

# Lifetime Maximization Strategy for Wireless Sensor Network Using Cluster-Based Method

Xiaoping Ma, Honghui Dong, Limin Jia and Ruhao Zhao

## 1 Introduction

As the development of the wireless sensor and communication technology, the Wireless Sensor Networks (WSNs) are applied widely to detect and estimate the service condition of the equipment and devices. However, the energy capacity, processing ability and communication bandwidth of the tiny sensor nodes are all limited, in order to make full use of the limited energy and guarantee the stability and reliability of the wireless sensor network, it is necessary to design an energy-efficiency protocol to extend the lifetime of the system.

There are various wireless sensor networks adopting the cluster-based protocol to cope with energy-efficiency issues [1, 2]. In the cluster-based structures, all the sensor nodes are divided into several clusters, and select one sensor node in each cluster as the cluster head (CH). The CH received the sensor information of others in the cluster and then transmitted to the base station (BS). The CH sensor nodes will consume large more energy than No-CH nodes. The clustering and CHs selection schemes as the most important factors had become the focus of study. As one of the effective cluster-based methods, The CHs selection strategy proposed in LEACH [3] and LEACH-C [4] is based on the probabilistic model, which just consider whether the node had been selected as the CH in recent rounds and the sensors are divided into several clusters uniformly while ignoring the location information of the sensors. In the practical monitoring field, the sensors are deployed unevenly, the distances among the sensor are different from each other. To reduce the energy waste caused by the unreasonable cluster solutions, K-means [5, 6] is used to gather the sensors close to each other into one group and save the energy as more as possible. However, the selection of the initial CHs will affect the

---

X. Ma · H. Dong · L. Jia (✉) · R. Zhao  
State Key Lab of Rail Traffic Control & Safety,  
Beijing Jiaotong University, Beijing 100044, China  
e-mail: jialm@vip.sina.com

optimization solutions mostly, to avoid the local optimal result, the K-means ++ is adopted in this paper [7]. Furthermore, the number of clusters will affect the total energy consumption of the system, which is determined by the scales of the sensors and the distances to the base station, the optimization method is used to calculate the optimal number of the clusters [8]. The lifetime of the system is determined by the communicate rounds of the first sensor which fails because of energy exhaust, the balance of the energy consumption is another very important factor to be considered in this paper [9, 10].

The structure of the sensor network is shown in Fig. 1, it is consisted of sensor layer and communication layer. In the sensor layer, there are a large number of sensor nodes in charge of sensing the service condition information, while the communication layer ensuring the information could be sent to the base station successfully. The energy will be consumed in the communication process mostly.

In this paper, we proposed an effective method to optimal the clusters to minimize the energy consumption. First, we calculate the optimal number of clusters to decline the total energy consumption. Second, we divided the sensor nodes into K clusters base on K-means ++, to save the energy consumption from the No-CH nodes to CH node. The strategy proposed in this paper will extend the lifetime of the network greatly.

## 2 Sensor Network Energy Consumption Model

To meet energy-efficiency demands in sensor network, based on the random distribution of the sensor nodes in the monitoring region, we adopt the K-means ++ strategy to divide the nodes close to each other into the same cluster. It is benefitted to minimize the energy consumption.

