A Data Fusion Algorithm Based on Improved LEACH Clustering Routing for Wireless Sensor Network

Shiyue Zhang[1], Jiande Wu[1,2], Xiaodong Wang[1]
[1]Faculty of Information Engineering and Automation,
Kunming University of Science and Technology,
Kunming, China
[2]Engineering Research Center for Mineral Pipeline Transportation YN,
Kunming, China
E-mail: zhangshinee@163.com, wjiande@gmail.com

***Abstract*—Energy limited of wireless sensor network limits the practical application. Therefore, in industrial applications we need to take into account introducing data fusion to reduce the data redundancy, improve performance, reduce energy consumption, and prolong the network life. This research idea: data collected by sensoring nodes is preprocessed firstly, and then based on the LEACH cluster routing algorithm we design one kind of improved algorithm for wireless sensor network data fusion, thereby reducing the amount of data communication and the energy consumption. Improved LEACH algorithm can reduce more energy consumption compared with the typical algorithm and has certain superiority.**

Keywords: Wireless Sensor Network; Improved LEACH; Data Fusion; SOM neural network

I. INTRODUCTION

A wireless sensor network (WSN) is a research hotspot after entering twenty-first century and has many applications [1]. An amount of WSN nodes are deployed randomly in the monitor area around the object fleetly, self-organized and flexibly, and they apperceive the object collaboratively and can adapt to bad environments, nodes are deployed in the monitoring region randomly. Perception, collection, processing and transitional information and data will be sent to the Sink node via multi-hop network, and then transmitted to the remote control center for monitoring personnel to view conveniently.

Although WSNs have many application prospects[2], there are also many constrains, where energy consumption is the most important thing. Since the distribution of nodes is close to each other, information obtained from the same object usually has a lot of correlations, the data obtained exists a lot of redundancy. Literature [3] has pointed out that the amount of power consumed by the radio transceiver is far greater than the power consumption in the sleep state. The devices, namely nodes, powered by batteries have a short lifetime. They can survival for 5 months to 1 year if they are not used frequently, but they will be Premature failure if they need to send and receive messages frequently. Literature [4] has pointed out that the energy consumption of 1 byte data transmission is equal to the consumption of thousands of CPU instruction in less time.

II. DATA FUSION ALGORITHM IN WIRELESS SENSOR NETWORK

WSNs usually use multi-sensors to collect information. However, increasing the number of nodes will lead to increasing amount of data and nodes' communication consumption. The data fusion algorithms existed solve the data redundancy problems [5, 6], but exist some problems more or less, such as reducing the bandwidth instead of reducing energy consumption. Some algorithms reduce the amount of data by increasing the number of data pretreatment [7,8], and process the data by using fuzzy theory of each sensor measurement data which have a close degree, measure the comprehensive approaching degree, fusion weight of distribution of data of sensor data through proximity matrix, then draw characteristic index. Similar to the common method such as weighted average, Kalman filtering, fuzzy logic neural network method and so on. These methods need a pair of data to be processed, or train the nodes and distribute weight threshold information in the beginning stage of network established by Sink nodes, the implementation is relatively complex. Some discuss the clustering fusion methods [9, 10], they analyze cluster data fusion algorithm error causes and put forward cluster data fusion after cluster measurement correction. Data fusion protocol based on the cluster has the advantages of convenient management, reducing the coordination complexity, suitable for large-scale network, but maintenance of the cluster costs more. Some of the algorithms [11, 12] consider of data security and data fusion in two factors can not only ensure the reduction of redundant data, and the data security. Literature [11] proposed ESPART algorithm, can reduce network communication data traffic based on data fusion tree structure characteristics, and can be used in other wireless sensor network data fusion privacy preserving algorithms. Literature [12] mainly analysis through the reputation of nodes, undertake quantitative credibility analysis using the research on the problem of uncertainty of set pair analysis theory to the sensor node.

The data fusion discussed above are located on the host computer. The improved LEACH data fusion algorithm put forward in this paper will preprocess the data when collected and then transfer the data to the cluster head, which can reduce the data traffic and the burden of the Sink node.

III. DATA FUSION BASED ON IMPROVED LEACH CLUSTERING ROUTING ALGORITHM

A. *Classical LEACH Clustering Routing Algorithm*

LEACH is the name of Low Energy Adaptive Clustering Hierarchy. As a result of sending, receiving and handling large amounts of data, a node will consume lots of energy, which can lead to premature failure of node, in particular those nodes closer to Sink nodes, which power consumption is bigger than further ones, and more susceptible to death. The idea of the LEACH algorithm is to constitute a cluster structure in each round with random selection of cluster head nodes, so the energy load is assigned to each cluster head node, the entire network performance is improved. Each round is divided into two stages: cluster establishment and data transmission stable stage. The general stability of the phase duration is longer than the create nodes in order to save energy.

The typical LEACH exists some problems, such as no provision for cluster head number and the nodes distributed on average throughout the network, thus likely lead to the congestion of election of cluster head node, and part of the nodes are isolated with no cluster head nodes around. Without considering the residual energy of the nodes, if some nodes are frequently elected to be cluster head, it may result in premature death of the node.

