IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 20, 2018, accepted April 30, 2018, date of publication May 3, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2832626

Compressive Data Gathering Based on Even Clustering for Wireless Sensor Networks

JIANHUA QIAO[1,2] AND XUEYING ZHANG[1], (Member, IEEE)
[1]School of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China
[2]School of Electronic and Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China

Corresponding author: Xueying Zhang (tyzhangxy@163.com)

This work was supported in part by the Natural Science Foundation Project of Shanxi Province, China, under Grant 2013011019-1, and in part by the Characteristic Discipline Fund of Shanxi Province, China, under Grant 80010302010053.

ABSTRACT Compressive data gathering (CDG) based on compressed sensing (CS) theory for wireless sensor networks (WSNs) greatly reduces the amount of data transmitted compared with the traditional acquisition method that each node forwards the collected data directly to the next node. CDG combined with sparse random projection can further reduce the amount of data and thus prolong the lifetime of the WSN. The method of randomly selecting projection nodes as cluster heads to collect the weighted sum of sensor nodes outperforms the non-CS (without using CS) and hybrid-CS (applying CS only to relay nodes that are overloaded) schemes in decreasing the communication cost and distributing the energy consumption loads. However, the random selection of projection nodes causes the overall energy consumption of the network to be unstable and unbalanced. In this paper, we propose two compressive data gathering methods of balanced projection nodes. For WSN with uniform distribution of nodes, an even clustering method based on spatial locations is proposed to distribute the projection nodes evenly and balance the network energy consumption. For WSN with unevenly distributed nodes, an even clustering method based on node density is proposed, taking into account the location and density of nodes together, balancing the network energy and prolonging the network lifetime. The simulation results show that compared with the random projection node method and the random walk method, our proposed methods have better network connectivity and more significantly increased overall network lifetime.

INDEX TERMS Cluster head, compressed sensing (CS), compressive data gathering (CDG), even clustering, random projection, sensor node, wireless sensor networks (WSN).

I. INTRODUCTION

The most immediate goal for wireless sensor networks (WSN) is to collect data. Because the data gathered by the sensor nodes in WSN has spatio-temporal correlation, it satisfies the condition that the signal is sparse or compressible in the application of compressed sensing (CS) theory. The sensor nodes have limited resources and the sink node has strong performance, which is suitable for the simple coding and complex decoding of compressed sensing theory. Therefore, the technology of WSN data collection based on compressed sensing has been gradually and extensively studied and developed.

Compressive data gathering (CDG) is based on the compressed sensing theory. Each row of entries of the $M \times N$ dimension measurement matrix Φ are projected on N sensor nodes. Each node multiplies the projection coefficient by the locally collected data and passes it to the next node. The next node also multiplies the collected data by the projection coefficient, then pluses the weighted data from the previous node, and transfers the weighted sum to the next node. In this way, each node calculates and transmits a weighted sum along the route. Eventually, the weighted sum of all nodes is transmitted to the sink node, thus the sink gets a measured value. After each node transfers M times, the sink can get M measurements. Then the reconstruction algorithm is adopted to reconstruct the original signals using M measurements and the sensing matrix θ. Compressive data gathering converts the N samples needed by the sink into collecting M weighted sums of the local samples, which reduces the amount of data transmitted, decreases energy consumption, and extends the network lifetime.

If the measurement matrix is sparse, there will be some zeros for its each row of entries. Thus, the weights of the samples for these nodes will also be zero, and the nodes will not

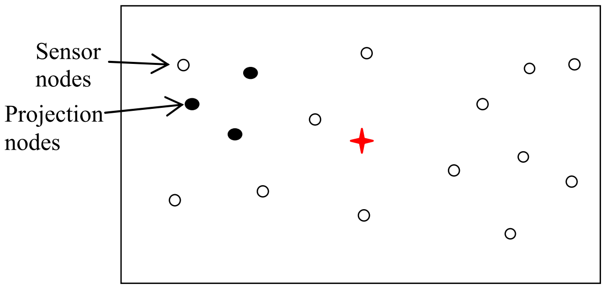


FIGURE 1. **Uneven distribution of projection nodes.**

need to transmit data. Therefore, for the sparse measurement matrix, only the weighted result of the nodes corresponding to non-zero coefficients needs to be transmitted. According to the CS theory, assuming that the measurement matrix is a sparse random projection matrix, as the non-zero entries are randomly distributed, and the sensor nodes corresponding to the projection are also random, therefore the projection procedure is called random projection. For the random projection the amount of data transferred becomes smaller, and the network's lifetime longer.

Ebrahimi and Assi [1] introduced a compressive data gathering method using random projection. Firstly, M nodes were selected randomly in the network as projection nodes to collect M weighted sums. The projection nodes are equivalent to cluster heads. The nodes corresponding to non-zero elements of each row of measurement matrix Φ are assigned to a projection node, and they transmit the weighted sum to the projection node through the minimum spanning tree route. The projection node (cluster head) will send the all weighted sum including itself (if its measurement coefficient itself is not zero) to the sink. Similarly, M projection nodes collect and send M weighted sums to the sink. Thus, M measurements are formed, completing the entire compressed data collection.

The first major problem of this scheme is that the selection of projection nodes is random. Because of its randomness, the selected projection node has no advantages at all. It is easy to face the situation that the distribution of the projection node is not ideal, it is far away from the sink node, or the node's remaining energy will quickly be depleted. Furthermore, there is also no guarantee that M projection nodes will be exactly chosen because each node has a M/N probability selected. Therefore, it is possible to randomly choose a little more or less nodes. Second, since the locations of non-zero coefficients in each row of the measurement matrix are also random, a group of nodes with different distances will transmit the weighted sum to the projection nodes with indefinite positions. Such network is bound to be very uneven energy consumption. Third, routing from a non-zero node to a projection node through a tree structure requires a large overhead. In addition, the authors does not explain the network's running situation after some nodes died.

As shown in Fig.1 is a schematic diagram of the random selection of the projection nodes, a total of 16 sensor nodes, the central star for the sink node, the black solid dots for the projection nodes with a probability of 1/5 randomly generated. It can be seen that randomly generated projection nodes may be unevenly distributed, which results in large energy consumption when collecting data. Therefore, how to choose M projection nodes is a key issue.

There are different ways to choose M projection nodes: one is random selection, such as the scheme in [1]. In N network nodes, M projection nodes are randomly selected with probability M/N. This method allows each node to rotatory projection, the load will be evenly distributed on the network. However, there is no guarantee that an ideal projection node of a suitable location, high energy and accurate quantity will be selected. Another is condition selection, according to certain conditions choose a satisfactory projection node to meet the requirements of balance and high efficiency. In addition, the projection nodes may also be fixed in position with higher-performance devices so that they do not need to continually change their role. Obviously, the position of the projection node in the network will affect the efficiency of the aggregation algorithm. Therefore, we can try to find the optimal projection node.

In view of the above problems, according to whether the distribution of WSN nodes is even, we proposes two compressive data gathering methods of even projection nodes. Aiming at WSN with uniform distribution of nodes, we propose an even clustering method based on spatial location. In order to make the projection nodes evenly distributed, the grids are uniformly divided in the monitoring area. Each grid selects the projection nodes according to the energy and the distance to ensure the uniform distribution of the projection nodes. Aiming at WSN with uneven distribution of nodes, an even clustering method based on the distribution density of nodes is proposed in this paper. As the nodes in some places are dense, while the nodes in some places are scarce, if the projection nodes are selected by location only, the equalization of the projection can not be guaranteed. Therefore, the spatial location and the node density are considered together to select the projection nodes, which ensures the balanced distribution of projection nodes. The validity of the proposed method is verified by comparison with the method of random projection nodes and the method of random walk.

The rest of the paper is organized as follows: Section II presents an overview of CDG research and the basic theory of CS. Section III presents the basic principle of the CDG method based on even projection. Section IV describes the specific implementation process and simulation results of the CDG method based on spatial location for WSN with uniformly distributed nodes. Section V describes the method of CDG based on node density for WSN with non-uniform distribution nodes, and compares it with the random projection node method and the random walk method in different scenarios. Section VI concludes the paper.

