

**ITC 3/47**

Journal of Information Technology  
and Control  
Vol. 47 / No. 3 / 2018  
pp. 406-418  
DOI 10.5755/j01.itc.47.3.20565  
© Kaunas University of Technology

**Low Energy Consumption and Data Redundancy Approach of  
Wireless Sensor Networks with Bigdata**

Received 2018/04/12

Accepted after revision 2018/04/22

<http://dx.doi.org/10.5755/j01.itc.47.3.20565>

# Low Energy Consumption and Data Redundancy Approach of Wireless Sensor Networks with Bigdata

**Xunli Fan**

School of Information Science & Technology, Northwest University, Xi'an 710127, China, e-mail: xunlfan@nwu.edu.cn

**Wei Wei**

School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China;  
Qilu University of Technology (Shandong Academy of Sciences); Shandong provincial Key Laboratory of Computer  
Network, e-mail: weiwei@xaut.edu.cn

**Marcin Wozniak**

Institute of Mathematics, Silesian University of Technology, Kaszubska 23, 44-100 Gliwice, Poland;  
e-mail: Marcin.Wozniak@polsl.pl

**Ye Li**

Qilu University of Technology (Shandong Academy of Sciences); Shandong provincial Key Laboratory of Computer  
Network

---

**Corresponding author:** WeiWei@xaut.edu.cn

---

To reduce the data redundancy and energy consumption caused by the explosive growth of data in wireless sensor networks, this paper presents a node data image based mean filtering algorithm. In the provided approach, the nodes are divided into active and de-active nodes to eliminate the data of the sleep nodes and data redundancy. The nodes of wireless sensor network are clustered, and the data obtained from the sensor network are modeled. Simulation results show that the proposed algorithm is more effective than that of *k-medoids*, which can reduce the redundant data. Moreover, it reduces the energy consumption, and improves the life cycle of wireless sensor networks.

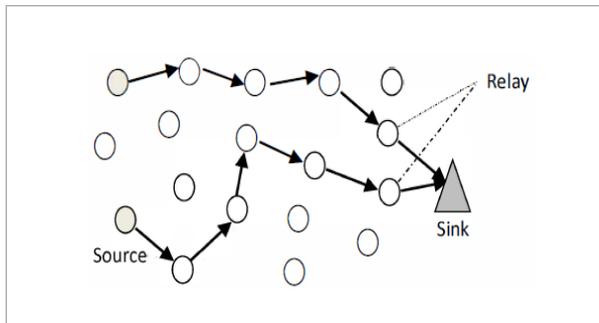
**KEYWORDS:** wireless sensor network, node image, data redundancy, energy consumption, big data.

## 1. Introduction

Wireless sensor network (*WSN*) is the combination of sensor, communication and computer technology. In *WSN* (Fig. 1), a source node provides the required information to the network. With the rapid development of Internet and the «Internet plus» concept, wireless sensor network technology has been developed rapidly again in recent years. More and more sensors come into our life, such as temperature sensor, humidity sensor, gas sensor and so on.

**Figure 1**

Wireless network topology



A large number of sensors are deployed in the monitor area to take a collaborative approach for real-time monitoring, sensing and gathering information in a variety of environment, including digital, text, pictures, voice, video, etc. The data volume with different types of data, low density value, and commercial value is high [3]. The characteristics of sensor network data reflect the key elements of the big data 3V model [18]: capacity (volume), speed (velocity) and various (variety). According to Oracle reports [1], the amount of data generated by a sensor network and related equipment will reach Giga bytes. Smart sensors and network topology are prerequisites to achieve the maximum network lifetime. In this case, any Sensor randomly selects one of its neighbors as the destination node and transmits its data to destination node. The data packets are sending from one node to another until being received at the base station.

The increasing demand for applications in wireless sensor network has made the quality of service an interesting and hot research topic. Quality of service requirements of wireless sensor network raises the significant challenges. While providing quality of service

guarantee, the network protocols need to deal with energy constraints. With the consideration of the properties of sensor networks such as limited energy, dynamic topology, high network density and large scale deployments have posed many challenges in the design, implement, and management of wireless sensor network. These challenges have demanded energy awareness and robust protocol design at all layers of the network protocol [24].

Efficient energy and network lifetime were and still are the main design considerations for the most proposed protocols and algorithms for sensor networks and have dominated most of the research in *WSN*. In fact, such a huge information data generated or transmitted by the nodes will consume large amounts of energy. Also since the sensor node is unstable, it designs the node redundancy to prevent the failure of some nodes. However in data gathering, the redundant nodes will bring a huge amount of redundant data, and the redundant data will consume a lot of network energy and waste data processing time [25, 24, 35] correspondingly. As a result, the sensor network life cycle will be decreases greatly. Faced with these problems, the environment is difficult to achieve effective on-site intervention. Because of energy exhausted, the node is unable to work and leads to the change of the network topology and the deterioration of network performance. Eventually it leads to the splitting of network and paralysis. Therefore, the energy effective problem has become a factor that can not be neglected in the design of wireless communication system. How to deal with the energy consumption and big data processing problem of wireless sensor network has become a new research field.

In recent years, some scholars have carried out the energy saving research in this field. Rios and Diguez describe three important algorithms (*LEACH*, *PEGASIS*, *KOCA*) to reduce the energy consumption of specific implementation process in wireless sensor network [23]. They verified the energy consumption of the above three algorithms and comprehensively compared the advantages and disadvantages respectively. Younis presented a distributed clustering to achieve a hybrid of energy and communication cost named Hybrid Energy Efficient Distributed clustering (*HEED*) [19]. The *HEED* protocol calculates the number of iterations,

independent of the network diameter. Each node individually decides to become cluster head or not by calculating the chance including its residual energy. Wang and Yang establish the sliding window to ensure that the number of cluster heads in each round is always maintained at an optimum range [17]. The sliding window is used to make the interval that generates random number and it can change with the residual energy. The initial energy information of the nodes and the other is the average energy information of those that have not already been cluster heads in the network. The number of living nodes is used to dynamically adjust the number of optimal cluster head in the whole network. The sliding window approach has solved the problem that the number of cluster heads reduces with the increase of the number of rounds. Also the sliding window approach can dynamically adjust the optimal cluster heads number of the whole network, avoiding the problem that too many cluster heads will elect themselves to be cluster heads after some nodes run out of energy.