B. *Improved LEACH clustering routing algorithm*

Aiming at the existing problems of typical routing protocols, this paper made some improvement. We can get the node coordinates with node localization algorithm when deployed of wireless sensor network at first. In case of knowing the geographical situation of all the nodes, divided the nodes into several areas according to the coordinate value of the node partition, the size and node numbers of each area equals each other roughly, and the parameters are set by management artificial. Each region according to the number of nodes and the size of the region or other practical requirements can select one cluster head or more, algorithm introduced in this paper selects one node as the cluster head.

The steps of improved LEACH data fusion algorithm clustering are listed as follows.

Step 1 The network is established at first, each node deployed in the object area computes its location, and then sent to the Sink node. Based on the position information of nodes routing search range can be limited in certain areas, in order to reduce the overhead of algorithm. Network topology and routing table are established according to the location information.

Step 2 Sink node broadcasts information to other nodes to campaign cluster head, according to the node energy, node with a maximum residual energy can be selected as the cluster head node in each region.

Step 3 If the number of nodes' residual energy are the same, we select a cluster head node according to the distance to the Sink node.

Step 4 The ordinary nodes will be added to the cluster of its region after cluster head selection is completed.

Step 5 Enter the stable stage after establishment accomplished of the cluster.

Step 6 Data collection and transmission after data fusion processing.

Step 7 A new round of cluster building and stable phase will start again after a period of time.

First of all, we should select some cluster heads using equation below. We consider both the distance between node and the Sink node and the residual energy, the probability for a node which residual energy is lower than average or the spacing distance is further to be a Sink node is lower than others.

The nodes entering the stable stage are required a certain treatment, it is the SOM neural network training mentioned later in this article, after training the node can process the data collected, convert the data into a small amount of state identification, which is stored into the local memory, and then the state identification data will be uploaded to the cluster head periodically, the cluster head node will transmit the data processed to the Sink node after a certain treatment.

$$T(n)=\begin{cases}\dfrac{p}{1-p\times[r\bmod(1/p)]}\times\dfrac{E_{i\_current}}{E_0}\times(1-\dfrac{D_i}{D_{\max}}), n\in G \\ 0, n\in Others\end{cases} \quad (1)$$

The nodes deployed in the target region apperceive object and collect data regularly, using neural network to convert the initial data into a corresponding state identification. Because a plurality of nodes supervises the same object, the data collected by neighbor nodes are associated with each other. In the same moment data collected form neighbor nodes consist a matrix A.

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (2)$$

Each column of matrix data are stored in a local memory of nodes, it is measurements of one node in the moment of φ; each row represents data collected by different sensor at the same time. Because the data have great similarity, the algorithm in this paper sent the data after preprocessing by neural network fusion. a_{1j} is the first state data of node j. A data collected should be compared with the $i-1$ data before.

Transmission data rules: when each column of data's difference between any two identifies exceeds the threshold value δ, or acquisition times more than T, the front $i-1$ data will be transferred to cluster head node, data collected now is remembered as i -th data.

If the difference of data received by cluster head node this time and front time is greater than the threshold δ, it should forward data to the Sink node immediately, otherwise accumulate receive data frequency, if the frequency number is greater than a predetermined threshold T0, then upload the saved data to the Sink node.

The data accessed is still stored in a matrix B. Row data is moment θ, columns are state identification uploaded to the cluster head accessed from each sensor at θ time.

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{i1} & b_{i2} & \dots & b_{in} \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (3)$$

In finally, we simulate both LEACH and improved LEACH proposed in this paper using MATLAB. The performance comparison of LEACH and improved LEACH is shown in Figure 1. From this figure we can see that improved LEACH data fusion algorithm can survive longer than LEACH. The processed data periodically uploaded to the cluster head, or when the data deviation is bigger than predetermined threshold. Then cluster head make further judgment and data integration; finally send them to the Sink node through multiple hops. The performance comparison of LEACH and improved LEACH is shown in Figure 1. The simulation parameters are listed in table 1.

TABLE I. THE PERFORMANCE COMPARISION OF LEACH AND IMPROVED LEACH

The number of nodes	100
Length[m]* Width[m]	10*10
BS Coordinate	（50,175）
k	5
l [bit]	4000
Initial Energy[J]	0.2
d_0 [m]	87.7058
E_{elec} [nJ/bit]	50
E_{DA} [nJ/bit/signal]	5
ε_{fs} [pJ/bit/m^2]	10
ε_{mp} [pJ/bit/m^2]	0.0013

From this figure we can see the first dead node of classical LEACH occurs in about 460th round, whenever of the improved LEACH occurs in about 540th round. What is more, all the nodes are dead in about 700th round in classical LEACH, when the nodes in improved LEACH are dead in about 840th round. The figure illustrates us that improved LEACH data fusion algorithm proposed im this paper can save the energy of each node, and then prolong the life time of network.

