A Data Collecting Strategy for Farmland WSNs using a Mobile Sink

Y.Q. Zhang, J.Y. Lin, H. Zhang

Yaqiong Zhang*

School of Information Engineering, Yulin University
Yulin 719000, China
*Corresponding author: yqzhang2208@163.com

Jiyan Lin

School of Information Engineering, Yulin University
Yulin 719000, China
linjiyan1018@163.com

Hui Zhang

School of Information Engineering, Yulin University
Yulin 719000, China
myhui521541@163.com

Abstract

To the characteristics of large number of sensor nodes, wide area and unbalanced energy consumption in farmland Wireless Sensor Networks, an efficient data collection strategy (GCMS) based on grid clustering and a mobile sink is proposed. Firstly, cluster is divided based on virtual grid, and the cluster head is selected by considering node position and residual energy. Then, an optimal mobile path and residence time allocation mechanism for mobile sink are proposed. Finally, GCMS is simulated and compared with LEACH and GRDG. Simulation results show that GCMS can significantly prolong the network lifetime and increase the amount of data collection, especially suitable for large-scale farmland Wireless Sensor Networks.

Keywords: farmland WSNs, data collection, virtual grid, mobile sink

1 Introduction

The development of intelligent agriculture and precision agriculture requires positioning, timing and quantitative monitoring and management of farmland parameters. As an accurate and real-time information acquisition method, Wireless Sensor Network has the characteristics of self-organization, low energy consumption, convenient configuration and low cost. It has a broad application prospect in the monitoring and control of farmland habitat parameters and crop growth monitoring [8]. A Wireless Sensor Network (WSN) consists of a large number of spatially distributed, wirelessly connected, self-governing sensor nodes, which are generally deployed in monitoring environments. These sensor nodes need energy to sense, process and transmit data, but their energy (battery) is limited. Therefore, it

is necessary to design an energy efficient data collection strategy to save energy consumption in order to extend the lifetime of a WSN [2].

A farmland Wireless Sensor Network has the characteristics of wide range and large number of sensors. In traditional Wireless Sensor Network, sensor nodes send data to the fixed sink through multi hop, which easily leads to the death of nodes near the sink due to excessive energy consumption, thus forming an energy hole. In recent years, the introduction of mobile sink can reduce the density of network nodes, improve the energy utilization of nodes, and extend the running time of the network [1, 3, 4, 13, 14, 21, 28]. Under the practical background of farmland application, how to monitor and collect various environmental parameters and crop parameters is a key problem. In this paper, the strategy using a single mobile sink node, combined with appropriate clustering and moving path, to improve the efficiency of energy utilization and data collection in farmland Wireless Sensor Network.

The movement modes of mobile sink nodes include fixed path, random movement, controlled motion and mobile path algorithms based on swarm intelligence algorithm or mathematical programming (linear programming path strategy) [17, 25, 26, 27, 29]. The movement of a mobile sink is modeled as a TSP problem with overlapping circular neighborhood, and the firefly algorithm is used to solve its optimal moving path [5, 19]. The heuristic algorithm [15] and the OPDG algorithm [11] aim to minimize the delivery delay and improve the real-time performance of the network by solving the TSP path. An algorithm is proposed using virtual force to determine the target position of sink node movement [22]. The mobile trend of sink nodes was predicted by combining the sleep and wake-up strategies to prolong the network lifetime [24]. A heuristic algorithm WRP is proposed, which can effectively reduce congestion and delay by giving higher priority to busy nodes in the network, at the same time, SMT is used to introduce virtual access points to optimize the access path of sink nodes [18]. These algorithms can shorten the delivery delay, prolong the network lifetime and reduce congestion in different scenarios, but there are some problems such as the complexity of the algorithm and poor scalability. Moreover, these algorithms are not suitable for large-scale Wireless Sensor Networks.

The farmland application of WSN is a kind of delay tolerant network, which has the characteristics of large range, many nodes and types. In the farmland WSN, CATS algorithm selects the most covered stops one by one until all nodes are covered, and then plans a shortest path in the selected rendezvous points [9], but the algorithm is complex and requires higher computing power of mobile sink. The GRDG (Grid based Data Gathering) algorithm proposed can effectively collect farmland data by dividing virtual grid in the network, establishing two-level routing and cooperating with a randomly moving mobile sink, which improves the scalability of the network and the energy utilization of nodes [12], but the delay is larger and the amount ofdata collected is not large.

In the farmland WSNs, a large number of sensor nodes are needed to form a sensor network, which is combined with the farmland meteorological station to realize the dynamic and high-precision monitoring of the farmland ecological environment and seedling, soil moisture, disease and disaster [23]. In this paper, how to use mobile sink to collect data from large-scale farmland effectively is studied, and the solution is given. Combined with the characteristics of hierarchical routing in Wireless Sensor Networks, a clustering scheme based on grid and a mobility strategy of a sink node is proposed.

2 Network model

2.1 Node model

It is assumed that a farmland WSN is composed of many sensor nodes which have same initial energy; the sensor nodes are non-mobile and randomly deployed in the monitoring area. Each node in WSN can acquire its own geographical position and perceive its own residual energy. Suppose the node model is as follows:

(1) Sensor nodes are non-mobile and each has its own unique ID, the set of ID numbers of N sensor nodes and one mobile sink node is $M = \{n_1, n_2, n_3, \ldots, n_N, n_{sink}\}$, where n_i is the i -th node and n_{sink} is the mobile sink node;

(2) All sensor nodes are isomorphic and can fuse data;

(3) The wireless communication link is bidirectional and symmetrical. The node has the function of ranging, and the transmitting power can be adjusted at any time according to the communication

distance;

(4) When there is no data collection task, the cluster member nodes can sleep, and the cluster head can wake up the member nodes regularly and request them to transmit the sensed data;

(5) The storage capacity of sensor node is limited. When the amount of unsent sensing data exceeds the maximum storage capacity, the node will save the latest data and discard the earliest data.