II. RELATED RESEARCH

This section introduces the related work of compressive data gathering (CDG) and the basic theory of compressed sensing (CS).

A. OVERVIEW OF CDG

Bajwa and Haupt first applied CS theory to data collection in WSN [2]. For self-organizing WSN formed by tens of thousands of small, inexpensive wireless sensor nodes, each node can generate and send data. It is a big challenge for WSN to ensure its effective transmission and information sharing.

Haupt *et al.* [3] describe the CS-based WSN data acquisition process. CS has two very good characteristics for a network data analysis, one is decentralization, which means that distributing data to fusion center (FC) does not need to be encoded by a central controller; the other is universal, sampling does not require prior knowledge. That is so-called universal sampling and decentralized coding. Moreover, the measurement matrix can be conveniently implemented using a network projection method. For example, a Toeplitz matrix can be generated by each node using a random number generator and a seed value, where each node generates its own random sequence with its unique integer identifier, then CS performs projection observations by calculating and communicating. In [3], two methods for transmitting k random projections from the network node to the FC by wireless style are given. One is direct transmission, the steps are as follows: (1) n sensors use their own network addresses as the seeds of the pseudo-random generator to generate k random projection vectors. (2) The sensor at position j multiplies its measured value by the projection vector to obtain a k-dimensional array $v_j = (A_{1,j}x_j, \ldots, A_{k,j}x_j)^T$, j = 1, 2, ..., n. All sensors transmit their respective v_j to FC with k time slots, then the FC receives the corresponding signal after k transmissions

$$y = \sum_{j=1}^{n} v_j = Ax \tag{1}$$

The above step of transmitting k random projections to the FC through k times by the sensor nodes is a completely decentralized way.

Another way to achieve the same goal is to assume that the sensor has only local communication capabilities, and to establish a spanning tree route that reaches a given cluster head. Then each sensor node can compute v locally, and these values are passed through the aggregation tree to the cluster head to get $v = Ax$, then the cluster heads passes these vectors to the FC. The main feature of the wireless transmission method described above is that it can be implemented without any complicated routing information and may be a suitable upgradeable option in many sensor network applications.

Reference [3] converted the observational projection of CS theory into the weighted operation and transmission of sensor nodes in WSN, and established the basic framework of WSN data collection based on CS. Moreover, it transforms the transmission amount of N nodes O(N^2) into the transmission amount of K (K $\ll$ N) times O(NK), which greatly reduces the transmission and reduces the energy consumption, and becomes a kind of important way of WSN data collection. However, the measurement matrix in this paper is a random number generated by network address and seed value. Its formation is complicated, which increases the amount of storage and computation. And k measurements generated by each sensor are sent directly to the cluster head or the sink node, thus increasing the storage capacity of the cluster head or the sink node.

However, [4] disagrees with this opinion, discussing whether CS actually increases the throughput of WSN. Three cases were analyzed: (1) non-CS acquisition using conventional methods; (2) plain-CS acquisition; and (3) hybrid-CS acquisition. In the part near the leaf nodes, the CS is not used because of the less amount of transmission by the traditional method. The closer to the root node, the larger the increase of the transmission amount, the CS is employed. This indicates that WSN with hybrid-CS has the highest throughput.

For the first time, Luo *et al.* [5] proposed a compressive data gathering (CDG) scheme for large-scale WSN. For a large number of nodes with spatially correlated readings in a densely distributed WSN, instead of adopting the method of each node separately transmission in [3], the sink node obtains the weighted sum of all the readings. The CDG schematic diagram shown in Fig.2.

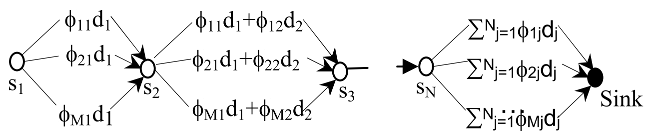


FIGURE 2. Compressive data gathering in a multi-hop route.

For example, a sensor node s_1 multiplies its reading d_1 by the projection coefficient φ_1, and transmits the product v_1 to the next node s2. The s2 multiplies its reading d_2 by the projection coefficient φ_2, then transmits the sum of the product v_2 and v_1 to the next node s3. In this way, eventually the sink gets the weighted sum of the readings of all nodes. Thus there is no complicated calculation and transmission control, reducing the cost of the global communications, making the load balancing and extending the network lifetime. Therefore, the collection scheme becomes the basic scheme for multi-hop routing collection.

However, Luo *et al.* [6] proposed two key issues of the CDG framework. First, how to generate the measured values that satisfy the RIP, taking into account the multi-hop communication consumption. Second, although the sparsity of sensor readings is common, it is quite complicated to fully apply it. At the same time, it shows that CDG scheme for large scale monitoring sensor network using CS principle can effectively reduce the communication cost and prolong the network lifetime, confirming that the network capacity increases proportionally with the sparseness of sensor reading. Due to the inherent flexibility of the CS rules, the proposed CDG framework can be applied to a variety of sparse modes, whether it be simplified or combined data collection processes.

Therefore, the establishment of the CDG framework mainly uses CS theory to reduce the transmission of sensing nodes, thereby reducing energy consumption and prolonging the network lifetime of the WSN.

In WSN, the communication energy consumption is much higher than other aspects of energy consumption. How to reduce traffic is an important factor for WSN to reduce the energy consumption. The key to the amount of data to be transmitted in a CS-based WSN is determined by the measurement matrix. For the dense Gaussian random matrices, each sensor node is involved in transmitting the weighted sum. However, most of the elements of sparse random projection matrix are 0, such that the corresponding nodes need not transmit data, greatly reducing the traffic. Therefore, the CDG based on sparse random projection becomes a widely used method of the compressive data gathering.

Haupt and Nowak first presented that a relatively small number of random projection of a signal can contain most of its salient information. Therefore, if a signal is compressible on some orthonormal basis, the reconstruction from random projection can be very accurate [7]. Moreover, this 'compressive sampling' method can recover accurately from random projections of noise pollution, and in many cases it may be more accurate than using a conventional method of sampling the same number of points, and apply it to remote wireless sensing [8]. Wang *et al.* [9] also proposed a distribution algorithm based on sparse random projection. The sparseness of random projection greatly reduces the communication cost.

Reference [1] stated that in order to increase the lifetime of the network, it is necessary to reduce the energy consumption of the entire network and distribute the energy load more evenly across the network. A method of data acquisition using CS and random projection to improve the lifetime of large-scale WSN - minimum spanning tree projection (MSTP) is proposed. The MSTP creates a minimum-spanning-trees (MSTs). Each root randomly selects projection nodes and uses the CS to gather the sensor's data. And further expand into eMSTP. That is, the sink node is added to each MST, and the sink node is taken as the root of each tree. Simulation results show that the MSTP and the eMSTP outperform existing data collection schemes in terms of reducing communication consumption and balancing energy consumption, thereby improving the overall network lifetime.

The problem addressed by [10] is to recover sparse signals observed by resource-limited WSN under channel fading, reducing the communication cost of information forwarding to FC by using sparse random matrix. in [11], for the large-scale WSN of measuring compressible data, the authors took into account that the sparsity of random projections affected the mean square error (MSE) and the system delay, and presented an adaptive sparse random projection algorithm to achieve better trade-off between MSE and system delay.

The topology and routing mechanism of WSN is also an important issue for data collection using CS. Data collection combining the CS and cluster structures has been proven to be an effective way to reduce WSN energy consumption [12]. Nguyen *et al.* [13] applied clustered-base CS scheme, pointed out that there are two ways to transmit data from the cluster heads to the sink or the base station. One is direct method and the other is multi-path routing through middle cluster heads. Wu *et al.* [14] proposed a high-performance clustered routing data collection scheme for large-scale WSN. By determining an energy consumption model to get the optimal number of clusters, an effective deterministic dynamic clustering scheme is designed to ensure that all cluster heads are basically uniformly distributed. Abbasi-Daresari and Abouei [15] utilized the advantages of sparse random measurement matrices to reduce energy consumption, focused on the energy control of sensor nodes, and formed an energy-efficient routing tree to distribute load balancing problems. Furthermore, a cluster-based weighted compressive data aggregation (CWCDA) is proposed to reduce the energy consumption and reduce the number of sensor nodes through minimum spanning tree projection. Xu *et al.* [16] proposed a new hierarchical data aggregation compressed sensing (HDACS), which combines the CS and the hierarchical network structure. Multiple compression thresholds are adaptively set based on the cluster size at different levels of the data aggregation tree to optimize the amount of data transfer.