To improve the performance of *WSN*, raising the sensor nodes' energy efficiency is one of the most important methods. "Hot Spot" and "Energy Hole" are two main problems to be resolved to improve energy efficiency. Based on an energy aware clustering method which is improved from hierarchical agglomerate clustering, the sensor nodes' residual energy would be taken into account in clustering operation, and a special packet head is defined to help update nodes' energy information when transmitting message among the nodes. The network lifetime can be improved by dynamic clustering in which clusters are reformed after every round, which increases the clustering overhead. To optimize the parameters, including clustering overhead, network lifetime, energy hole, the network is divided into layers and clusters of various sizes. The cluster heads are selected based on available energy, the distance to the sink and the number of neighbors [14].

The above algorithms can reduce the energy consumption to a certain extent, but for the big data wireless sensor network, it will be limited. Wu and Tan put forward the facilitate data collection and reduce communication distance by moving the sink node according to the distributed sensor network in large amount of data acquisition [33]. They reduced the energy consumption of information transmission by the mobility of the sink node. However, the algorithm cannot determine the path of the Sink node and the

data quantity before the cluster formation, and the energy consumption of the Sink node is not considered. Li and Liu proposed an energy efficient data gathering algorithm based on regression [11]. In this algorithm, the sensor node set is divided into active nodes and energy efficient nodes. The active node is at the cluster center and uses the regression approach to predict the effective energy during data transferring. However, they did not consider how to reduce the redundancy problem and the amount of transferring data in sensor network. To maximize the network lifetime, Taheri introduces a multi-hop clustering algorithm using the fuzzy logic improvement methods [4]. In this approach, fuzzy logic is used to select the best nodes which they became cluster-head. The cluster head election is based residual energy as primary parameter, node proximity to its neighbors distance to base station and node concentration. A multi-hop method, which is used for inter-cluster and intra-cluster communication distance, and the message should be send through, is divided into several shorter paths. As a result, the multi-hop communication in cluster nodes and between cluster heads reduces the consumption of energy in the network.

The increasing popularity of wireless networks and mobile devices has taken cloud computing to new heights because the data processing capability, data storage capacity, and the lifetime of each device are limited [22]. Big data and its analysis are at the centre of modern science and business. These data are generated from online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science data, sensors and mobile phones and their applications [6]. For data processing of the wireless sensor network, Bodik et al. [39] proposed integration tool for the big data and used Hadoop applications to collect, store and monitor the air pollution with wireless sensor network. Big data is eliciting attention from the academia, government, and industry. Big data is a term utilized to refer to the increase in the volume of data that are difficult to store, process, and analyze through traditional database technologies. The nature of big data is indistinct and involves considerable processes to identify and translate the data into new insights. The characters of big data are: (1) data are numerous, (2) data cannot be categorized into regular relational databases, and (3) data are generated, captured, and processed rapidly. Combined with the research and

practical applications, they analyzed and processed the data collected by the sensor network effectively. However, the implementation process is complicated with some limitations and does not take into account the energy consumption of the big data in wireless sensor network. Bergelt proposed data aggregation strategy for large data sensor network [3], but they did not consider the data redundancy problem which could not effectively extract valuable data information. Because of the limitations of the existing algorithms for wireless sensor networks with large data background, the existing algorithms are not effective in solving the problem of energy consumption and data redundancy. Wireless sensor network is utilized to efficiently deliver the monitoring data of the photo-voltaic (PV) modules from power stations to the monitoring center located in Cloud data center [30, 32, 29]. With the aim of detecting the problems of PV modules from the monitoring big data, a two-class data fusion method is developed to integrate the monitoring data at sensor nodes of WSNs. Then a semi-supervised support vector machine (SVM) classifier is designed and trained by existing solar irradiance big data at the monitor center [31]. To evaluate the performance of the given methods, they setup a comprehensive experimental platform. The experimental results show that the predicted values match well with the theoretical values of power generation [7, 21, 26, 40, 15, 5, 10].

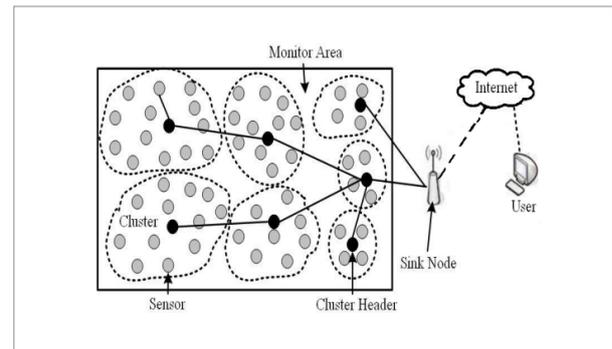
In this paper, we put forward a kind of node data image based mean filtering algorithm. The proposed algorithm models the node data image to find out the redundant sensor nodes; it does not only collect data from the non redundant nodes, and but also collects it from redundant nodes. It predicts the node data without distortion while reducing the sensor network energy consumption [41, 38, 9, 37, 18, 20].

## 2. Network Model and Definition

A set of sensor nodes are randomly deployed in the square field to continuously monitor the phenomenon under inspection as shown in Fig. 2.

Each node is decided to become a cluster head (CH) randomly. Once a node is decided to become a cluster head, it aggregates the data received from various nodes inside the cluster and sends to the base station. However, completely independent random cluster

**Figure 2**  
WSN Architecture



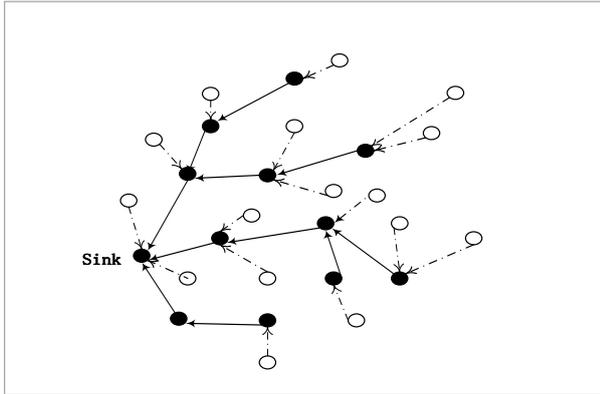
head selection can't guarantee the number nodes in the cluster and the distribution of cluster head in each round. It may select a node which is far away from the base station and has low remaining energy to become the CH, which will cause the uneven energy loss for nodes in the network and form monitoring blind spot, even will influence the network's whole performance [16, 2, 27, 28, 12, 13].

