

A Multi-Chain Based Hierarchical Topology Control Algorithm for Wireless Sensor Networks

Hong Tang, Hui-Zhu Wang*

Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts and Telecommunications (CQUPT)
Chongqing 400065 – P. R. China

[e-mail: tangh@cqupt.edu.cn, wangyajane@foxmail.com]

*Corresponding author: Hui-Zhu Wang

*Received January 6, 2015; revised April 10, 2015; accepted July 8, 2015;
published September 30, 2015*

Abstract

In this paper, we present a multi-chain based hierarchical topology control algorithm (MCHTC) for wireless sensor networks. In this algorithm, the topology control process using static clustering is divided into sensing layer that is composed by sensor nodes and multi-hop data forwarding layer that is composed by leader nodes. The communication cost and residual energy of nodes are considered to organize nodes into a chain in each cluster, and leader nodes form a tree topology. Leader nodes are elected based on the residual energy and distance between themselves and the base station. Analysis and simulation results show that MCHTC outperforms LEACH, PEGASIS and IEEPB in terms of network lifetime, energy consumption and network energy balance.

Keywords: Wireless sensor networks, multi-chain, hierarchical topology, energy balance, network lifetime

This research was supported by Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1299), Chongqing Science and Technology Research Funds (No. CSTC2012jjA40044, No. cstc2013yykfA40010), Chongqing University of Posts and Telecommunications Research Funds (No. A2011-51), and special fund of Chongqing key laboratory (CSTC).

1. Introduction

Wireless sensor networks (WSNs) are self-organizing networks composed by a large number of wireless sensor nodes. Nodes cooperatively collect information and transmit report messages to the target user [1][2]. Most of the wireless sensor devices at present have certain limitations on energy, power and communication capabilities, and the energy consumption directly has an impact on the lifetime of WSNs [3][4][5][6]. Topology control is an important energy-saving technology which ensures network coverage and connectivity by reducing redundant links in the original topology. It reduces network energy consumption, communication interference and Media Access Control (MAC) layer competition, improves the efficiency of routing protocols and provides a basis for data fusion [7][8].

Hierarchical topology is a simple structure, which is beneficial to the design of MAC layer protocol and easy to control. It avoids direct communication of sensor nodes and the base station, reduces communication distance between nodes and the number of data packets transmitted in the network. Therefore, hierarchical topology can significantly reduce energy consumption, which is very important to the energy-constrained WSNs [1][9][10][11].

Existing hierarchical topology control algorithms were mostly evolved based on LEACH. There is a relatively small number of topology control algorithms based on chain structure. The existing chain topologies were an improvement of LEACH. Node with the farthest distance from the base station started to establish a chain throughout the entire network, and the chain formed based on greedy algorithm. However, there was long-chain problem between some nodes, which would cause premature death of nodes. Selecting the nearest node as the next node waiting to join the chain when new node added in chain, and node waiting to join the chain selected the nearest node added in chain as the parent node can avoid the problem as mentioned above. Some works used static clustering that divided the network into subareas, and a chain formed in each subarea. The short chains formed in the network may reduce the transmission delay.

The static clustering methods most existing papers used only divided the network into several subareas, and did not consider factors associated with nodes or the network. Apart from this, constructing topology based on the distance between nodes which was measured by the received signal strength only reflected the positional relationship between nodes in the space, but the energy factor when nodes communicate with each other was ignored. In this paper, we propose a multi-chain based hierarchical topology control algorithm (MCHTC) for wireless sensor networks. The topology control process of MCHTC using static clustering includes sensing layer that is composed by sensor nodes and multi-hop data forwarding layer that is composed by leader nodes. When dividing the network into subareas, we consider the total number of nodes in the network as well as the probability of nodes becoming the leader nodes. The communication cost and the residual energy of nodes are considered when organizing nodes into a chain in each cluster, and leader nodes form a tree topology. It consists of two phases to reduce the energy consumption of the network when electing leader nodes. First, electing candidate leader nodes based on the residual energy threshold, and then electing the final leader nodes based on the residual energy and distance between themselves and the base station.

Performance of the proposed algorithm was evaluated via theoretical analysis and simulations and it was compared with performance of LEACH, PEGASIS and IEEPB. The

simulation results show that the proposed algorithm can reduce energy consumption and prolong network lifetime based on certain network energy balance.

The rest of this paper is organized as follows: Section 2 is an overview of related works. Section 3 describes the theoretical models and related concepts. The proposed algorithm (MCHTC) is presented in Section 4. Section 5 is the theoretical analysis of MCHTC. The simulation results are presented in Section 6. Section 7 is the summary of this paper.

2. Related Work

A number of hierarchical topology control algorithms have been proposed for wireless sensor networks. In this section, some of the related works are reviewed.

One of the most popular hierarchical topology control algorithm in wireless sensor networks is LEACH. The network is divided into several clusters, and the topology setup phase is initiated by cluster heads. Cluster heads collect data packets which are transmitted from sensor nodes, and transmit data packets to the base station. The operation of LEACH is divided into rounds. It randomly selects a few nodes as cluster heads and rotates this role to balance energy consumption in each round [12][13].

LEACH-Centralized [13] is a centralized clustering algorithm. Base station collects the location, energy information, and broadcasts the information of clusters and cluster heads to the sensor nodes. The topology derived by this algorithm is the same as LEACH. LEACH-C considers the residual energy of nodes and the average energy of the entire network when elects the cluster heads to avoid premature death of nodes that repeatedly become cluster heads.

The residual energy and the communication cost of nodes in each cluster are considered in HEED [14]. The clustering process terminates in $O(1)$ iterations and does not depend on the network topology or size. It defines the Average Minimum Reachability Power (AMRP) as the communication cost in each cluster. AMRP is the average minimum power of sensor nodes communicating with cluster heads. AMRP provides all nodes, including the cluster heads a unified clustering mechanism, therefore, it is superior to elect a cluster head based on the distance.

TopDisc [15] is a classical algorithm based on minimum dominating set in graph theory. It defines nodes' status based on different colors to solve the problem of topology setup of backbone network. Topology setup phase is initiated by a node in the network sending a query message for discovering neighbors. A color marker is designated in turn for each node with the query message spreading in the network. The dominant nodes are determined based on the color marker, and the communication links between dominant nodes are established by looking through the reverse propagation path of query messages. Dominant nodes manage the sensor nodes within their scope.

GAF [16] is a clustering algorithm based on nodes' location. The backbone network is formed by selecting the region and cluster heads. This algorithm assumes that the location information of nodes is known, all nodes have the same communication radius, and the monitoring area is divided into a plurality of virtual square cells. Any two nodes in adjacent cells are able to communicate with each other directly. Each cell elects a cluster head, and the backbone network is formed by cluster heads.

EEUC [17] defines a competition radius for each node. It elects candidate cluster heads based on the cluster head selection formula in LEACH. Each candidate cluster head establishes their neighbor information list, and determines the final cluster head according to the energy information in the neighbor information list. A multi-hop topology is formed by

cluster heads. Several qualified neighbor cluster heads are identified based on the distance between themselves and the base station and the competition radius, the neighbor cluster head with the maximal residual energy is selected as the next hop neighbor.

PEGASIS [18] is an improvement of LEACH. This is an algorithm aiming at reducing the communication distance between nodes. Node with the farthest distance from the base station starts to establish a chain throughout the entire network. After forming a chain, nodes serve as the leader node in turn, leader node transmits the fused data packets to the base station and informs node in the end of chain transmitting data packets to it. In PEGASIS, the load of leader node is small, and the communication distance between nodes is reduced. It has a better performance in terms of network energy efficiency and energy balance.

EEPB [19] is an improvement of PEGASIS. When it is a large-scale network, a long chain will form, which leads to large latency of data transmission, and the long distance between nodes will cause premature death of nodes. Therefore, EEPB avoids long distance between nodes by introducing a distance threshold. The residual energy of nodes and the distance between nodes and the base station are considered when electing a leader node, which to some extent balances the network energy consumption.

The formation of the chain structure and the leader node election are improved in IEEPB [20], because the uncertainty of distance threshold in EEPB may be unable to avoid generating a long chain between nodes, and the proportion of the residual energy and the distance between nodes and the base station are not considered when electing a leader node. In IEEPB, the topology setup phase is initialized by node with the farthest distance from the base station. New node added in chain selects the nearest node as the next node waiting to join the chain. Node waiting to join the chain selects the nearest node which has been added in chain as the parent node. Different proportions are designated to the residual energy and the distance between nodes and the base station when electing a leader node.

MIEEPB [21] reduces energy consumption by the mobile station. The network is divided into four subareas, a chain is formed in each subarea the same as PEGASIS. A primary leader node is elected in each subarea which is the same as EEPB. When electing the secondary leader node, it needs to compare the distance between nodes and their parent nodes with the distance between nodes and the base station. If the former is larger than the latter, node will be elected as the secondary leader node, and transmit the fused data packets to the base station directly. The base station moves periodically in the four subareas and collects data packets transmitted from leader nodes.

3. Theoretical Models and Concepts

In this paper, the following assumptions are assumed for wireless sensor networks.

- Nodes are isomorphic and energy limited.
- Nodes obtain the specific position information in the distribution area by existing positioning technologies or the relationship between the received signal strength and the distance between nodes.
- Nodes are randomly distributed in the monitoring area.
- The location of nodes in the network and the base station is fixed.
- Data link is symmetric, which means if node i can communicate with node j , then node j can also communicate with node i .

3.1 Network Model

Wireless sensor networks can be abstracted as an undirected simple graph $G(V, E)$ in the plane, where $V(G)$ is the set of nodes in the network and $E(G)$ is the edge set of $G(V, E)$. Assume that r_{max} is the common transmission range when the maximal communication transmission power is p_{max} . For any two nodes $i \in V(G)$ and $j \in V(G) - \{i\}$, the distance between them can be presented as $d(i, j)$, therefore, $E(G)$ satisfies $E(G) = \{(i, j) : d(i, j) \leq r_{max}, i, j \in V(G)\}$. A unique id is assigned to every node. We define the *Visible Neighborhood* and *Visible Neighborhood Set* of node i as follow:

Definition 1 (Visible Neighborhood and Visible Neighborhood Set): Node j that can correctly receive the messages sending with p_{max} from node i is a *Visible Neighborhood* of node i . The *Visible Neighborhood Set* is a set of all the *Visible Neighborhood*, which satisfies $NV_i(G) = \{j \in V : d(i, j) \leq r_{max}\}$.

3.2 Determination of Transmission Power

Assume that the maximal transmission power p_{max} is the same to all nodes. By measuring receiving power of messages, every node can determine the specific power level p_t it needs to communicate with each neighbor. We first describe two commonly-used propagation models, and then introduce the method for determining transmission power.