2.2 Energy consumption model

Energy consumption model of Wireless Sensor Network is based on that of wireless communication system [10, 16].

The energy consumption of the transmitting end includes the energy consumption of the radio frequency module and the signal amplifier, and the energy consumption of the receiving end only includes the energy consumption of the receiving circuit. Signal amplifier energy consumption is on the basis of distance between the transmitter and receiver where free space fading model and multi-path fading model can be used [20].

The energy consumed by transmitting L bits of data is shown in the formula (1).

$$E_{Tx}(L, D) = \begin{cases} E_{RF} * L + \xi_{fs} * L * D^2, & D < D_c \\ E_{RF} * L + \xi_{tworay} * L * D^4, & D \geq D_c \end{cases} \tag{1}$$

In the above formula (1), ξ_{fs} is the amplifier multiple of free space model, ξ_{tworay} is the amplifier multiple of multiple path attenuation model, Dc is a constant distance, and E_{RF} is the energy consumption of transmit/receive circuit.

The energy consumed by receiving L bit data is not related to the transmission distance, as shown in formula (2)[6, 7]

$$E_{Rx}(L) = E_{RF} * \text{L} \tag{2}$$

It can be seen from formula 1 and 2 that once the communication distance is greater than Dc, the energy consumption will increase rapidly.

3 GCMS algorithm

In the farmland WSN, a strategy based on grid clustering and using a mobile sink (GCMS) to collect data is proposed. The clustering structure can avoid the sensor sending a large amount of redundant data to sink. In this paper, virtual grid is used for clustering to balance the payload and energy consumption. Each grid is a cluster, and the cluster head (CH) is selected in the cluster. The cluster head is responsible for collecting the sensing data of the nodes in the cluster and forwarding the data to the mobile sink after data fusion. When the sink node moves to a communication area of a CH, sink needs to inform the CH of its moving. When sink moves to a rendezvous point, the new location is reported to the cluster head in the communication range.

The network still runs periodically as rounds, and each round includes three steps, as shown in Fig. 1. First, update CH, the node with the most residual energy and closest to the grid center will be selected as the new CH. Next, CH collects the data in the cluster, and finally CH sends the data to the mobile sink. All sensor nodes have the same hardware structure and function. They are randomly deployed in the monitoring area. Each node knows its own location and remains static all the time. Before the normal operation of the network, each node already knows the number of grids divided, so nodes can determine their position in the grid.
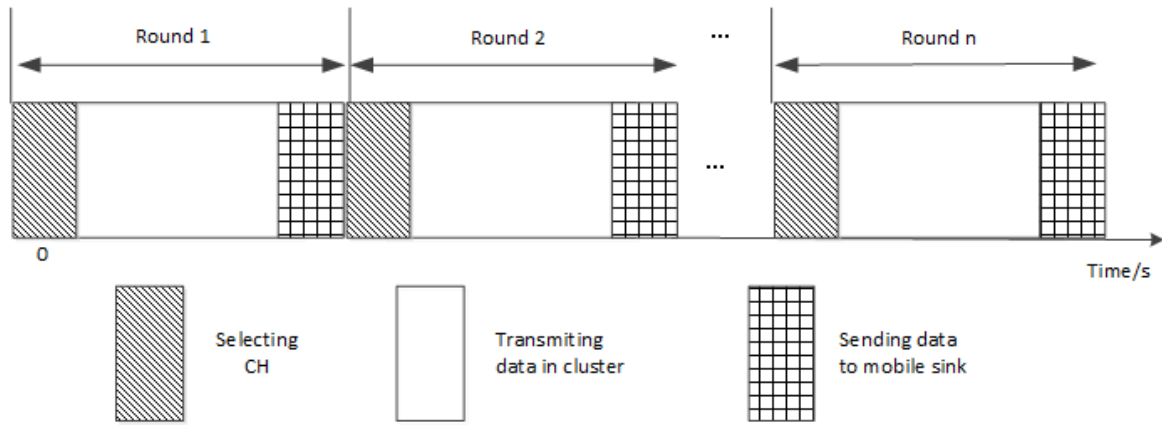
Figure 1: Operating round of GCMS algorithm

3.1 Clustering based on grid

It is supposed that the farmland Wireless Sensor Network is an M*M rectangular area, N sensor nodes are randomly deployed in the network. The network is divided into several equal area grids, as shown in Fig. 2. The total number of grids C_{num} is determined by N, as shown in formula (3). The width C_1 of each virtual grid can be expressed by formula (4)

$$C_{num} = \lfloor N/50 \rfloor^2 \tag{3}$$

$$C_l = M/\sqrt{C_{num}} \tag{4}$$

The coordinates of reference origin are (x_0, y_0), and the position coordinates of sensor node n_i are (x_i, y_i). If the logical coordinates $\left(\text{Gr}_x^i, \text{Gr}_y^i \right)$ denote the number of virtual grid in which node n_i is located, the grid to which the node belongs can be obtained according to the following formula (5).

$$Gr_x^i = \left\lfloor \frac{x_i - x_0}{c_l} \right\rfloor, \quad Gr_y^i = \left\lfloor \frac{y_i - y_0}{c_l} \right\rfloor \tag{5}$$
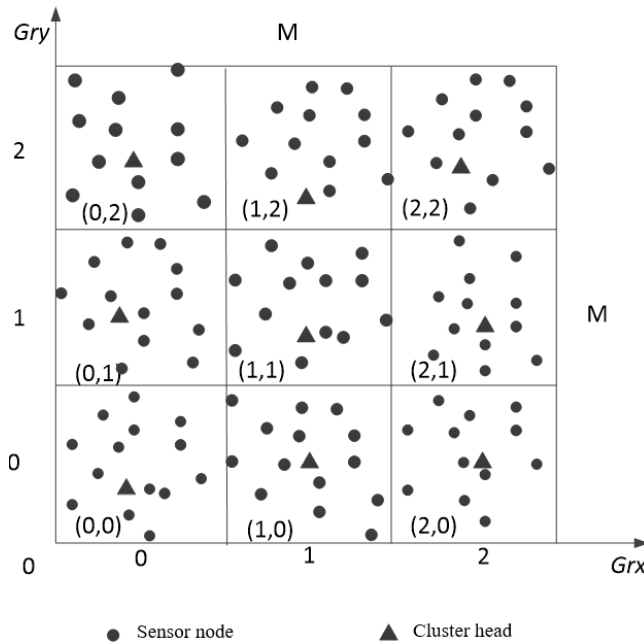