### 2.1. Network Model

A few assumptions are made about the sensor nodes and the network model:

- 1 There are  $N$  sensor nodes distributed in an  $M \times N$  field randomly. Once deployed, each node's location information is determined.
- 2 All nodes are homogeneous. The wireless sensor network nodes are constructed as hierarchy. In a cluster, the nodes are divided into cluster head and ordinary nodes, and the ordinary nodes are in the same level while the cluster head is in the higher level, as shown in Fig. 3. The sensor nodes have the same structure, capabilities of receiving and sending data.
- 3 A sensor node has the same performance such as the initial energy, energy consumption and parameters.
- 4 The sink is outside of the sensor field. It can be reached by all sensors and has sufficient energy. The CH can communicate directly with Sink node.
- 5 CH can perform data aggregation and compression. All nodes' transmission power can be adjusted, and the ordinary node can become a cluster head node. The ordinary nodes are in the same level while the cluster head is in the higher level. The cluster head can communicate directly with Sink node.

**Figure 3**  
Node's location collection process in forward iteration



In wireless transmission, attenuation of sending power decreased exponentially with the increasing transmission distance. Here we add  $E_{DA}$  (the energy consumption of the node data fusion) to energy model. Both the free space ( $d^2$  power loss) and the multipath fading ( $d^4$  power loss) channel models are used, depending on the distance between the transmitter and receiver. The electronics energy,  $E_{elec}$ , depends on factors such as the digital coding, and modulation, whereas the amplifier energy,  $\epsilon_{fs}d^2$  or  $\epsilon_{mp}d^4$ , depends on the transmission distance and the acceptable bit-error rate.

$$E_{TX}(l, d) = \begin{cases} l^*(E_{elec} + E_{DA}) + l^*\epsilon_{fs}*d^2, & d < d_0 \\ l^*(E_{elec} + E_{DA}) + l^*\epsilon_{mp}*d^4, & d \geq d_0 \end{cases} \quad (1)$$

$$E_{RX}(l) = E_{RX-elec}(l) = l^*E_{elec}, \quad (2)$$

where,  $E_{elec}$  is the energy consumption per bit in the transmitter and receiver circuitry;  $\epsilon_{fs}$  is free space model's amplifier energy consumption;  $\epsilon_{mp}$  is multiple attenuation model's amplifier energy consumption;  $d_0$  is a constant which relies on the application environment [8, 34].

According to (1), the distance is the key factor influencing energy consumption because in most WSNs all sensor nodes have the same hardware parameters and the information packet formation would be stable. So decreasing the distance of every hop would conserve the nodes' energy. However, it cannot ensure to prolong the network lifetime just through conserving nodes' energy.

It is assumed that the sensed information is highly correlated, thus the cluster head can always aggregate the data gathered from its members into a single length-fixed packet. This assumption is impractical because the correlation degree of sensed data from different clusters is comparatively low. In this work, relay nodes don't aggregate the incoming packets. We assume that a cluster head consumes  $E_{DA}$  (nJ/bit/signal) amount of energy for data aggregation.

## 2.2. Definitions

Energy efficiency is one of the key factors of improving WSNs' performance and the concept of energy efficiency is introduced through the following two definitions.

**Definition 1. Network lifetime:** The network lifetime is the period from WSNs activated to the failure for some sensor nodes dying for the energy exhausting.

**Definition 2. Energy efficiency:** It is the ratio of network lifetime to the sensor node's energy. If the sensor node's energy is stable, the longer network lifetime, the more energy efficiency. Coverage control algorithms not only require lowest energy consumption in a single monitoring task, but also maintain energy balance of the network in a series of monitoring tasks.

**Definition 3. Number of active nodes:** In the case of meeting the coverage requirements, the fewer number of active nodes are, the larger effective coverage area will be.

**Definition 4. Network scalability:** Coverage control algorithm should be able to adapt to both the scale of different WSNs and the network topology dynamically changed.

In WSNs, the node's energy is consumed to sense, calculate and transmit the message. Among these operations, energy consumption in transmitting is much more than others, and there are many factors influencing energy consumption in transmitting such as node hardware parameters, the length of message and the distance from start node to end node. In addition, in multi-hop WSNs, the inner sensor nodes which near the sink node would forward a lot of message from the outer layer nodes, so these nodes' energy would be exhausted much faster than the outer ones.

In the digital image using mean filter method, the removal of the image of the noise, let  $S_{xy}$  denotes the set

of a set of coordinate center  $(x, y)$  is in the spot, the size of the rectangle window  $M \times M$ , the mean filter in the defined area calculation is noise pollution image  $\hat{f}$  of the average value  $g(x, y)$ . The value of the denoising image obtained at the center point  $(x, y)$  is to calculate the mean value of the pixels in the defined area:

$$\hat{f}(x, y) = \frac{1}{mn} \sum_{(s,t) \in S_y} g(s, t). \tag{3}$$

Through the above description, the mean filter algorithm is the core idea of the field average method. The mathematical meaning is shown in (4):

$$g(x, y) = \frac{1}{M} \sum f(x, y). \tag{4}$$

The idea of the neighborhood average method is to remove the pixels of the pixel by the average of the pixels in the neighborhood, and then filter out the noise. Based on the above analysis, in the image of the node data, there is the same noise, which is generated by the redundant design of the sensor network. In order to remove these noises, a mean value model of the node data image is proposed. Assuming that the node pixel set is a set of pixels in the corresponding digital image, the mean filter algorithm is used to filter out the noise of the node in the set, so as to eliminate the data redundancy.