In the Free Space propagation model, assume that the transmission power of a node is p_t , p_r is the power to receive packets. The relationship between p_t and p_r is $p_r = p_t G_t G_r \lambda^2 / (4\pi d)^2 L_0$, where G_t is the antenna gain of the transmitter, G_r is the antenna gain of the receiver, λ is the wave length, d is the distance between the transmitter and receiver, and L_0 is the system loss.

In the Two-Ray Ground propagation model, the relationship between p_t and p_r is $p_r = p_t G_t G_r h_t^2 h_r^2 / d^4 L_0$, where h_t is the antenna height of the transmitter, h_r is the antenna height of the receiver, the meaning of other parameters are the same as mentioned above.

The relationship between p_t and p_r can be generally expressed as $p_t = p_r G_f$, where G_f is a function of $G_t, G_r, h_t, h_r, \lambda, d, \alpha, L_0$, and α is the path loss exponent. Before topology setup phase, nodes need to collect information of neighbors with p_{max} . When node A receives messages from node B, by measuring the receiving power p_r , we can obtain $G_f = p_r / p_{max}$. Therefore, node A needs to communicate with node B on the condition of $p_{th} G_f = p_{th} p_r / p_{max}$ so that node B can receive messages, where p_{th} is the power threshold to correctly receive the message. A power level that can reach the farthest neighbor is required when nodes broadcast messages [22].

3.3 Radio Energy Dissipation Model

The energy dissipation model adopted in this paper is the same as LEACH [23]. Fig. 1 is the energy consumption model of wireless sensors.

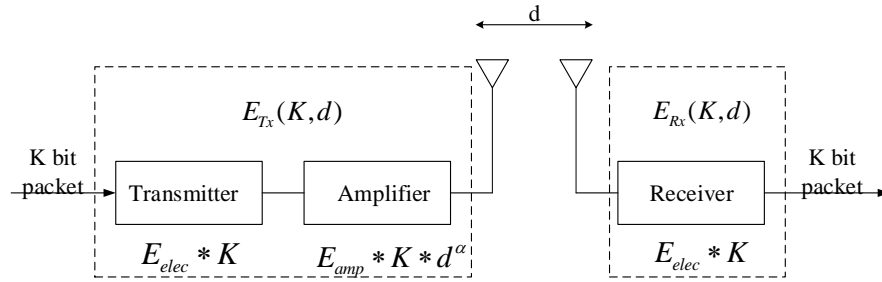


Fig. 1. Energy consumption model of wireless sensors

The energy consumed in sending k bit data packet over a distance d for each node is as in Eq. (1).

$$E_{Tx}(k, d) = \begin{cases} kE_{elec} + k\varepsilon_{fs}d^2, & d < d_0 \\ kE_{elec} + k\varepsilon_{mp}d^4, & d \geq d_0 \end{cases} \quad (1)$$

where E_{elec} is the energy consumed in electronics, d is the distance between the transmitter and receiver, d_0 is the reference distance satisfying $d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$. If $d < d_0$, power amplifier uses the free space loss model; If $d \geq d_0$, power amplifier uses multi-path fading model. ε_{fs} and ε_{mp} are the energy consumed in amplifiers.

The energy consumed in receiving k bit data packet is as in Eq. (2).

$$E_{Rx}(k, d) = kE_{elec} \quad (2)$$

The energy consumed in aggregating x data packets to a single packet is as in Eq. (3).

$$E_{fuse}(x, k) = xkE_{DA} \quad (3)$$

3.4 Energy Consumption Model of Chain Topology

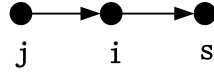
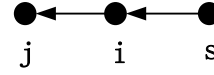
In PEGASIS, nodes in chain transmit data packets to the next neighbor after aggregating data packets, and leader node transmits the fused data packets to the base station. Assume that the total number of nodes in the network is N . The location of node i in chain is n , that is there are $n-1$ nodes in front of node i . The location of node j , node s in chain are $n-1$ and $n+1$ respectively. Assume that the energy consumed in transmitting data packets to the base station is E_{toBS} .

Fig. 2-a shows data packets transmission when leader node behind node i , and node i is not in the end of chain. Based on Eq. (1), Eq. (2), Eq. (3), the energy consumption of node i is as in Eq. (4).

$$E_{nor}(i) = E_{Rx}(j, i) + 2E_{fuse} + E_{Tx}(i, s) \quad (4)$$

Fig. 2-b shows data packets transmission when leader node in front of node i , and node i is in chain. The energy consumption of node i is as in Eq. (5).

$$E_{nor}(i) = E_{Rx}(s, i) + 2E_{fuse} + E_{Tx}(i, j) \quad (5)$$

Fig. 2-a. Leader node behind node i Fig. 2-b. Leader node in front of node i

If node i is in the end of chain, and is not the leader node, the energy consumption of node i is as in Eq. (6).

$$E_{nor}(i) = E_{Tx}(i, j) \quad (6)$$

If node i is in chain, and is the leader node, the energy consumption of node i is as in Eq. (7).

$$E_{ld}(i) = E_{Rx}(j, i) + E_{Rx}(s, i) + 3E_{fuse} + E_{toBS} \quad (7)$$

If node i is in the end of chain, and is the leader node, the energy consumption of node i is as in Eq. (8).

$$E_{ld}(i) = E_{Rx}(j, i) + 2E_{fuse} + E_{toBS} \quad (8)$$

4. MCHTC- the Proposed Algorithm

In PEGASIS, node with the farthest distance from the base station starts to establish a chain throughout the entire network. However, when it is a large-scale network, a long chain will form, which leads to large latency of data transmission, and the problem of long communication distance exists between some nodes. In EEPB, it adopts a distance threshold to avoid the problem of long communication distance between nodes. The distance threshold can be adjusted by a constant a . However, the uncertainty of the distance threshold has a negative impact on the formation of a chain. When a is large, the distance threshold will lose its role, which is unable to avoid generating the long communication distance between nodes. When a is infinite, the topology derived by EEPB is the same as PEGASIS. When a is small, the distance between nodes is always higher than the threshold, which is also unable to avoid the problem of long communication distance between nodes. Therefore, it is difficult to determine an appropriate value of a [20]. In IEEPB, the purpose of controlling distance between nodes is obtained by comparing distance between nodes twice. However, some nodes have multiple child nodes, which has a negative impact on the network energy balance.

In this paper, we propose a multi-chain based hierarchical topology control algorithm (MCHTC) for wireless sensor networks. In the proposed algorithm, the network is divided in the horizontal direction and vertical direction based on the *Region Partition Parameter*. Several subareas are formed, and each subarea is a cluster. These clusters, once formed, will not change in the entire life cycle of the network [24]. Node in each cluster with the farthest distance from the base station starts to establish a chain. New node added in chain selects the nearest neighbor based on *weight function*. The process repeats until all nodes in each cluster have been added in chain. After chains are formed in each cluster, leader nodes are elected in turn among nodes. Therefore, the operation of MCHTC is divided into rounds. When electing leader heads, the residual energy of nodes which are higher than the residual energy threshold of the current round are elected as the candidate leader nodes. Candidate leader nodes with the maximal value of Q will be elected as the leader nodes. Since leader nodes are relatively distant, to avoid the long communication distance between leader nodes, a tree topology is

formed by leader nodes. The formation of tree topology starts from the root leader node, which is the nearest node from the base station.

In MCHTC, nodes are randomly distributed in the monitoring area. The *Region Partition Parameter* t is as in Eq. (9).

$$t = \lceil Np/2 \rceil \quad (9)$$

where N is the total number of nodes in the network, p is the *Region Partition Factor* satisfying $0 < p < 1$. It is the probability of nodes becoming leader nodes, which determines the number of subareas. Fig. 3 shows subareas when $p = 0.05$. If $p = 0.05$, $t = 3$, network is divided into three portions in the horizontal direction and vertical direction respectively. Therefore, 9 subareas are formed. The number of subareas can be changed by adjusting the *Region Partition Factor* p based on the actual requirements of the distribution area.

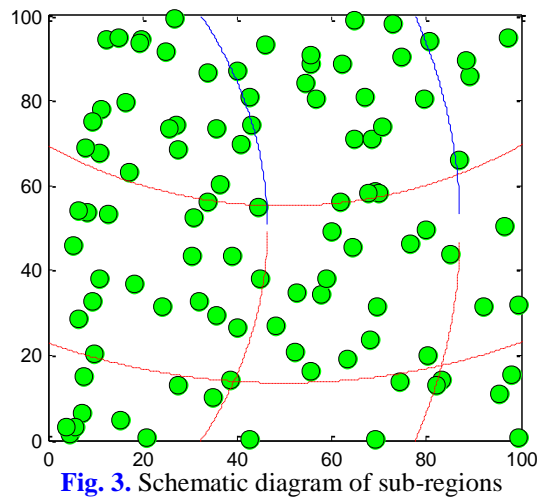


Fig. 3. Schematic diagram of sub-regions

The operation of MCHTC is organized as rounds. Each round of this algorithm consists of the following phases:

- Topology setup phase
- Data transmission phase

4.1 Topology Setup Phase

Topology setup phase consists of three stages.

4.1.1 Topology Setup Phase in Clusters

When designing multi-hop transmission topology of wireless sensor networks, it needs to take full account of node itself, rather than only look for the shortest path from a global point of view. Global path planning does not fully consider whether the transmission path of a node is the best or not. For example, a path is the shortest in the network, but a node needs to transmit data packets through a long communication distance simultaneously, which is unfair for the node [25].

Most of the existing topology control algorithms commonly use Euclidean Distance as the communication cost between nodes, which merely reflects the relationship between nodes in space, but ignores the energy factor when nodes communicate with each other. Therefore, if only considering the spatial distance between nodes, the energy consumption of nodes,

especially the leader nodes may be too large, and cause premature death of nodes. Therefore, in this paper, the communication cost is defined as *Energy Distance* between nodes. The energy consumption when nodes communicate with each other and the residual energy of the destination node are considered in *Energy Distance*, which is as in Eq. (10).

$$ED(i, j) = Ploss(i, j) * \frac{E_0(j)}{E_r(j)} \quad (10)$$

where $E_0(j)$ is the initial energy of the destination node j , $E_r(j)$ is the residual energy of the destination node j in current rounds, $Ploss(i, j)$ is the path loss from the source node i to the destination node j satisfying $Ploss(i, j) = p_t(i) - rssi(j)$, and $rssi(j)$ is the received signal strength of the destination node j .

Energy Distance plays a significant role when designing topology control algorithm. If node j is the next hop neighbor of node i , it shows that the *Energy Distance* between node i and node j is the smallest one, which indirectly shows that the path loss between node i and node j is small, or the residual energy of node j in the current round is large.

To ensure the uniqueness of weight between nodes, in this paper, the *Energy Distance* and nodes' *id* are considered when defining *weight function*.