Figure 2: Grid division

When each grid is regarded as a cluster, the whole network is divided into C_{num} clusters of equal size. Which cluster a sensor node belongs to is determined by its logical coordinates. Let the k grid

cluster be represented as $CU_k(k = 1, 2, \ldots, n)$, and CH_k is the cluster head of cluster CU_k, and the number of nodes in the cluster is N_K. It is called the gravity center of cluster CU_k to find the location which has the smallest sum of distance squares with all nodes in the cluster. Obviously, the gravity center of cluster CU_k satisfies formula (6).

$$\sum_{i=1}^{N_k} \left[(x_{g-k} - x_j)^2 + (y_{g-k} - y_j)^2 \right] = \min \sum_{i=1}^{N_k} \left[(x - x_j)^2 + (y - y_j)^2 \right] \tag{6}$$

In formula (6),(x_{g-k}, y_{g-k}) is the coordinates of gravity center of cluster CU_k, and (x_i, y_i) is the coordinates of any sensor node in the cluster.

According to formula (6), the center of gravity coordinates of cluster CU_k can be obtained as formula (6).

$$x_{g-k} = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i, y_{g-k} = \frac{1}{N_k} \sum_{i=1}^{N_k} y_i \tag{7}$$

According to the wireless communication energy consumption model, the communication energy consumption between cluster head and member nodes is mainly related to distance. Therefore, in order to reduce and balance the communication energy consumption in the cluster, the node acting as the cluster head should be close to the center of gravity of the cluster. In addition, because the cluster head needs to forward data and manage the cluster, the node as the cluster head should also have more residual energy. So, the closer the node to the center of gravity and the more residual energy, the more likely it is to be selected as the cluster head. Therefore, for the k grid, the weighted sum function for cluster head selection is constructed as shown in formula (8).

$$f_{j-k} = \omega \frac{C_l - d_{gj-k}}{C_l + d_{gj-k}} + (1 - \omega) \frac{E_{rés}^{n_j}}{E_0} \qquad \omega \in (0, 1) \tag{8}$$

In formula (8), ω is a weight coefficient, $E_{rés}^{n_j}$ is the residual energy of node n_j, E_0 is the initial energy of sensor node, and d_{gj-k} represents the distance between the center of gravity of the cluster CU_k and the node n_j in the cluster, as shown in formula (9).

$$d_{gj-k} = \sqrt{(x_{g-k} - x_j)^2 + (y_{g-k} - y_j)^2} \tag{9}$$

According to formula (8), calculate the f f_{j-k} of each node in the cluster CU_k, and select the node with the largest f_{j-k} as the cluster head.

In formula (8), the value of weight ω is determined according to the actual needs. If the value of ω is large, the cluster head selection focuses more on the distance between the node and the center of gravity; otherwise, it focuses on the residual energy of the node.

At the end of each round, each node in the cluster reports its remaining energy and distance to the center of gravity to the cluster head. The cluster head calculates and compares the f_{j-k} value of each node according to formula (8). The node with the largest f_{j-k} will be selected as the cluster head of the next round.

After clustering based on virtual grid, cluster head manages the whole cluster, including determining the time interval of data collection, reporting sink location, data fusion and forwarding, etc.

3.2 Data transmission in cluster

After clustering and selecting the cluster head, the other nodes in the cluster send joining request to the cluster head, the request contains its own ID. On the basis of the number of member nodes, the cluster head divides time into several slots, that is, the TDMA mechanism is adopted to assign the corresponding transmission time slots to each member node. Each cluster member node collects and transmits data to the cluster head in its own time slot, and is dormant to save energy in other time slots. Cluster head collects information and fuses the data, then sends information to the mobile sink. In this way, the information is forwarded to the mobile sink.

3.3   Path planning of the mobile sink

The time interval between two moves of the sink to the same point can be regarded as the delay of data acquisition of sensor nodes, which is recorded as T_p. Therefore, the maximum delay limit means that there is a T_{limit} and formula (10) must be satisfied.

$$T_p < T_{\text{limit}} \tag{10}$$

On the other hand, the mobile sink can be considered as moving at a constant speed, and the time consumption of data transmission can be ignored; if the moving speed of the sink is V. as shown in formula (11), the maximum delay limit of mobile sink data collection can be transformed into the path length limit of mobile sink.

$$\begin{cases} v * T_p < v * T_{limit} \\ L_p < L_{limit} \end{cases} \tag{11}$$

The goal of GCMS algorithm is to plan a periodic mobile data collection loop for mobile sink in the node deployment area, which contains several rendezvous points. To achieve this goal, GCMS is divided into three steps. The first step is to set up a virtual grid in WSN. All grid line intersections become the candidate rendezvous points (RP) of GCMS. The second step is to determine the first rendezvous point from the candidate rendezvous points; the first rendezvous point will be set in the area with the most neighbor nodes. The third step is to add the new rendezvous points to the path one by one after setting up the first rendezvous point to realize the path expansion. This process is heuristic, and its heuristic function mainly considers two factors: the distance from the existing rendezvous point, the better, and the higher the density of the neighbor nodes, the better. Among them, "distance from the existing rendezvous point" is Euclidean distance. For solving the traveling salesman problem (TSP), the moving path and its length of mobile sink for data collection are obtained. Then, whether the length exceeds the limit L_{limit} of the total length of the path is detected. If it exceeds the limit, the point will be discarded; otherwise it will be added to the path to start a new round of rendezvous point generation process. The path planning of sink is finished until the list of candidate residency points is empty.