In the mean value model of the node data image, the mean value of the node data is defined based on (3) and (4). The mean value of the axis data of the nodes in the first cluster is the mean value of the distance between the nodes in the cluster. Then:

$$Ave_x = \frac{1}{M_x} \sum D(x), \tag{5}$$

$$Ave_y = \frac{1}{N_y} \sum D(y). \tag{6}$$

**Definition 5. Node pixel distance:** The relationship between the expression of the Euclidean distance and the location of the points in the two-dimensional image. In the image of the node data, the Euclidean distance is also applicable, so that the distance between the nodes, the node 1 of the coordinates  $(x_1, y_1)$ , the coordinates  $(x_2, y_2)$  of the node 2. The distance of the node is expressed as:

$$d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{7}$$

**Definition 6. Data Circles:** Contains node of the pixel point of circular definitions for data round. (5) and (6) in the value data circle  $C$  is the center of a circle  $(Ave_x, Ave_y)$ , the radius of the circle in the implementation process of algorithm in a dynamic change process, of which there is a critical value. The radius at this point is defined as the radius of the circle that exactly can contain all the ordinary nodes within the cluster.

**Definition 7. Fidelity:** The fidelity is the degree of the data average of the nodes in the data circle. The fidelity of data is distorted through the intuitive expression of the data value. The mathematical meaning is as follows:

$$Pro = \frac{\frac{1}{m} \sum_1^m D(k)}{Ave_k}, Pro \in (0, 1], \tag{8}$$

where  $D(k)$  is the value of the distance between the nodes in the cluster.

### 3. Node Data Image Model Description

#### 3.1. Node Data Image

According to the data obtained by the sensor network node, the two-dimensional image described in the node pixel mode is called a node pixel point. The horizontal and vertical coordinates of the nodes are 1 and 2 of the data obtained at the same time. Assuming that the image  $V(m \times m)$ , which  $S_{(x,y)}$  represents a node,  $x = \{x_i / i = 1, 2, \dots, N\}$  and  $y = \{y_i / i = 1, 2, \dots, N\}$  were the two node in the data set respectively. The data values  $x_i$  and  $y_i$  are collected at different times from the nodes, and  $\begin{cases} 0 \leq x_i \leq m \\ 0 \leq y_i \leq n \end{cases}$ .

Through the above definition, we can get the node data image after image. At this point, the image is in the two kinds of pixels, the pixels of the node and the non node pixels.

#### 3.2. Structural Clustering and Algorithm Description

The key idea of this paper is divided into two stages: the structure of the sensor network and the data processing in the cluster. During the first phase, we need

to consider how best to create each sub cluster. In this section, we will first introduce the principle of cluster construction and how a cluster is clustered. Then, a detailed implementation step of the large data processing in a single cluster is illustrated.

### 3.2.1. Construction and Clustering of Node Clusters

In the first stage, the construction of node cluster, that is how to deploy the sensor nodes in WSN, has a certain impact on the data collected by nodes. It helps to find out the redundant design of the sensor network according to the geographical location of the nodes to build cluster. Through the node's position information, the distance among the nodes is proportional to the actual geographic distance, which can fully reflect the close relationship among the nodes. After the constructing the node cluster, the nodes are divided into clusters using *k-medoids* algorithm. The partitioning method is based on the principle of minimizing the degree of dissimilarity between the  $p$  of all objects and the corresponding representative objects. We use the absolute error criterion to measure the clustering effect which is defined in (9).

$$E(w) = \sum_{i=1}^k \sum_{p \in C_j} |p - o_i|^2, \quad (9)$$

where,  $p$  is the object of cluster  $O_i$ ,  $C_j$  is cluster center,  $O_i$  are multidimensional vector.

The typical clustering algorithm is described as follows:

**Algorithm 1:** *k-medoids* clustering algorithm

Input: The number of input clusters  $k$ , including the  $N$  object data set  $S_n$

Output:  $k$  optimal cluster set  $S_k$

Step 1: Dealing with the data set  $S_n$ , randomly select  $k$  objects  $O_i$ , as the original center point, denoted as  $w$ .

Step 2: Dividing the remaining objects into the cluster that is represented by the nearest center point of the object  $E(w)$ .

Step 3: Selecting a non representative object or instead of the representative object  $O_i$  randomly. Then updating each cluster center, re-clustering into  $w'$  and calculating  $E(w')$ .

Step 4: Calculating  $\Delta E = E(w') - E(w)$ , if  $\Delta E < 0$ , we use  $O_r$  to replace  $O_i$ . In this case,  $w'$  replaces  $w$ .

Step 5: Repeating step 3 to step 4, until the object in the cluster collection does not change.

Algorithm 1 is the traditional *k-medoids* algorithm, which is one of the commonly used algorithms in data processing. In this paper, we construct the node image  $V(m, n)$  of node cluster based on the geographical coordinates of nodes to cluster, and form a  $K$ -cluster of images  $V'(m, n)$ . Analysis and processing of data is based on given algorithm within  $V'(m, n)$ .

### 3.2.2. Implementation Steps of Node Data Image Based Mean Filtering Algorithm

After Cluster construction and cluster completion, the node data image based mean filtering (*NDIMF*) algorithm is used to cluster the nodes during the data processing phase. The key idea of *NDIMF* algorithm is based on the image processing to design the mean filter. The node image data is the core component of the given algorithm. At the start of *NDIMF*, each cluster node image data is used to search the minimum radius  $R$  that is determined according to the elements within a cluster dynamically. In this way, it can guarantee the cluster size differently, and the searching process can cover most of the nodes to ensure that the data achieves the highest fidelity. Once the minimum searching radius  $R$  is determined, the data fidelity of the corresponding data circle is calculated, which is called *Pro*. The searching process that obtains the minimum radius  $R$  of the corresponding data circle will not stop until *Pro* is less than 1 or equal to 1. Through the difference among the circles, it is determined that each node is active or dormant. The specific performance is as follows: If the node is inside the circle, it is active node which provides effective data for the cluster. Otherwise if the node is outside the circle, it is a dormant node, and the data is redundant, which can be in sleep mode. When the data is processed, the data is discarded, and the sensor network does not transmit the node data so as to reduce the energy consumption.

**Algorithm 2:** *NDIMF* Algorithm

The node data set for each cluster is  $S$ .

The number of dormant nodes is  $m$ .

Step 1: Calculate the size of the data set  $(m, n)$ , and drawing the node data image  $V(m, n)$ .

Step 2: Determine the center point  $(Ave_{ix}, Ave_{iy})$  and calculating the distance between the center point and all other points. We select the maximum distance

*Larg\_R*. Check all of the sensors if there are some not controlled by any clusters. If so, then define the nearest cluster to those sensors, goto step 3, otherwise goto step 6.

