A Novel Grouping-Based WSN Clustering Data Transmission Protocol

Jiequ Ji, Xianjin Fang(✉), Dai Chen, Yafei Xiao, and Yanting Wu

School of Computer Science and Engineering,
Anhui University of Science and Technology, Huainan, China
xjfang@aust.edu.cn

Abstract. The Grouping-based Clustering Data Transmission (GCDT) protocol is proposed in this paper, which aims at the problem of data loss easily during transmission process and the defects of transmission path after clustering on wireless sensor network (WSN). By calculating the reach probability of data packets, the original data packets are grouped to guarantee data integrity before transmission. Creating multi-path between cluster heads and the base station is to ensure reliable delivery of packets. The simulation experiment results show that the packet loss ratio using GCDT protocol can decrease about 15% to 28% compared with the existing routing protocols. The GCDT protocol proposed in this paper improved the reliability of the data transmission and the network survivability effectively.

Keywords: Grouping-based clustering · Data transmission · Wireless sensor network

1 Introduction

With the development and integration of sensors, computers, wireless communication and micro-electromechanics in recent years, the Wireless Sensor Network (WSN) has been paid more and more attention. Its concept was put forward in the 1970s. Consisting of low-cost, low-power, multi-functional and wireless-capable sensor nodes that can be deployed in random form at any location and at any time, enabling new applications and services in a wide range of areas, including environmental monitoring, target tracking, natural disaster prediction, smart home, intelligent traffic control and military situations etc [1].

WSN usually includes sensor nodes, sinks and base stations (BSs) and management nodes. A large number of sensor nodes are usually deployed in a random manner, which can achieve data transmission of short distance, low power consumption and low rate. Sensor nodes are limited in battery, memory, computing power and communication bandwidth, and they often operate in harsh environments, which all lead to the failure of data transmission of sensor nodes easily. Therefore, the survivability has become a key issue in the study of WSN [2].

ARRIVE proposed by Karlof et al. [3] is a topology-based, tree-like probabilistic algorithm to achieve fault-tolerant routing to prevent link failure and

F. Chen and Y. Luo (Eds.): Industrial IoT 2017, LNICST 202, pp. 152–162, 2017.
DOI: 10.1007/978-3-319-60753-5_16

node-mode failure. This method relies on periodic network flooding to prevent malicious or an orbiting nodes. Because flooding consumes a large amount of traffic and consumes a lot of resources, Braginsky et al. Proposed that the query-based route is no longer flooding the entire network, but uses a random path to send a query protocol to the target area [4].

The intrusion-tolerant wireless sensor network routing proposed by Deng et al. [5] adopts multi-path and combines the cryptographic mechanism to realize the WSN routing intrusion, it can achieve intrusion detection by bypassing the malicious nodes instead of intrusion detection. In order to facilitate the communication between the sensor nodes and the base station, INSENS constructs a forwarding table at each node.

Younis et al. Proposed a hybrid clustering protocol HEED (Hybird Energy-Efficient Distributed clustering) [6]. HEED points out that extending network life cycle, scalability and load balancing are the three most important requirements in WSN and extend the life cycle of the network by distributing the energy consumption evenly across the network. The disadvantage of HEED is that energy balance between cluster heads is not considered.

The survivability of the data transmission is studied by several algorithms above. In the wireless sensor network, all the data can not be lost in the transmission process [7]. In view of how to ensure the reliable transmission of data to establish multi-path between the CH and BS, this paper presents a grouping-based clustering data transmission protocol (GCDT) for WSN, and adopts distance based cluster head selection algorithms (DBCH) [8] to cluster each node. Under the premise of multi-path, the packets transmitted via WSN are divided into a series groups, then these are sent to the destination through multiple independent paths. In this paper, we propose a multi-path generation algorithm between cluster heads and base stations, which can ensure the transmission of cluster information in order, without loss or error, provide reliable data transmission services to the base station, and save some energy of the cluster head through multi-hop communication between cluster head and base station. Thus, the network's survival time is extended.

