH-members}^2] = \iint (x^2 + y^2) d(x, y) dx dy = \frac{L^2}{2\pi k}. \quad (7)$$

The energy consumption of one cluster is as follows.

$$E_{cluster} = E_{CH} + \left\lfloor \frac{N}{k} \right\rfloor E_{CH-members}. \quad (8)$$

Therefore, the energy consumption in one round is the following.

$$\begin{aligned} E_{round} &= k \cdot E_{cluster} \\ &= l \cdot (2 \cdot N \cdot E_{elec} + N \cdot E_{da} + k \cdot \varepsilon_{fs} \cdot d_{CH-Sink}^2 + N \cdot \varepsilon_{fs} \cdot d_{CH-members}^2). \end{aligned} \quad (9)$$

Then, we can obtain the optimal number of cluster heads.

$$N_{opt-CH} = \left\lceil \frac{\sqrt{N}}{\sqrt{2\pi}} \cdot \frac{L}{d_{CH-Sink}} \right\rceil. \quad (10)$$

Similarly, the optimal number of cluster heads in the multi-channel attenuation model space can be obtained as follows.

$$N_{opt-CH} = \left\lceil \frac{\sqrt{N}}{\sqrt{2\pi}} \cdot \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \cdot \frac{L}{d_{CH-Sink}^2} \right\rceil. \quad (11)$$

3.3. The Scheme of Cluster Head Selection

In this paper, the selection of cluster heads takes into account the comprehensive consideration of the remaining node energy, node degree of v_i (i.e., the number of nodes that can communicate with node v_i), the distances from nodes to Sink and the CR. CR is a ratio of the nodes in the constructed data transmission topology to the total nodes, and we use CR to represent it hereinafter. It's worth noting that the selection of cluster heads in this paper is different from that of the LEACH algorithm, which randomly selects sensor nodes as cluster heads according to the preset probability in the network.

Based on the above considerations, we first calculate the candidate value of node v_i being a cluster head as follows:

$$CH_{Candidate-i} = \alpha \cdot \left(\frac{E_{Remain-i}}{E_{Initial}} \right) + \beta \cdot \left(\frac{Deg_{max}}{Deg_i} \right) + \gamma \cdot \left(\frac{Dis_{i-Sink}}{Dis_{max-Sink}} \right), \quad (12)$$

where α, β, γ represent the parameters and satisfy $0 < \alpha, \beta, \gamma < 1, \alpha + \beta + \gamma = 1$; $E_{Remain-i}$ represents the remaining energy of node v_i ; $E_{Initial}$ represents the initial energy of all the nodes; Deg_i represents the degree of node v_i ; Deg_{max} represents the maximum degree of all the nodes; Dis_{i-Sink} represents the distance between node v_i and Sink; $Dis_{max-Sink}$ represents the maximum distance from source nodes to Sink.

When the candidate values are sorted from high to low according to the above equation, the candidate set $Candidate_CH\{\}$ is obtained. Then, the optimal cluster head is selected from the candidate sets when the minimum CR of the network is equal or greater than 90%. The set of cluster heads $List\ CH\{\}$ is also established.

The cluster head selection method is shown as Algorithm 1.

Algorithm 1 The cluster head selection algorithm.

Require:

N // number of alive nodes
 $V(i, j), i, j = 1, 2, \dots, N$ // location of alive nodes
 $S(a, b)$ // location of Sink
 $Neighbor_{v_i}\{\}$ // neighbor information of node v_i
 r // node transmission range

Ensure:

$List\ CH\{i \mid i = 1, \dots, N_{opt-CH}\}$
1: Declare $List\ CH\{\} = NULL$
2: **for** $i = 1$ **to** N **do**
3: Calculate the candidate value of node v_i as a cluster head
 $CH_{Candidate-i} = \alpha \cdot \left(\frac{E_{Remain-i}}{E_{Initial}} \right) + \beta \cdot \left(\frac{Deg_{max}}{Deg_i} \right) + \gamma \cdot \left(\frac{Dis_{i-Sink}}{Dis_{max-Sink}} \right)$
4: **end for**
5: **for** $i = 1$ **to** N **do**
6: $n = +1$
7: **for** $j = i + 1$ **to** N **do**
8: **if** $n \neq N_{opt-CH}$ **then**
9: **if** $CR < 90\%$ **then**
10: Select next node v_j
11: **end if**
12: **end if**
13: **end for**
14: **end for**