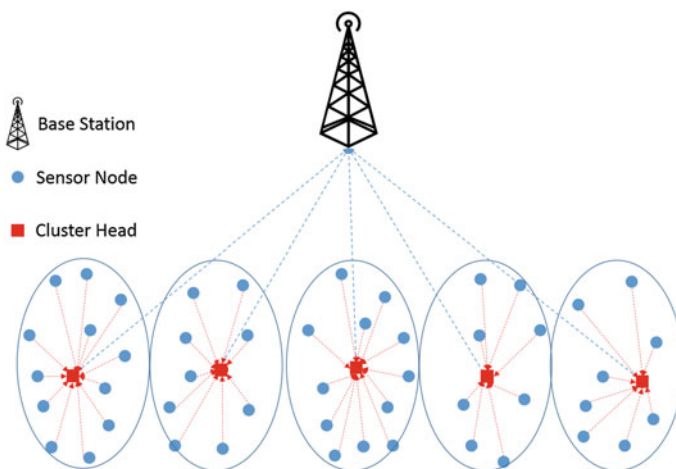


Fig. 1 The communication structure of the sensor network

There are lots of sensors deployed in the monitoring area to detect the condition. The sensor nodes are all with sufficient transmission power to communicate with the nearest base station directly. A large quantity of energy will be wasted in this direct transmission model because the transmission energy consumption is proportional to the square or four powers of the distance between source and destination nodes. The energy consumption for  $l$ -bit data is shown in Eq. 1:

$$E = E_R + E_D + E_T \quad (1)$$

where  $E_R$  is the energy consumption by received  $l$ -bit data, it is defined as:

$$E_R = L * E_{ele} \quad (2)$$

where  $E_{ele}$  is the energy consumption to process  $l$ -bit data, and  $E_D$  is the energy consumption to aggregate  $l$ -bit data, which is defined as,

$$E_D = l * E_{DA} \quad (3)$$

where  $E_{DA}$  represents the energy consumption in  $l$ -bit data aggregation. And  $E_T$  is the energy consumption to transmit  $l$ -bit data, which is defined as:

$$E_T = \begin{cases} l * E_{ele} + l * \zeta_{fs} * d^2, & d < d_0 \\ l * E_{ele} + l * \zeta_{mp} * d^4, & d > d_0 \end{cases} \quad (4)$$

where  $\zeta_{fs}$  and  $\zeta_{mp}$  are the energy coefficient related to the distance between the source and the destination. The reference distance  $d_0$  is defined as:

$$d_0 = \sqrt{\frac{\zeta_{fs}^2}{\zeta_{mp}}} \quad (5)$$

The Eq. (4) shows that the transmission energy consumption go hand in hand with the distance, which imply that the information transmitted from the sensor nodes to the base station will waste larger number of energy, it is a wise way to select one sensor node to complete the mission on behalf of the cluster. In this paper, we divided the sensor nodes into several clusters, and then select one node in each cluster as the cluster head (CH), other sensor nodes as the cluster member called Non-CH nodes sent their information to the CHs, and then transmitted to the base station.

### 3 The Optimal Number of Clusters

As shown in last section, the energy consumption is mostly related to the distance, declining the valid transmitting distance is an effective way [5, 6]. Dividing the sensor nodes close to each other into the same group will be benefit to the energy efficiency of the nodes. As an effective strategy to generate the clusters based on the distance, K-means ++ would complete the objective in the paper. However, before this we should calculate the number of the clusters  $K$ .

The optimal number of the clusters  $k$  is the prerequisite to design the cluster-based protocols based on K-means ++, and the energy-efficiency is the fine criterion. We assume that the base station is located at the ordinate origin  $B(0, 0)$ , there are  $N$  sensor nodes deployed in the monitoring region  $M \times M$  following uniform distribution  $S(x, y)$ , where  $x \in [-M/2, M/2]$  and  $y \in [-(1 + \alpha)M, -\alpha M]$ , which means that nearest distance between the sensor node to the base station is  $\alpha M$  as shown in Fig. 2.

The sensor nodes would be given two different identities (CH nodes and Non-CH nodes), and the energy consumption of the sensor system is also generated by the two parts [7]. The Non-CH nodes just transmit their information to CH nodes, the transmission energy consumption follows the free-space model because of the small distance among them. As shown in Eq. (5):

$$E_{Non-CH} = l * E_{ele} + l * \xi_{fs} * d_{toCH}^2 \tag{6}$$

The CHs in charge of the data reception, aggregation and transmission, the energy consumed by CHs is correlated to the above parts. Due to the BS located far away from the sensor nodes, the multipath model is adopted in the transmission