Rendezvous point is the location of mobile sink for data collection. Whenever a rendezvous point is reached, the mobile sink collects the data cached by the nearby cluster head nodes. The mobile path of sink in the network is a loop which is obtained by solving the traveling salesman problem for all the rendezvous points in the network, along which the data is collected periodically.

In order to generate residences, a list of candidate residences should be generated first. The network deployment area is continuous, in which there are infinite points. The computer algorithm cannot detect infinite points, and it is obviously unnecessary to repeatedly detect the candidate rendezvous points close to each other. Therefore, a grid like structure is introduced, in which the intersection of horizontal and vertical lines of the grid like structure is taken as the candidate rendezvous point. As shown in Fig. 3, the black points are the candidate rendezvous points. After the candidate rendezvous points are generated, the first rendezvous point can be selected first, and then the path can be expanded step by step. GCMS will select the candidate rendezvous point with the highest density of neighbor cluster head nodes as the first rendezvous point. If more than one residency point satisfies the condition, the point closest to the sink is selected as the first residency point. In order to describe the density of cluster head nodes, a covering function $c(n, i)$ is introduced. For any candidate rendezvous point nand a cluster head node i, the covering function $c(n, i)$ is shown in formula (12).

$$c(n, i) = \begin{cases} 0, & d(n, i) > R_c \\ 1, & d(n, i) \le R_c \end{cases} \tag{12}$$

In formula (12), $d(n, i)$ is the Euclidean distance between i and n, and R_C is called the coverage radius of the candidate rendezvous point, which is a parameter. Therefore, the node density C_{num} of the candidate rendezvous point n is defined as the total number of cluster head nodes covered by the candidate rendezvous point, as shown in formula (13), where n is the total number of cluster head nodes in the network.

$$N_n = \sum_{i=1}^{C_{num}} c(n, i) \tag{13}$$
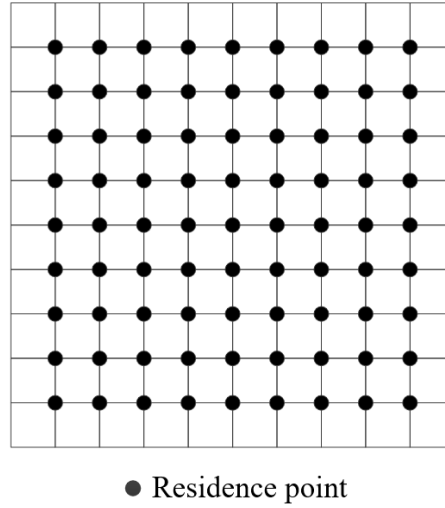
● Residence point

Figure 3: Candidate rendezvous points

After selecting the first rendezvous point, GCMS will gradually expand from the remaining candidate rendezvous points, and then form a complete loop by solving the traveling salesman problem (TSP). In order to evaluate the priority of the candidate rendezvous points, GCMS will evaluate the fitness of the candidate rendezvous points, and try to select a rendezvous point with the highest fitness to join the loop in each round. The fitness evaluation function of candidate rendezvous point n is shown in formula (14).

$$f(n) = \frac{N_n}{C_{num}} * \frac{D_{n-\min}}{L_{limit}} \tag{14}$$

In formula (14), D_{n-min} is the minimum distance from the candidate rendezvous point n to all established rendezvous points, which is obtained by Euclidean distance. L_{limit} is the total path limit of mobile sink. It can be seen from this function that GCMS will preferentially select the candidate rendezvous points with higher density of peripheral sensor nodes and far away from the existing rendezvous points. If there are more than one candidate residency point satisfying the condition, one of them is selected randomly.

Whether the candidate rendezvous point can be added to the rendezvous point list needs to be checked whether the length of the new loop exceeds the limit of the total path length of the mobile sink. There are two possible results: $L > L_{\text{limit}}$ or $L <= L_{\text{limit}}$. If it is the former, it will not be considered next; if it is the latter, it will be added to the path and become the rendezvous point. The candidate residency point is then removed from the list of candidate residences. Therefore, in GCMS, each candidate rendezvous point is checked once and only once. The generation process of each new rendezvous point is a main cycle of GCMS. Until all the candidate residences have been checked, that is, when the list of candidate residences is empty, the loop ends, and the final result is a data collection path of the mobile sink.

3.4 Allocation of residence time for the mobile sink

In order to further reduce the network delay, the cache data of cluster head node is obtained in the process of data transmission, and then the dwell time of sink is reserved according to the data size. At the same time, the residency point of the next hop cluster head is broadcast to the cluster head in the next hop cluster to prepare for data transmission. Let L_p denote the total moving distance of sink, T_p denotes the time interval of each round of data collection, T and T_k are the total residence time of each round and the residence time of each residency point respectively, and $\boldsymbol{v}$ represents the move rate of sink. The total residence time T and residence time of each cluster head T_k can be described as formula (15).

$$\begin{cases} T = T_p - \frac{L_p}{v} = \sum_{k=1}^{C_a - num} T_k \\ T_k = \frac{1}{N(k)\sigma} \sum_{j=1}^{N(k)} \mu_{kj} \end{cases} \tag{15}$$

In the formula (15),C_{a-num} is the total number of residency point accessed, $N(k)$ is the number of neighbor nodes of the kth residency point, μ_{kj} is the size of data packets collected by cluster head j, and σ is the transmission rate of data packets when cluster head communicates with mobile sink.

According to the above time allocation mechanism, sink can improve the data transmission balance between cluster heads within a certain network delay time. Considering the communication ability of cluster head nodes, sink can change the residence time according to the total amount of data cached by cluster head within the communication range. Therefore, the allocation mechanism can be used for fixed slot TDMA mechanism.

DMA communication mechanism includes the following steps:

Step 1: mobile sink sends broadcast packets to notify cluster head nodes in the adjacent area of its location coordinates and communication radius.

Step 2: the neighbor cluster head nodes requests to establish a connection with sink after receiving the broadcasting packet and the request data contains the amount of data of cluster heads.