Definition 2 (Weight Function): Assume that (i_1, j_1) and (i_2, j_2) are any two node pairs in the network, the *weight function* is defined as:

$$\begin{aligned} W(i_1, j_1) < W(i_2, j_2) &\Leftrightarrow ED(i_1, j_1) < ED(i_2, j_2) \\ \text{Or } (ED(i_1, j_1) = ED(i_2, j_2) \\ &\& \& \max\{id(i_1), id(j_1)\} < \max\{id(i_2), id(j_2)\}) \\ \text{Or } (ED(i_1, j_1) = ED(i_2, j_2) \\ &\& \& \max\{id(i_1), id(j_1)\} = \max\{id(i_2), id(j_2)\} \\ &\& \& \min\{id(i_1), id(j_1)\} < \min\{id(i_2), id(j_2)\}) \end{aligned}$$

Fig. 4 is the flow chart of topology setup phase in clusters. Nodes exchange neighbor information list by broadcasting helloMSG with p_{max} . The neighbor information list is as follow:

$CID(i)$	$NID(i)$	$E_r(i)$	$RSSI(i)$
----------	----------	----------	-----------

where $CID(i)$ is the cluster *id* of neighbor node i , $NID(i)$ is the node *id* of node i , $E_r(i)$ is the residual energy of node i , $RSSI(i)$ is the received signal strength of node i . Neighbor information list is contained in helloMSG. If the received helloMSG of sensor nodes is unrelated with the cluster nodes they belong to, nodes will discard it directly.

In each cluster, node with the farthest distance from the base station is the first node added in chain. New node added in chain calculates the weight between neighbor nodes which are not added in chain, and selects node with minimum weight as the next node added in chain. The process repeats until all nodes are added in chain.

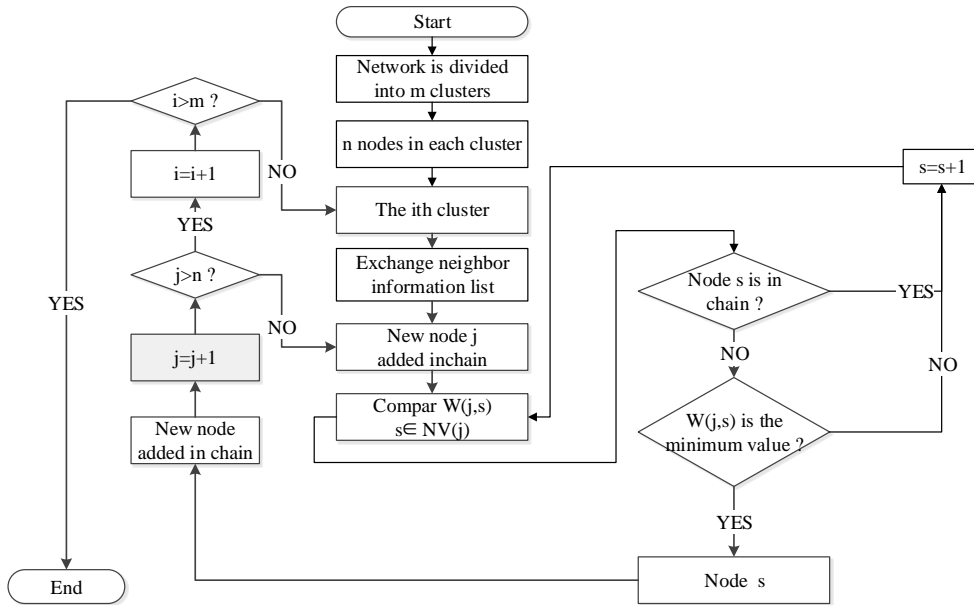


Fig. 4. Flow chart of topology setup phase in clusters

4.1.2 Selection of Leader Nodes

Due to that wireless sensor networks are energy-constrained networks, and leader nodes have a larger energy consumption than sensor nodes, in this stage, a residual energy threshold E_{th} is defined in this paper to reduce the energy consumption of leader nodes and balance the energy consumption of the network. The definition of residual energy threshold is as in Eq. (11).

$$E_{th} = (r_{max} - r_{cur})E_0 / r_{max} \tag{11}$$

where r_{cur} is the current working round, E_0 is the initial energy of nodes, r_{max} is the predicted maximal network working rounds satisfying $r_{max} = E_{ini} / E_{each}$, where E_{ini} the sum of E_0 , E_{each} is the energy consumption of each round, which can be predicted based on the reference network topology. Four critical reference network topologies are considered, which are as shown in Fig. 5-a, 5-b, 5-c, 5-d. In case 1 and case 2, leader nodes form a chain structure, which is a special kind of tree structure. In case 3 and case 4, we consider a special case of tree structure, where there are only the root node and the leaf nodes.

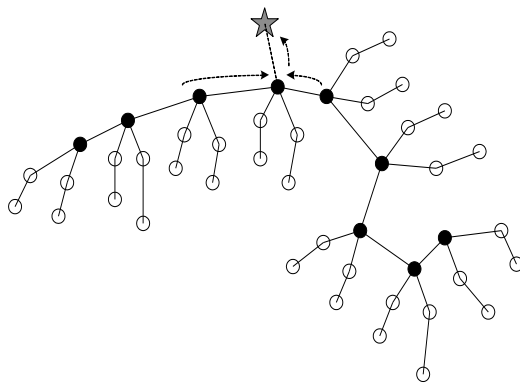


Fig. 5-a. Reference topology of case 1

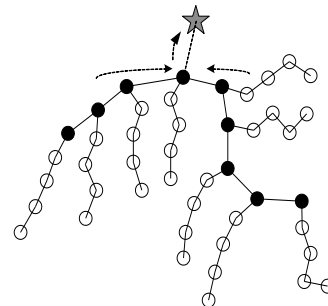


Fig. 5-b. Reference topology of case 2

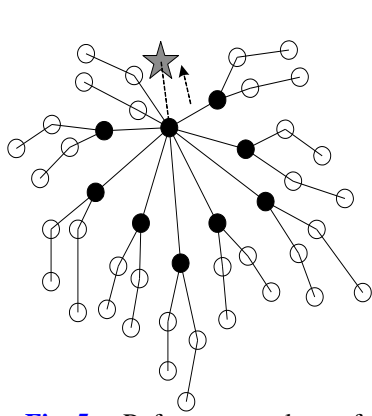


Fig. 5-c. Reference topology of case 3

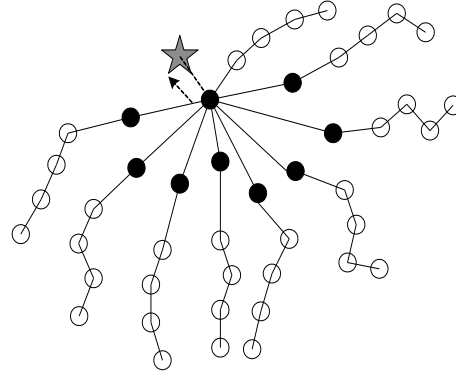


Fig. 5-d. Reference topology of case 4

Based on the radio energy dissipation in section 3.3 and the energy consumption model of chain topology in section 3.4, we analyze the energy consumption of each round as follows, and the relevant parameters are as shown in **Table 1**.

Table 1. Parameters of analyzing energy consumption

Parameters	Meaning of the expression
d_{in}	distance between nodes in each cluster
d_{ld}	distance between leader nodes
d_{bs}	distance between leader nodes and base station
m	number of leader nodes
E_{Tx}^{la}	energy consumed by leader nodes transmitting data packets
E_{Rx}^{la}	energy consumed by leader nodes receiving data packets
E_{Tx}^{bs}	energy consumed by leader nodes transmitting data packets to the base station
E_{Tx}^{in}	energy consumed by sensor nodes transmitting data packets
E_{Rx}^{in}	energy consumed by sensor nodes receiving data packets
E_{fuse}	energy consumed by fusing data packets

- Case 1: as is shown in **Fig. 5-a**, the energy consumption E_{each}^1 of each round is:

$$E_{each}^1 = 2(E_{Tx}^{ld} + 3E_{fuse} + 2E_{Rx}^{in}) + (2E_{Rx}^{ld} + 2E_{Rx}^{in} + 5E_{fuse} + E_{Tx}^{bs})$$

$$+ (m-3)(E_{Rx}^{ld} + E_{Tx}^{ld} + 2E_{Rx}^{in} + 4E_{fuse}) + 2mE_{Tx}^{in} + (N-3m)(E_{Rx}^{in} + E_{Tx}^{in} + 2E_{fuse})$$

$$= (m-1)E_{Tx}^{ld} + (2N-2m-1)E_{fuse} + (N-m)E_{Rx}^{in} + (m-1)E_{Rx}^{ld} + E_{Tx}^{bs} + (N-m)E_{Tx}^{in}$$
- Case 2: as is shown in **Fig. 5-b**, the energy consumption E_{each}^2 of each round is:

$$E_{each}^2 = 2(E_{Rx}^{in} + E_{Tx}^{ld} + 2E_{fuse}) + (E_{Rx}^{in} + 2E_{Rx}^{ld} + E_{Tx}^{bs} + 4E_{fuse})$$

$$+ (m-3)(E_{Rx}^{in} + E_{Tx}^{ld} + E_{Tx}^{ld} + 3E_{fuse}) + mE_{Tx}^{in} + (N-2m)(E_{Rx}^{in} + E_{Tx}^{in} + 2E_{fuse})$$

$$= (N-m)E_{Rx}^{in} + (m-1)E_{Tx}^{ld} + (2N-m-1)E_{fuse} + (m-1)E_{Rx}^{ld} + E_{Tx}^{bs} + (N-m)E_{Tx}^{in}$$
- Case 3: as is shown in **Fig. 5-c**, the energy consumption E_{each}^3 of each round is:

$$\begin{aligned}
E_{each}^3 &= \left[E_{Tx}^{bs} + (m-1)E_{Rx}^{ld} + 2E_{Rx}^{in} + (m+2)E_{fuse} \right] + (m-1)(2E_{Rx}^{in} + E_{Tx}^{ld} + 3E_{fuse}) \\
&\quad + 2mE_{Tx}^{in} + (N-3m)(E_{Rx}^{in} + E_{Tx}^{in} + 2E_{fuse}) \\
&= E_{Tx}^{bs} + (m-1)E_{Rx}^{ld} + (N-m)E_{Rx}^{in} + (2N-2m-1)E_{fuse} + (m-1)E_{Tx}^{ld} + (N-m)E_{Tx}^{in}
\end{aligned}$$

- Case 4: as is shown in **Fig. 5-d**, the energy consumption E_{each}^4 of each round is:

$$\begin{aligned}
E_{each}^4 &= \left[E_{Tx}^{bs} + (m-1)E_{Rx}^{ld} + E_{Rx}^{ld} + (m+1)E_{fuse} \right] + (m-1)(E_{Tx}^{ld} + E_{Rx}^{in} + 2E_{fuse}) \\
&\quad + mE_{Tx}^{in} + (N-2m)(E_{Tx}^{in} + E_{Rx}^{in} + 2E_{fuse}) \\
&= E_{Tx}^{bs} + (m-1)E_{Rx}^{ld} + (N-m)E_{Rx}^{in} + (2N-m-1)E_{fuse} + (m-1)E_{Tx}^{ld} + (N-m)E_{Tx}^{in}
\end{aligned}$$

Based on the above analysis, case 1 and case 3, case 2 and case 4 have the same energy consumption. In summary, E_{each} is as in Eq. (12).

$$E_{each} = (E_{each}^1 + E_{each}^2)/2 \quad (12)$$

Nodes in each cluster compare the residual energy with E_{th} . For any node i in the network, if $E_r(i) \leq E_{th}$, node i gives up the competition of leader nodes. If $E_r(i) > E_{th}$, node i becomes the candidate leader node.