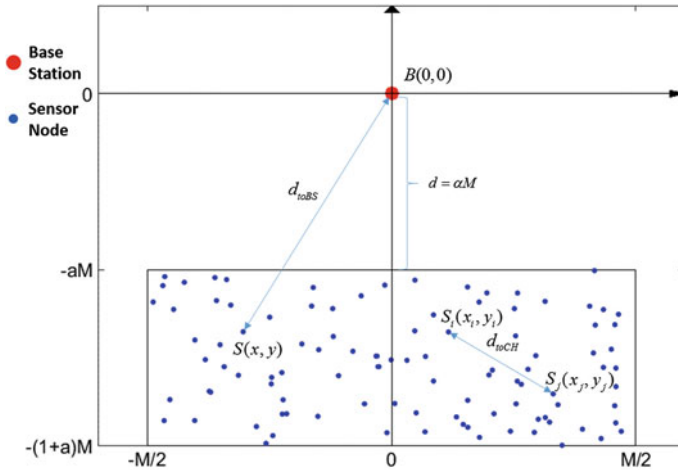


Fig. 2 The cluster structure of the sensor network

energy consumption, we assume that all the sensor nodes are divided into  $k$  clusters equally, and there are  $n = N/k$  nodes in each cluster. As shown in Eq. (7):

$$\begin{aligned} E_{CH} &= (n-1) * l * E_{ele} + n * l * E_{DA} + l * E_{ele} + l * \zeta_{mp} * d_{toBS}^4 \\ &= l * (n * E_{ele} + n * E_{DA} + \zeta_{mp} * d_{toBS}^4) \\ &= l * \left( \frac{N}{k} * E_{ele} + \frac{N}{k} * E_{DA} + \zeta_{mp} * d_{toBS}^4 \right) \end{aligned} \quad (7)$$

The energy consumed in one cluster in 1 rounds is calculated as:

$$E_{Cluster} = E_{CH} + \sum_{i=1}^{n-1} E_{Non-CHi} \quad (8)$$

It means in 1 round, there is one sensor node acts as CH while other  $n - 1$  nodes as Non-CH. The total energy consumption in the cluster in the sum of both.

The probability density of the sensor nodes is  $\rho(x, y)$ , where the distribution of  $x$  and  $y$  are all uniformed. The location of the sink node is  $B(0, 0)$ , the distance between the CH and the base station is defined as:

$$d_{toBS} = \sqrt{x^2 + y^2} \quad (9)$$

The expected power distance from the CH to the base station is given as:

$$\begin{aligned} E[d_{toBS}^4] &= \iint (\sqrt{x^2 + y^2})^4 \rho(x, y) dx dy \\ &= \int_{-(\alpha+1)M}^{-\alpha M} \int_{-M/2}^{M/2} (x^2 + y^2)^2 \rho(x, y) dx dy \end{aligned} \quad (10)$$

We assume that the sensor nodes in the area follows a uniform distribution, the probability density is defined as:

$$\rho(x, y) = 1/M^2 \quad (11)$$

And the expectation is calculated as:

$$E[d_{toBS}^4] = (0.0125 + \frac{(\alpha+1)^3 - \alpha^3}{18} + \frac{(\alpha+1)^5 - \alpha^5}{5}) * M^4 \quad (12)$$

The distance from the Non-CH I to CH j is defined as:

$$d_{toCH} = \sqrt{(x(i) - x(j))^2 + (y(i) - y(j))^2} \quad (13)$$

The expected squared distance from the Non-CH to the CH is expressed as:

$$\begin{aligned}
E[d_{wCH}^2] &= \frac{1}{2} \int \int \int \int (\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2})^2 \rho(x_1, x_2, y_1, y_2) dx_1 dx_2 dy_1 dy_2 \\
&= \frac{1}{2} \rho(x_1, x_2, y_1, y_2) * \left( \int \int \int_0^{M/\sqrt{k}} \int_0^{M/\sqrt{k}} (x_1 - x_2)^2 dx_1 dx_2 dy_1 dy_2 \right. \\
&\quad \left. + \int \int \int_0^{M/\sqrt{k}} \int_0^{M/\sqrt{k}} (y_1 - y_2)^2 * dy_1 dy_2 dx_1 dx_2 \right)
\end{aligned} \tag{14}$$