Step 3: mobile sink selects the next rendezvous point according to the path planning, and sends ACK messages containing dwell time information.

Step 4: after the neighbor cluster heads of the next rendezvous point receive the ACK message, they collect the data of the member nodes in the cluster, and pack and fuse the data.

Step 5: the mobile sink node collects the data of the current access cluster heads, and then the cluster heads are identified to enter the sleep mode, and the sleep slot is T.

Step 6: if this point is the last rendezvous point, then finish; otherwise, the sink node moves to the next rendezvous point and return to step 1.

4　Simulation Analysis

4.1　Simulation parameters setting

The simulation runs in MATLAB r2018a environment. In the simulation experiment, 100 -600 sensor nodes are randomly distributed in the specified range for multi round data collection performance. The initial energy of each node is 1J, the data generation rate is random, and there is redundancy in the data sent by nodes in the sensor network. The ERF of 1 bit data processing circuit is $50*10^{-9}$J/bit, and the maximum data storing capacity of a single node is 32kbit. Considering the residual energy and distance, ω is chosen as 0.5. The simulation parameters are shown in Table 1.

Table 1: Simulation Parameters

Parameters	Values
Area of Network (m)	1000m*1000m
Number of sensor nodes	$100 - 600$
Length of packet (bit)	4000
Initial energy of sensor node E_0(J)	1
Transmission radius of cluster head (m)	70
Communication range of sin k(m)	120
D_c(m)	87.7
Moving speed of mobile sin k(m/s)	20
Cache of sensor node (k bit)	32
Sensing speed of sensor node (bit/s)	1000
Energy consumption of transmit/receive circuit (E_{RF}) (nJ/bit)	50
Amplifier multiple of free space model (ξ_{friss}) (pJ/bit/m^2)	10
Amplifier multiple of Multiple path attenuation model (ξ_{tworay}) (pJ/bit/m^2)	0.0013
Energy consumption of data fusion (nJ/bit/ packet)	5nJ/bit/ packet
Weight coefficient ω in CH election formula (8)	0.5

4.2   Results and analysis

In order to analyze the effectiveness of the proposed algorithm, GCMS is compared with LEACH which is based on fixed sink and GRDG which is based on mobile sink and grid clustering. The number of nodes alive, network lifetime, the amount of data collected and delay will be compared and analyzed.

4.2.1   Comparison of the number of nodes alive

Considering the farmland Wireless Sensor Network with 400 sensor nodes randomly deployed, the mobile rate of sink is set at 10m/s. With the operation of the network, some nodes die after some rounds due to the depletion of energy. The number of nodes alive corresponding to the 3 methods in different rounds is shown as Fig. 4.
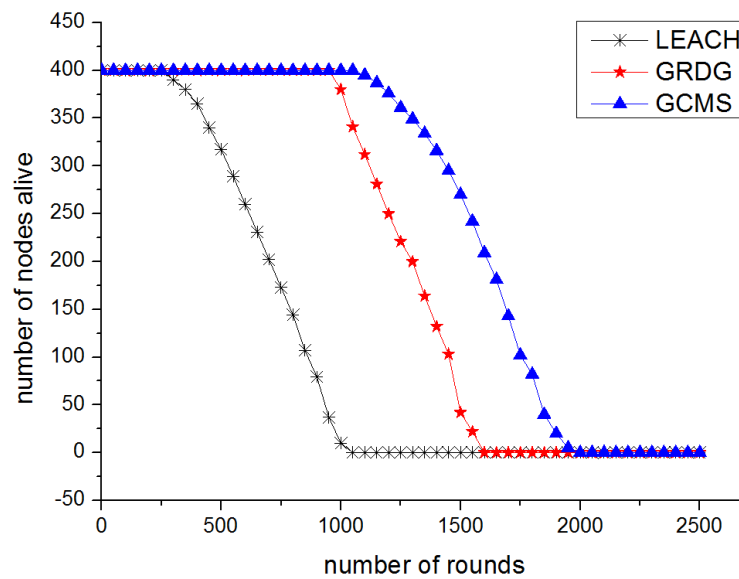
Figure 4: Comparison of number of nodes alive varies with the number of rounds

Network life cycle can be divided into stable period and unstable period. Before the FND (the time of first node death) of the network, the network is in the stable period; after the FND of the network, the network is in the unstable period. It can be seen from Fig. 4 that the stable period in GCMS and GRDG is significantly longer than that in LEACH. And, the number of nodes alive in GCMS and GRDG in unstable period is always larger than that in LEACH. This is because that GCMS and GRDG use mobile sink for data collection, which outperforms LEACH in energy consumption balance of network nodes, so FND appears later. Moreover, due to the mobility of sink, the network node can communicate with sink in a short distance, thus saving energy consumption, so the survival time of the node is longer. In terms of the two mobile sink data collection methods, the proposed GCMS is better than GRDG in terms of the number of nodes alive in the stable and unstable periods of the network. This is because the mobile sink in GCMS moves according to the shortest path, which makes the network energy consumption more balanced, while the mobile sink in GRDG moves randomly. It can also be seen from Fig. 4 that after entering unstable period, the number of surviving nodes in GCMS decreases most rapidly, which means that once a node dies, the energy of other nodes will also be exhausted quickly. It shows that the network energy consumption in GCMS is more balanced further.

Therefore, in terms of network stability and running time, the proposed GCMS is better.

