A K-medoids based Clustering Algorithm for Wireless Sensor Networks

Jin Wang, Kai Wang, Junming Niu, Wei Liu

School of Information Engineering, Yangzhou University, Yangzhou , China

jinwang@yzu.edu.cn, kennwong99@163.com, njm_2016@163.com, yzliuwei@126.com

***Abstract*—Wireless sensor networks (WSNs) have become a hot research topic in recent years, due to their wide potential applications such as military surveillance, industrial and agricultural monitoring, smart home with multimedia sensors to collect visual data like image or video. Generally, energy consumption is a challenging research issue for WSNs since the tiny sensors cannot be easily recharged after random deployment. In this paper, an efficient energy saving scheme is proposed to reduce energy consumption and to prolong network lifetime. One of the popular scenarios to reduce energy consumption for WSNs is clustering technology. The main idea of this method is to reduce the communication distance of sensor nodes by using clustering technique. To get the ideal result of cluster, we use K-medoids algorithm to calculate the optimal medoids between sensor nodes. Then, we can select the appropriate cluster heads. By balancing network load among the clusters, energy efficiency can be improved and network lifetime can get extended effectively. Simulation results show the outperformance of our proposed algorithm than other popular algorithms in terms of energy consumption and network lifetime.**

Keywords—wireless sensor networks; clustering; K-medoids; energy consumption; multimedia

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are composed of large number of tiny and low-cost sensor nodes. These sensors are self-organized, which can form a multi-hop network adaptively and transmit the compressed data to base station [1, 2]. With the development of WSNs, multimedia wireless sensor networks have been widely applied. Transmitting Multimedia data like image or video is the characteristic of multimedia wireless sensor networks, and more visual information can be collected in military surveillance, industrial and agricultural monitoring, healthcare and smart home etc.

As we all know, the size of multimedia data is normally large and sensor nodes may consume more energy. Some scholars try to study and solve this problem in [3, 4]. Power saving is one of the most important factor for sensor nodes to extend their lifetime in wireless sensor networks. Most energy was consumed by data packets transmitting and receiving. In wireless sensor networks, the sensor node gets widespread use of battery. Due to a large number of sensor nodes, battery cannot be recharged easily, and the capacity becomes the most valuable resource for each sensor node. Thus, the energy saving becomes a vital concern in wireless sensor networks. In order to maximize the network lifetime and increase energy efficiency, new optimal algorithm must be developed.

In this paper, we proposed a K-medoids based Clustering Algorithm (KCA) to provide efficient energy saving scheme for wireless sensor networks. In order to get the ideal result of cluster, we use K-medoids algorithm to avoid negative influence of the outliers and then calculate the optimal medoids between sensor nodes. By balancing the network workload, our proposed scheme can improve the energy efficiency and extend network lifetime effectively.

We organize the rest of this paper as follows. Section II introduce several classic algorithms of wireless sensor networks. Section III presents the relevant system model. Section IV illustrate our proposed scheme in detail. Section V present simulation result of our proposed scheme. Section VI concludes this paper.

II. RELATED WORK

In order to optimize the network topology, improve the energy utilization rate and reduce the energy consumption of sensor nodes, many scholars have carried out a lot of researches on WSNs routing algorithm, and gained some achievements [5, 6, 7]. These researches can be divided into two categories: tree-based routing algorithm and cluster-based routing algorithm. In tree-based routing algorithms, construction of optimal routing tree is a difficult problem. And cluster-based routing algorithms have the advantages of efficient node management and low node collaboration, while the cost of the maintenance of all clusters is very large.

Low Energy Adaptive Clustering Hierarchy (LEACH) is a classical cluster-based algorithm. In LEACH, all sensor nodes are divided into several clusters. Each cluster has one cluster head node selected by probability. Other common sensor nodes choose the nearest cluster head node to form clusters. The common sensor nodes monitor the environment and transmit data directly to their cluster head node. Cluster head nodes aggregate all data collected from common sensor nodes and transmit compressed data to BS. LEACH can reduce energy consumption in WSNs. However, the random selection of cluster head nodes may obtain a poor clustering result. An uneven distribution of cluster head nodes will break the network load balance and decrease network efficiency.