4. Our Proposed ODTD-LI Algorithm

In this section, we construct an optimal hybrid routing from each cluster head to the Sink after clustering. In order to reduce the complexity of the algorithm, the transmission within one cluster is a one-way direct transmission, which means that the nodes within the cluster transmit the collected data to the cluster head by one hop directly. When the data within the cluster are collected, the cluster head needs to transfer the data to Sink. Researchers [10,16,27,35] often employ the following four strategies: (1) The cluster heads transmit the collected data to Sink directly through one hop. (2) The Sink uses the greedy algorithm to construct a multi-hop routing to the cluster heads, and there are only cluster heads in the routing. (3) The Sink uses the greedy algorithm to construct the multi-hop routing to the cluster heads. There are only non-cluster heads in the routing, and all the

cluster heads are eaf nodes, which are not used as relay nodes again. (4) The hybrid routing is used to construct the optimal routing from the cluster head to the Sink. Both cluster heads and non-cluster heads nodes can be used as relay nodes.

In Strategy (1), the energy of a few cluster heads is likely to be drained because they need to transfer much more data than other nodes. In Strategy (2), the nodes near the Sink also need to transfer much more data than other nodes, and the energy holes are easily appear near Sink. In Strategy (3), all the cluster heads are eaf nodes, and they are no longer responsible for forwarding data from other cluster nodes, which results in a waste of node resources. In Strategy (4), all nodes in the network can be used for data transmission, and the data traffic is effectively averaged across all nodes. This strategy can effectively balance the data load on nodes. In this way, the life of the node can be extended.

In order to enhance the load balance of the nodes, we chose Strategy (4) for data transmission and constructed a hybrid integer linear mathematical model in order to solve the aforementioned optimal problem. Assume that the number of the network rounds is R . Therefore, the mathematical programming problem of maximum CR and minimum energy consumption is established as follows.

$$\max CR = \frac{\sum_{i=1}^{N_{opt-CH}} |ListCH[i]|}{N}, \quad (13)$$

$$\min E = \sum_{i=1}^R E_{round-i}, \quad (14)$$

s.t.

$$C_1: E_{round-k} \geq 0, \forall k = 1, 2, \dots, R, \quad (15)$$

$$C_2: Candidate_CH\{i\} \neq NULL, \forall i = 1, 2, \dots, N, \quad (16)$$

$$C_3: List\ CH\{i\} \neq NULL, \forall i = 1, 2, \dots, N_{opt-CH}, \quad (17)$$

$$C_4: 0 < Deg_{CH-i} < \left\lceil \frac{N}{N_{opt-CH}} \right\rceil, i = 1, 2, \dots, N_{opt-CH}, \quad (18)$$

where $E_{round-k}$ represents the energy consumption of the network in round k ; R is the network running rounds; Deg_{CH-i} is the number of nodes with cluster head v_i ; and $|ListCH[i]|$ is the number of nodes within the set of cluster $ListCH[i]$.

Equation (13) represents the maximum number of nodes in the network topology, which meets the requirement of the maximum CR. Equation (14) represents the lowest energy consumption target of the whole network. Equation (15) represents the energy consumption constraint of each round. Equations (16) and (17) are non-empty constraints on the set of the representative cluster heads. Equation (18) is a constraint on the number of nodes in each cluster, ensuring that the members in each cluster are appropriate for meeting the load balance of data.

There are two objectives (13) and (14) at the same time here, so we need to further combine them into a single objective function. Here, we introduce an error value ε to represent the combined single objective function as follows:

$$F = \kappa_1 \cdot \frac{\sum_{i=1}^{N_{opt-CH}} |ListCH[i]|}{N} + \kappa_2 \cdot \frac{1}{\sum_{k=1}^R E_{round-k}} + \varepsilon, \quad (19)$$

where κ_1 and κ_2 are the assumed linear correlation coefficients.