Step 3: Let the initial defined search radius be  $R$ , we compute the fidelity of data 1 which is  $R\_T\_Pro$ , and the fidelity of data 2 which is  $R\_H\_Pro$ . Update all sensors by locating them to the nearest clusters using minimum Euclidean distance

Step 4: while  $R\_T\_Pro$  or  $R\_H\_Pro$  is greater than or equal to 1

if the distance between points to the center point  $< R$

Calculating  $R\_T\_Pro$  and  $R\_H\_Pro$  in the circle.

end if

end while

Update the cluster by averaging closest sensor locations.

Step 5: Repeats step 4 until  $R\_T\_Pro$  and  $R\_H\_Pro$  meet the requirements or searching the radius to reach the maximum radius. When both  $R\_T\_Pro$  and  $R\_H\_Pro$  are equal to 1, there is no redundant design.

Step 6: Output the number of dormant nodes

### 3.3. Analysis of Energy Consumption

The energy consumption of the introduced new algorithm is calculated step by step:

If the average message length is defined with  $k$  byte packet,  $n$  is the number of nodes and Agent broadcasts *ADVERTISEMENT* message to the nodes in the network area. The consumed energy while the node receiving this message is defined as

$$R = k * n * E_{elec} \quad (10)$$

*REPLY\_ADVERT* message is sent to the Agent when the nodes receive *ADVERTISEMENT* messages. During sending *REPLY\_ADVERT* message, the consumed energy is defined as

$$T = k * E_{elec} + k * n * d_0 * C_{fs} \quad (11)$$

After *CH* runs *NDIMF* and calculates the clusters, it starts backward iteration and broadcasts *ALERT*

message to the nodes. While receiving this message, the consumed energy is defined as

$$R = k * n * E_{elec} \quad (12)$$

After the nodes receive *ALERT* message, they send *REPLY* messages to the *CH*. While sending this message, the consumed energy is defined as

$$T_x = k * E_{elec} + k * n * d_0 * C_{fs} \quad (13)$$

After the *CH* receives *REPLY* message, it sends *INFORMATION* messages back to the nodes. The energy consumption definition during the nodes receive this message is

$$R = k * n * E_{elec} \quad (14)$$

## 4. Experiment Analysis

### 4.1. Experimental Data Set

The experimental data set use Intel Labs Berkeley through the deployment of 50, 100, 200 and 500 sensors respectively to collect real data sets [7]. The parameters are shown in Table 1.

**Table 1**  
Simulation Parameter

Parameter	Value
$E_{elec}$	50nJ/bit
$C_{fs}$	10pJ/(bit·m <sup>2</sup> )
$C_{mp}$	0.0013pJ/(bit·m <sup>4</sup> )
$E_{DA}$	5nJ
Network monitoring area	500×500m <sup>2</sup>
Initial node energy $E_0$	0.5J
Node number	100

*NDIMF* and *k-medoids* are implemented on sensor network and compared with each other. All algorithms are simulated using Matlab 12.0. In each sim-

ulation, the number of 50, 100, 200, and 500 sensors are random located in a  $500 \times 500\text{m}$  square location. The number of total *CH* needed for each algorithm is found for the radius 25, 50, 75, 100m in order to cover all sensors. The proposed algorithm has not only the minimum center number for all radius, but also has the maximum coverage capacity, which means the maximum efficiency. Since there is a large amount of data, the information abundance is rich and suitable for the application of data analysis and processing in *WSNs*. The information obtained from each sensor includes temperature, humidity, light, and voltage. Here we choose the temperature and humidity information for the experimental analysis.

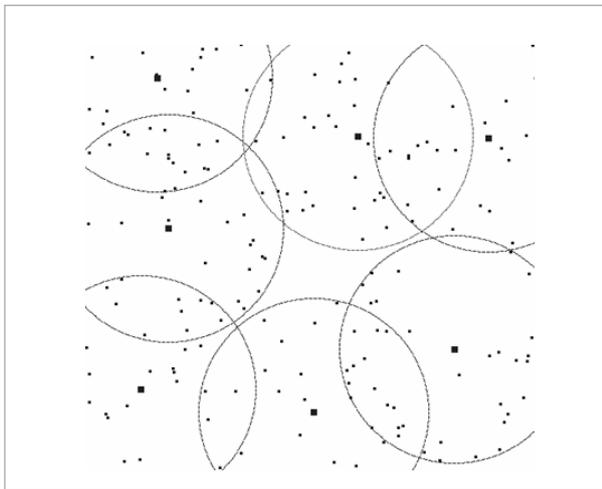
## 4.2. Node Clustering Analysis

Based on Algorithm 1, the nodes are clustered into clusters using *k-medoids* algorithm. The node coordinates described in Fig. 4 are divided into 7 clusters according to the optimal center point.

In Fig. 4, we see that the *k-medoids* algorithm has been divided into 7 clusters, and the number of nodes in each cluster is different. In the next section, we will use Algorithm 2 to analyze and process the data in each cluster.

**Figure 4**

Node coordinates



## 4.3. Experimental Results Analysis

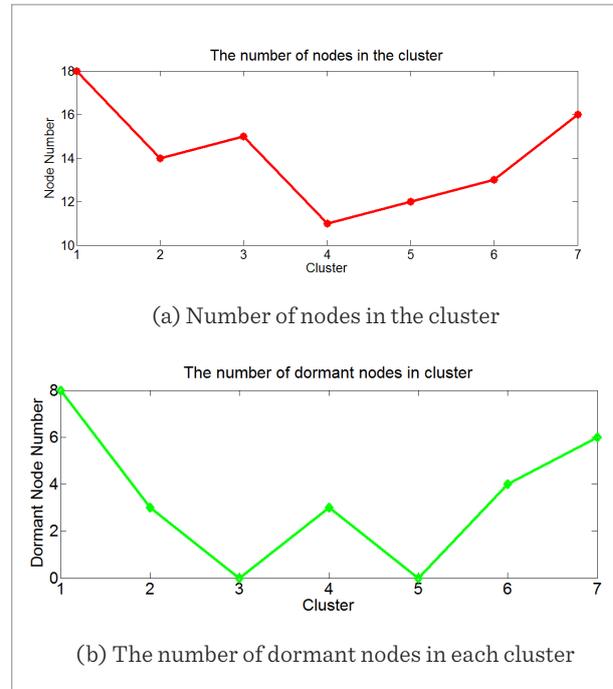
### 4.3.1. Dormant Nodes and Data Fidelity

In this section, we will introduce some simulation results based on *NDIMF* algorithm. In the clustering re-

sults of Fig. 4, the model construction for each cluster node data image. After the model is constructed, the node data image is processed by Algorithm 2. We illustrate 100 nodes and the number of dormant nodes separated from each cluster is clearly described in Fig. 5.