When electing leader nodes, the residual energy of candidate leader nodes and the distance between candidate leader nodes and the base station are considered. The two factors are defined as Q of each node, which is as in Eq. (13).

$$Q(i) = E_r(i)/d_{toBS}(i) \quad (13)$$

where $d_{toBS}(i)$ is the distance between node i and the base station. The candidate leader nodes with the maximal value of Q will become leader node by competition. **Fig. 6** is the flow chart of leader node selection.

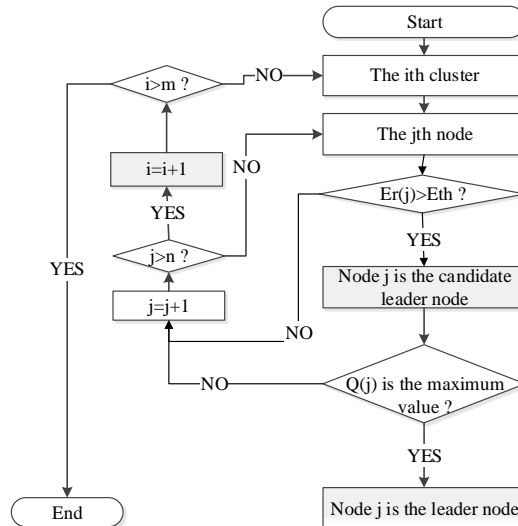


Fig. 6. Flow chart of leader nodes selection

4.1.3 Topology Setup Phase of Leader Nodes

In this stage, a tree topology is formed among leader nodes to avoid generating long communication distance between leader nodes. The tree topology is the minimum cost tree which is formed based on *weight function*. Fig. 7 is the flow chart of topology setup phase of leader nodes. Sensor nodes temporarily close the communication module and stay in sleeping state. Every leader node calculates weights with neighbor leader nodes, and stores the weights in ascending order. The nearest leader node from the base station will be the root node of the tree. Root leader node starts to establish the minimum cost tree based on the *weight function*. The next leader node added in tree will be determined by leader nodes which have been added in tree.

Assume that leader node s_1 and leader node s_2 are any two leader nodes that have been added in tree. $V(T)$ is the set of leader nodes that are still not added in tree. $W(s_1, g_1)$ is the minimum weight of s_1 and its neighbor leader node g_1 , and $g_1 \in V(T)$. $W(s_2, g_2)$ is the minimum weight of s_2 and its neighbor leader node g_2 , and $g_2 \in V(T)$. If $W(s_1, g_1) > W(s_2, g_2)$, s_2 will be added in tree with priority. On the contrary, s_1 will be added in tree with priority. The process repeats until the optimal leader node with minimum weight is selected as the next leader node added in tree. The phase ends when all leader nodes have been added in tree.

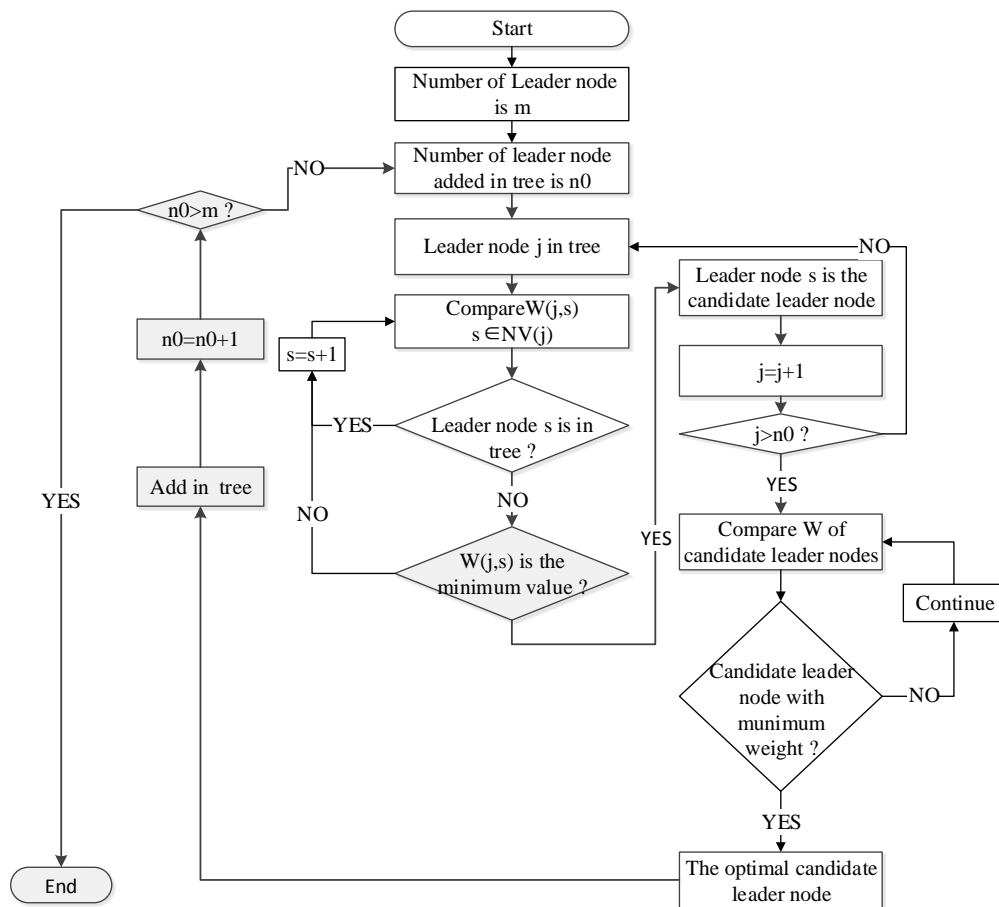


Fig. 7. Flow chart of topology setup phase of leader nodes

4.2 Data Transmission Phase

In this paper, the data fusion percentage is 100%. If the design of time sequence is reasonable, data packets transmitted in the network will be reduced in multi-hop topology. It needs to consider the relationship between nodes when data packets are transmitted over multi-hops.

4.2.1 Data Transmission in Each Cluster

When data packets are transmitted in each cluster, leader node informs node in the end of chain starting to transmit data packets. Sensor nodes receive data packets transmitted from their neighbors, and transmit the fused data packets to the next hop neighbor.

4.2.2 Data Transmission among Leader Nodes

Fig. 8 shows data transmission among leader nodes. Node 1 is the root node. Node 9 transmits data packets to node 7, and node 7 transmits the fused data packets to node 5 in its time slot. Node 5 needs to wait for the data packets transmitted from node 8, and then transmits the fused data packets to node 3. node 3 and node 2 execute the same operation as node 5. Node 1 is the last node to receive and process data packets, and transmits the fused data packets to the base station.

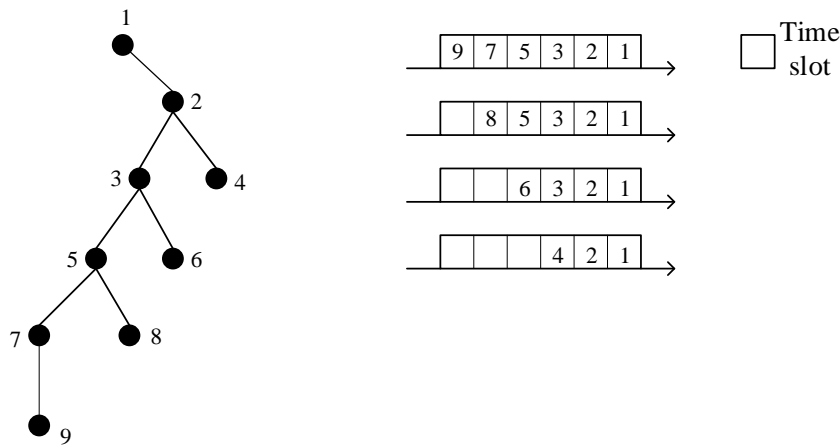


Fig. 8. Schematic diagram of data transmission among leader nodes

Energy is saved when the number of data packets in the network is reduced. In addition, using slot mechanism to transmit data packets can avoid nodes transmitting data at the same time, and reduce the probability of packet conflict.

5. Performance Analysis of MCHTC

In this section, we state and prove several properties of the network topology derived by MCHTC.

5.1 Network Connectivity

We prove that the topology G_0 , derived by MCHTC preserves the network connectivity of G . For any two nodes $i, j \in V(G_0)$, node i is said to be connected to node j (denoted $i \leftrightarrow j$) if there exists a path $l = (w_0 = i, w_1, w_2, \dots, w_{n-1}, w_n = j)$ from node i to node j such that node w_g and

$w_{g+1}(g = 0, 1, \dots, n-1)$ is the neighbor node of each other, and $w_g \in V(G_0)(g = 0, 1, \dots, n)$. It follows that $i \leftrightarrow w$ if $i \leftrightarrow j$ and $j \leftrightarrow w$.

Lemma 1. For any node pairs (i, j) , $i, j \in V(G_0)$, if $d(i, j) \leq r_{max}$, then $i \leftrightarrow j$.

Proof: For any node pairs (i, j) , $i, j \in V(G_0)$ satisfying $d(i, j) \leq r_{max}$, sort them in ascending order, i.e., $W(i_1, j_1) < W(i_2, j_2) < \dots < W(i_n, j_n)$. We prove lemma 1 based on induction.

- Basis: $n = 1$, $W(i_1, j_1)$ satisfying $W(i_1, j_1) = \min\{W(i_1, j_1), W(i_2, j_2), \dots, W(i_n, j_n)\}$, and $d(i_1, j_1) \leq r_{max}$. Therefore, node i_1 and j_1 is the neighbor of each other satisfying $i_1 \leftrightarrow j_1$.
- Induction: Assume that lemma 1 holds for all pairs $(i_g, j_g)(g = 0, 1, \dots, n-1)$. We need to prove lemma 1 also holds for the pair (i_n, j_n) . Node j_n satisfying $j_n \in NV(i_n)$ when $d(i_n, j_n) \leq r_{max}$, and there exists a path $l = (w_0 = i, w_1, w_2, \dots, w_{n-1}, w_n = j)$ from node i to node j satisfying $(w_g, w_{g+1}) \in E(G_0)(g = 0, 1, \dots, n-1)$. Applying the induction hypothesis to each node pair $(w_g, w_{g+1})(g = 0, 1, \dots, n-1)$, we have $w_g \leftrightarrow w_{g+1}$. Therefore, $i_n \leftrightarrow j_n$.