In Equation (19), the mathematical expectation of ε is expected to be 0, that is:

$$E(\varepsilon) = 0. \quad (20)$$

Therefore, Equations (13)–(18) are changed to the following:

$$\max F = \kappa_1 \cdot \frac{\sum_{i=1}^{N_{opt-CH}} |ListCH[i]|}{N} + \kappa_2 \cdot \frac{1}{\sum_{k=1}^R E_{round_k}} + \varepsilon, \quad (21)$$

such that

$$C_1: E_{round-k} \geq 0, \forall k = 1, 2, \dots, R, \quad (22)$$

$$C_2: Candidate_CH\{i\} \neq NULL, \forall i = 1, 2, \dots, N, \quad (23)$$

$$C_3: List\ CH\{i\} \neq NULL, \forall i = 1, 2, \dots, N_{opt-CH}, \quad (24)$$

$$C_4: 0 < Deg_{CH-i} < \left\lceil \frac{N}{N_{opt-CH}} \right\rceil, i = 1, 2, \dots, N_{opt-CH}. \quad (25)$$

However, in practice, due to the nonlinear relationship between these two optimization goals, it is difficult to combine them, which means the optimization problem is NP-hard. Thus, we need to use the heuristic algorithm to approximate the solution.

Algorithm 2 provides the optimal tree-clustering energy-efficient algorithm for secure big data transmission. Firstly, the network is divided into several clusters according to the LEACH algorithm. Secondly, the optimal number of cluster heads is calculated, and the network is divided into a certain number of clusters according to the parameters such as the distances, degrees and energy of nodes. Then, we perform security detection within the cluster. Finally, an optimal tree-clustering energy-efficient data transmission model is constructed from the cluster heads to Sink.

In order to better describe the proposed algorithm, a specific flow chart is shown in Figure 3.

Figure 3. The flow chart of the proposed algorithm.

Algorithm 2 Optimal data transmission for WSNs with data-location integration.**Require:**

N // number of alive nodes
 $V(i, j), i, j = 1, 2, \dots, N$ // location of alive nodes
 $S(a, b)$ // location of Sink
 $Neighbor_{v_i}\{\}$ // neighbor information of node v_i
 r // node transmission range
 $v_i \in V$

Ensure:

$List\ CH\{i\}$.

```

1: Declare Data transmission routing Set  $\{\} = NULL$ 
2: for each node do
3:   Calculate  $T(n) = \begin{cases} (\frac{p}{1 - p \cdot (r \cdot \text{mod}(1/p))}), & n \in G \\ 0, & n \notin G \end{cases}$ 
4:   if  $rand(v_i) < T(v_i)$  then
5:     elect  $v_i \in List\ CH\{1, 2, \dots, N_{opt-CH}\}$ 
6:   end if
7: end for
8: after run 1 round do
9:   while  $rounds < R$  do
10:    //Cluster election phase
11:    for  $i = 1$  to  $N$  do
12:      if node  $v_i \in List\ CH\{1, 2, \dots, N_{opt-CH}\}$  then
13:        declare node  $v_i$  as a CH node
14:      end if
15:    end for
16:    //Cluster construction phase
17:    for  $i=1$  to  $N_{opt-CH}$  do
18:       $CHv_i$  broadcasts its information to its cluster members
19:      for  $j = 1$  to  $N$  do
20:        if  $Dis_{j-CH(i)} < Dis_{j-CH(others)}$  then
21:          Node  $v_j$  attach with  $CHv_i$ 
22:        end if
23:      end for
24:    end for
25:    // Tree-clustering based multi-routing data
26:    for  $i = 1$  to  $N_{opt-CH}$  do
27:      // Data transmission phase
28:      for  $j = 1$  to  $N$  do
29:        use greedy algorithm to construct a minimum distance-based transmission tree
        from node  $v_i$  to Sink
30:        if the relay node of  $v_i$  is  $v_j$  then
31:          node  $v_i.parent \rightarrow$  node  $v_j$ 
32:        end if
33:      end for
34:    end for
35:     $Rounds = +1$ 
36:  end while

```

5. Experimental Results and Analysis

In order to verify the effectiveness of Algorithm 2, we perform the simulation experiments by comparing with the other four algorithms in this section: LEACH, Greedy Incremental Tree Algorithm (GIT), Power Efficient Data Gathering and Aggregation Protocol (PEDAP) and EEICS. The simulation environment is VC6.5 and MATLAB 2017B. The simulation parameters are shown in Table 1.