**Figure 5**

The number of nodes in each cluster and the number of dormant nodes



From Fig. 5, it is clear that less the number of nodes in the cluster and dormancy of existence of a linear relationship among nodes, the more complex the node number relative to the redundant nodes and the smaller cluster redundancy, and the less the corresponding node sleep. Through the experiment, we can know that the large data in the sensor network can be established through the model of the data image, and the analysis of Algorithm 2 has achieved good results. The algorithm is fast and accurate to find out the redundant part of the sensor network, which greatly reduces the amount of data to be processed.

Table 2 describes the mean data after each cluster is removed from the sleep nodes.

According to (9), the data fidelity can be very good to the change of the data value. The greater value of the

**Table 2**

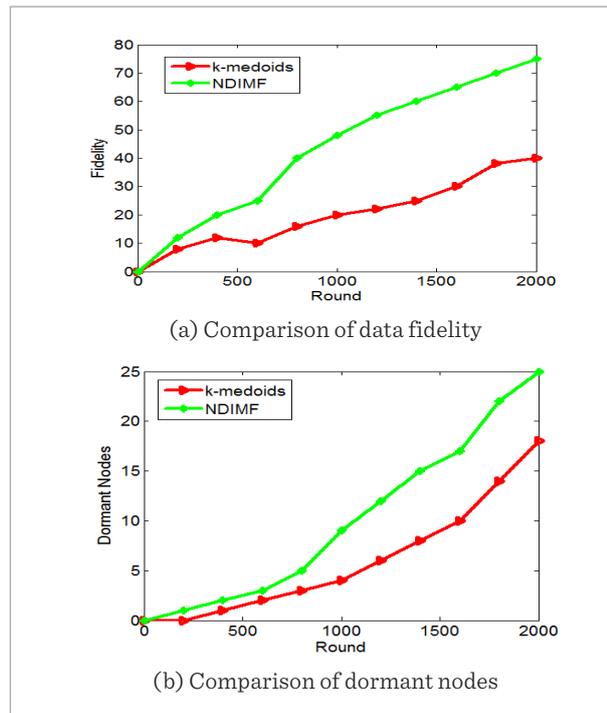
Data fidelity of each cluster after node data image processing

Cluster No	Temperature		<i>R_T_Pro</i>	Humidity		<i>R_H_Pro</i>
	Before	After		Before	After	
1	18.27	18.14	99.43%	40.69	40.48	99.88%
2	19.98	19.78	99.85%	37.82	37.52	99.86%
3	23.86	23.66	99.78%	31.72	31.42	99.68%
4	23.33	23.09	99.71%	31.62	31.46	99.56%
5	22.82	22.62	99.69%	32.78	32.68	99.89%
6	19.96	19.75	99.33%	36.85	36.58	99.42%
7	20.96	20.86	99.63%	34.66	34.46	99.32%

fidelity, the smaller the data change is. In Table 2 we can see that although access to some of the nodes in each cluster of data have been removed, but the entire cluster in the data (temperature and humidity) fidelity remains very high which is shown in Fig. 6, and in the absence of Hugh sleep node in the cluster, the fidelity of the data for 1, algorithm shows that without the original data damage.

**Figure 6**

Comparison of data fidelity and dormant nodes



Through the elimination of dormant nodes in each cluster, the amount of data generated in WSN will also be reduced. We assume that each node in the laboratory generated the same amount of data which is  $D_{Ave}$ .

The total amount of data within the total active number of nodes in the cluster is  $D_{sum}=m*D_{Ave}$ . While the sleep node number is  $n$ , the amount of data generated by the sleeping node is  $D_{sl\_sum}=n*D_{Ave}$ .

$D_{Ave}$  is the data reduction ratio and  $D_{Ave} \in (0,1]$ . Data reduction ratio is described in (15):

$$D\_Rate = \frac{D\_sl\_sum}{D\_sum} \tag{15}$$

Fig. 7 shows the amount of data reduction for each cluster, which shows that the amount of data in a single cluster can be reduced by 67%. This greatly reduces the amount of data needed to be transmitted in the sensor network.

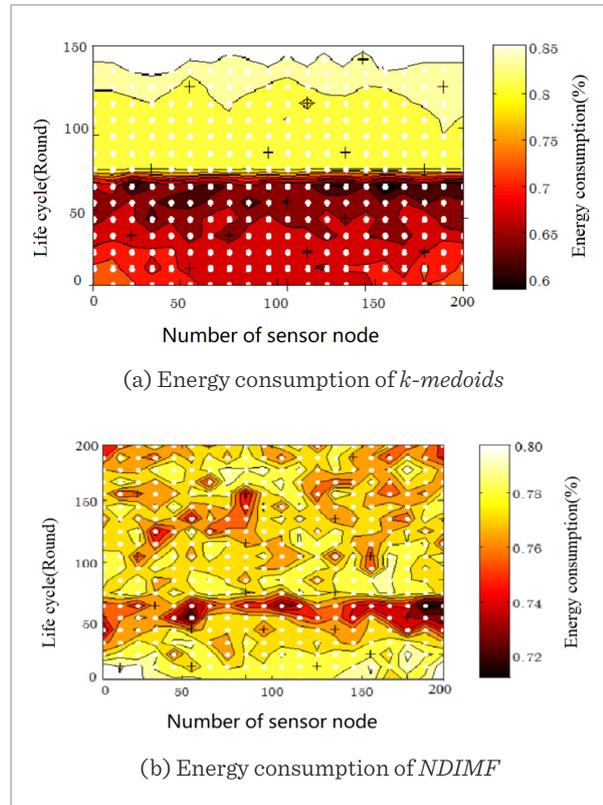
Assuming that the energy consumption of the transmission is  $E$ ,

$$E = Ne(T_{end} - T_{str}), \tag{16}$$

where,  $E$  is the energy consumption of data transmission between two nodes,  $N$  is the number of nodes in each cluster. By (16), the transmission time  $T_{str}$  and the stop transmission time  $T_{end}$  are the same in the start time. The greater the amount of energy required by the  $N$ , the greater the amount of energy  $E$ . In the cluster, it reduces the amount of data of WSN, and the energy consumption can also be reduced at the same time as shown in Fig. 7.