LEACH-centralized (LEACH-C) is proposed to improve LEACH algorithm by using a centralized clustering algorithm. In LEACH-C, BS receives the data of position and residual energy from all sensor nodes. Using the data, BS calculates the number of cluster and set-up network. In [8], a hybrid, energy-efficient, distributed (HEED) clustering algorithm is proposed

978-1-5386-2615-3/18/$31.00 ©2018 IEEE

to prolong network lifetime and support scalable data aggregation. In HEED, the cluster head nodes are selected based on their residual energy probabilistically and sensor nodes choose cluster according to their residual energy. The saving energy clustering algorithm (SECA) in [9] is proposed to provide efficient energy consumption for sensor nodes and to prolong the lifetime of wireless sensor networks. The uniform cluster is created by the centralized method in SECA. The distance from common sensor nodes and cluster head nodes is suitable [10]. Therefore, the energy consumption is reduced and the network lifetime is extended. However, the centralized method has some drawbacks. The method of selecting cluster head nodes will lead to the ignorance of outliers.

In consideration of these disadvantages above, this paper proposes an K-medoids based clustering algorithm to provide efficient energy consumption in wireless sensor networks. In order to reduce the sensitivity of outliers, our proposed scheme uses K-medoids algorithm to optimize the selection of cluster head nodes and set up a more suitable network topology. I this paper, we compare our proposed scheme with some cluster-based routing algorithms such as LEACH-C, HEED and SECA

III. System Model

A. Network Model

The system infrastructure is composed of a large number of sensor nodes and one base station (BS). All sensor nodes are divided into two categories. The one is common nodes and the other is cluster head nodes. The task of common nodes is monitoring environment information and transmitting sensing data to cluster head node. The cluster head node is systematically selected from common nodes and receives data from common nodes, aggregates these data and transmits to BS.

B. Energy Model

The first order radio model is used as energy model. We only consider the energy consumption in communication phase. The total energy consumption consists of the energy dissipation generated by data transmission, receiving and aggregation. In this model, a common node and a cluster head node exchange L-bit data, the energy consumption can be calculated by Equation 1.

$$E_{Tx}(L,d) = E_{elec} \times L + \mathcal{E}_{amp} \times L \tag{1}$$

$$E_{Rx}(L) = E_{elec} \times L \tag{2}$$

where d is the distance from transmission node to receiving node, $E_{Tx}(L,d)$ is the energy consumption during the phase of transmitting L-bit packet, and $E_{Rx}(L)$ is the energy consumption during the phase of receiving. E_{elec} is the electronics energy consumption per bit in transmitter and receiver sensor nodes. ε_{amp} is the amplifier energy consumption during the phase of transmission, which can be calculated by Equation 3.

$$\mathcal{E}_{amp} = \begin{cases} \mathcal{E}_{fs} \times d^2, \ when\ d \le d_0 \\ \mathcal{E}_{mp} \times d^4, when\ d > d_0 \end{cases} \tag{3}$$

where d_0 is a threshold valuse about the transmission model of sensor node. If distance d is less than or equal to d_0, the sensor node will use free-space propagation model. On the contrary, the multipath fading channel model is used. ε_{fs} and ε_{mp} are communication energy parameters. d_0 is calculated by Equation 4.

$$d_0 = \sqrt{\frac{\mathcal{E}_{fs}}{\mathcal{E}_{mp}}} \tag{4}$$

The initial energy of sensor node is set as $E_{init} = 2\,J$, and the aggregative energy of a cluster head node is set as $E_{DA} = 5\ nJ/bit/signal$.

IV. Our Propoesd Scheme

In order to reduce energy consumption and extend network lifetime, we proposed a K-medoids based Clustering Algorithm (KCA) to get a universal clustering method. In our proposed scheme, we use base station to gather all nodes sensing data via cluster head nodes. Firstly, we harvest the node coordinate and residual energy information, and then calculate k of cluster number. We optimize the K-medoids algorithm to reduce the iteration time by calculating central circle mean points and residual energy. In our proposed scheme, the algorithm contains two phases: set-up phase and communication phase.