Theorem 1. G_0 is connected if G is connected.

Proof: Suppose G is connected. For any two nodes $i, j \in V(G)$, there exists a path $l = (w_0 = i, w_1, w_2, \dots, w_{n-1}, w_n = j)$ from node i to node j , where $(w_g, w_{g+1}) \in E(G)(g = 0, 1, \dots, n-1)$ and $d(w_g, w_{g+1}) \leq r_{max}$. Since w_g and w_{g+1} satisfy $w_g \leftrightarrow w_{g+1}$ by lemma 1, we have $i \leftrightarrow j$.

5.2 Average Connectivity of Nodes in Each Cluster

Node connectivity refers to the number of connecting edges of each node. In the network, the connecting edges of each node can be divided into edges for receiving data packets and edges for transmitting data packets. There is only one edge for transmitting data packets, and multiple edges can be used to receive data packets. Therefore, node connectivity refers to the number of edges for receiving data packets.

Fig. 9 is the illustration of node connectivity. Based on the definition of node connectivity, node connectivity of node 1 is 2, node 4 is 1, node 2 and node 3 are 0. The average connectivity of nodes refers to the ratio of the total number of edges for receiving data packets and the total number of nodes in the network.

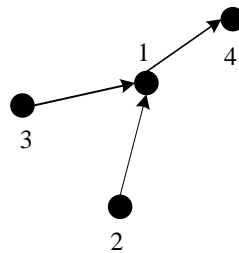


Fig. 9. Illustration of node connectivity

Assume that the number of nodes in each cluster is n_0 . Since there is only one leader node in each cluster, the number of sensor nodes is n_0-1 . Based on the multi-hop transmission principle, nodes transmit data packets to the next hop neighbor. Therefore, there is only one edge for transmitting data packets of each sensor node, which means that the total edges for transmitting data packets in each cluster are n_0-1 . Based on the principle that connecting edge is a bidirectional edge, edges for transmitting data packets of a node are also the edges for

receiving data packets of another node. Therefore, the total edges for receiving data packets are n_0-1 , and the average connectivity of nodes is $(n_0 - 1)/n_0$.

5.3 The Time Complexity of MCHTC

Since the number of leader nodes is small compared with the total number of nodes in the network, the time complexity of MCHTC is mainly determined by the time complexity of topology setup phase in clusters. The average number of nodes in each cluster can be expressed as N/m , and m is the number of leader nodes in section 4.1.2.

In the phase of topology setup in clusters, if new node added in chain needs to traverse all nodes which are waiting to be added in chain, the time complexity is the maximal time complexity. The number of nodes which the first node added in chain needs to traverse is $N/m-1$, the number of nodes which the second node added in chain needs to traverse is $N/m-2$ Similarly, the number of nodes which the penultimate node needs to traverse is 1, and the last node will be added in chain directly. Therefore, the maximal time complexity T_{max}^{in} of topology setup in clusters is as follow:

$$\begin{aligned} T_{max}^{in} &= m\{(N/m-1) + (N/m-2) + \dots + [N/m - (N/m-1)]\} \\ &= m[(N/m-1)(N/m - N/2m)] = N^2/2m - N/2 \end{aligned}$$

Therefore, the maximal time complexity T_{max}^{in} of topology setup in clusters is $O(N^2)$.

If new node added in chain only needs to traverse one neighbor, the time complexity is the minimum time complexity T_{min}^{in} , which is as follow:

$$T_{min}^{in} = m(N/m-1) = N - m$$

Therefore, the minimum time complexity T_{min}^{in} of topology setup in clusters is $O(N)$.

In the phase of topology setup among leader nodes, it needs to execute $m-1$ outer loops to obtain the node pairs with the minimum weight, and m inner loops need to be executed to update the leader nodes which have been added in tree. Therefore, the time complexity of topology setup among leader nodes is $m(m-1)$.

In summary, the maximal time complexity of MCHTC is $O(N^2)$, and the minimum time complexity is $O(N)$.

5.4 The Number of Leader Nodes and Network Energy Consumption

Based on the schematic diagram of data transmission among leader nodes in section 4.2.2, leaf leader nodes only need to transmit data packets to their parent nodes, inner leader nodes not only need to receive and aggregate data packets transmitted from child nodes, but also need to transmit the fused data packets to the parent nodes, and root leader node only needs to transmit the fused data packets to the base station.

In the topology setup phase among leader nodes, there is only one root leader node. Assume that the number of leader nodes that are in the end of chain is m_1 , the number of leader nodes that are in chain is m_2 , which satisfies $m_1 + m_2 = m$. The number of leaf nodes in m_1 is n_1 , and the number of leaf nodes in m_2 is n_2 . Based on the radio energy dissipation model in section 3.3 and the energy consumption model of chain topology in section 3.4, we analyze the relationship between the number of leader nodes and the network energy consumption as follows, and meanings of the parameters are the same parameters as in section 4.1.2.

- The energy consumption of root leader node in chain is:

$$E_{root} = 2E_{Rx}^{in} + 4E_{fuse} + E_{Tx}^{bs} + a_1 E_{Rx}^{ld}$$
- The energy consumption of leaf leader nodes which are in the end of chain is:

$$E_{leaf}^1 = n_1 (E_{Rx}^{in} + 2E_{fuse} + E_{Tx}^{ld})$$
- The energy consumption of inner leader nodes which are in the end of chain is:

$$E_{ld}^1 = (m_1 - n_1)(E_{Rx}^{in} + 3E_{fuse} + E_{Tx}^{ld} + a_2 E_{Rx}^{ld})$$
- The energy consumption of leaf leader nodes in chain is:

$$E_{leaf}^2 = n_2 (2E_{Rx}^{in} + 3E_{fuse} + E_{Tx}^{ld})$$
- The energy consumption of inner leader nodes in chain is:

$$E_{ld}^2 = (m_2 - n_2 - 1)(2E_{Rx}^{in} + 4E_{fuse} + E_{Tx}^{ld} + a_3 E_{Rx}^{ld})$$
- The energy consumption of sensor nodes which are in the end of chain is:

$$E_{nor}^{end} = 2m_2 E_{Tx}^{in} + m_1 E_{Tx}^{in}$$
- The energy consumption of sensor nodes in chain is:

$$E_{nor}^{in} = (N - m - m_1 - 2m_2)(E_{Rx}^{in} + 2E_{fuse} + E_{Tx}^{in})$$

where a_1, a_2, a_3 are the number of child nodes of corresponding leader nodes. Therefore, the network energy consumption is as in Eq. (14).

$$\begin{aligned}
E_{cost}(m) &= E_{root} + E_{leaf}^1 + E_{ld}^1 + E_{leaf}^2 + E_{ld}^2 + E_{nor}^{end} + E_{nor}^{in} \\
&= [2N - m + a_3(m_1 - n_1) + a_2(m_2 - n_2) + (a_1 - a_2)]kE_{elec} \\
&\quad + (2N - 2m + m_1 - n_1 - n_2)kE_{DA} \\
&\quad + (m - 1)k\varepsilon_{fs}d_{ld}^2 + (N - m)k\varepsilon_{fs}d_{in}^2 + k\varepsilon_{mp}d_{bs}^4
\end{aligned} \tag{14}$$

where $m_1 - n_1 = \beta_1 m$, $m_2 - n_2 = \beta_2 m$, β_1 is the proportion of inner leader nodes which are in the end of chain, β_2 is the proportion of inner leader nodes which are in chain, and satisfies $0 < \beta_1 < 1, 0 < \beta_2 < 1$. $m_1 = \alpha m$, where α is the proportion of leader nodes which are in chain satisfying $0 < \alpha < 1$. Therefore, Eq. (14) can be expressed as Eq. (15).

$$\begin{aligned}
E_{cost}(m) &= [2N + (a_3\beta_1 + a_2\beta_2 - 1)m + (a_1 - a_2)]kE_{elec} + [2N + (\alpha + \beta_1 + \beta_2 - 3)m]kE_{DA} \\
&\quad + (m - 1)k\varepsilon_{fs}d_{ld}^2 + (N - m)k\varepsilon_{fs}d_{in}^2 + k\varepsilon_{mp}d_{bs}^4
\end{aligned} \tag{15}$$

The derivation of Eq. (15) is:

$$E'_{cost}(m) = (a_3\beta_1 + a_2\beta_2 - 1)kE_{elec} + (\alpha + \beta_1 + \beta_2 - 3)kE_{DA} + k\varepsilon_{fs}d_{ld}^2 - k\varepsilon_{fs}d_{in}^2$$

We can obtain that the number of leader nodes and the network energy consumption are unrelated. The same conclusion can be obtained when root leader node is in the end of chain.

5.5 Energy Consumption of “Long Distance” in Each Cluster

Assume that the distribution area is a square, and the side length is L . The subareas are approximately small squares. The area of distribution area is L^2 , and the area of each cluster is approximately L^2/m . The average number of nodes in each cluster is N/m . The diagonal length is l_0 , distance between nodes in each cluster satisfies $d \leq l_0$. Based on the relationship

between diagonal length and area of square, we can obtain $l_0 = \sqrt{2L}/\sqrt{m}$. The diagonal length l_0 decreases with the increasing of the number of subareas.