2 Model Building

2.1 Network Model

Reference [8,9], this paper on the wireless sensor network has the following settings.

(1) The nodes in the network remain the same location after the completion of the layout, and the nodes know themselves location information;
(2) A node can make damage around a small part of itself, but not the whole network;
(3) Prior to data transmission, there are already multiple-path established by the routing protocol throughout the network;
(4) Each path has its own key, which is independent of each other;

(5) Neighboring clusters can communicate with each other, and the cluster heads have data sent to the base station so that the base station can more accurately estimate whether the data is reliably transmitted to the destination;
(6) Base station is rich in resources, and can be trusted.

Based on the idea of divide and conquer, this paper divides the data transmission problem of large-scale network into small units of data transmission in the form of clustering. In the initial stage of the network, reference [10], using DBCH algorithm for cluster division and cluster head selection, and making the cluster head selection threshold $T(n)$ to the following improvements:

$$T(n) = \frac{p}{1 - p * [r \ mod(i/p)]} + (1-p)\frac{D_{max} - D_{i\ to\ BS}}{D_{max} - D_{min}}(\frac{D_R}{D_0}) \qquad (1)$$

Where, p is the optimal percentage of cluster heads, which determines the number of cluster heads per round; r is the election round number, $r\ mod\ 1\ p$ represents the elected cluster head nodes in a round, D_{max} and D_{min} represent respectively the maximum and minimum distance from the base station, E_R is remaining energy of nodes in the current round, E_0 is the initial energy of nodes. This improvement needs to take the residual energy and distance factors into consideration, cluster to take the residual energy and distance factors into consideration, cluster head selection is based on the distance between the node and the base station, and the node near the base station will be elected as the cluster head.

2.2 Data Transmission Model

Wireless sensor network data transmission is usually to save energy as the goal, to maximize the extension of network life. The GCDT is a transmission protocol which makes data packet into groups and through multi-paths transmission to improve the transmission reliability, each node calculate the data packet reachable probability based on local channel bit error rate and hop count information. In this protocol, firstly it determines the expected successful transmission rate on the basis of the importance of the data packet before the source node sends a packet, and then determines the number of packets to be sent and the next cluster head node. The number of grouping N can be calculated from the locally estimated channel bit error rate, the number of hops from the source node to the cluster head node, and the expected success rate, which is calculated as follows [11]:

$$N = \frac{\log(1-r)}{\log[1-(1-e)^h]} \qquad (2)$$

Where, r indicates the expected successful transmission rate, e indicates the local channel bit error rate, h represents the hoc count of source node to the cluster head node. After the grouping is completed, the data packets in each group are transmitted to the base station through individual independent paths.

2.3 Energy Consumption Model

In this paper, only the energy consumption of communication is calculated according to the simple energy consumption model, and the energy consumption of nodes in the process of storage and calculation is neglected. The formula for the energy consumption of the transmitter is [12]:

$$E_{Tx}(ld) = E_{Tx\_elc}(k) + E_{Tx\_amp}(l, d) = \begin{cases} l \times E_{elc} + l \times \varepsilon f_s \times d^2, d < dl_1 \\ l \times E_{elc} + l \times \varepsilon amp \times d^4, d \geq dl_1 \end{cases} \quad (3)$$

The formula for the energy consumption of the receiver when acquiring information is:

$$E_{Rx}(1) = E_{Rx\_elec}(1) = 1 \times E_{elec} \quad (4)$$

$$d1_1 = \frac{\varepsilon_{fs}}{\varepsilon_{amp}} \quad (5)$$

Where, l is the number of bits of transmitted packets, d is the distance between transmitter and receiver, dl_1 is a distance threshold, E_{elec} is energy consumption of each bit data between transmitting circuit and receiving circuit, ε_{f_s} is the energy consumption which adopts the definition of free space model, ε_{amp} is the energy consumption of multi-path amplifying circuit.