4.2.2   Comparison of network lifetime

The lifetime of the network is compared by the time of the first node death. Fig. 5 shows the lifetime of the network at different scales. The abscissa represents the number of sensor nodes, and

the ordinate represents the life cycle of the network. Here, the number of rounds of the FND is taken as the measurement standard. With the increase of the number of sensor nodes, the network life cycle of LEACH algorithm shows a downward trend, on the contrary, the network lifetime of GCMS and GRDG strategies shows an upward trend. LEACH algorithm uses fixed sink for data gathering, and the energy consumption from cluster head to sink node is large. When the number of sensors increases, the amount of data forwarded by cluster head increases, so it leads to the rapid depletion of energy consumption and the death of nodes. In the GCMS and GRDG algorithms using mobile sink for data gathering, due to the less energy consumption of uniform clustering and data transmission from cluster head to mobile sink, the energy consumption does not increase sharply with the increase of the number of sensors. Therefore, the number of rounds of the FND is not significantly advanced, and the performance of GCMS algorithm is slightly better than that of GRDG. Therefore, mobile sink is used for data collection is more suitable for large-scale Wireless Sensor Networks.
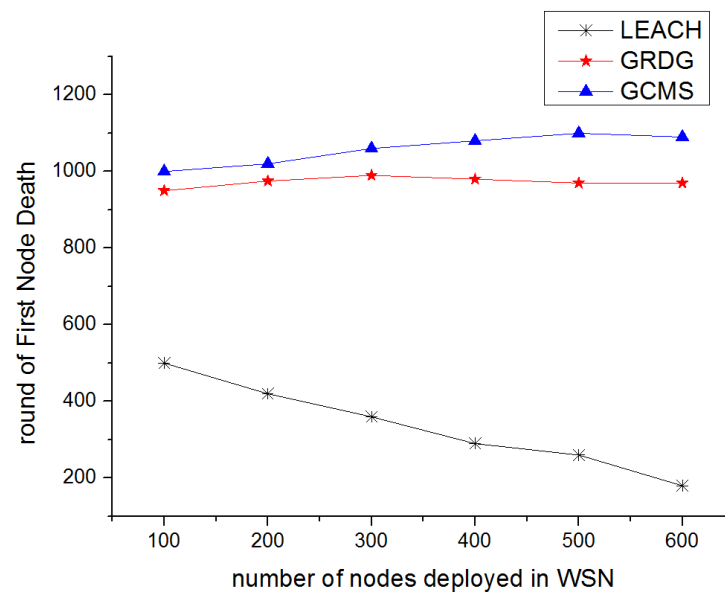

Figure 5: Comparison of FND with different number of nodes deployed in WSN

4.2.3 Comparison of amount of data collected

As shown in Fig. 6, the total amount of data collected in the network lifetime is compared when different numbers of sensor nodes are deployed in the Wireless Sensor Network. It can be seen that the amount of data collected by LEACH algorithm is significantly less than that of the other two algorithms using mobile sink. This is because the introduction of mobile sink can improve the network lifetime, so the amount of data collected is larger. The GCMS algorithm in this paper collects more data because of the longer network lifetime. However, with the increase of the number of nodes in the network, the amount of data collected by LEACH algorithm does not increase significantly. This is because with the increase of node density, it will lead to data transmission collision in LEACH, making some time slots wasted and not transmitting effective data. But, In GCMS proposed in this paper and GRDG, the communication between nodes in the cluster adopts TDMA mechanism instead of CSMA mechanism. In other words, CH allocates a fixed communication time slot to each cluster member node, so there is no collision and the amount of data collected will increase with the extension of network lifetime.
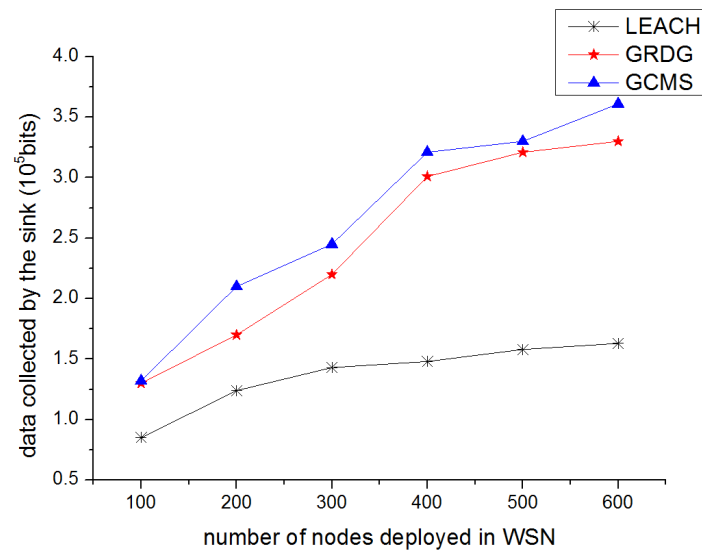
Figure 6: Comparison of the amount of data collected by sink node

In short, with the increase of the number of nodes, the amount of data collected by various methods increases approximately linearly. In theory, this is because the increase in the number of nodes makes the perceived data more, so the amount of data sent to sink increases significantly. However, in the actual data transmission process of WSNs, when the network runs to a certain round, if some nodes die prematurely, the network will not work and the data collection performance will be greatly damaged. Therefore, network life cycle (network lifetime) is the premise to improve the performance of data collection. If the network lifetime is long enough, nodes can work longer and collect more data.

In summary, compared with LEACH algorithm, the GCMS algorithm proposed in this paper has a slight advantage in data collection. That is, the GCMS algorithm proposed in this paper is suitable for data collection of large-scale sensor networks.

4.2.4   Comparison of network delay

Finally, the effectiveness of the proposed method is verified from the perspective of data transmission delay. Considering that the speed of wireless transmission is far greater than that of sink, only GCMS and GRDG using mobile sink technology are simulated and compared in terms of data transmission delay in this section. Assuming that the movement rate of mobile sink is 20m/s and 400 sensor nodes are deployed in the WSN, it shows the average delay comparison of data transmission with different rounds (i.e. from 0 to 1200 rounds) in Fig. 7.
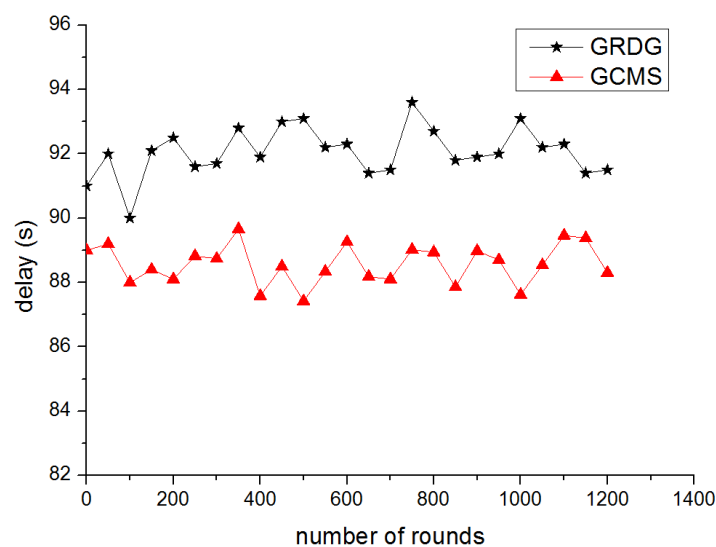