The distribution of the sensor nodes is independent to each other, so the joint probability density function is defined as:

$$\rho(x_1, x_2, y_1, y_2) = \rho(x_1) * \rho(x_2) * \rho(y_1) * \rho(y_2) = 1/(M^4/k^2) \tag{15}$$

And the expectation is calculated as:

$$E[d_{wCH}^2] = \frac{M^2}{6k} \tag{16}$$

And then the expectation of total energy consumption in one node is given as:

$$E_{cluster} = E_{CH} + \left(\frac{N}{k} - 1\right)E_{Non-CH} \approx E_{CH} + \frac{N}{k} * E_{Non-CH} \tag{17}$$

The total energy consumption of all nodes is defined as:

$$\begin{aligned}
E_{total} &= k * E_{cluster} = k * E_{CH} + N * E_{Non-CH} \\
&= l * (2N * E_{ele} + N * E_{DA} + k * \zeta_{mp} * d_{toBS}^4 + N * \zeta_{fs} * d_{toCH}^2)
\end{aligned} \tag{18}$$

The optimal number of clusters is calculated by setting the derivative to zero:

$$k_{opt} = \sqrt{\frac{\zeta_{fs} * N * M^2}{6\zeta_{mp} * d_{toBS}^4}} \tag{19}$$

## 4 Communication Protocol Base on K-Means ++

The energy consumption is related to the information transmission distance, declining the transmission distance is an effective way to decrease the energy consumption and prolong the lifetime of the sensor network. The location of each

node and the distance between the nodes are all different. The cluster-based protocol proposed in this paper based on K-means ++ select the nodes close to each other in one cluster, which could decline the energy consumption of the Non-CH nodes. The initial optimal number of clusters  $k$  is determined by the number of all the sensor nodes and the distance to the sink node as shown in (19).

#### 4.1 The Process of K-Means ++ Based Cluster Strategy

The proposed strategy based on K-means ++ classify the sensor nodes close to each other into the same cluster. The objective function is minimizing the total distance among the Non-CH nodes to the CHs. The cluster generation method based on K-means ++ is carried out as follows.

Assume that there are  $N$  sensor nodes deployed in the monitoring region, and the optimal number of clusters is  $k$ . Due to the uniform distribution of the sensor nodes, in order to minimize the energy consumed during the communication among the Non-CH to CH, ideally, the nodes close to each other should be divided into the same cluster. To avoid obtaining partial optimal solutions only as the cluster strategy adopted K-means in, the K-means ++ method is implemented to initial the clusters as Table 1:

In step 3, in order to avoid the effect of noise, the node with larger but not largest distance will be selected as the new CH. The allocation solution by K-means ++ probably not ideal enough to make full use of the energy. The clusters should be optimized in the following phases.

#### 4.2 The Cluster Head Generation

The CHs (cluster heads) are the most important roles in the wireless network systems. First, the cluster heads as the leaders in their own cluster are in charge of collecting the information from other Non-CH nodes and transmitting forward to the sink nodes, which will consume more energy than others. Second, the clusters

**Table 1** The clusters generation steps

The clusters generated based on K-means ++
Step 1: 1 node is selected from $N$ sensor nodes as the first CH;
Step 2: Calculate the distance to the nearest selected CH for every sensor nodes $D(x)$ ;
Step 3: Select a new sensor node as another new CH. The larger the $D(x)$ the greater the probability to be selected;
Step 4: Repeat the step2 and step3 until $k$ CHs are all selected;
Step 5: The standard K-means algorithm is operated based on the $k$ CHs selected.

generation in this paper by the generic algorithm are mostly based on the conditions of the cluster heads. Third, the rotation of cluster heads is necessary to balance the energy consumption among all nodes. The design of cluster heads selection and rotation criterion is the vital challenges in the monitoring network. At first, we should select the first group of CHs based on the K-means ++ solution, and the nodes nearest to them will be selected as the first-round CHs.

$$CH_{S_{virtual}} = \left( \frac{\sum_{i=1}^n x(i)}{n}, \frac{\sum_{i=1}^n y(i)}{n} \right) \quad (20)$$

where,  $n$  is the sensor number in the cluster. Once the CHs are selected, the system will communicate based on the structure. After these phases, the CHs will be updated and rotated in the new round to balance the energy consumption.