Table 1. Simulation parameters.

Parameters	Values
Network Size	300 m × 300 m
Sink Location	(0, 0)
Number of nodes	300
Initial Energy	2
Link Delays	1×10^{-12}
Data Packet Size	1000 bits
Data Packet rate	1 packet/s
r	32 m
Channel width	250 kbps
Second aggregation data numbers	200 kbps
E_{elec}	50 nJ/bit
E_{da}	5 nJ/bit/unit
p	5%
ε_{fs}	10 pJ/bit/m ²
ε_{mp}	0.0013 pJ/bit/m ⁴

5.1. The Lifetime of the Network

We usually use the survival nodes in the network with change in the network running time in order to reflect the ifetime of the network.

In Figure 4, we can observe that the node death speed becomes faster and faster when the running time of the network increases.

Figure 4. The ifetime of the network.

This phenomenon is caused by the oad increase in the node ink, which renders the node unable to bear the subsequent data transmission tasks. Particularly for the LEACH algorithm, the ratio of dead nodes is higher than ratio in the other four algorithms because LEACH only considers the node distances in the construction of the data transmission path.

The number of alive nodes can reflect the variation of the network performance. As shown in Figure 4, the proposed ODTD-LI algorithm improves the network ifetime effectively. When the first node dies, the network ifetimes of ODTD-LI, EEICS, PEDAP, GIT and LEACH reach 8015 rounds, 6760 rounds, 4632 rounds, 4520 rounds and 4026 rounds, respectively. When there are 10 nodes dead, the network ifetimes of ODTD-LI, EEICS, PEDAP, GIT and LEACH reach 9268 rounds, 8074 rounds, 8106 rounds, 7835 rounds and

7726 rounds, respectively. When there are 10 nodes dead, the performance improvement rates of the network lifetime in ODTD-LI can reach up to 30.36% and 34.34%, respectively, compared with EEICS and PEDAP because the proposed ODTD-LI algorithm is better in the construction of the optimal hybrid routing. In particular, the number of optimal cluster heads, energy, distances and degrees of nodes is taken into consideration in our scheme, which improves the network load balance in the tree maintenance phase.

5.2. Average Delay of the Network

Real-time data transmission is one important measure of network overhead. In this paper, the latency time (the network average delay) for data transmission represents the average data transmission time from all nodes to Sink. From Figure 5, we can observe that the proposed ODTD-LI algorithm has lower delay compared with EEICS, PEDAP, GIT and LEACH.

Figure 5. Average delay of the network.

Unfortunately, the average delay of LEACH is the lowest in all the algorithms. This is because relay nodes handed down in LEACH number the least, and the network topology is single, which is basically point-to-point data transmission. However, in the proposed ODTD-LI algorithm, the communication time of sink and the computing time of nodes increased correspondingly because when the number of network nodes increases, the number of sinks and the amount of data increase.

For the hybrid topology with 300 nodes, the delays are 20.63 ms, 21.8 ms, 22 ms, 22.18 ms and 22.35 ms for ODTD-LI, EEICS, GIT, PEDAP and LEACH, respectively. The performance improvement ratios almost reach 6.9% and 7.7%, respectively, compared with GIT and PEDAP. Of course, we can observe from Figure 5 that the overall delays of all the algorithms grow with the increase in network nodes. This is because the density of nodes results in the increase in data and sub-routes that, in return, results in an increase in data transmission delays from the source nodes to Sink. Fortunately, we construct an optimal clustering model at the beginning, and then construct an optimal routing from each cluster head to Sink after clustering. In this manner, the delay is effectively reduced.