Figure 7: comparison of delay of different rounds

As can be seen from Fig. 7, the average transmission delay of GCMS is significantly lower than that of GRDG. This is because mobile sink can choose the shortest route from multiple mobile paths and determine different pause times according to the data size in GCMS. In GRDG, the mobile sink moves randomly, but it is not necessarily the optimal path. The longer the path may lead to the longer data collection delay.

The average delay comparison of different nodes in the sensor network is shown in Fig. 8. It can be seen that the average delay increases slightly with the increase of the number of nodes in GRDG algorithm, but it is not obvious. In GCMS algorithm, with the increase of the number of nodes, the delay increases obviously. When the number of nodes increases to 400, the delay increase is no longer obvious. This is because in GCMS, with the increase of the number of nodes, the movement path of sink will increase, which will lead to the increase of delay. However, the growth rate slows down when the number of nodes increases to a certain value, which proves that GCMS algorithm is more suitable for large-scale and delay insensitive networks.
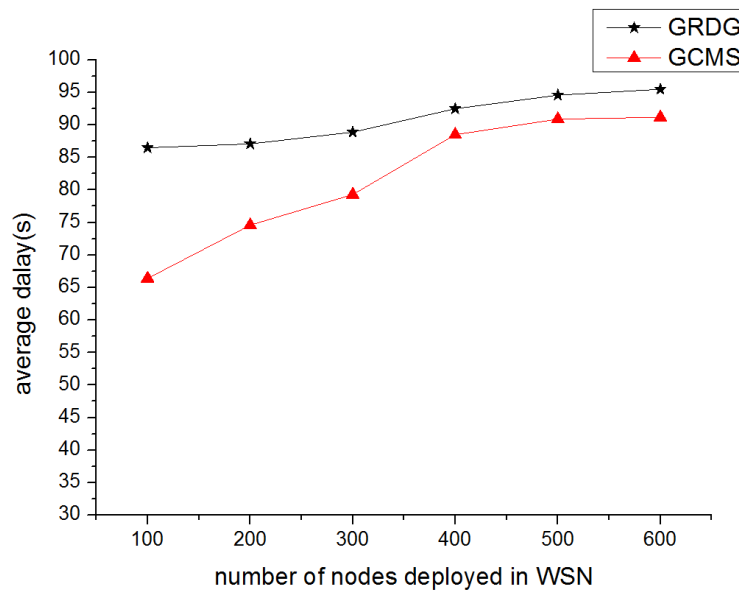

Figure 8: comparison of average delay of different number of nodes

5   Conclusions

A data collection algorithm (GCMS) based on virtual grid and mobile sink is proposed in the paper. The farmland Wireless Sensor Network is divided into several grids, that each grid is a cluster, in which the geographical location and residual energy are considered to select the cluster head. In order to improve the efficiency of data collecting of mobile sink, different access probabilities are set according to the distance between cluster head nodes, the number of neighbor nodes in cluster head. So the shortest path of sink can be selected from multiple mobile paths. Considering the impact of data delay and different data size collected by cluster heads, different pause times can be determined according to the amount of data to be collected. The residence time of sink in each location is set to a different value to ensure that each cluster head can send all its cache data to sink within a given time. Finally, the GCMS algorithm in this paper is simulated and compared with LEACH and GRDG methods. The results show the algorithm proposed in this paper has great advantages in prolonging the network lifetime and improving the amount of data collection. Compared with GRDG algorithm, the amount of data collected is increased by 10% on average and the network life cycle is prolonged by about 14% in GCMS algorithm. However, the network considered in this paper is simple and sensor nodes are isomorphic. Heterogeneous nodes and sink nodes moving at variable speed should be considered in next step for practical application.

Acknowledgements

References

[1] Ang, K.L.M.; Seng, J.K.P.; Zungeru, A.M. (2017). Optimizing energy consumption for big data collection in large-scale wireless sensor networks with mobile collectors. *IEEE Systems Journal*, 12(1), 616-626, 2017.

[2] Dong, M.; Ota, K.; Liu, A. (2016). RMER: Reliable and energy-efficient data collection for large-scale wireless sensor networks. *IEEE Internet of Things Journal*, 3(4), 511-519, 2016.

[3] Dutta, P.K.; Banerjee, S. (2019). Monitoring of aerosol and other particulate matter in air using aerial monitored sensors and real time data monitoring and processing, *Journal of System and Management Sciences*, 9(2), 104-113.

[4] Fan, P.F.; Shang, Z. (2019). Application of wireless sensor network in monitoring of weapon and equipment production. *Instrumentation Mesure Metrologie*, 18(1), 37-41, 2019.

[5] Ha, I.; Djuraev, M.; Ahn, B. (2017). An optimal data gathering method for mobile sinks in WSNs. *Wireless Personal Communications*, 97(1), 1401-1417, 2017.

[6] Hu, W.; Yao, W.; Hu, Y.; Li, H. (2019). Selection of Cluster Heads for Wireless Sensor Network in Ubiquitous Power Internet of Things. *International Journal of Computers Communications & Control*, 14(3), 344-358, 2019.