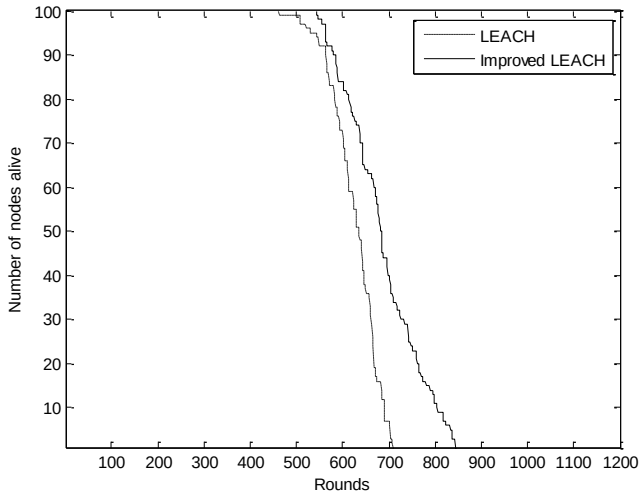

Fig. 1 The performance comparison of LEACH and improved LEACH

IV. SUMMARY

This paper presents an improved LEACH clustering routing data. The vast majority of cases of the wireless sensor network applications, the data collected is not the ultimate information needed, through preprocessing the different data representations for a smaller identification data to save, it can save the memory while reducing the burden on the energy consumption of the wireless fusion algorithm, it can improve network vitality. But the algorithm still has some problems; it is not applied to an actual system to verify the communication of the entire network, which is also our future research direction.

ACKNOWLEDGMENTS

*(Jiande Wu) is corresponding author. This work is supported by the National Natural Science Founder of China (No. 51169007), Science & Research Program of Yunnan province (No.2011CI017 & No.2011DA005 & 2011FZ036 & 2012CA022).

REFERENCES

[1] Akyildiz I.F., Su W., Sankarasubramaniam Y., Cayirci, E.: Wireless Sensor Networks: A Survey[J]. Computer Networks, Elsevier, vol. 38(2002), p.393-422.

[2] Machado, R., Ansari, N., Wang, G., Tekinay, S.: Adaptive Density Control in Heterogeneous Wireless Sensor Networks with and without Power Management[J]. IET Commun. , Vol 4, (7)(2010), p. 758-767.

[3] D.F. Larios, J. Barbancho, G. Rodrı´guez, et a.l. Energy Efficient Wireless Sensor Network Communications based on Computational Intelligent Data Fusion for Environmental monitoring[J]. IET Commun. , Vol. 6(2012), Iss. 14, pp. 2189-2197.

[4] Min R, Bhardwaj M, Seong-hwan Cho, et a.l. Energy-Centric Enabling Technologies for Wireless Sensor Networks[J]. IEEE Wireless Communications, Vol 9(4)(2002), p.28-39.

[5] Huang Q, Liu C, Guo C. Reliable Aggregation Routing for Wireless Sensor Networks based on Game Theory[C], Int. Conf. on Innovative Technologies, Prague, Czech Republic, InTech (2010). ISBN 978-953-307-132-9.

[6] Guo Y., Hong F., Guo Z., Jin Z., Feng Y.: EDA: Event-oriented Data Aggregation in Sensor Networks. IEEE 28th Int Performance

Conf.[C]. On Computing and Communications (IPCCC), (2009), ISSN 1097-2641, p. 25-32.

[7] Jiao Zhuqing, Xiong Weili, Zhang Lin, Xu Baoguo. Multi-sensor Data Fusion Method based on Belief Degree and Its Applications[J]. Journal of Southeast University (Natural Science Edition), Vol 38(2008), A01: p.253-257.

[8] Lee M F R, Stanley K, Wu Q M J. Implementation of Sensor Selection and Fusion Using Fuzzy Logic[C]. IFSA World Congress and 20th NAFIPS International Conference. Vancouver, BC:(2001), p.328-333.

[9] Martalo M, Ferrari G. A Simple Information-theoretic Analysis of Clustered Sensor Networks with Decentralized Detection[J]. IEEE Communications Letters, Vol 14(6)(2010), p.560-562.

[10] Manjunatha P, Verma A, Srividya K. Multi-sensor Data Fusion in Cluster Based Wireless Sensor Networks Using Fuzzy Logic Method[C]. IEEE Region 10 and the third International Conference on Indus-trial and Information Systems(ICIIS), (2008), p.1-6.

[11] Yang Geng, Wang An-qi, Chen Zheng-yu, Xu Jian, Wang Hai-yong. An Energy-saving Privacy-preserving Data Aggregation Algorithm[J]. Chinese journal of computers, Nanjing, Vol 34(5)(2011), p.792-800.

[12] Ma Shou-ming , Wang Ru-chuan , Ye Ning. Secure Data Aggregation Algorithm Based on Reputations Set Pair Analysis in Wireless Sensor Networks [J]. Journal of Computer Research and Development, Vol 49(9)(2011), p.1652-1658.