5.3. Node CR

Node CR represents the transmission capability of the network. The higher the node CR is, the higher the correlation between the nodes and the higher the accuracy of the collected data. Figure 6 provides the node CRs of the network when the network runs 10,000 and 20,000 rounds, respectively. When the network running rounds increase, the

node CR of all the five algorithms decrease because the dead nodes gradually appear in the network. We can observe from Figure 6 that the proposed ODTD-LI algorithm has higher node CR than the other compared algorithms, which explains the advantage of the proposed ODTD-LI algorithm in maintaining the capability of data collection and transmission from another aspect.

Figure 6. Node CR.

5.4. Load Balance

Load balance can reflect the data transmission capacity of each node and each routing, and we always use the variance of energy consumption to represent the load balance of the network. Figure 7 shows the variance of energy consumption under 300 nodes, and we can observe that the proposed ODTD-LI algorithm has lower variance than the compared algorithms. Remarkably, ODTD-LI maintains a small and stable variance of energy consumption with the increase in the network rounds.

Figure 7. The variance of energy consumption.

The LEACH, GIT and PEDAP algorithms have higher variances of energy consumption with the increase in rounds. This is because the algorithms do not maintain the constructed network topology dynamically. For the above three algorithms, data congestion easily occurs in part of the network when there are dead nodes in the network. As a result, the data load cannot be balanced for the data transmission routing. However, the EEICS and ODTD-LI maintain the network topology dynamically, and the data load of each node is balanced. In addition, compared with the EEICS algorithm, the proposed ODTD-LI algorithm improves the data transmission routing based on data-location integration. Therefore, the data sizes in each node and routing are balanced.

6. Conclusions and Future Work

This paper mainly improves the clustering phase, the data transmission path and the communication mechanism. Firstly, the number of the optimal cluster heads is calculated according to the geographic location information. Then, the optimal cluster heads are selected according to energy, node degree and other information. Lastly, we construct an optimal hybrid routing for big data transmission. The simulation results show the good performance of the proposed ODTD-LI algorithm in terms of the lifetime, delay and network CR, and load balance. This research mainly optimizes the network topology and data transmission scheme from the view of the network layer. Compared with the EEICS algorithm, the proposed ODTD-LI algorithm achieves 2.59%, 14.38% and 10.74% improvement in key performance indicators (KPIs) of the network average delay, lifetime and node CR, respectively. The proposed ODTD-LI algorithm also has good stability in energy consumption, which can be reflected from Figure 7.

In future works, we plan to design a scalable and secure big data transmission algorithm by combining big data transmission and the modern intelligent optimization algorithm with the multi-layer heterogeneous network. This could provide new ideas and methods for improving the use ratio of the network, and the proposed algorithm could meet the network QoS better and offer the theoretical and technical support for secure transmission of big data.

Author Contributions: Methodology, S.W.; Validation, X.M.; Writing original draft, J.L.; Review and editing, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Natural Foundation of Xianyang Normal University under Grant (12XSYK011).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Areg, K.; Chi-Kin, C.S.; Khaled, E. Multisensor Adaptive Control System for IoT-Empowered Smart Lighting with Oblivious Mobile Sensors. *ACM Trans. Sens. Netw.* **2020**, *11*, 1–12.
2. Wang, H.; Hxu, Z.; Fujita, H.; Liu, S. Towards felicitous decision making: An overview on challenges and trends of big data. *Inf. Sci.* **2016**, *367*, 747–765. [\[CrossRef\]](#)
3. Shen, D.L.J.; Wang, A.X. Lightweight and practical node clustering authentication protocol for hierarchical wireless sensor networks. *Int. J. Sens. Netw.* **2018**, *27*, 95–102.
4. Lin, D.; Wang, Q.; Min, W. A survey on Energy-Efficient Strategies in Static Wireless Sensor Networks. *ACM Trans. Sens. Netw.* **2020**, *17*, 1–38. [\[CrossRef\]](#)
5. Vijayalakshmi, K.; Ananda, P. Global evy flight of cuckoo search with particle swarm optimization for effective cluster head selection in wireless sensor network. *Intell. Autom. Soft Comput.* **2020**, *26*, 303–311.
6. Miao, Y.; Wu, G.; Lin, C.; Hossain, M.; Muhanmad, G. Green Cognitive Body Sensor Network: Architecture, Energy Harvesting, and Smart Clothing-Based Applications. *IEEE Sens. J.* **2019**, *19*, 8371–8378. [\[CrossRef\]](#)
7. Manisha, R.; Sushil, K.; Amir, G. Ant Colony Optimization Based Quality of Service Aware Energy Balancing Secure Routing Algorithm for Wireless Sensor Networks. *IEEE Trans. Eng. Manag.* **2021**, *68*, 170–182.