**Figure 7**

Comparison of energy consumption



The redundant data eliminated is transmitted in sense of the life cycle that has been improved. To ensure the structural integrity of *WSN*, it makes the *WSN* more complex for a long time.

## 5. Conclusion

In this paper, the energy consumption of the wireless sensor network with large data environment is

## References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., Cayirci, E. Wireless Sensor Networks: A Survey. *Computer Networks*, 2002, 38(4), 393-422. [https://doi.org/10.1016/S1389-1286\(01\)00302-4](https://doi.org/10.1016/S1389-1286(01)00302-4)
2. Alya, G. B., Shekh, F. A. L. The Direction of Lightweight Ciphers in Mobile Big Data Computing. *Procedia Computer Science*, 2015, (72), 469-476.
3. Bergelt, R., Vodel, M., Hardt, W. Energy Efficient Handling of Big Data in Embedded, Wireless Sensor Networks. 2014 IEEE Symposium of Sensors Applications (SAS), 2014, 53-58. <https://doi.org/10.1109/SAS.2014.6798916>
4. Bodik, P., Hong, W., Guestrin, C. Intel Lab Data. <http://db.csail.mit.edu/labdata/labdata.html>.

studied. To reduce the energy consumption and data redundancy, a mean filter algorithm based on the node data image is proposed, which provides a new direction for solving the problem of large data sensor networks. Comparing with existing algorithms, the provided algorithm can not only solve the problem of generation, transmission of large data which brings huge energy consumption of sensor network, it also can effectively reduce the data redundancy, improve the density of the data value and data post-processing. The provided algorithm is validated through a real data set. The experiment results show that under the premise of attaining high data fidelity, the nodes are effectively divided into active nodes and dormant nodes. The dormant nodes turn into sleep mode, which reduces the redundant data and energy consumption, and increase the life-cycle of wireless sensor network. For the future work, we will study the multi-variable, multi-dimension data model. At the same time, we need to optimize the algorithm in a timely manner to ensure that the algorithm has been in an efficient mode.

## Acknowledgment

This work was supported by Natural Science Foundation of China (61503300) and the Open Research Fund from Shandong provincial Key Laboratory of Computer Network, Grant No.SDKLCN-2017-02 and the Key Research and Development Program of Shaanxi Province(No.2018ZDXM-GY-036). This work is also supported by Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No.2013JK1139). This work is supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No.20136118120010).

5. Fazio, M., Celesti, A., Puliafito, A., Villari, M. Big Data Storage in the Cloud for Smart Environment Monitoring. *Procedia Computer Science*, 2015, 52, 500-506. <https://doi.org/10.1016/j.procs.2015.05.023>
6. Guo, J., Zhao, N., Richard, F. Y., Liu, X., Leung V. C. M. Exploiting Adversarial Jamming Signals for Energy Harvesting in Interference Networks. *IEEE Transactions on Wireless Communications*, 2016, 16(2), 1267-1280. <https://doi.org/10.1109/TWC.2016.2643658>
7. Hashem, I. A. T., Yaqoo, I., Anuar, N. B. The Rise of "Big Data" on Cloud Computing: Review and Open Research Issues. *Information Systems*, 2015, 47, 98-115. <https://doi.org/10.1016/j.is.2014.07.006>
8. Hu, T., Zheng, M. H., Tan, J. J., Zhu, L., Miao, W. Intelligent Photovoltaic Monitoring Based on Solar Irradiance Big Data and Wireless Sensor Networks. *Ad Hoc Networks*, 2015, (35), 127-136. <https://doi.org/10.1016/j.adhoc.2015.07.004>
9. Khan, A. N., Kiah, M. L. M., Ali, M., Madani, S. A., Rehman Khan, A. U., Shamshirband, S. BSS: Block-Based Sharing Scheme for Secure Data Storage Services in Mobile Cloud Environment. *The Journal of Supercomputing*, 2014, 70(2), 946-976. <https://doi.org/10.1007/s11227-014-1269-8>
10. Li, Q., Liu, X., A K-Medoids. Clustering Algorithm with Initial Centers Optimized by a P System. *Human Centered Computing*, Springer International Publishing, 2015, 488-500.
11. Liu, Y., Li, X., Richard Yu, F., Ji, H., Zhang, H., Leung, V. C. M. Grouping and Cooperating Among Access Points in User-Centric Ultra-Dense Networks with Non-Orthogonal Multiple Access. *IEEE Journal on Selected Areas in Communications*, 2017, 35(10), 2295- 2311. <https://doi.org/10.1109/JSAC.2017.2724680>
12. Malathi, L., Gnanamurthy, R., Chandrasekaran, K. Energy Efficient Data Collection Through Hybrid Unequal Clustering for Wireless Sensor Networks. *Computers & Electrical Engineering*, 2015, 48, 358-370. <https://doi.org/10.1016/j.compeleceng.2015.06.019>
13. Manikandan, K., Kanmani, P., Munnira Sulthana, M. Energy Efficient Algorithms for Wireless Sensor Network. *International Journal of Advanced Research in Computer and Communication Engineering*, 2015, 4(1), 342-346.
14. Omar, Y. A., Paul, D. Y., Sami, M., George, K. K., Kamal, T. Efficient Machine Learning for Big Data: A Review. *Big Data Research*, 2015, 2, 87-93. <https://doi.org/10.1016/j.bdr.2015.04.001>
15. Padmavathy, C., Geethanjali, S., Jeba, S. J., Mohana Priya, G. Self-Organized Energy Efficient Algorithm for Wireless Sensor Network. *International Journal of Computer Science and Mobile Computing*, 2015, 4(4), 373-378.
16. Qiang, Y., Zhang, J. A Bijection Between Lattice-Valued Filters and Lattice-Valued Congruences in Residuated Lattices. *Mathematical Problems in Engineering*, 2013, 36(8), 4218-4229.
17. Richard, F. Y. *Advances in Communications-Based Train Control (CBTC) Systems*. ISBN 978-1-4822-5743-4, CRC Press, 2015.
18. Richard, F. Y. *Cognitive Radio Mobile Ad Hoc Networks*. ISBN: 978-1-4419-6171-6, Springer, 2011.
19. Richard, F. Y., Leung, V. C. M. *Advances in Mobile Cloud Computing Systems*. ISBN 978-1-4987-1509-6, CRC Press, 2015.
20. Richard, F. Y., Zhang, X., Leung, V. C. M. *Green Communications and Networking*. ISBN 978-1-4398-9913-7, CRC Press, 2012.
21. Rios, L. G., Diguez, J. A. I. Big Data Infrastructure for Analyzing Data Generated by Wireless Sensor Networks. 2014 IEEE International Congress on Big Data, 2014, 816-823. <https://doi.org/10.1109/BigData.Congress.2014.142>
22. Sabarina, K., Priya, N. Lowering Data Dimensionality in Big Data For The Benefit of Precision Agriculture. *Procedia Computer Science*, 2015, 48, 548-554. <https://doi.org/10.1016/j.procs.2015.04.134>
23. Song, H., Li, W., Shen, P., Vasilakos, A. Gradient-Driven Parking Navigation Using a Continuous Information Potential Field Based on Wireless Sensor Network. *Information Sciences*, 2017, 408, 100-118. <https://doi.org/10.1016/j.ins.2017.04.042>
24. Srivastava, H. M., Zhang, Y. Y., Wang, L., Shen, P. Y., Zhang, J. A Local Fractional Integral Inequality on Fractal Space Analogous to Anderson's Inequality. *Abstract and Applied Analysis*, Hindawi Publishing Corporation, 2014, 46(8), 5218-5229.
25. Taheri, M., Kavian, Y. S. Energy Efficient Clustering Algorithm for Wireless Sensor Networks Using Fuzzy Logic. *International Journal of Computer Applications*, 2014, 89(14), 1-5.
26. Takaishi, D., Nishiyama, H., Kato, N., Miura, R. Toward Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks. *IEEE Transactions on Emerging Topics in Computing*, 2014, 2(3), 388-397. <https://doi.org/10.1109/TETC.2014.2318177>