The tree structure is also a typical WSN routing structure, commonly used is the minimum spanning tree. The entire network uses the sink node as the root node and the source node as the leaf node to construct an aggregation tree [17], [18]. Xiao *et al.* [19] the joint minimum spanning tree and the nodes of interest to form a tree. Projection nodes are randomly selected, and the projection node is used as the root node to construct the minimum spanning tree for connecting the nodes of interest. The projection node aggregates the data from the sensing nodes with the CS and sends it to the sink through the shortest path, allowing the sink to participate in the process of constructing a minimum spanning tree. Reference [20] proposes a three layer hierarchy for distributed clustering. Reference [21] applies distributed compressive sensing and quantization configuration to reduce energy consumption. In [22], the proposed clustering method is to collect data within a cluster without CS, and the cluster heads use CS to send data to the sink. This method makes a large number of sensor nodes do not use CS, a small number of cluster head nodes use CS, which can not effectively use CS.

Random walk (RW) has been effectively used for data collection in wireless sensor networks [23]–[27]. It does not require global information as the shortest path route, in addition, it enables network load balancing. Because sparse random projection has been proven to be as effective as a dense Gaussian matrix, the combination of RW and CS becomes a powerful routing method that helps to save energy and extend network life effectively [24]. Fletcher [23] introduces RW to WSN based on distributed compression sensing, which proves that RW performs better than shortest path routing.

References [28] and [29] introduced a fuzzy C-means clustering algorithm to make the application of WSN more flexible. Zheng *et al.* [30] the authors studied the impact of sink position on the number of transmissions and network lifetime, and applied a weighted routing tree to transmit data projections from the cluster heads to the sink, and fairly distributed energy consumption among different nodes. The secure data collection scheme [31] is also the focus of future WSN research.

These papers show the effective implementation of compressed data collection for WSN and some methods of reducing network energy consumption, and illustrate that the application of sparse random projection together with routing structure can greatly reduce the amount of data transmitted. The compressive data gathering method of balanced projection nodes proposed in this paper is to solve the problem of how to effectively collect data through the projection nodes.

B. COMPRESSED SENSING THEORY

Compressed sensing [32] (also called compressive sensing [33]) is a new theory of information acquisition proposed by David L. Donoho et al. in 2004. The theory states that for sparse or compressible signals, the data can be sampled much below the Nyquist sampling frequency and then perfectly reconstructed by a nonlinear reconstruction algorithm [34]. That is to say, the sparse N-dimensional signal x can be sparsely decomposed into under an N×N dimensional sparse transformation matrix Ψ as:

$$x = \Psi\theta. \tag{2}$$

where θ is a sparse N×1 column vector with only K non-zero entries, and K≪N. And then projection under $M \times N$ dimensional measurement matrix (observation matrix) Φ, we can get M observation values y, and $M \ll N$. That is

$$y = \Phi x = \Phi\Psi\theta = T\theta, \tag{3}$$

where T is called sensor matrix. It can be seen that $y = T\theta$ is the underdetermined equation or ill-posed equation, and in general there is no definite solution.

However, Candès et al. have developed the well-known Restricted Isometry Property (RIP) [35], [36] and given sufficient conditions for the existence of definite solutions to (3). RIP theory states that for a matrix T, the matrix T satisfies the RIP if it exists $\delta \in (0, 1)$ such that all k sparse signals θ satisfy the following (4) [35], [37].

$$(1-\delta)||\theta||_2^2 \le ||T\theta||_2^2 \le (1+\delta)||\theta||_2^2. \tag{4}$$

Therefore, if the signal θ is sparse and the sensing matrix T satisfies the RIP, the reconstructed signal can be obtained by solving the l_1 norm or the l_0 norm of (5) [33], [34].

$$\hat{\theta} = \arg\min ||\theta||_1 \text{ or } \hat{\theta} = \arg\min ||\theta||_0 \quad \text{s.t. } y = T\theta \tag{5}$$

The above equation can be used to reconstruct sparse signals or reconstruct compressible signals with high probability [32], [36] by using linear programming such as Basis Pursuit (BP) algorithms. It is also possible to recover the sparse transform coefficient $\hat{\theta}$ using other reconstruction algorithms such as Orthogonal Matching Pursuit (OMP) greedy algorithm [38], [39], and then the reconstructed original signal $\hat{x}$ can be obtained by $\hat{x} = \Psi\hat{\theta}$.

However, judging whether a given T has RIP property is a combinatorial complexity problem. It is difficult to directly construct a sensing matrix T such that $T = \Phi\Psi$ satisfies the RIP, that is, any 2K columns in the matrix are guaranteed to be irrelevant. And RIP is a sufficient, not a necessary, condition for signal reconstruction. Baraniuk gives the equivalent condition of RIP that the measurement matrix Φ and sparse representation base Ψ are incoherence [33]. He points out the M×N dimensional independent and identically distributed (iid) Gaussian random matrices satisfies the RIP with high probability when M≥cKlog(N/K) (c is a small constant) and is uncorrelated with most of the orthogonal bases Ψ, which is of universal [40]. The correlation is to describe the maximum degree of association between any two elements of base Φ and Ψ. The correlation coefficient of the matrix Φ and Ψ is

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{i,k} |<\Psi_i, \Phi_i>|, \tag{6}$$

where $<\Phi, \Psi>$ represents the inner product of Φ and Ψ, and $\mu(\Phi, \Psi) \in [1, \sqrt{N}]$ [34]. The smaller μ indicates that the greater the incoherence of the measurement matrix Φ and the transform base Ψ, the more the coefficients required to represent each other, the less the number of measurements required [41]. Otherwise, the correlation is stronger.

III. BASIC PRINCIPLE OF EVEN PROJECTION METHOD

According to CS theory, the measurement matrix (or projection matrix, observation matrix) is a matrix of M×N dimensions, and x is a vector of N×1 dimensions.

$$y = \Phi x = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} & \cdots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} & \cdots & \varphi_{2N} \\ & & \vdots & & \\ \varphi_{N1} & \varphi_{N2} & \varphi_{N3} & \cdots & \varphi_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix} \tag{7}$$

$$y_i = \sum_{j=1}^{N} \varphi_{ij} x_j, \quad (i = 1, 2, \cdots, M) \tag{8}$$

From (7), we can see that for N nodes x_j (j = 1, 2, ⋯, N), the result of multiplying each line in Φ by x can be calculated once to obtain M measurement values. N nodes can also be divided to deal with, assuming N nodes are divided into h segments, each segment can be a different number of points. The measurement matrix Φ may also be correspondingly divided into h blocks, each of which is a matrix of M rows, and the number of columns is equal to the number of corresponding segments of x. Each block matrix is multiplied by each segment of the corresponding x to obtain an M×1 column vector, and then the h column vectors are summed to obtain M measurement values. Expressed as a formula

$$\begin{aligned} y = \Phi x &= [\Phi_1, \Phi_2, \cdots, \Phi_h] \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_h \end{bmatrix} \\ &= \Phi_1 X_1 + \Phi_2 X_2 + \cdots + \Phi_h X_h = \sum_{j=1}^{h} \Phi_j X_j \end{aligned} \tag{9}$$

Where X_j is a segment obtained by dividing x of N points, and the length of each segment is $\mathrm{N_j}$, $\mathrm{j} = 1, 2, \ldots, \mathrm{h}$; each Φ_j is an $\mathrm{M}\times\mathrm{N_j}$ matrix, $\mathrm{j} = 1, 2, \ldots, \mathrm{h}$. In this way, each segment x can be called a cluster, each cluster can choose a projection node as a cluster head. Each cluster head collects the data of each cluster and sends it to the sink. The sink accordingly adds all the data transmitted from the cluster heads to obtain the final measurement. With these measurements and sensing matrix, the original signal can be reconstructed. If only some areas are of interest, only some of them need be taken. Due to the randomness of the measurement matrix, part of the random matrices still satisfy the RIP, the signal can still be recovered. As the measurement ratio increases, a higher reconstruction performance is obtained.