A. Set-up Phase

The set-up phase is to divide all the nodes into appropriate cluster and find cluster head nodes. During this phase, we calculate the number of clusters, select initial cluster head nodes and gain the final clustering result. Firstly, we let k be the number of clusters, and k can be calculated by Equation 5.

$$k = \sqrt{\frac{N}{2}} \tag{5}$$

where N represent the number of all nodes. Next, we calculate the initial mean points. Let O be the center location for all nodes. O can be calculated by Equation 6.

$$O = \frac{\sum_{i=1}^{N} X_i}{N} \tag{6}$$

where X_i represents the coordinate of sensor *i.*

Let R be the average distance between O and all sensor nodes. R can be calculated by Equation 7.

$$R = \frac{\sum_{i=1}^{N} |X_i - O|}{N} \tag{7}$$

According to O and R, a central circle can be calculated. We select k points on central circle uniformly as the initial mean point I_i, it can be calculated by Equation 8.

$$\begin{cases} I_{i_x} = R \times cos\left(\frac{360}{k} \times (i-1) \times \frac{\pi}{180}\right) + O_x \\ I_{i_y} = R \times sin\left(\frac{360}{k} \times (i-1) \times \frac{\pi}{180}\right) + O_y \end{cases} \tag{8}$$

where I_{i_x} and I_{i_y} represents the mean point coordinate of cluster *i*. k is the number of clusters.

According to the K-medoids algorithm, the initial mean points are selected randomly, so we always get inferior results that two or more mean points are very close, and it result in a large amount of iteration time. These steps above are very important, because we get the uniform initial mean points by calculating central circle.

We use K-medoids algorithm to divide all sensor nodes into k clusters. Different from K-means algorithm, K-medoids algorithm uses real node to represent cluster head node rather than theoretical mean point. Thus, the partition method is based on minimizing the distance between common node and represent node. Exactly, the absolute-error criterion E can be represented by Equation 9.

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} dist(x, r_i) \tag{9}$$

where x is common node in cluster C_i, and C_i represents cluster i. E represents the sum of distance between all common nodes and represent nodes. The main idea of K-medoids is to minimize E, however, it is a *NP*-hard problem. Therefore, a certain greedy method should be used to solve this problem.

Let $I = \{ I_1, I_2, \dots, I_j, \dots, I_{k-1}, I_k \}$ as the set of cluster head nodes. We randomly select a common node I_{random} to replace cluster head node I_j, and the residual energy of this random node must be higher than average residual energy of all nodes. The replace method shows in Equation 10.

$$I^{(t+1)} = \begin{cases} I^* & , E^* - E^{(t)} < 0 \\ I^{(t)} & , otherwise \end{cases} \tag{10}$$

where $I^* = \{ I_1, I_2, \dots, I_{random}, \dots, I_{k-1}, I_k \}$, the network should be redistricted into k clusters temporarily. According to Equation 9 and new clusters, the new absolute-error criterion E^* can be calculated. Compared with original absolute-error criterion $E^{(t)}$ of clusters during t iteration, if E^* is less than $E^{(t)}$, $I^{(t+1)}$ will be replaced by I^*. Otherwise, nothing will be done, and the next iteration will be started. The flow chart of our proposed KCA is shown in Fig. 1

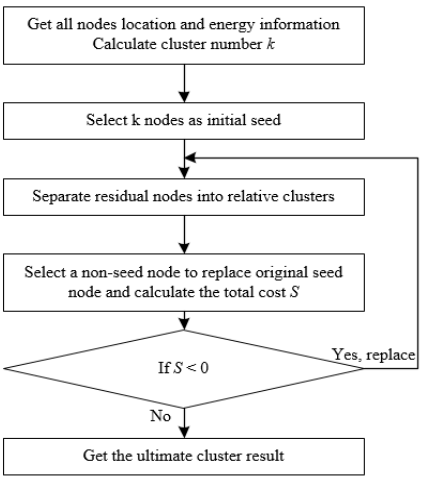

Fig. 1. Flow chart of our proposed KCA

During iteration time, the residual energy of cluster head nodes must be considered. At the beginning of one iteration, if the residual energy of a cluster head node is less than average residual energy of all sensor nodes, this cluster head node must be replaced. By executing Equation 10 repeatedly, we can get the ultimate clustering result.