27. Wang, A., Yang, D., Sun, D. A Clustering Algorithm Based on Energy Information and Cluster Heads Expectation for Wireless Sensor Networks. *Computers and Electrical Engineering*, 2012(38), 662-671. <https://doi.org/10.1016/j.compeleceng.2011.11.017>
28. Wei, W., Fan, X., Song, H., Fan, X., Yang, J. Imperfect Information Dynamic Stackelberg Game Based Resource Allocation Using Hidden Markov for Cloud Computing. *IEEE Transactions on Services Computing*, 2016, 11(1), 78-89. <https://doi.org/10.1109/TSC.2016.2528246>
29. Wei, W., Fan, X., Song, H., Wang, H. Video Tamper Detection Based on Multi-Scale Mutual Information. *Multimedia Tools & Applications*, 2017, 1-18. <https://doi.org/10.1007/s11042-017-5083-1>
30. Wei, W., Qi, Y. Information Potential Fields Navigation in Wireless Ad-Hoc Sensor Networks. *Sensors*, 2011, 11(5), 4794-4807. <https://doi.org/10.3390/s110504794>
31. Wei, W., Song, H., Wang, H., Fan, X. Research and Simulation of Queue Management Algorithms in Ad Hoc Network under DDoS Attack. *IEEE Access*, 2017, 5 27810-27817. <https://doi.org/10.1109/ACCESS.2017.2681684>
32. Wei, W., Sun, Z., Song, H., Wang, H., Fan, X., Chen, X. Energy Balance-Based Steerable Arguments Coverage Method in WSNs. *IEEE Access*, 2017, 6, 33766-33773. <https://doi.org/10.1109/ACCESS.2017.2682845>
33. Wei, W., Yang, X. L., Shen, P. Y., Zhou, B. Holes Detection in Anisotropic Sensornets: Topological Methods. *International Journal of Distributed Sensor Networks*, 2012, 21(9), 3216-3229. <https://doi.org/10.1155/2012/135054>
34. Wei, W., Yang, X.-L., Zhou, B., Feng, J., Shen, P.-Y. Combined Energy Minimization for Image Reconstruction from Few Views. *Mathematical Problems in Engineering*, 2012, 16(7), 2213-2223. <https://doi.org/10.1155/2012/154630>
35. Winter, R. Big Data: Business Opportunities, Requirements and Oracle's Approach, 2011.
36. Wu, J., Guo, S., Li, J., Zeng, D. Big Data Meet Green Challenges: Big Data Toward Green Applications. *IEEE Systems Journal*, 2016, 10(3), 888-900. <https://doi.org/10.1109/JSYST.2016.2550530>
37. Wu, J., Guo, S., Li, J., Zeng, D. Big Data Meet Green Challenges: Greening Big Data. *IEEE Systems Journal*, 2016, 10(3), 873-87. <https://doi.org/10.1109/JSYST.2016.2550538>
38. Wu, M., Tan, L., Xiong, N. A Structure Fidelity Approach for Big Data Collection in Wireless Sensor Networks. *Journal of Sensors*, 2015, 15(1), 248-273. <https://doi.org/10.3390/s150100248>
39. Xu, Q., Wang, L., Hei, X. H., Shen, P., Shi, W., Shan, L. GI/Geom/1 Queue Based on Communication Model for Mesh Networks. *International Journal of Communication Systems*, 2014, 27(11), 3013-3029.
40. Younis, O., Fahmy, S., HEED: A Hybrid Energy-Efficient Distributed Clustering Approach for Ad-Hoc Sensor Networks. *IEEE Transactions on Mobile Computing*, 2004, 3(4), 660-669. <https://doi.org/10.1109/TMC.2004.41>
41. Zhao, N., F. Richard, Y., Chen, Y., Leung, V. C. M. Collusive Eavesdropping in Interference Alignment (IA)-Based Wireless Networks. *IEEE Transactions on Wireless Communications*, 2017, 16(8), 5549-5561. <https://doi.org/10.1109/TWC.2017.2712644>