We analyze the energy consumption of “long distance” based on the maximal energy consumption of a node. Assume that leader node and a sensor node are located in the two vertices of the diagonal respectively, and the sensor node is not in the end of chain. The distance between the sensor node and the leader node is $d_{in}^* = l_0 = \sqrt{2L}/\sqrt{m}$, distance between other nodes is $d_{in} = cd_{in}^*$ ($0 < c \leq 1$). The energy consumption of the sensor node is as in Eq. (16).

$$E_{nor}^{long} = E_{Rx}^{in} + E_{Tx}^{in} + 2E_{fuse} = 2kE_{elec} + 2kE_{DA} + 2k\varepsilon_{fs}L^2/m \quad (16)$$

We analyze the average energy consumption of nodes in each cluster as follows. Assume that leader node is in the end of chain, but is not the root leader node, and it is the leaf leader node of the minimum cost tree. The energy consumption of the leader node is E_{ld} , the energy consumption of sensor nodes in the end of chain is E_{nor}^{end} , the energy consumption of sensor nodes in chain is E_{nor}^{in} . Therefore, the average energy consumption of a node in each cluster is as in Eq. (17).

$$\begin{aligned} E_{ave} &= (E_{ld} + E_{nor}^{end} + E_{nor}^{in})/(N/m) \\ &= [(E_{Rx}^{in} + 2E_{fuse} + E_{Tx}^{ld}) + E_{Tx}^{in} + (N/m - 2)(E_{Rx}^{in} + 2E_{fuse} + E_{Tx}^{in})]/(N/m) \\ &= [(2N/m - 1)kE_{elec} + 4kE_{DA} + 2Nk\varepsilon_{fs}c^2L^2/m^2]/(N/m) \end{aligned} \quad (17)$$

Based on Eq. (16) and Eq. (17), the difference between E_{nor}^{long} and E_{ave} is as in Eq. (18).

$$E_{diff}(m) = E_{nor}^{long} - E_{ave} = mkE_{elec}/N + (2 - 4m/N)kE_{DA} + 2L^2(1 - c^2)k\varepsilon_{fs}/m \quad (18)$$

The derivation of Eq. (18) is:

$$E_{diff}^i(m) = kE_{elec}/N - 4kE_{DA}/N - 2L^2(1 - c^2)k\varepsilon_{fs}/m^2.$$

If $E_{diff}^i = 0$, we can obtain $m = L\sqrt{2N(1 - c^2)\varepsilon_{fs}/(E_{elec} - 4E_{DA})}$, which means that there exists an optimal number of leader nodes to ensure a minimum value of $E_{diff}(m)$.

Based on the Eq. (10) in section 4.1.2, the residual energy threshold of round i is $E_{th}^i = (r_{max} - r_i)E_0/r_{max}$, and the threshold of round $i+1$ is $E_{th}^{i+1} = (r_{max} - r_{i+1})E_0/r_{max}$. Therefore, the energy consumption threshold E_{th}^c of each round is as in Eq. (19).

$$E_{th}^c = E_{th}^i - E_{th}^{i+1} = (r_{i+1} - r_i)E_0/r_{max} = E_0/r_{max} \quad (19)$$

If $E_{diff} \leq E_{th}^c$, the impact of “long distance” on the energy consumption of nodes in clusters is within an acceptable range, while if $E_{diff} > E_{th}^c$, the “long distance” has a negative impact on the energy consumption of the network.

6. Simulations and Results

To evaluate performance of MCHTC discussed in the previous section, simulations are presented by MATLAB and its performance is compared with LEACH, PEGASIS, and

IEEPB. We describe performance metrics, simulation setup and simulation results in this section.

6.1 Performance Metrics

The following metrics are used to capture performance of the proposed algorithm and to compare with other algorithms.

Network lifetime: Wireless sensor networks is a data collection network, failure of any node may lead to the incompleting data and unconnected network. Therefore, in this paper, the metric of network lifetime is the round that appears the first death node.

Energy consumption: The total energy consumed by nodes receiving and sending the data packets.

Network energy balance: In this paper, we use the same definition of network energy balance as in [26]. Based on the definition of network lifetime, to prolong network lifetime, we need to prolong the round that appears the first death node. Assume that round appears the first death node is r_i , and the energy consumption of any node j is E_j . To prolong network lifetime, we can obtain Eq. (20).

$$\sum_{i=1}^N E_0 = \sum_{i=1}^N E_j \quad (20)$$

Eq. (20) presents that when the first death node appears in the network, the energy consumption of all nodes in the network is equal to the total energy consumption of the network, which means that the network energy is not wasted.

Assume that E_r^i is the residual energy of node i when the first death node appears in the network, the energy load factor is as in Eq. (21).

$$load_i = E_r^i / E_0^i \quad (21)$$

The average energy load factor is as in Eq. (22).

$$load_{ave} = \sum_{i=1}^N load_i / N \quad (22)$$

Based on Eq. (21) and Eq. (22), the *Network Energy Balance* (EB) is as in Eq. (23).

$$EB = \sqrt{\sum_{i=1}^N (load_i - load_{ave})^2} / N \quad (23)$$

We can obtain that the smaller the difference between $load_i$ and $load_{ave}$, the more balanced the energy consumption of nodes.

6.2 Simulation Setup

100 nodes are randomly distributed in the monitoring area with the size of $100m \times 100m$ in the simulation. The coordinates of the base station is (50,150). The basic parameters for the simulation are as shown in [Table 2](#).

Table 2. Simulation parameters

Parameters	Values
Network size	100m×100m
Number of nodes	100
Base station location	(50,150)
Data packet size	4000bit
Control packet size	200bit
Initial energy of nodes	0.5J
Region partition factor	0.01, 0.03, 0.05, 0.07, 0.09
E_{elec}	50nJ/bit
ϵ_{fs}	100pJ/bit/m ²
ϵ_{mp}	0.0013pJ/bit/m ⁴
E_{DA}	5nJ/bit/signal

6.3 Simulation Results

6.3.1 The Topology of Network

Fig. 10 is the network topology that does not execute topology control. If topology control is not executed, nodes will communicate with neighbors with the maximal transmission power, a lot of redundant links will exist in the network. The energy consumption of a node is large, and leads to a short network lifetime.

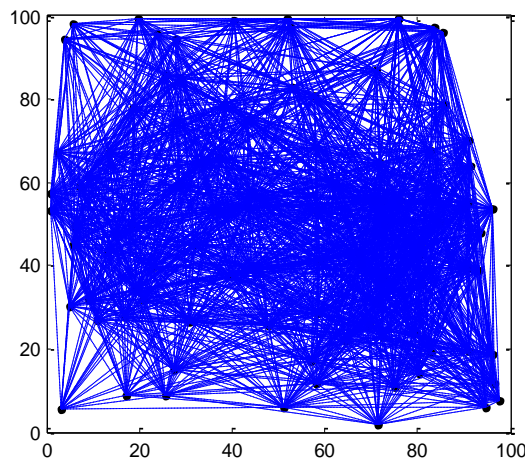


Fig. 10. Topology that does not execute topology control

Fig. 11-c and **Fig. 11-d** are the topologies derived by MCHTC when $p = 0.05$. If $p = 0.05$, the *Region Partition Parameter* satisfies $t = 3$, which means that node distribution area is divided into 9 subareas. Therefore, 9 short chains are formed in the network. The solid blue dots in figures are the leader nodes elected in each cluster. The solid red lines are the links

between leader nodes, which presents that a tree topology is formed by leader nodes. Compared with Fig. 10, the topology derived by MCHTC reduces a lot of redundant links between nodes, the transmission power of nodes only needs to satisfy that nodes can reach the farthest neighbor when broadcasting messages. Compared with the topology derived by PEGASIS, the multi-chain topology derived by static clustering not only reduces the transmission delay of data packets, but also preserves the superiority that multi-hop topology can reduce the communication distance between nodes and energy consumption of nodes.

In MCHTC, according to the actual requirements of the distribution area, the *Region Partition Factor* p can be adjusted to change the number of clusters and the length of chains formed in clusters. Fig. 11-a, 11-b are the topologies derived by MCHTC when p equals to 0.03, 0.07 respectively and nodes have the same distribution. p is independent of the distribution of the nodes. No matter whether nodes have the same distribution or not, executing MCHTC will form a hierarchical topology as mentioned above, and changing p will change the number of chains and the tree topology formed by leader nodes. Fig. 11-c, 11-d are the topologies derived by MCHTC when p equals to 0.05 and nodes have different distribution.

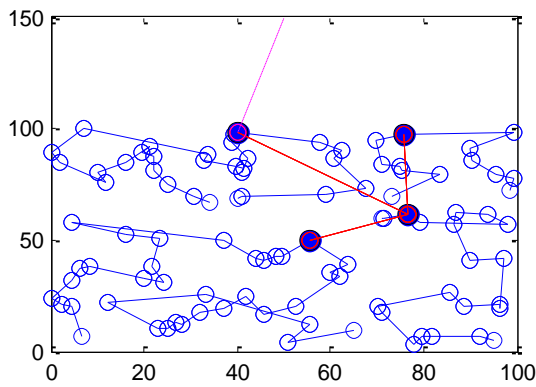


Fig. 11-a. Topology derived by $p = 0.03$

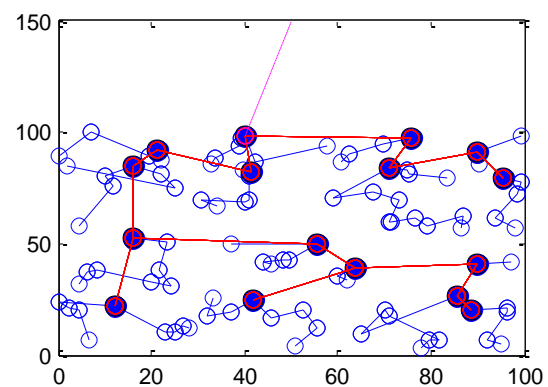


Fig. 11-b. Topology derived by $p = 0.07$

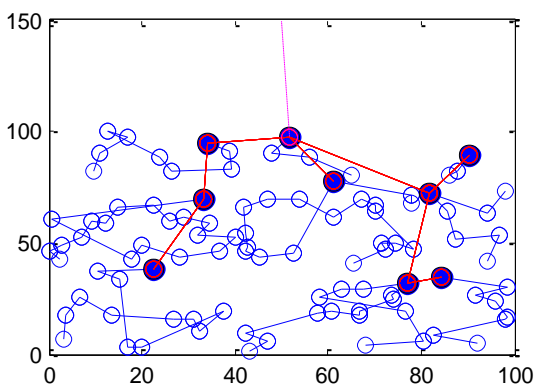


Fig. 11-c. Topology derived by $p = 0.05$

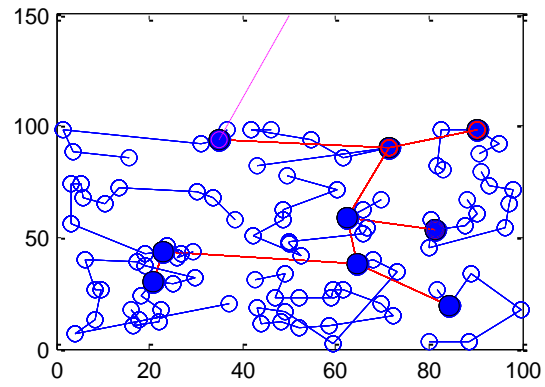


Fig. 11-d. Topology derived by $p = 0.05$

6.3.2 Network Lifetime

Fig. 12-a shows the trend of total number of nodes that remain alive over simulation runs. Fig. 12-b is the network lifetime corresponding to different number of leader nodes. Table 3 shows the number of rounds when 1%, 10%, 50%, and 90% of nodes die. Table 4 shows the percentage of prolonged network lifetime with different *Region Partition Factor*.