8. Mehto, A.; Tapaswi, S.; Pattanaik, K.K. A review on rendezvous based data acquisition methods in wireless sensor networks with mobile sink. *Wirel. Netw.* **2020**, *26*, 2639–2663. [\[CrossRef\]](#)
9. Anamika, S.; Siddhartha, C. Sensor Fusion for Distributed Detection of Mobile Intruders in Surveillance Wireless Sensor Networks. *IEEE Sens. J.* **2020**, *20*, 15224–15231.
10. Lin, C.; Han, D.; Deng, T. P2S: A Primary and Passer-by Scheduling Algorithm for On-demand Charging Architecture in Wireless Rechargeable Sensor Networks. *IEEE Trans. Veh. Technol.* **2017**, *66*, 8047–8058. [\[CrossRef\]](#)
11. Lin, C.; Zhou, J.; Guo, C. TSCA: A Temporal-Spatial Real-Time Charging Scheduling Algorithm for On-Demand Architecture in Wireless Rechargeable Sensor Networks. *IEEE Trans. Mob. Comput.* **2018**, *17*, 211–224. [\[CrossRef\]](#)
12. Heinzelman, W.; Chandrakasa, A.; Balakrishnan, H. Energy efficient communication protocol for wireless microsensor. In Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui, HI, USA, 7 January 2000.
13. Heinzelman, W.B.; Chandrakasan, A.P.; Balakrishnan, H. An application-specific protocol architecture for wireless micro-sensor networks. *IEEE Trans. Wirel. Commun.* **2002**, *1*, 660–670. [\[CrossRef\]](#)
14. Younis, O.; Fahmy, S. Heed: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Trans. Mob. Comput.* **2004**, *3*, 366–379. [\[CrossRef\]](#)
15. Behera, T.; Mohapatra, S.; Samal, U.; Khan, M.; Daneshmand, M.; Gandomi, A. I-SEP: An Improved Routing Protocol for Heterogeneous WSN for IoT-Based Environmental Monitoring. *IEEE Internet Things J.* **2020**, *7*, 710–717. [\[CrossRef\]](#)
16. Yu, J.; Qi, Y.; Wang, G. An energy-driven unequal clustering protocol for heterogeneous wireless sensor networks. *J. Control. Theory Appl.* **2011**, *9*, 133–139. [\[CrossRef\]](#)
17. Mazumdar, N.; Om, H. Coverage-aware unequal clustering algorithm for wireless sensor networks. *Procedia Comput. Sci.* **2015**, *57*, 660–669. [\[CrossRef\]](#)
18. Amgoth, T.; Jana, P.K. Energy-aware routing algorithm for wireless sensor networks. *Comput. Electr. Eng.* **2015**, *41*, 357–367. [\[CrossRef\]](#)
19. Nabajyoti, M.; Hari, O. Distributed fuzzy logic based energy-aware and coverage preserving unequal clustering algorithm for wireless sensor networks. *Int. J. Commun. Syst.* **2017**, *20*, 1–12.
20. Nabajyoti, M.; Hari, O. Distributed fuzzy approach to unequal clustering and routing algorithm for wireless sensor networks. *Int. J. Commun. Syst.* **2018**, *31*, 12.
21. Zhang, H.; Zhou, X.; Wang, Z.; Wang, H.; Sun, J. Adaptive Consensus-Based Distributed Target Tracking with Dynamic Cluster in Sensor Networks. *IEEE Trans. Cybern.* **2019**, *49*, 1580–1591. [\[CrossRef\]](#)
22. Shi, F.; Yang, S.; Tuo, X.; Ran, L.; Huang, Y. A Novel Rapid-Flooding Approach with Real-Time Delay Compensation for Wireless-Sensor Network Time Synchronization. *IEEE Trans. Cybern.* **2020**, 1–14. [\[CrossRef\]](#)