### 4.3 The Cluster Head Rotation

In general, the energy consumed by the CHs is far more than the Non-CH, equality of CHs duty opportunity is conducive to system harmony. However, everyone's situation is different, as we make the selection, each factor should be considered.

To decide on the CH rotation, we should synthesize each kind of factor into consider, the probability on CHs candidacy, the residual energy and the energy consumption prediction. The importance of each factor is various in different IA (Industrial Application), the weighting coefficients are introduced to adjust the impact of the above factors to the selection of CHs. The whole probability to select the CHs is defined as:

$$P = \lambda_1 p_1 + \lambda_2 p_2 + \lambda_3 p_3 \quad (21)$$

where  $\lambda_i$  is the weighting coefficient, the selection and rotation of the CHs based on the probability take all the factors account, the optimal solution is benefitted to generate clusters.  $p_1$  represents the probability on CHs candidacy, which means that the nodes had not been the CHs will be selected as CHs with high probability;  $p_2$  represents the ratio of residual energy of the nodes, which means that the nodes with more residual energy will be selected as the CHs with higher probability;  $p_3$  represents the energy consumption prediction in the following round, the less the energy consumption the more the chance to be selected.

The system lifetime is determined by the sensor who fails at first. CHs rotation scheme could avoid the energy excessive consumption for the CHs and balance the energy consumption among all the sensors to extend the system lifetime.



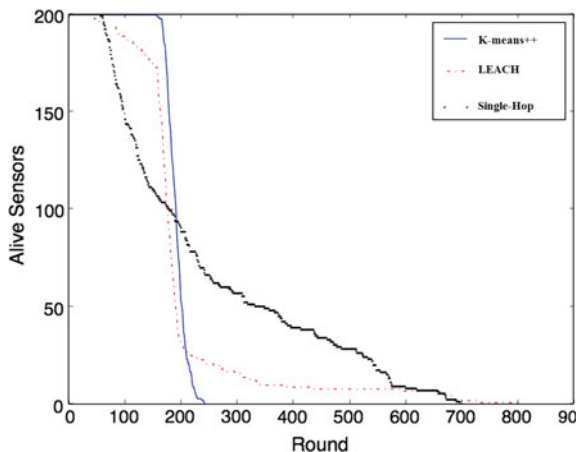
## 5 Performance Evaluation

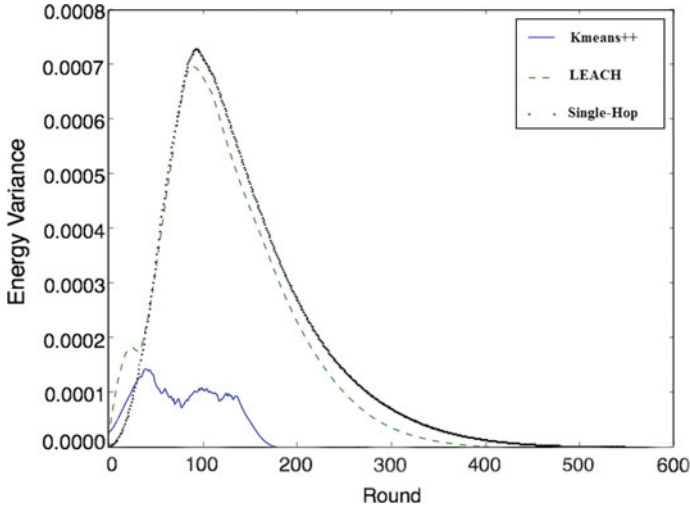
In this section, the validity and reliability of the communication protocols proposed in this paper are tested via computer simulation with Python. We calculate and compare the simulation results with other protocols used widely in the sensor networks.