Select a projection node in each segment, the projection node can be distributed evenly, so called even projection based compressive data gathering method. Moreover, according to certain conditions to choose, you can choose a projection node of higher energy and superior location. The design of this method mainly consider as the following:

(1) Clustering routing algorithm should be distributed as much as possible, which is beneficial to save network energy and improve network reliability.

(2) The method of selecting randomly projection nodes is not enough to ensure load balancing and avoid excessive energy consumption of the nodes, and the network lifetime can not be extended more effectively.

(3) In order to achieve energy balance, the projection nodes should be evenly distributed throughout the network. If the distribution of projection nodes is not uniform, the number of nodes joining each cluster in clustering is very different. Therefore, the clusters with fewer nodes consume more energy in each round, resulting in uneven energy consumption between clusters, and such the network lifetime can not reach the longest.

(4) In the case of a fixed network topology, an effective clustering algorithm should generate a relatively stable number of cluster heads. The energy consumed by a cluster head accounts for a major part of the energy consumption in the network, thus effectively controls the number of cluster heads, thereby reducing energy consumption per round and prolonging the network lifetime.

(5) The energy cost of clustering algorithm itself should be as small as possible.

This paper chooses the projection nodes according to the above principles and proposes two methods of balanced clustering according to whether the distribution of network nodes is uniform.

IV. EVEN CLUSTERING METHOD BASED ON SPATIAL LOCATION (LEC)

Aiming at the network with uniform density distribution of nodes, an even clustering method based on spatial location is proposed. In order to uniformly distribute the projection nodes, the monitoring area is divided into grids of equal size, and the node with the best performance is selected as the projection node in each grid. Then the projection node collects the weighted sum of the cluster and sends it to sink. The specific implementation process is as follows:

Select the first round of projection nodes and clustering:

(1) Determine the number of clusters.

(2) Divide the entire monitoring area into the same size of grids and calculate the number of clusters.

(3) In each cluster, select the higher energy nodes around the cluster center as projection nodes.

(4) Send the information of the projection node to each cluster's projection node.

(5) Each sensor node chooses its corresponding projection node based on the shortest distance (least hop). After a round of clusters is established, the data can be collected. The data in the cluster is transmitted to the intra-cluster projection node via the route. The data collected by the projection nodes of all the clusters are transmitted to the sink, and then the collected data are added up correspondingly, that is, all the measurement values. Then the original data can be recovered through reconstruction.

Then you can proceed to the next round of projection node selection and the establishment of a new cluster, the process is:

(1) The intra-cluster nodes transmit their residual energy values to the projection node.

(2) The projection node selects the high energy node at the cluster center as the new projection node.

(3) Each original projection node transmits its information to the new projection node.

(4) Each sensor node looks for its new projection node to form a new cluster.

A. DETERMINE THE NUMBER OF CLUSTER HEADS

As the first factor to be considered in the design of routing algorithms for networks is the energy problem, the design of routing algorithms is closely related to the channel energy loss model in WSN. Fig.3 is a simplified network channel loss model [42].

For transmitting or receiving a k bit message data to a distance of d,the radio dissipates energy as following, respectively [43].

$$E_{Tx}(k, d) = \begin{cases} kE_{elect} + kE_{fs}d^2, & d < d_0 \\ kE_{elect} + kE_{mp}d^4, & d \geq d_0 \end{cases} \tag{10}$$

$$E_{Rx}(k) = kE_{elect} \tag{11}$$

Where k is the number of bits of information per data, E_{elect} is the energy consumed per bit of data in the transmit or receive circuit. E_{fs} and E_{mp} are the energy dissipation values to run the amplifier for close and far distances. Depending on the distance between the transmitter and the receiver, free space (E_{fs}) or multi-path fading (E_{mp}) channel models is used, and d_0 is the critical value between the two models.

Suppose there are N sensor nodes randomly distributed in an area with radius R, where the number of the cluster heads is h, then the average coverage area of cluster head nodes

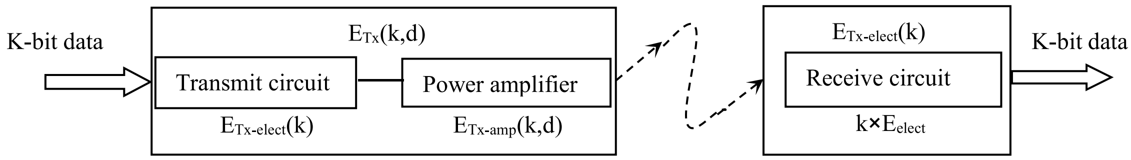


FIGURE 3. **Channel loss mode.**

is $\pi R^2/h$. Suppose that all cluster head nodes in the network send data in a multi-hop manner, and the transmission distance per hop is assumed to be D. The energy consumption of each cluster head of the network consists of receiving information of all cluster members, merging the data, and transmitting the fused information to the sink node. The energy consumed by the cluster head node in one frame is:

$$\begin{aligned} E_{CH} &= kE_{elect}(\frac{N}{h}-1)+kE_{DA}\frac{N}{h}+kE_{elect}+kE_{fs}D^2 \\ &= \frac{N}{h}kE_{elect}+\frac{N}{h}kE_{DA}+kE_{fs}D^2 \end{aligned} \tag{12}$$

Where E_{DA} is energy consumption for data fusion.

Assume that the distance between the nodes in the cluster and the cluster head does not exceed the critical value of the free-space model. Thus, the energy consumption of a non-cluster head node is:

$$E_{non-CH} = kE_{elect}+kE_{fs}d_{toCH}^2. \tag{13}$$

d_{toCH} is the distance between cluster members and cluster heads. Given that the coverage area of the cluster head is $\pi R^2/h$. Assuming the area is circular, the distribution density of nodes is $\rho(x,y)$ and the cluster heads are located in the center of the cluster. Thus, the square of the distance from the node to the cluster head is [2]:

$$d_{toCH}^2 = \frac{\rho\pi R^4}{2h^2}. \tag{14}$$

If the density of the nodes in the entire cluster is uniform, then $\rho = \frac{N}{\pi R^2}$, $d_{toCH}^2 = \frac{NR^2}{2h^2}$, then

$$E_{non-CH} = kE_{elect}+kE_{fs}\frac{NR^2}{2h^2} \tag{15}$$

The energy consumption of the entire clusters in sending a frame is:

$$\begin{aligned} E_{cluster} &= E_{CH}+(\frac{N}{h}-1)E_{non-CH} \\ &= \frac{N}{h}kE_{elect}+\frac{N}{h}kE_{DA}+kE_{fs}D^2+(\frac{N}{h}-1)kE_{elect} \\ &\quad +(\frac{N}{h}-1)kE_{fs}\frac{NR^2}{2h^2} \\ &= \frac{N}{h}kE_{elect}+\frac{N}{h}kE_{DA}+kE_{fs}D^2+\frac{N}{h}kE_{elect} \\ &\quad +\frac{N}{h}kE_{fs}\frac{NR^2}{2h^2} \end{aligned} \tag{16}$$

The total energy consumption of sending one frame per cluster in the R region (ignoring the energy consumption for data forwarding) is:

$$\begin{aligned} E_{total} = hE_{cluster} &= 2NkE_{elect}+NkE_{DA}+hkE_{fs}D^2 \\ &\quad +Nk\frac{NR^2}{2h^2}E_{fs} \end{aligned} \tag{17}$$

Solve E_{total}'s derivative to make it equal to zero, and get the optimal cluster head number:

$$h = \sqrt[3]{\frac{N^2R^2}{D^2}} \tag{18}$$

If the size of the monitoring area is set as 100m×100m, 100 sensor nodes are deployed, and the transmission distance D of per hop is $50\sqrt{2}$ m, then the radius R is 50m, and such according to equation (18), got h≈17. For convenience, the area can be divided into 16 grids.