We can obtain that the network lifetime is the shortest under the topology derived by LEACH. Single-hop communication between cluster heads and the base station results in a rapid energy consumption of cluster heads. In PEGASIS and IEEPB, the death rate of nodes is relatively slow at first, but accelerates when the network enters into the decline phase. Due to that there exists long-chain problem between some nodes in PEGASIS, and long communication distance leads to premature death of some nodes, the network lifetime is shorter than IEEPB and MCHTC. In IEEPB, although the method of constructing chain topology could avoid the long-chain problem in PEGASIS, it has a negative effect on the network energy balance, because some nodes may have more child nodes, while others may have fewer or no child nodes. Nodes with more child nodes have a larger energy consumption, which will shorten the lifetime of the network. The network lifetime is the longest under the topology derived by MCHTC, and the death rate of nodes is relatively slower when network enters into the decline phase. When all nodes die in other three kinds of algorithms, there are certain number of nodes alive in MCHTC. Therefore, MCHTC not only prolongs the network lifetime, but also balances the energy consumption of nodes. It is because the communication cost, the residual energy of nodes and distance are considered when nodes look for the next hop neighbor. The simulation result in Fig. 12-a is obtained when $p = 0.05$.

Because there is only one leader node in PEGASIS and IEEPB, the network lifetime of PEGASIS and IEEPB are reference values which are presented by the dotted lines in Fig. 12-b. Compared with LEACH, MCHTC significantly prolongs the network lifetime when the number of leader nodes is the same in the network. Therefore, multi-hop topology can reduce the energy consumption of nodes compared with single-hop topology. It is very important for the energy-constrained wireless sensor networks.

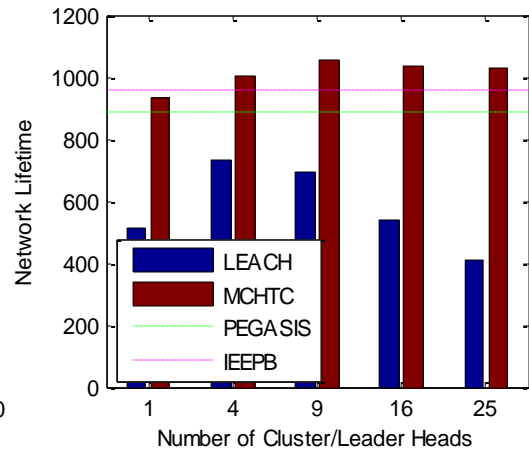
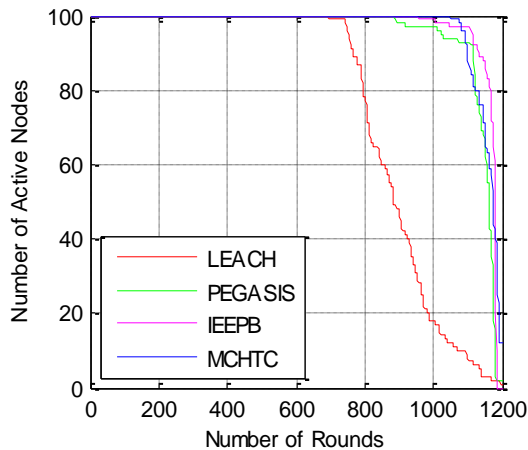


Fig. 12-a. Number of active nodes per round

Fig. 12-b. Comparison of the network lifetime

We can obtain from Table 3 that the network lifetime is the longest when $p = 0.05$. When $p = 0.01$, the topology derived by MCHTC is the same as PEGASIS. In MCHTC, nodes select next hop neighbor based on *Energy Distance*, the network lifetime increases compared with PEGASIS. However, the problem of long distance between nodes leads to a slightly lower network lifetime compared with IEEPB. But we can not ignore that the negative impact of long distance on the network lifetime is reduced when nodes look for the next hop neighbor based on *Energy Distance*.

Table 3. Number of rounds when different proportions of nodes die

Algorithm	No. of rounds (1%)	No. of rounds (10%)	No. of rounds (50%)	No. of rounds (90%)
LEACH	695	724	899	1172
PEGASIS	885	1115	1160	1176
IIEPB	957	1136	1182	1190
MCHTC($p=0.01$)	933	1115	1159	1174
MCHTC($p=0.03$)	1003	1099	1173	1189
MCHTC($p=0.05$)	1052	1076	1181	1256
MCHTC($p=0.07$)	1036	1097	1174	1242
MCHTC($p=0.09$)	1028	1111	1165	1255

Table 4 shows the percentage of prolonged network lifetime with different *Region Partition Factor*. When $p=0.01$, MCHTC extends network lifetime 34.24% and 5.42% respectively compared with LEACH and PEGASIS. However, the network lifetime is shorter than IIEPB, because when $p=0.01$, the network is not divided into subareas, a chain is formed in the network, and there still exists long-chain problem between some nodes. We can obtain from the **Table 4** that when $p=0.05$, MCHTC has the best performance of network lifetime, it extends network lifetime 51.37%, 18.87% and 9.93% respectively compared with LEACH, PEGASIS and IIEPB. When $p=0.03, 0.05, 0.07$ and 0.09 respectively, several short chains are formed in the network, and a tree topology is formed by leader nodes. Compared with IIEPB, MCHTC extends network lifetime 4.81%, 9.93%, 8.25% and 7.42% respectively, it shows that multi-chain topology and multi-hop transmission can prolong the network lifetime as well as avoid the condition that some nodes have more child nodes while others have fewer or no child nodes.

Table 4. Percentage of prolonged network lifetime with different *Region Partition Factor*

p	LEACH	PEGASIS	IIEPB
0.01	34.24%	5.42%	-2.51%
0.03	44.32%	13.33%	4.81%
0.05	51.37%	18.87%	9.93%
0.07	49.06%	17.06%	8.25%
0.09	47.91%	16.16%	7.42%

6.3.3 Energy Consumption

Fig. 13 demonstrates the residual energy of the network during the simulation rounds. LEACH consumes more energy because the cluster heads collect data packets from sensor nodes and transmit the fused data packets directly to the base station. Compared with PEGASIS, IIEPB and MCHTC only slightly reduce the energy consumption, and the performance of IIEPB and MCHTC is nearly the same. It is because the topology derived by PEGASIS has significantly reduces the distance between nodes, and has a better performance in terms of energy consumption and network energy balance. Based on the theoretical analysis in section 5.4, in MCHTC, the number of leader nodes is unrelated with the energy consumption, therefore, the number of leader nodes does not have an impact on the network energy consumption. At the same time point, the network residual energy in PEGASIS, IIEPB and MCHTC is higher than LEACH, which means that the energy consumption of each round in multi-hop topology is smaller, and the energy efficiency is higher.

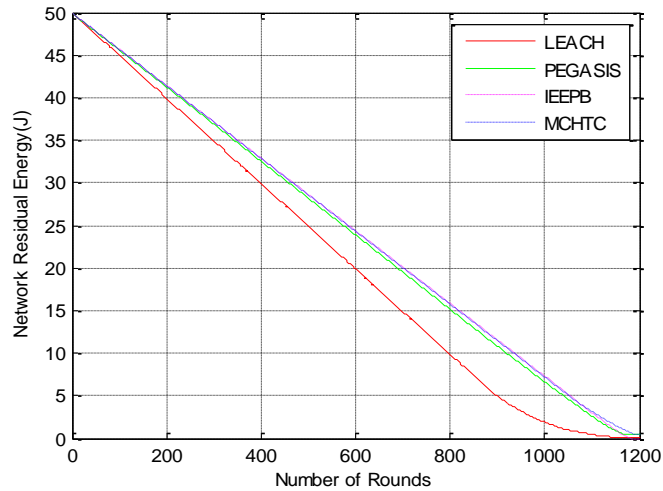


Fig. 13. Energy consumption of the network

6.3.4 Network Energy Balance

Fig. 14-a is the network energy balance with different *Region Partition Factor* p . The parameter p is the theoretical percentage of cluster nodes in LEACH. In MCHTC, parameter p is the *Region Partition Factor* which determines the number of clusters. The energy balance of PEGASIS and IEEPB are the reference values which are presented by dotted lines.

Based on the definition of EB in section 6.1, the smaller the EB, the more balanced the energy consumption of nodes in the network. LEACH has a poor performance in terms of EB because sensor nodes communicate with cluster heads directly, and the energy consumption of nodes which are far from cluster heads is large. On the contrary, the energy consumption of nodes which are near the cluster heads is small. Cluster heads transmit data packets to the base station directly, which leads to a large energy consumption of cluster heads.

The performance of PEGASIS is better because there is only one leader node elected to transmit data packets directly to the base station. The way PEGASIS electing leader node can ensure that all nodes in the network are elected as leader node after N rounds and nodes only communicate with the nearest neighbor.

The chain formed in IEEPB can avoid long distance between nodes, however, increase the number of comparisons and some nodes have multiple child nodes, which has a negative impact on the network balance. Therefore, the performance of IEEPB is poorer than PEGASIS. In MCHTC, when $0.01 \leq p \leq 0.02$, the topology derived by MCHTC is the same as PEGASIS, and has a better performance in terms of EB compared with PEGASIS and IEEPB. However, with the increasing of p , the number of leader nodes increases, which means that the number of nodes with large energy consumption increases. Therefore, it has a negative impact on EB. When $0.03 \leq p \leq 0.06$, the performance of MCHTC is between PEGASIS and IEEPB, which means that the multi-chain topology not only can avoid long distance between nodes, but also will prevent nodes from having multiple child nodes. when $0.07 \leq p \leq 0.09$, MCHTC has a poorer performance compared with PEGASIS and IEEPB, but better than LEACH.

Fig. 14-b is the energy balance with different number of leader nodes. As mentioned above, the energy balance of PEGASIS and IEEPB are reference values which are presented by dotted lines. The performance of MCHTC is significantly better than LEACH when the number of leader nodes is the same in the network.

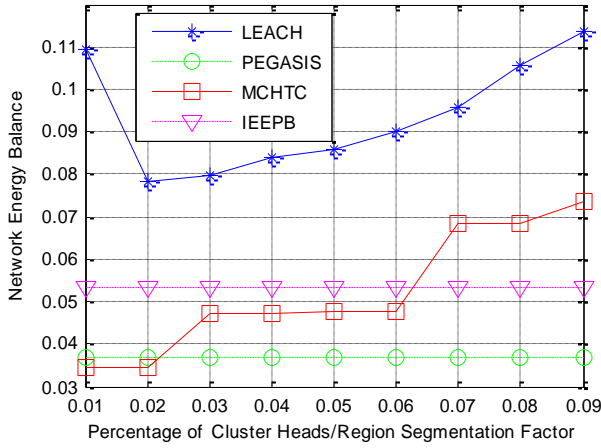


Fig. 14-a. Comparison of network energy balance

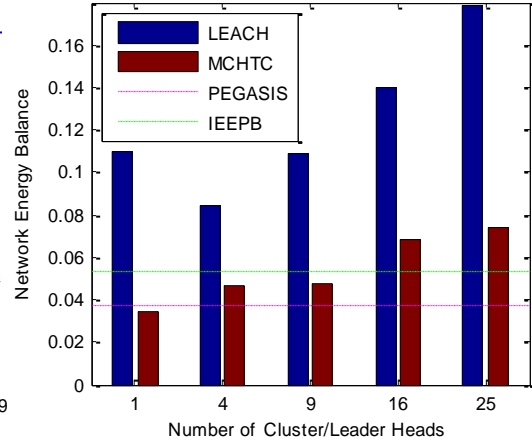


Fig. 14-b. Comparison of network energy balance with the same number of leader nodes

6.3.5 Energy Consumption Analysis of “Long Distance” in Clusters

Based on the theoretical analysis in section 5.5, Fig. 15 is the comparison of E_{Nor}^{long} and E_{ave} . To ensure the uniform distribution of the abscissa, we use *Region Partition Parameter* t to present the abscissa. The difference between E_{Nor}^{long} and E_{ave} decreases with the increasing number of clusters. We analyze whether E_{diff} is within an acceptable range as follows.