In our simulation scenarios, there is one BS located at  $(x = 0, y = 0)$  and the sensor nodes are distributed randomly in the rectangular region between  $(x = -50, y = -100)$  and  $(x = 50, y = -200)$ . The base station is in charge of collecting the sensor information in corresponding monitoring region. In the first communication phase, the sensor nodes sent the detecting information to the CHs; in the second phase, the CHs transmit the information to the base station. In the simulation phase, we assume that there are 200 sensor nodes distributed randomly in the separate monitoring area. The optimal initial number of clusters is selected as 6.

We compared the protocol proposed in this paper with other two representative methods LEACH and Single-Hop, with the same simulation initial conditions, same nodes deployment and initial energy. However, the gaps in simulation results are shown in Fig. 3, we can see that almost all the sensor nodes died at about 200 rounds in K-means ++, while in LEACH nodes died to begin at 50 rounds and decreased to 175 at 180 rounds, and then to zero quickly. Even worse in single-hop, the nodes died grandly from 60 rounds, and quickly to 100 about 110 rounds. As we all know that once the nodes are less than one-half of all, the stability of the system is destroyed. Figure 4 reveals that the variance of the residual energy is all fluctuating as the increase of communication rounds. Whereas, the variance of K-means ++ is less than the other two methods which imply that the residual energy is more balance among all sensors. It is benefited to extend the lifetime of the system.

**Fig. 3** The comparison of the alive sensors





**Fig. 4** The comparison of the energy variance

## 6 Conclusions

A lifetime maximum and energy consumption balanced protocol based on K-means ++ is proposed in this paper. The K-means ++ strategy is used to initialize the clusters to accelerate the optimal process and avoid the remaining local optima. Furthermore, the cluster heads are selected and rotated based on the energy information. The simulation results show that comparing to LEACH and Single-Hop, the K-means ++ is remarkably better at prolonging the lifetime and balancing the energy consumption. The two characteristics are of greatest importance to guarantee the stability and continuity of the wireless monitoring system. The accuracy of prediction based on the sensor information and the real-time of signal exchanges will be improved markedly, and then ensure the safety of the system.

**Acknowledgements** This work was supported by the National Science and Technology Major Project Grant No. 2016YFB1200100; State Key Laboratory under Grant No. RCS2016ZT018.

## References

1. Wei D, Jin Y, Vural S, Moessner K, Tafazolli R (2011) An energy—efficient clustering solution for wireless sensor networks. *IEEE Trans Wireless Commun* 10(11):3973–3983
2. Zhang DG, Wang X, Song XD, Zhang T, Zhu Y-N (2015) A new clustering routing method based on PECE for WSN. *EURASIP J Wirel Commun Networking* 2015(1):1–13

3. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocol for wireless micro-sensor networks, in System sciences. In: Proceedings of the 33rd annual Hawaii international conference on, IEEE, 2000, pp 1–10
4. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002) An application-specific protocol architecture for wireless micro-sensor networks. *IEEE Trans Wireless Commun* 1(4):660–670
5. Khan F (2012) An initial seed selection algorithm for k-means clustering of georeferenced data to improve replicability of cluster assignments for mapping application. *Appl Soft Comput* 12(11):3698–3700
6. Bholowalia P, Kumar A (2014) Ebk-means: a clustering technique based on elbow method and k-means in wsn. *Int J Comput Appl* 105(9)
7. Zimichev EA, Kazanskii NL, Serafimovich PG (2014) Spectral-spatial classification with k-means ++ particional clustering. *Computer Optics* 38(2):281–286
8. Liu F, Peng S (2010) On stabilization for stochastic control systems which the coefficient is time-variant. *J Syst Sci Complex* 23(2):270–278
9. Younis O, Fahmy S (2004) Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Trans Mob Comput* 3(4):366–379
10. Zhang H, Shen H (2009) Balancing energy consumption to maximize network lifetime in data-gathering sensor networks. *IEEE Trans Parallel Distrib Syst* 20(10):1526–1539