B. GRID COORDINATES OF NODES

Spatial projection nodes are evenly divided by grids. First consider the node density ρ is constant. Therefore, a 100m×100m area is divided into 16 grids, as shown in Fig.4 and Fig.5. The nodes in the adjacent grids can transmit information to each other. And nodes at both ends of the diagonal of both grids are within the communication range. Suppose the length of the edge of these grids is d and the maximum distance a node can transmit is Rs, then the distance between two nodes transmitting information to each other is $Rs \le 2\sqrt{2}\,d$.

In the division of the grid, we must first determine the node's grid coordinates. Assuming that the sink is located in the center of the region, the node's grid coordinates in the region are calculated according to their geographical position,

$$\begin{cases} T_x = floor(S(i).xd/(Xm/L_1)) \\ T_y = floor(S(i).yd/(Ym/L_2)). \end{cases} \tag{19}$$

Where T_x denotes the grid horizontal coordinates of the nodes, T_y denotes the grid vertical coordinates of the nodes, S(i).xd and S(i).yd respectively represent the abscissa and ordinate of the nodes themselves, Xm and Ym are the length and width of the monitoring area, L1 is the equally divided number of the length of the region, L2 is the equally divided number of the width of the region, and floor(x) represents an integer no greater than x. The nodes of the same grid coordinates are as a group. For example, in Fig.4 and Fig.5, according to formula (19), the coordinates of a grid starting from the lower left corner in turn are (0,0), (0,1), (0,2), (0,3), (1,0), . . ., (3,3). Set a number for each grid, corresponding to

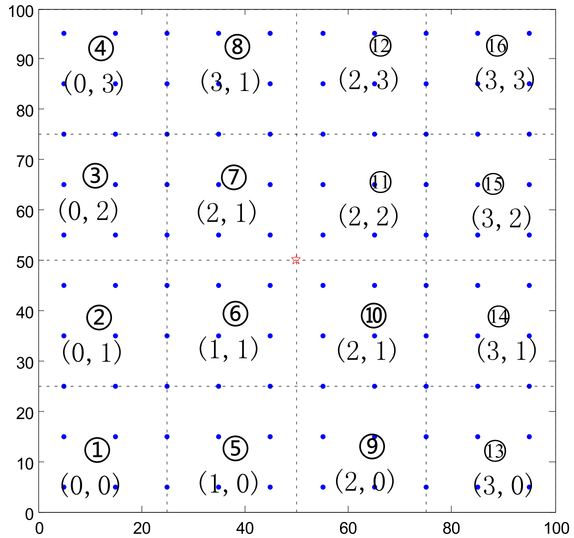


FIGURE 4. Grid with uniform distribution of nodes.

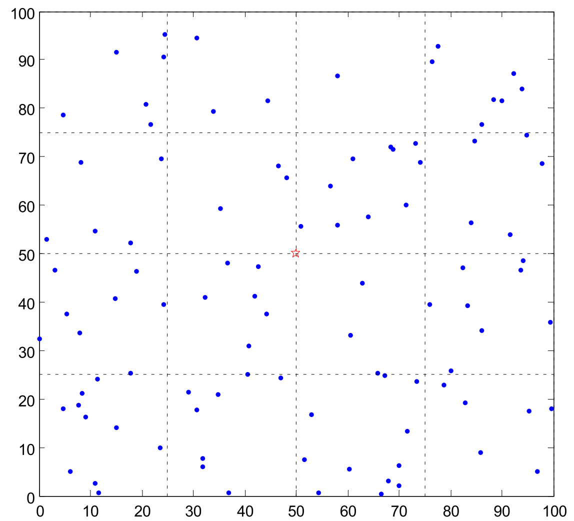


FIGURE 5. Grid with uneven distribution of nodes.

from 1 to 16 respectively. This makes it easy to determine the grid coordinates for each node.

C. CLUSTER HEAD ELECTION ANDCLUSTERING

In each grid, cluster heads are selected based on the remaining energy of each node and the distance from the node to the data collection center. Select the node with larger competition radius as the cluster head.The competitive radius of a node is expressed as

$$S(i) \cdot rc = (1 - b \cdot \frac{d_{\max} - dbs}{d_{\max} - d_{\min}} \cdot \frac{E0 - S(i) \cdot E}{E0 - Ed}) \cdot D. \quad (20)$$

Where, b is a random number in the range of (0,1) to ensure that the result is within a certain range. $d_{\max}$ represents the farthest distance between the sink and the node in a grid, and $d_{\min}$ represents the nearest distance between the sink and the node in a grid. dbs represents the distance from the sink to the current node. E0 represents the initial energy of each node, Ed is the minimum value of the energy that a node can transmit information with other nodes. S(i).E represents the node's current energy. D represents the maximum sensing radius. The competition radius takes into account the remaining energy of the node and the distance between the node and the sink. The election of the cluster head with the large contention radius is most beneficial for all nodes in the grid.

In each grid, the competition radius of each node in the grid is calculated, and the node with the largest competition radius is selected as the cluster head. After all the grids select the cluster heads, other nodes in the whole network except the cluster heads find the cluster head closest to it, and send a request to the cluster head to join the cluster created by the cluster head. After receiving the request from the node to join the cluster, the cluster head sends information to these nodes to allow them to join the cluster. After all the nodes have their own clusters, they form a wireless sensor network.

D. SIMULATION RESEARCH

In the simulation experiment, we make the following assumptions about WSN and sensor nodes:

(1) The sink node location fixed;

(2) The sensor nodes are static, not mobility;

(3) The sensor nodes can know its own location information;

(4) The cluster head fusing unit data has the same energy consumption;

(5) The unit area of the network needs to detect the same amount of data;

(6) The coverage area of each node is the same;

(7) Wireless channel symmetry.

In addition, for the even clustering based on spatial location, it is assumed that the sensor nodes are uniformly distributed in the monitoring area. MATLAB simulation platform is adopted. Simulation parameters in Table 1.

TABLE 1. Simulation related parameters.

Experimental parameters	Scene 0
Area size	100m×100m
Number of nodes	100
Sink location	（50,50）
Initial energy	0.01J
Data information length	4000b
Control information length	32b
Send energy E_{Tx}	10^{-10}J/b
Receive energy E_{Rx}	10^{-10}J/b
Fusion energy E_{DA}	10^{-11}J/b
Dissipated energy E_{fs}	2×10^{-11}J/b
Critical energy Ed	10^{-5}J
Communication radius D	71m

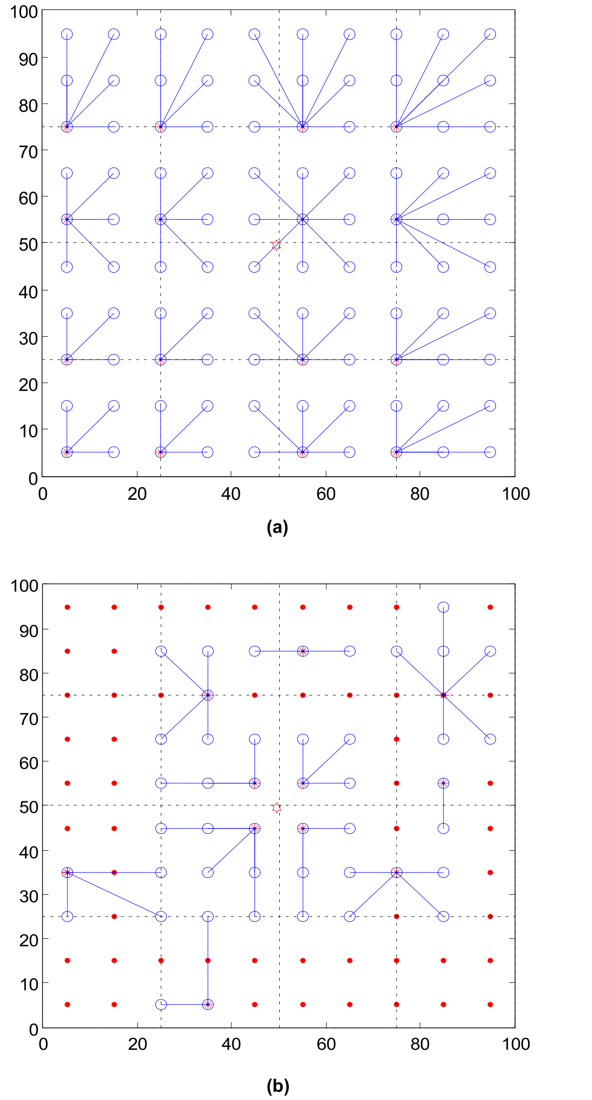


FIGURE 6. **Clustering properties of WSN with node even distributions based on the location clustering. (a) First-round cluster heads distribution and clustering effect of evenly distributed WSN nodes based on the location clustering. (b) The distribution and clustering of nodes at the 300th round.**