In this paper, $r_{max} = 1200$, $E_0 = 0.5J$. Based on Eq. (19), the energy consumption threshold of each round is approximately 4.17×10^{-4} . We can obtain that E_{diff} is not within the acceptable range when the distribution area is not divided. Therefore, the energy consumption of “long distance” has a negative impact on the energy consumption of the network. E_{diff} decreases with the increasing number of clusters, and satisfies $E_{diff} \leq E_{th}^c$. Therefore, the multi-chain topology derived by MCHTC can to some extent avoid the negative effect of “long distance” in clusters.

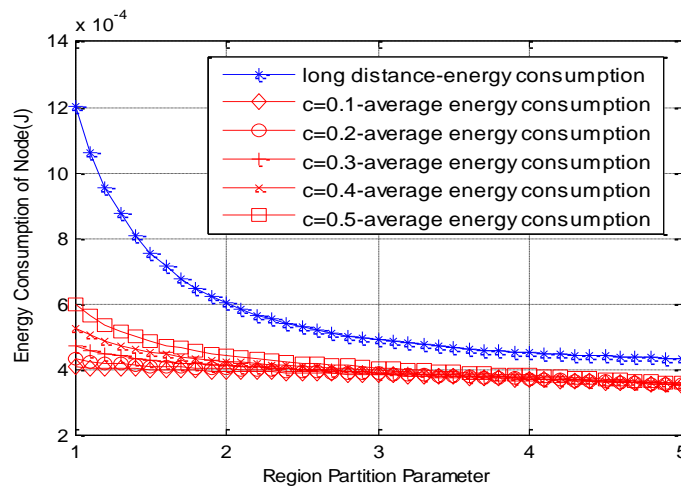


Fig. 15. Comparison of energy consumption

7. Conclusion

In this paper, we have proposed a Multi-chain Based Hierarchical Topology Control Algorithm (MCHTC) for wireless sensor networks. The main goal of MCHTC is to minimize energy consumption and prolong network lifetime based on certain network energy balance. MCHTC can meet the requirements for energy-saving applications. MCHTC divides the network area into several subareas based on Region Partition Factor by static clustering. The communication cost, the residual energy and distance are considered when nodes select the next hop neighbor to reduce the energy consumption of nodes. MCHTC organizes nodes in each cluster into a chain, and leader nodes form a tree topology to avoid long distance between them. When electing leader nodes, candidate leader nodes are elected based on the residual energy threshold. Leader nodes are elected based on the residual energy of candidate leader nodes and the distance between themselves and the base station. The number of clusters can be adjusted by changing the Region Partition Factor based on the actual requirements of distribution area. Simulation results show that compared with LEACH, PEGASIS and IEEPB, the proposed algorithm can reduce the energy consumption and prolong the network lifetime based on certain network energy balance.

References

- [1] Y. L. Chen, Y. N. Shih and J. S. Lin, "A four-layers hierarchical clustering topology architecture with sleep mode in a wireless sensor network," in *Proc. of 7th IEEE Int. Conf. on Complex, Intelligent, and Software Intensive Systems*, pp. 335-339, July 3-5, 2013. [Article \(CrossRef Link\)](#)
- [2] A. Chamam and S. Pierre, "A distributed energy-efficient clustering protocol for wireless sensor networks," *Computers and Electrical Engineering*, vol. 36, no. 2, pp. 303-312, March, 2010. [Article \(CrossRef Link\)](#)
- [3] Y. Chen and Y. Chen, "An energy efficient clustering algorithm based on residual energy and concentration degree in wireless sensor networks," in *Proc. of the Second Symposium International Computer Science and Computational Technology (ISCST '09)*, pp. 306-309, December 26-28, 2009.
- [4] S. Zairi, B. Zouari, E. Niel and E. Dumitrescu, "Nodes self-scheduling approach for maximising wireless sensor network lifetime based on remaining energy," *IET Wireless Sensor Systems*, vol. 2, no. 1, pp. 52-62, March, 2012. [Article \(CrossRef Link\)](#)
- [5] V. Tabus and J. Astola, "Maximizing the lifetime of energy constrained wireless sensor networks having tree topology," in *Proc. of 6th IEEE Int. Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 388-391, May 21-23, 2014. [Article \(CrossRef Link\)](#)
- [6] Q. Mamun, S. Ramakrishnan and B. Srinivasan, "Multi-chain oriented logical topology for wireless sensor networks," in *Proc. of 2nd IEEE Int. Conf. on Computer Engineering and Technology*, pp. 367-372, April 16-18, 2010. [Article \(CrossRef Link\)](#)
- [7] L. Liu, X. Qi, J. Xue and M. Xie, "A topology construct and control model with small-world and scale-free concepts for heterogeneous sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2014, article ID 374251, pp. 1-8, March, 2014. [Article \(CrossRef Link\)](#)
- [8] M. A. Labrador and P. M. Wightman, *Topology control in wireless sensor networks: with a companion simulation tool for teaching and research*, 1st Edition, Springer, USA, 2009.
- [9] L. Wen, B. Zhang, S. Chai and R. Zhao, "A connectivity-keeping hierarchical topology for mobile wireless sensor networks," in *Proc. of 24th Chinese Conf. on Control and Decision*, pp. 3378-3383, May 23-25, 2012. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6244538&tag=1>
- [10] K. S. Babulal and R. R. Tewari, "Cross layer cooperative clustering for performance enhancement in wireless sensor network," in *Proc. of IEEE Nirma University Int. Conf. on Engineering*, pp. 1-6, December 8-10, 2011. [Article \(CrossRef Link\)](#)

- [11] T. Zhang, J. Cao, Y. Chen, L. Cuthbert and M. ElKashlan, "A small world network model for energy efficient wireless networks," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1928-1931, October, 2013. [Article \(CrossRef Link\)](#)
- [12] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. of the 33rd Annual Hawaii Int. Conf. on System Science (HICSS'00)*, pp. 8020-8029, January 4-7, 2000. [Article \(CrossRef Link\)](#)
- [13] W. B. Heinzelman, A. P. Chandrakasan and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660-670, October, 2002. [Article \(CrossRef Link\)](#)
- [14] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366-379, October, 2004. [Article \(CrossRef Link\)](#)
- [15] B. Deb, S. Bhatnagar and B. Nath, "A topology discovery algorithm for sensor networks with applications to network management," *DCS Technical Report DCS-TR-411*, Rutgers University, May, 2001.
- [16] Y. Xu, J. Heidemann and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in *Proc. of 7th Annual Int. Conf. on Mobile Computing and Networking*, pp. 70-84, July 16-21, 2001. [Article \(CrossRef Link\)](#)
- [17] C. F. Li, G. H. Chen, M. Ye and J. Wu, "An uneven cluster-based routing protocol for wireless sensor networks," *Chinese Journal of Computers*, vol. 30, no. 1, pp. 27-36, January, 2007.
- [18] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proc. of IEEE conf. on Aerospace*, pp.1125-1130, March 9-16, 2002. [Article \(CrossRef Link\)](#)
- [19] Y. C. Yu and G. Wei, "An improved PEGASIS algorithm in wireless sensor networks," *Acta Electronica Sinica*, vol. 36, no. 7, pp. 1309-1313, July, 2008.
- [20] F. Sen, Q. Bing and T. Liangrui, "An improved energy-efficient pegasis-based protocol in wireless sensor networks," in *Proc. of 8th IEEE Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp. 2230-2233, July 26-28, 2011. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6020058
- [21] M. R. Jafri, N. Javaid, A. Javaid and Z. A. Khan, "Maximizing the lifetime of multi-chain pegasis using sink mobility," *World Applied Sciences Journal*, vol. 12, no. 9, pp. 1283-1289, March, 2013. [http://www.idosi.org/wasj/wasj21\(9\)13/5.pdf](http://www.idosi.org/wasj/wasj21(9)13/5.pdf)
- [22] N. Li, J. C. Hou and L. Sha, "Design and analysis of an MST-based topology control algorithm," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1195-1206, May, 2005. [Article \(CrossRef Link\)](#)
- [23] X. Fan, F. Du, J. GUO and L. Guan, "Lifetime optimization algorithm for mobile sink in wireless sensor network," *Journal of Computational Information Systems*, vol. 10, no. 5, pp. 1841-1848, March, 2014. <http://www.jofcis.com/showabstract.aspx?id=4327>
- [24] S. K. Chaurasiya, T. Pal and S. D. Bit, "An enhanced energy-efficient protocol with static clustering for wsn," in *Proc. of IEEE Int. Conf. on Information Networking (ICOIN)*, pp. 58-63, January 26-28, 2011. [Article \(CrossRef Link\)](#)
- [25] Q. Tang, B. Wang and Z. Dai, "MS-Leach: a routing protocol combining multi-hop transmissions and single-hop transmissions," in *Proc. of IEEE PACCS'09. Pacific-Asia Conf. on Circuits, Communications and Systems*, pp. 107-110, May 16-17, 2009. [Article \(CrossRef Link\)](#)
- [26] Q. Tang, B. Wang, Z. Dai and A. Yi, "Semi-centralized clustering protocol with energy balance and multi-hop transmissions," *Journal of Chinese Computer Systems*, vol. 31, no. 4, pp. 583-586, April, 2010.



Tang Hong obtained his PhD degree in Computer Software & Theory from the University of Chongqing in 2003. He received his Master degree on Photoelectric Signal Process from Sichuan University. He is currently a Professor at the Chongqing University of Posts and Telecommunications, China. His current research interests include computer networks and telecommunication technologies.



Wang Huizhu received her B.E. degree on Telecommunication Engineer from Chongqing University of Posts and Telecommunications, China in 2013. She is a postgraduate in Telecommunications & Information College, Chongqing University of Posts and Telecommunications. Her research interests include wireless sensor networks and complex network models.