3 GCDT Protocol Analysis

The goal of the GCDT protocol is to guarantee the reliability of data transmission in WSN. It has the following four assumptions [13]:

(1) Before the data transmission, the whole network has been divided into several clusters by clustering algorithm, CHs have been selected;
(2) Base station is rich in resources, and can be trusted;
(3) Reliable data transmission based on grouping can use NACK response mode;
(4) Before the WSN is configured, the base station and the cluster head node can send data according to the previous clustering algorithm, and send the CH_Msg message;
(5) In order to avoid data redundancy, the data packets sent by the common node to the cluster head node can be sent in groups. The transport layer uses the NACK packets for end-to-end packets loss recovery.

3.1 Data Packets Grouping Phase

The GCDT protocol first groups packets sent between nodes in each round [9], where the first three steps are performed at the sending node and the latter two steps at the receiving node. The procedure is as follows:

(1) The original data packets are divided into M blocks, and then re-assembled into N groups. N is the number of encryption algorithms that can be supported by the sensor node and is encrypted using a different encryption algorithm;

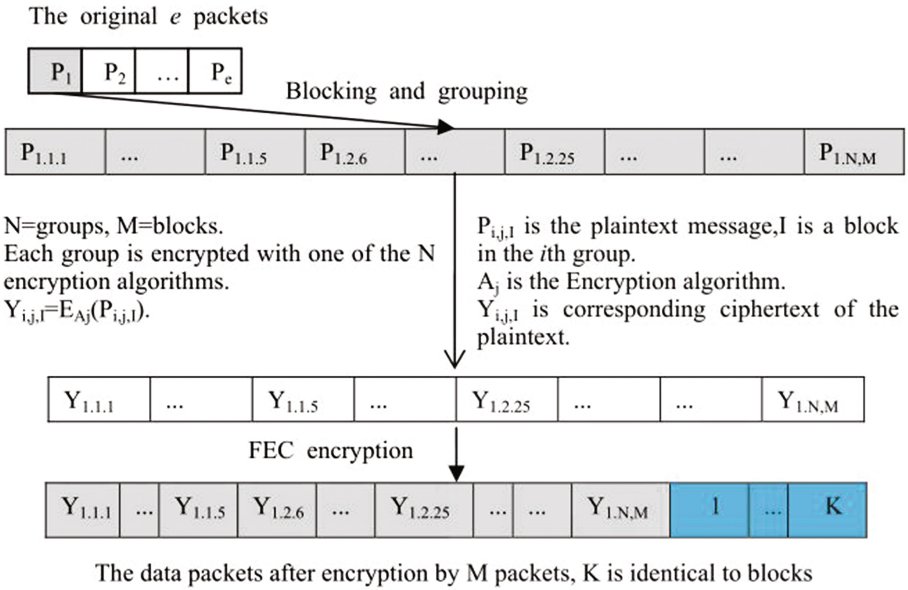

Fig. 1. GCDT protocol grouping process

(2) The data packets made up of M slices can be encrypted into M+K blocks using FEC error-correcting code before transmission, as shown in Fig. 1. It can be seen that only $M < K < M + K$, the packet can be used to reconstruct the encrypted data packet. The use of FEC error-correction codes allows the receiver to reconstruct the original packet even if several packets are lost during transmission;
(3) Packets encrypted with different encryption algorithms can be sent to the destination through multiple independent paths;
(4) When the receiver receives more than M encrypted blocks, and through the reliability of detection, the receiver will rebuild the packet, so you can save some of the energy and computing resources. The original packet can be obtained by decryption;
(5) If more than a certain period of time, or the last packet is not doing security and integrity testing, the data transmission system will re-request the transmitter to send a group desired by the receiver.

The pseudo-code for message receiving is described in Fig. 2.

3.2 Multi-paths Establishment Phase

According to the DBCH clustering algorithm, all the cluster heads of the network are generated, and then several transmission paths are established between the cluster head and the base station through the algorithm. Figure 3 is packet

```
MsgHandle(msg){
 switch(msg->type){
   case Layer_Msg:
UpdateParents(msg);
     forward(msg);
     break;
case DATA:
     forward(msg);
     break;
  case Request_Msg:
     UpdateChildrens(msg);
     if(IsMychild){
       UpdateT(msg);
       SendNotify();
     }break;
  case Confirm_Msg:
    UpdateParents(msg);
    break;
  case Notify_Msg:
    UpdateChildrens(msg);
    if(IsMychild){
       UpdateT(msg);
       SendNotify();
     }break;
   )
}
```

Fig. 2. Message receiving part of the pseudo-code

format used for the establishment of multi-paths in the GCDT protocol between cluster head and base stations.