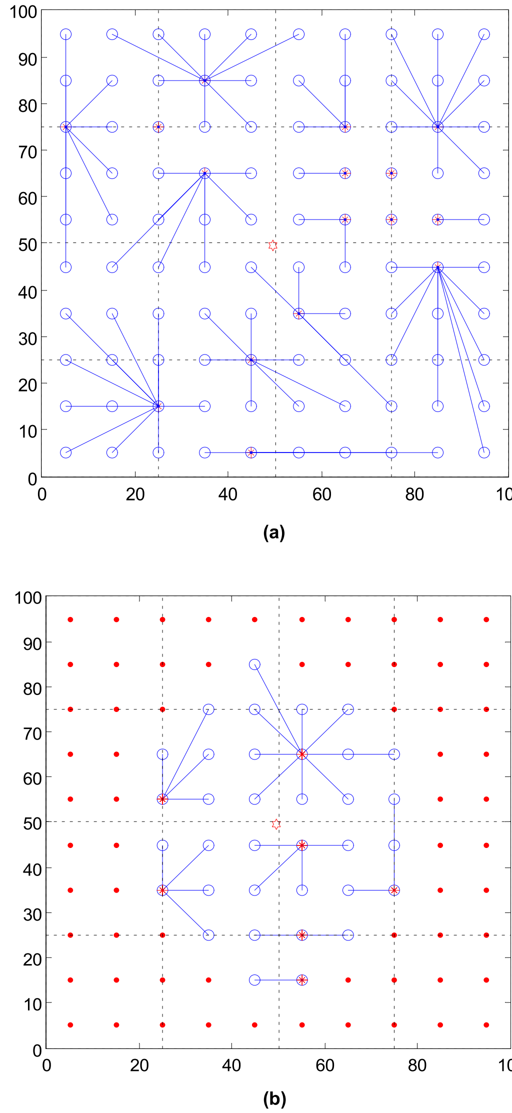


FIGURE 7. **Clustering properties of WSN with node even distribution based on random clustering. (a) First-round cluster heads distribution and clustering effect of WSN with uniform distribution of nodes based on random clustering. (b) The distribution and clustering of nodes at the 300th round.**

Fig.6(a) shows the location-based clustering effect in the nodes uniform network. It can be seen from the figure that through the grid, the nodes of the network are evenly distributed, and the selected cluster heads are also evenly distributed. The cluster head of each cluster, which is equivalent to the projection node, can be rotated by other nodes in the cluster. Fig.6(b) shows the node distribution and the clustering when the network is in 300 rounds. The red dots refer to the dead nodes.

Fig.7 is clustering effect based on random projection node method for the same node layout network. It can be seen from the figure that the randomly selected projection nodes are unevenly distributed and the clustering effect is also very unsatisfactory. Some isolated nodes are selected as the projection nodes, but do not form a cluster with other nodes. However, as a projection node, it consumes more energy than an ordinary node, which shows that the method of randomly selecting projection nodes consumes much more energy.

Fig.8 shows the comparison of the remaining nodes of two methods of based on location and based on random projection nodes. When the network starts to run, the effect based on location clustering is worse, the death nodes appear earlier and the number of remaining nodes is less, but that does not affect the network running. When the network runs to the remaining nodes number is 80% of the total nodes, the number of death nodes based on random clustering rapidly increases and quickly exceeds the number of death nodes based on location clustering, the number of remaining nodes rapidly decreases. At 385 rounds, the number of remaining nodes becomes 0 instantaneously from 20 and the

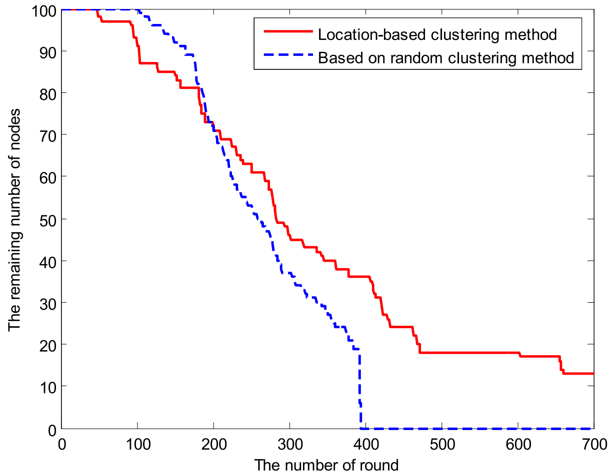


FIGURE 8. Comparison of the remaining nodes of WSN of the node even distribution based on the location clustering and the random clustering.

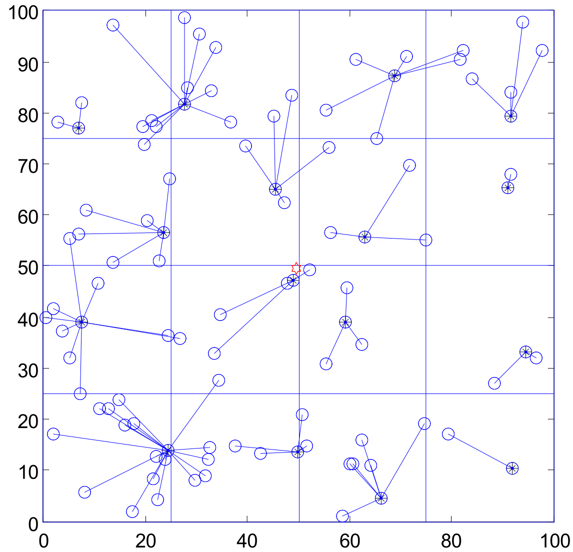
network is paralyzed. However, the network based on location clustering has still survived nearly 40% of nodes this moment and the network can still run. Supposed the remaining nodes number is 20, which is taken as the critical value for the normal operation of the network. This is to say that when the network nodes number is below 20, the network can not be connected. As a result, the network based on location clustering extends the lifetime of network by about 35% compared with that based on random projection nodes.

V. EVEN CLUSTERING METHOD BASED ON NODE DISTRIBUTION DENSITY (DEC)

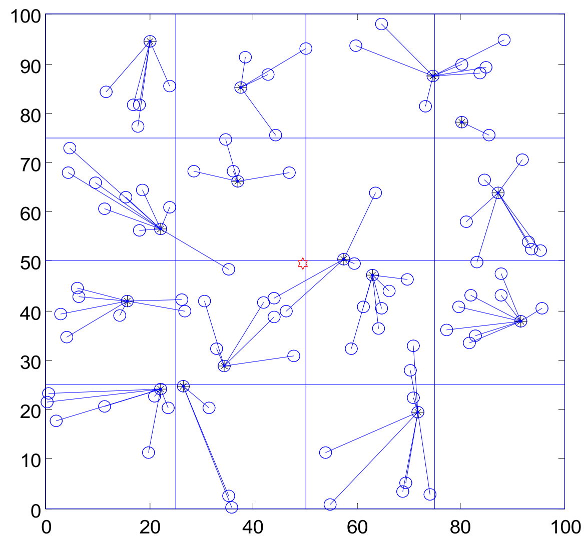
For the WSN with unevenly distributed nodes, the number of nodes in each grid varies. If the location-based clustering method is used to select one projection node in each grid, the amount of data collected by each projection node is also very different. Therefore, in order to adapt to the network with uneven nodes, an even clustering method based on node density is proposed.