23. Chen, Z.; Lin, Y.; Gong, Y.; Zhan, Z.; Zhang, J. Maximizing Lifetime of Range-Adjustable Wireless Sensor Networks: A Neighborhood-Based Estimation of Distribution Algorithm. *IEEE Trans. Cybern.* **2020**. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Zhang, Z.; Liu, S.; Yang, Y. A Link-adding strategy for improving robustness and traffic capacity in large-scale wireless sensor networks. *Clust. Comput.* **2019**, *22*, 7687–7694. [\[CrossRef\]](#)
25. Chowdhury, S.; Giri, C. EETC: Energy Efficient Tree-Cluster in Delay Constrained Wireless Sensor Networks. *Wirel. Pers. Commun.* **2019**, *109*, 189–210. [\[CrossRef\]](#)
26. Yao, B.K.; Gao, H.; Chen, Q.; Li, J.Z. Energy-Adaptive and Bottleneck-Aware Many-to-Many Communication Scheduling for Battery-Free WSNs. *IEEE Internet Things J.* **2021**, *8*, 8514–8529. [\[CrossRef\]](#)
27. Zhang, Z.; Li, J.; Yang, X. Data Aggregation in Heterogeneous Wireless Sensor Networks by using Local Tree Reconstruction Algorithm. *Complexity* **2020**, *2020*, 3594263.
28. Kim, T.; Vecchiotti, L.F.; Choi, K.; Lee, S.; Har, D. Machine Learning for Advanced Wireless Sensor Networks: A Review. *IEEE Sens. J.* **2021**, *21*, 12379–12397. [\[CrossRef\]](#)
29. Almaslukh, B. Deep Learning and Entity Embedding-Based Intrusion Detection Model for Wireless Sensor Networks. *CMC Comput. Mater. Contin.* **2021**, *69*, 1343–1360.
30. Karthik, S.S.; Kavithamani, A. Fog computing-based deep learning model for optimization of microgrid-connected WSN with load balancing. *Wirel. Netw.* **2021**, *27*, 2719–2727. [\[CrossRef\]](#)
31. Zhou, Q.W.; Chen, Y.H.; Li, B.Q.; Li, X.X.; Zhou, C.; Huang, J.C.; Hu, H.G. Training deep neural networks for wireless sensor networks using loosely and weakly labeled images. *Neurocomputing* **2021**, *427*, 64–73. [\[CrossRef\]](#)
32. Wang, A.; Shen, J.; Vijayakumar, P.; Zhua, Y. Secure big data transmission for energy efficient intra-cluster in WSNs. *Inf. Sci.* **2019**, *505*, 586–599. [\[CrossRef\]](#)
33. Muneeshwari, P.; Athisha, G. Extended artificial immune system-based optimized access control for big data on a cloud environment. *Int. J. Commun. Syst.* **2020**, *33*, e3947. [\[CrossRef\]](#)
34. Gharaei, N.; Bakar, K.A.; Hashim, S.Z.M. An energy-efficient mobile Sink-based unequal clustering mechanism for WSNs. *Sensors* **2017**, *17*, 1858. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Hu, Y.; Niu, Y.; Lan, Z. An energy-efficient adaptive overlapping clustering method for dynamic continuous monitoring in WSNs. *IEEE Sens. J.* **2017**, *17*, 834–847. [\[CrossRef\]](#)
36. Huang, J.; Qian, H.; Hu, R.Q. Secure and Efficient Privacy-Preserving Authentication Scheme for 5G Software Defined Vehicular Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 8542–8554. [\[CrossRef\]](#)