After the multi-path establishment between the cluster head and the base station and the data packet format is completed, the optimal path is selected for each data transmission, so as to avoid redundant data. The path generation algorithm is implemented in Fig. 4.

Type	Ows	Source_node	Path_Msg	Sig_Msg	Destination_Node

(a) Path request packet

Path_Msg	N_1	N_2	N_3	...

(b) Path information

Sig_Msg	S_1	S_2	S_3	...

(c) Signal strength information

Type	Ows	Source_Node	Forwarding_Table	Destination_Node

(d) Path feedback packet

Fig. 3. Packet format

```
Procedure Path Building
 Begin
  (1) Enter all the cluster head nodes into node V set,input one
node
to (2);
  (2) Calculates the number of all neighbor nodes of theinput
node,referred to as n;
  (3) If the number of neighbor nodes equals to 0, continue (4):
 Else repeat (3) .
End
  (4) The neighbor nodes are input to neigh V one by one;
Then do
Begin
 (4.1) While there is node vi do
If the node vi is the same as the parent node of the input node,
delete it;
  Else If energy of vi has been excessive consumed,delete it;
  Else If vi equals to Base node;
   Then outputs the path,de•lete the base node from
  neigh_V.
   End
 End
End
 (4.2) If neigh_V=0&&node_V=0, end the search;
Else turn to  (1) ;
If neigh V≠0;
  Then the next node is taken from the set, repeat (2) .
End
End
```

Fig. 4. The steps of Path generation algorithm

4 Protocol Simulation and Result Analysis

4.1 Simulation Environment

In order to evaluate the performance of GCDT protocol, 100 nodes were set up by MATLAB7.0 to simulate ARRIVE, INSENS, HEED and GCDT under the same condition. The nodes were randomly generated in 500 m × 500 m scene, and the energy consumption is respectively $1 \times 10 - 5$ J, $5 \times 10 - 5$ J and $1 \times 10 - 4$ J

Table 1. MATLAB network simulation parameters tables.

Parameter	Parameter description	Value
E_0	Initial energy of each node	50 J
l	Number of bits transmitted	20 bit
$d1_1$	Distance threshold	32 m
d	The distance between the transmitter and the receiver	18 m
Packet	Size Size of packet	10 bytes
n	Total number of nodes	100

for sensing, receiving and transmitting 1 byte data, and the other simulation parameters are shown in Table 1 [14].

4.2 Simulation Results Analysis

Figure 5 shows 100 nodes distributed randomly in the set area using MATLAB 7.0. Each node has the same initial energy E_0. Figure 6 shows the simulation of the selected 20 cluster head nodes (circle part) by the DBCH clustering algorithm.

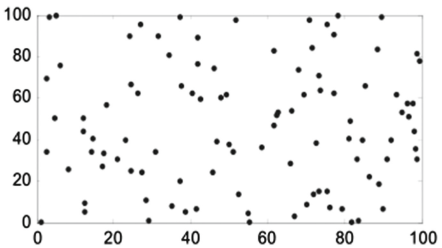

Fig. 5. Sensor node distribution diagram

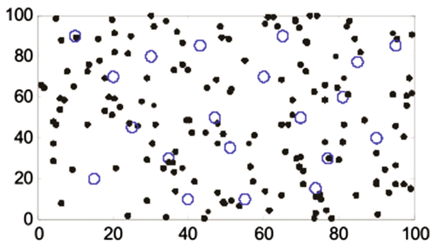

Fig. 6. Cluster heads distribution of DBCH algorithm

In the simulation experiment, the number of packets transmitted in each cluster is counted, and a performance index equivalent to the network throughput is obtained. Figure 7 shows the network throughput of three algorithms, it can be seen, GCDT protocol network throughput is greater than ARRIVE and INSENS. Compared with ARRIVE and INSENS, the network throughput of GCDT is increased by about 18% and about 15% respectively when the data transmission rate is high (packet generation rate is more than 1 packet/s), which means that the effect of data aggregation is better after GCDT protocol grouping, and makes the network throughput had improved significantly.