A. EVEN CLUSTERING BASED ON NODE DISTRIBUTION DENSITY

For the WSN with unevenly distributed nodes, as shown in Fig.5, the number of nodes in each grid is different in the entire area. In some grids, the number of nodes is slightly larger. After these nodes form a cluster, they can perform the data transmission functions between nodes and nodes, nodes and the cluster heads, the cluster heads and the sink well. It is also possible to allocate some nodes that collect the same kind of information into sleeping mode, saving the energy consumption of this node. However, there are only a few nodes in some grids. For example, the grids (3,0), (3,1) and (3,2) in Fig.9(a) have only two or three nodes. Such that a cluster formed by these few nodes, these nodes must work at all times, and the cluster head has been rotated between these nodes, thus making the energy consumption of the nodes within the grid faster than the other grids with many



(a)



(b)

FIGURE 9. Cluster head election and clustering with uneven distribution nodes. (a) Effect illustration 1 (a small number of nodes in the grid also have to choose cluster head, easy to make node premature death). (b) Effect illustration 2 (nodes with near distances between different grids also need to be clustered separately, increasing energy consumption).

nodes. As shown in Fig.9(b), the two projection nodes of the grid (0,0) and (1,0) are very close but clustered respectively, resulting in increasing energy consumption.

These situations have affected the entire network. The nodes in some areas die prematurely, and the energy of the nodes in other areas can still continue to work. Therefore, the entire WSN can not transmit the complete information to the sink, it is impossible to make a comprehensive and correct decision. Therefore, the best method is to reallocate the nodes in the grid of fewer nodes and maximize the lifetime of all the nodes in the entire network.

For the WSN with uneven distribution of nodes, we propose an even clustering method based on node distribution density. The distribution density of nodes is measured by the grid. The small number of nodes in the grid indicates that the density is small; on the contrary, the density is large. Therefore, the nodes in the grid are firstly adjusted, the number of nodes in each grid is calculated. If the number of nodes in the grid is less than or equal to the threshold, the nodes in these grids are assigned to other neighboring grids whose number of nodes is greater than the. threshold. According to the shortest distance algorithm, this node is classified into the grid where the cluster head closest to it. After all the nodes in the grid with the number of nodes less than or equal to the threshold are classified to the designated grid, a new pattern of clustering is formed. In this way, it is possible to avoid unbalanced energy consumption of nodes, thereby improving the lifetime of the entire network. Next, the cluster head election and formation of clusters are performed, clustering effect as shown in Fig.10.

In Fig.10(a), the grids with fewer nodes, such as grids (1, 3) and (2, 1), are not necessarily clustered separately due to the small number of nodes.They are assigned to neighboring clusters, thus avoiding energy consumption in the election of projection nodes and in separate clustering. Fig.10(b) shows the distribution of nodes at 300 rounds. Obviously, there are fewer dead nodes, and most of the nodes are still alive for data transmission and collection. Fig.10(c) shows the relationship between the number of rounds and the number of remaining nodes. It can be seen that the first dead node appears at about 60 rounds, and the number of dead node descends more slowly.

B. SIMULATION RESEARCH

In simulation experiments, the assumptions for the WSN and the sensor nodes are the same as in Section IV.D. In addition, for the even clustering method based on the distribution density of nodes, it is assumed that the sensor nodes are randomly and non-uniformly distributed in the monitoring area. For the sake of comparison, the simulations were carried out in two scenarios respectively. See Table 2 for the specific parameters. For scene 2, the size of the monitoring area is 200m$\times$200m, and 400 sensor nodes are deployed. The sensor node communication radius D is $50\sqrt{2}$m, the monitoring area radius R is equivalent to 100m. Then according to (18), h$\approx$68. The area can be divided into 64 grids for simulation convenience.

Now we test the performance of the clustering method based on the distribution density of nodes and the method in [1] using random projection nodes in the same node layout area.

Fig.11 shows the clustering effect of the even clustering method based on the distribution density of nodes in scene 1, and Fig.11(a) shows the first round of clustering of the scene.We can see that the nodes in the grids of low node density are redistributed to the adjacent clusters, and the grids no longer select the projection nodes. Fig.11(b) is the clustering effect at the 300th round with about 60% of the nodes still alive and the network is still running.

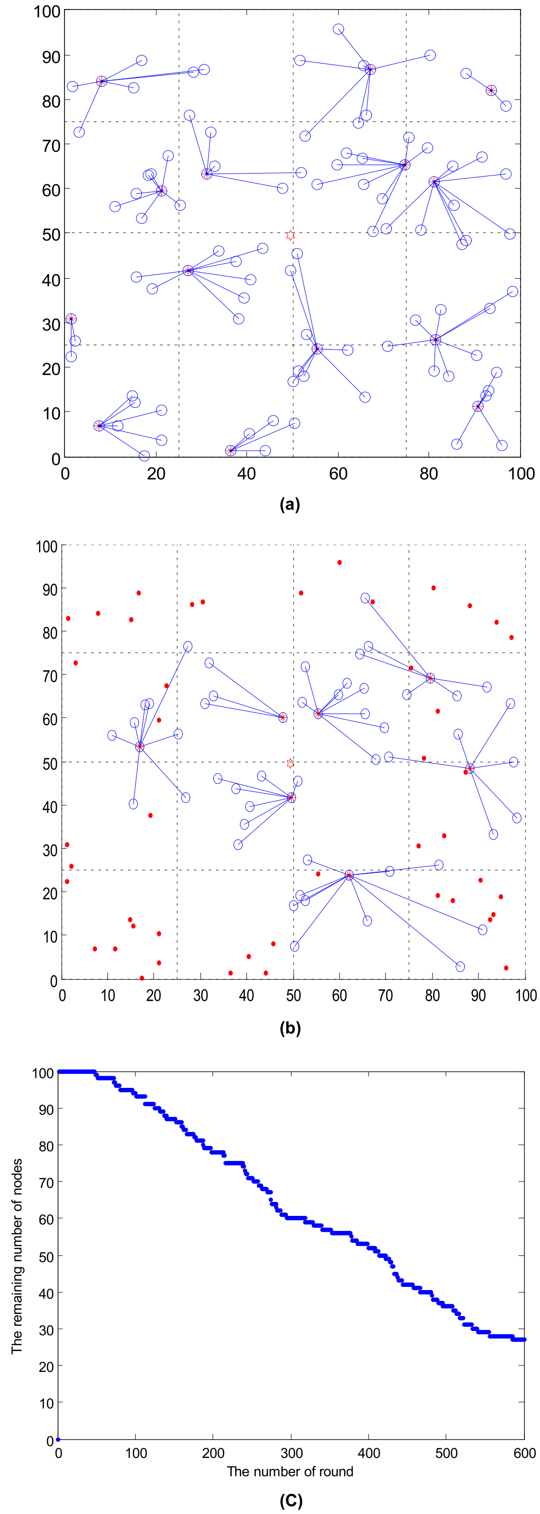


FIGURE 10. Clustering characteristics of the WSN with uneven nodes based on the density clustering. (a) The first round of cluster head distribution and clustering effect of unevenly distributed nodes based on the density clustering. (b) Node distribution and clustering at the 300th round. (c) The relationship between the number of rounds and the number of remaining nodes based on the density clustering.

Fig.12(a) is a method of [1] whose projection nodes are randomly generated. It can be seen from the figure, the locations of the projection nodes are very random, the sizes

TABLE 2. Simulation related parameters.

Experimental parameters	Scenes 1	Scenes 2
Area size	100m×100m	200m×200m
Number of nodes	100	400
Sink location	（50,50）	（100,100）
Initial energy	0.01J	0.01J
Data formation length	4000b	4000b
Control formation length	32b	32b
Send energy E_{Tx}	10^{-10}J/b	10^{-10}J/b
Receive energy E_{Rx}	10^{-10}J/b	10^{-10}J/b
Fusion energy E_{DA}	10^{-11}J/b	10^{-11}J/b
Dissipated energy E_{fs}	2×10^{-11}J/b	2×10^{-11}J/b
Critical energy Ed	10^{-5}J	10^{-5}J
Communication radius D	71m	71m

of the clusters formed are not balanced. Some grids have two or three projection nodes, some grids whose nodes are densely populated are without projection nodes. Fig.12(b) shows the clustering effect of the 300th round. Compared with the number of nodes in Fig.11(b), the number of nodes is significantly reduced. Most of the remote nodes died. Only the network near the sink is in operation and the information acquisition has been unbalanced.