B. Communication Phase

The BS receives the ultimate clustering data and sends them to all sensor nodes, then real network architecture will be set up. The common sensor nodes transmit data to relative cluster head node, and the cluster head node condenses all received data into a single signal in order to send sensing data to BS effectively. When BS receives all signal of cluster head nodes, this round is done, and the next round starts from set-up phase to communication phase repeatedly.

V. Performance Evaluation

In this section, we compare our proposed scheme with HEED, LEACH-C and SECA. To ensure the accuracy of the experimental result, the size of sensing area is $100 \times 100\ \mathrm{m}^2$, the initial energy of all nodes is $2J$ and these sensor nodes are deployed randomly in sensing area. The parameters for our experimental are listed in Table 1.

TABLE 1. Network Parameters

Parameter	Definition	Value
E_{elec}	Energy consumption on circuit	$50\, nJ/bit$
$\mathcal{E}_{fs}$	Free space model of transmitter amplifier	$10\, pJ/bit\ /m^2$
$\mathcal{E}_{mp}$	Multi-path model of transmitter amplifier	$0.0013\, pJ/bit\ /m^4$
E_{DA}	Energy for data aggregation	$5\, nJ/bit\ /signal$
E_{init}	Initial energy of node	$2J$
N	Number of nodes	50
Pk_size	Packet size	2000 *bits*
BS_coor	Coordinate of BS	(50,180)

The simulation environment is composed of a BS and 50 sensor nodes. The BS is fixed and located far from the sensing area. All sensor nodes are randomly deployed in sensing area. The common sensor node can transmit data to its cluster head node. The cluster head node can collect and compress data, transmit data to BS.

As we can see in Fig.2, the total energy of different cluster routing algorithms is recorded. The residual energy of our proposed scheme KCA is higher than that of SECA, LEACH-C and HEED schemes. Because of combining the advantage of calculate initial mean points in SECA and reduce the sensitivity of outliers, our proposed scheme has the superior performance. The data transmission distance from each common sensor nodes to its cluster head nodes is minimized. Thus, less energy is consumed and more lifetime is prolonged.

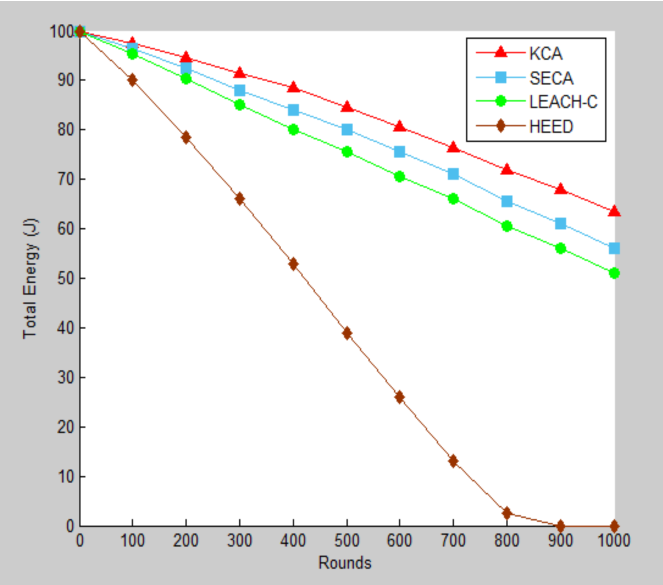

Fig. 2. Residual energy comparison

The number of alive nodes is shown in Fig.3. As we can see, the node begins to die around 600 rounds in HEED algorithm. After about 800 rounds, the number of alive nodes begins to die in LEACH-C algorithm and the number in SECA is about 900. However, the first dead node appears until about 1100 rounds in our proposed scheme KCA. As can be seen from different survival number of node in different wireless sensor networks, the survival time of sensor nodes in KCA scheme is the longest. Because of using K-medoids algorithm to reduce sensitivity of outliers, our proposed scheme can minimize absolute-error criterion and achieve optimal clustering result. These strategies can effectively reduce the energy consumption. So, the conclusion we have got is KCA can improve the wireless sensor network performance very efficiently.