We can get the packet loss rate after subtracting the total number of received packets from the total number of transmitted packets, divide by the total number

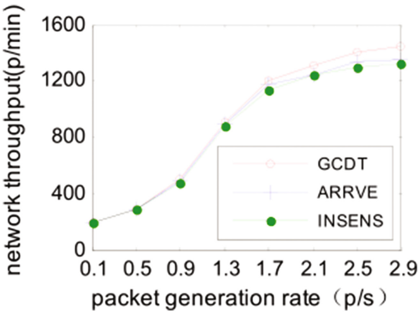

Fig. 7. Comparison of network throughput

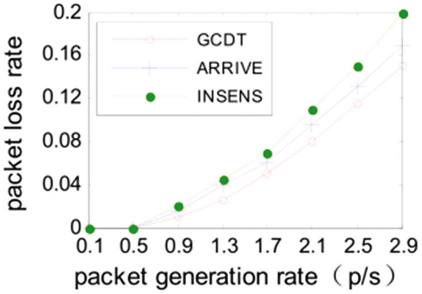

Fig. 8. Comparison of packet loss rate

of transmitted packets. Figure 8 shows the packet loss rate at different packet generation rates in each cluster [15]. It is calculated that, compared to ARRIVE and INSENS, the GCDT packet loss rate is reduced respectively by about 28% and about 15%, because the GCDT protocol determines the success rate according to the importance of the packet, and then determines the need to send the packet number of packets, reducing the probability of packet loss, better reliability

Figure 9 compares the effect of different paths on the rate of network adaptation under different node failure rate. GCDT protocol reduces the packet loss rate within the cluster, builds multi-path of energy more balanced, shares network load equally, without any external interference. The results show that with the increase of node failure rate, the network adaptation rate of GCDT decreases the most slowly, and the algorithm is more practical.

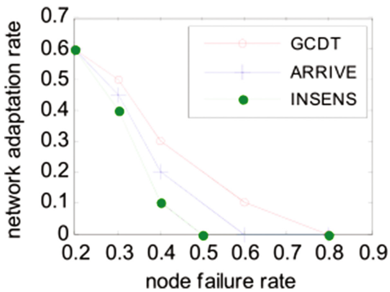

Fig. 9. Comparison of network adaptation rate

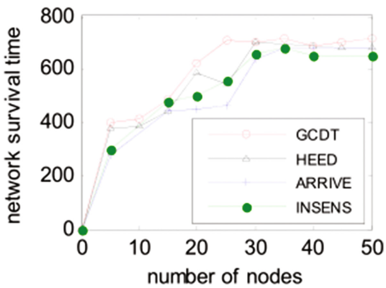

Fig. 10. Comparison of network survival time

Figure 10 shows the network survival time comparison of each algorithm [16]. Compared with the ARRIVE flooding method, which consumes a lot of traffic and much resources, the GCDT protocol establishes multi-path generation algorithm between the cluster head and the base station in the stable phase, and the cluster information can be reliably transmitted to the base station by multi-path. The multi-hop communication between the cluster head and the base station can save the energy of the cluster head, save the node's memory resources, thus prolong the network' s survival time.

5 Conclusion

In wireless sensor networks, the reliability of data transmission is the foundation of network performance. In order to achieve reliable transmission, this paper proposes a grouping-based wireless sensor network clustering data transmission protocol (GCDT), establishes a number of transmission paths before data transmission, access to the expected success rate of transmission and other parameters, determines the number of packet grouping, and then through the GCDT protocol for each round of the node to send the original packet between the packet, and finally these packets through multiple independent paths to send to the receiving end. Then grouping for the original data packets sent by nodes of each round through the GCDT protocol, and finally these groups are sent to the receiver through a number of independent paths. Through the GCDT protocol for each round of the node to send the original data packets are grouped, and finally these packets through a number of independent path to send to the receiver. Simulation results show that the GCDT protocol can effectively guarantee the communication reliability and improve the overall performance of the network compared with the existing survivable routing ARRIVE and INSENS.