[7] Hu, W.; Li, H.H.; Yao, W.H.; Hu, Y.W. (2019). Energy Optimization for WSN in Ubiquitous Power Internet of Things. *International Journal of Computers Communications & Control*, 14(4), 503-517, 2019.

[8] Javaid, N.; Rasheed, M.B.; Imran, M.; Guizani, M.; Khan, Z.A.; Alghamdi, T.A.; Ilahi, M. (2015). An energy-efficient distributed clustering algorithm for heterogeneous WSNs. *EURASIP Journal on Wireless communications and Networking*, 2015(1), 151, 2015.

[9] Khan, A.W.; Abdullah, A.H.; Anisi, M.H.; Bangash, J.I. (2014). A comprehensive study of data collection schemes using mobile sinks in wireless sensor networks. *Sensors*, 14(2), 2510-2548, 2014.

[10] Khan, A.W.; Abdullah, A.H.; Razzaque, M.A.; Bangash, J.I.; Altameem, A. (2015). VGDD: A virtual grid based data dissemination scheme for wireless sensor networks with mobile sink. *International Journal of Distributed Sensor Networks*, 11(2), 890348, 2015.

[11] Kinalis, A.; Nikoletseas, S.; Patroumpa, D.; Rolim, J. (2014). Biased sink mobility with adaptive stop times for low latency data collection in sensor networks. *Information Fusion*, 15, 56-63, 2014.

[12] Kumar, A.K.; Sivalingam, K.M.; Kumar, A. (2013). On reducing delay in mobile data collection based wireless sensor networks. *Wireless Networks*, 19(3), 285-299, 2013.

[13] Kumar, I.; Sachan, V.; Shankar, R.; Mishra, R.K. (2018). An investigation of wireless S-DF hybrid satellite terrestrial relaying network over time selective fading channel. *Traitement du Signal*, 35(2), 103-120, 2018.

[14] Kumar, R.V.K.; Naik, G.M.; Murali, G. (2019). Wireless nano senor network (WNSN) for trace detection of explosives: The case of RDX and TNT. *Instrumentation Mesure Metrologie*, 18(2), 153-158, 2019.

[15] Lee, K.; Kim, Y.H.; Kim, H.J.; Han, S. (2014). A myopic mobile sink migration strategy for maximizing lifetime of wireless sensor networks. *Wireless Networks*, 20(2), 303-318, 2014.

[16] Lee, K.; Kim, Y.H.; Kim, H.J.; Han, S. (2014). A myopic mobile sink migration strategy for maximizing lifetime of wireless sensor networks. *Wireless Networks*, 20(2), 303-318, 2014.

[17] Lin, T.; Wu, P.; Gao, F.M.; Wang, L.H. (2019). A secure query protocol for multi-layer wireless sensor networks based on internet of things. *Revue d'Intelligence Artificielle*, 33(2), 145-149, 2019.

[18] Liu, W.; Fan, J.; Zhang, S.; Wang, Y.; Chu, Y. (2013). Grid-based real-time data gathering protocol in wireless sensor network with mobile sink. *2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing*, 857-864, 2013.

[19] Mehrabi, A.; Kim, K. (2015). Maximizing data collection throughput on a path in energy harvesting sensor networks using a mobile sink. *IEEE Transactions on Mobile Computing*, 15(3), 690-704, 2015.

[20] Misbahuddin, M.; Ratna, A.A.P.; Sari, R.F. (2018). Dynamic Multi-hop Routing Protocol Based on Fuzzy-Firefly Algorithm for Data Similarity Aware Node Clustering in WSNs. *International Journal of Computers Communications & Control*, 13(1), 99-116, 2018.

[21] Ren, J.; Huang, S.Y.; Song, W.; Han, J. (2019). A novel indoor positioning algorithm for wireless sensor network based on received signal strength indicator filtering and improved Taylor series expansion. *Traitement du Signal*, 36(1), 103-108, 2019.

[22] Salarian, H.; Chin, K.W.; Naghdy, F. (2013). An energy-efficient mobile-sink path selection strategy for wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 63(5), 2407-2419, 2013.

[23] Salarian, H.; Chin, K.W.; Naghdy, F. (2013). An energy-efficient mobile-sink path selection strategy for wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 63(5), 2407-2419, 2013.

[24] Srbinovski, B.; Magno, M.; O'Flynn, B.; Pakrashi, V.; Popovici, E. (2015). Energy aware adaptive sampling algorithm for energy harvesting wireless sensor networks. *2015 IEEE Sensors Applications Symposium (SAS)*, 1-6, 2015.

[25] Talmale, R.; Bhat, M.N.; Thakare, N. (2019). Energy attentive pre-fault detection mechanism with multilevel transmission for distributed wireless sensor network. *Revue d'Intelligence Artificielle*, 33(2), 97-103, 2019.

[26] Wang, F.F.; Hu, H.F. (2019). An improved energy-efficient cluster routing protocol for wireless sensor network. *Ingénierie des Systèmes d'Information*, 24(4), 419-424, 2019.

[27] Wen, W.; Zhao, S.; Shang, C.; Chang, C.Y. (2017). EAPC: Energy-aware path construction for data collection using mobile sink in wireless sensor networks. *IEEE Sensors Journal*, 18(2), 890-901, 2017.

[28] Zhang, R.; Pan, J.; Xie, D.; Wang, F. (2015). NDCMC: A hybrid data collection approach for large-scale WSNs using mobile element and hierarchical clustering. *IEEE Internet of Things Journal*, 3(4), 533-543, 2015.

[29] Zhou, H.; Yu, K.M. (2019). A novel wireless sensor network data aggregation algorithm based on self-organizing feature mapping neutral network, *Ingenierie des Systemes d'Information*, 24(1), 119-123, 2019.

C O P E

Member since 2012
JM08090

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Zhang, Y.Q.; Lin, J.Y.; Zhang, H. (2020). A Data Collecting Strategy for Farmland WSNs using
a Mobile Sink, *International Journal of Computers Communications* & *Control*, 15(6), 4024, 2020.
https://doi.org/10.15837/ijccc.2020.6.4024