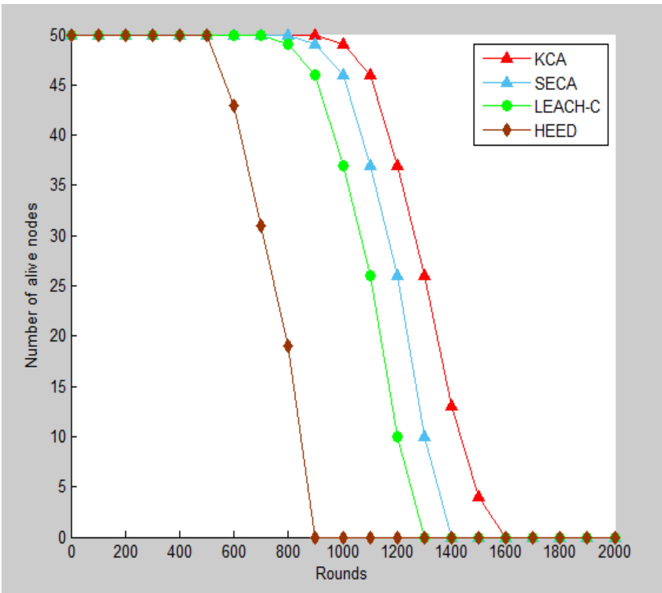

Fig. 3. Number of alive nodes

VI. Conclusion

The size of multimedia sensing data is large. So we need compress these data as form of image or video, and we also need provide efficient routing algorithm to reduce energy consumption as much as possible [11, 12]. In order to reduce the amount of data transmission and improve the performance of wireless sensor networks, we proposed a K-medoids based clustering algorithm (KCA) in this paper. According to the experimental results and analysis, it can be concluded that our proposed scheme KCA can reduce the energy consumption and can prolong the network lifetime than other three popular algorithms.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, et al. Wireless sensor networks: a survey, Computer Networks, 2002, pp.393–422.

[2] W. Wang, S. Zhang, G. Duan, et al. Security in wireless sensor networks, IEEE Wireless Communications, 2008, vol. 15, pp.60-6

[3] Budhewar A S, Thool R C. Improving performance analysis of multimedia Wireless sensor network: A survey[C]// International Conference on Advances in Computing, Communications and Informatics. IEEE, 2015:1441-1448.

[4] Jeličić V, Bilas V. Reducing power consumption of image transmission over IEEE 802.15.4/ZigBee sensor network[C]// Instrumentation and Measurement Technology Conference. IEEE, 2010:1211-1215.

[5] L. Villas, A. Boukerche, H. S. R. Filho, et al. DRINA: A Lightweight and Reliable Routing Approach for In-Network Aggregation in Wireless Sensor Networks [J]. IEEE Transactions on Computers, 2013, 62(4):676-689.

[6] Z. Yang, M. R. Chen, W. Wu. Algorithm for Wireless Sensor Network Data Fusion Based on Radial Basis Function Neural Networks [J]. Applied Mechanics & Materials, 2014, 577(577):873-878.

[7] A. Kalis, A.G. Kanatas, G P. Efthymoglou. A co-operative beam forming solution for eliminating multi-hop communications in wireless sensor networks [J]. Communications IEEE Journal on Selected Areas, 2010, 28(7):1055-1062.

[8] O Younis, S Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans Mob Comput. 3(4), 366–379 (2004).

[9] J.Y. Chang and P.H. Ju, An efficient cluster-based power saving scheme for wireless sensor networks", EURASIP Journal on Wireless Communications and Networking, May 2012.

[10] A. W. Khan, A. H. Abdullah, M. A. Razzaque, et al. VGDRA: A Virtual Grid-Based Dynamic Routes Adjustment Scheme for Mobile Sink-Based Wireless Sensor Networks[J]. IEEE Sensors Journal, 2015, 15(1):526-534.

[11] Kakoyiannis C G, Stamatiou G, Constantinou P. Small square meander-line antennas with reduced ground plane size for multimedia WSN nodes[C]// European Conference on Antennas and Propagation. IEEE Xplore, 2009:2411-2415.

[12] Pham C, Lecuire V, Moureaux J M. Performances of multi-hops image transmissions on IEEE 802.15.4 Wireless Sensor Networks for surveillance applications[C]// IEEE, International Conference on Wireless and Mobile Computing, NETWORKING and Communications. IEEE, 2013:477-484.