Acknowledgement. The authors would like to thank our colluges and students in Anhui University of Science and Technology. This work is supported by the National Natural Science Foundation of China (Grant no. 61402012), and the National Natural Science Foundation of China (Grant no. 61572034).

References

1. Heinzelman, W.R., Balakrishnan, C.A.: Energy efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Annual Hawaii International Conference on System Science, Hawaii, pp. 10–15 (2000)
2. Huangfu, W., Zhang, Z.S., Chai, X.M., Long, K.: Survivability-oriented optimal node density for randomly deployed wireless sensor networks. Sci. China Inform. Sci. **57**(2), 1–6 (2014)
3. Karlof, C., Li, Y., Polastre, J.: Arrive: Algorithm for Robust Rounting in Volatile Environments. T R UCBCSD-03-1233, Computer Science Dept, Univ of California at Berkeley (2003)
4. Braginsky, D., Eslytin, D.: Rumor rounting algorithm for wireless sensor networks. In: The First ACM International Workshop on Wireless Sensor Networks and Applications (WSNA 2002) (2002)
5. Deng, J., Han, R., Mishra, S.: Intrusion tolerance and anti-traffic analysis strategies in wireless sensor networks. In: IEEE International Conference on Dependable Systems and Networks(DSN 2004), Florence, Italy (2004)

6. Younis, O., Fahmy, S.: HEED. A hybrid, energy-efficient distributed clustering approach for adhoc sensor networks. IEEE Trans. Mobile Comput. **3**(10), 366–379 (2004)
7. Arumugam, G.S., Ponnuchamy, T.: EE-LEACH: development of energy-efficient LEACH Protocol for data gathering in WSN. Wireless Communications Networking, 76 (2015)
8. Papadimitratos, P., Haas, Z.: Secure data transmission in mobile Ad Hoc networks. In: ACM Workshop on Wireless Security (2003)
9. Cao, J., Chen, Y., Ren, Z., Li, Q.: Multi-round clustering based multi-hop clustering routing protocol for wireless sensor networks. Comput. Sci. **40**(7), 67–70 (2013)
10. Sharma, R., Mishra, N., Srivastava, S.: A proposed energy efficient distance based cluster head (DBCH) Algorithm: An Improvement over LEACH. Procedia Comput. Sci. **57**, 807–814 (2015)
11. Deb, B., Bhatnagar, S., Nath, B.: Reinform: reliable information forwarding using multiple paths in sensor networks. In: Proceedings of IEEELCN03, Bonn, Germany, pp. 406–415 (2003)
12. Kumar, N., Ghanshyam, C., Sharma, A.K.: Effect of multi-path fading model on T-ANT clustering protocol for WSN. Wireless Netw. **21**, 1155–1162 (2015)
13. Wan, C.-Y., Campbell, A.T., Krishnamurthy, L.: PSFQ: a reliable transport protocol for wireless sensor networks. In: Proceedings of ACM WSNA 2002, Atlanta, GA, pp. 1–11, September 2002
14. Mahadevaswamy, U.B., Shanmukhaswamy, M.N.: Delay aware and load balanced multipath routing in wireless sensor networks. Int. J. Wireless Inform. Networks **19**, 278–285 (2012)
15. Sun, B., Gui, C., Song, Y., Chen, H.: A novel network coding and multi-path routing approach for wireless sensor network. Wireless Pers. Commun. **77**, 87–99 (2014)
16. Luo, J., Di, W., Pan, C., Zha, J.: Optimal energy strategy for node selection and data Relay in WSN-based IoT. Mobile Netw. Appl. **20**, 169–180 (2015)