Fig.13 shows the comparison the number of remaining nodes in scenario 1 between the proposed density clustering method and the random clustering method. It can be seen that the dead node of random clustering method appears relatively late, but the node's death rate declines faster. When the number of remaining nodes is 20 the network suddenly paralyzed. However, the number of death nodes based on node density clustering method dropped relatively slowly, making the network running longer, and the network did not abruptly terminate. Considering the number 20 of remaining nodes as the critical value of the normal operation of the network, the random clustering method has a life expectancy of nearly 400 rounds, the network lifetime based on density clustering is about 550 rounds, and the network lifetime based on the density clustering method is about 27% longer than that based on random clustering method.

In addition, as shown in Fig.8 and Fig.13, we can see that nodes based on random clustering method all die faster in different networks and will be paralyzed when there are about 20 surviving nodes. It is because that the randomly selected projection nodes have not any advantage and can not balance the energy consumption. As a result, the remaining little energy of the nodes is depleted immediately, causing the network to collapse instantly and all the nodes to die. Therefore it causes the curve to plummet. However, the number of death nodes in this paper's method drops more slowly, and the network can maintain a longer time.

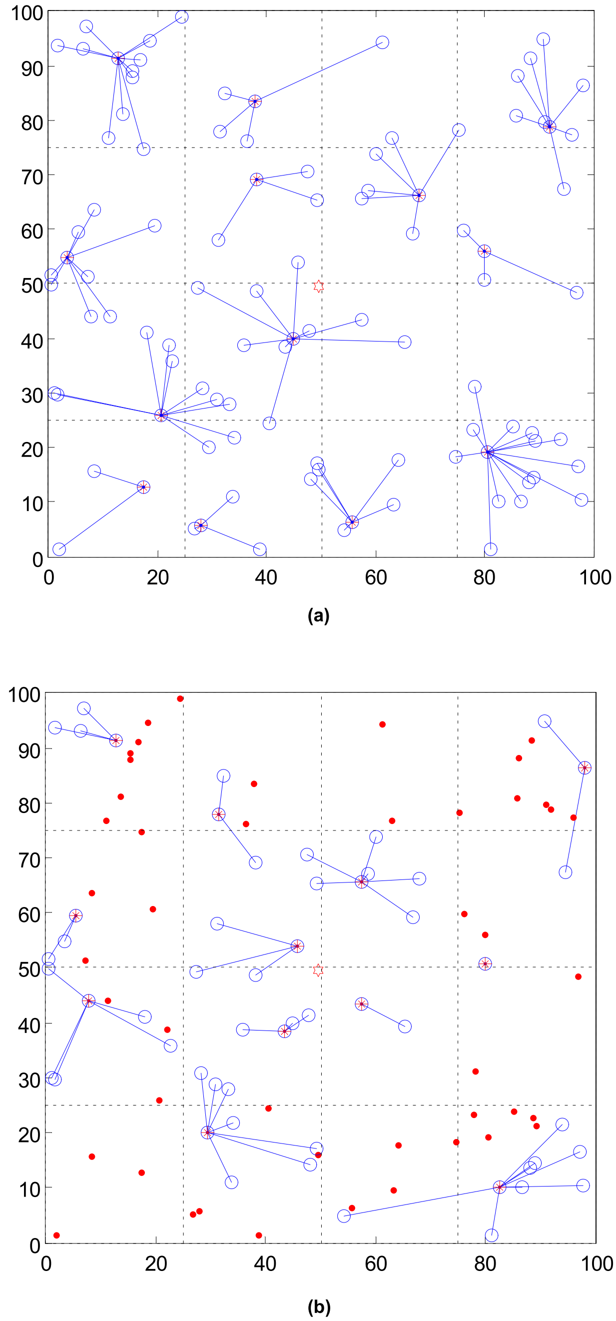


FIGURE 11. Clustering characteristics of the WSN with uneven nodes based on the density clustering (scene 1). (a) The first round of clustering effect based on the node density clustering method. (b) Cluster head distribution and clustering at the 300th round.

For scenario 2, the same non-uniform nodes are clustered based on node density clustering method and random clustering method. The cluster heads distribution and the clustering effect based on node density clustering method are shown in Fig.14.

Fig.15 compares the number of remaining nodes based on the node density clustering method and the random clustering

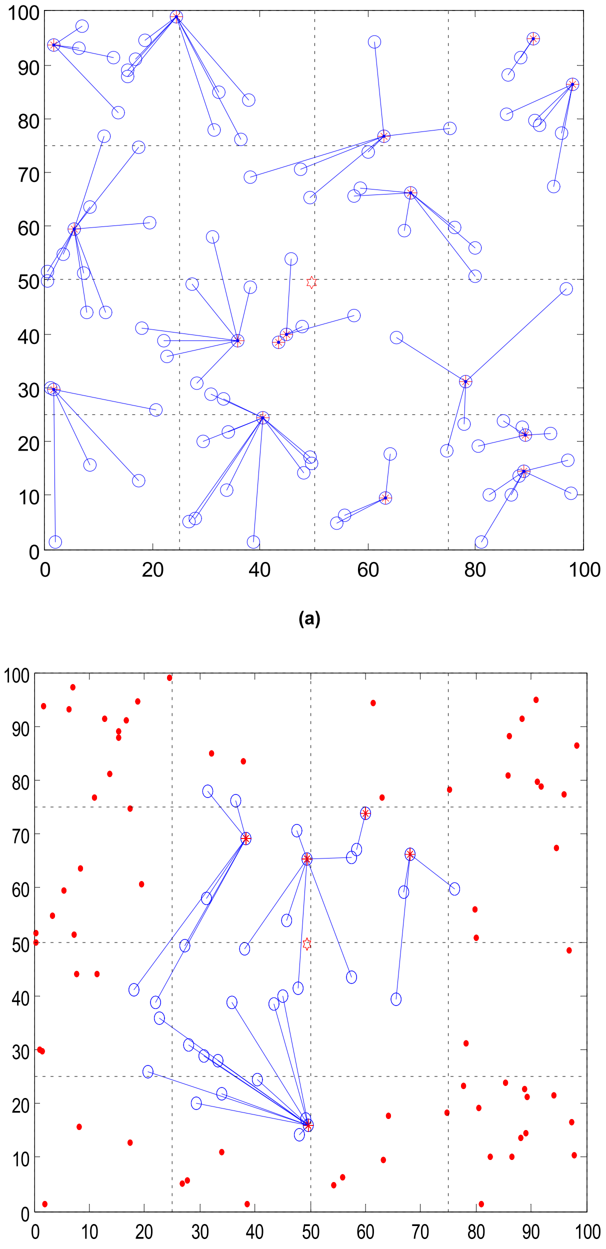


FIGURE 12. Clustering characteristics of the WSN uneven nodes based on the random projection nodes clustering (scene 1). (a) The first round of clustering based on the random projection nodes method. (b) Nodal distribution and clustering at the 300th round.

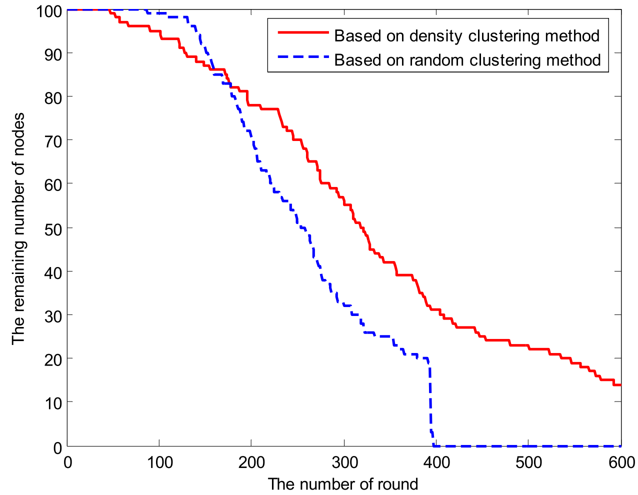


FIGURE 13. Comparison of the number of remaining nodes based on density-based clustering method and random clustering nodes method in WSN with uneven nodes (scene 1).

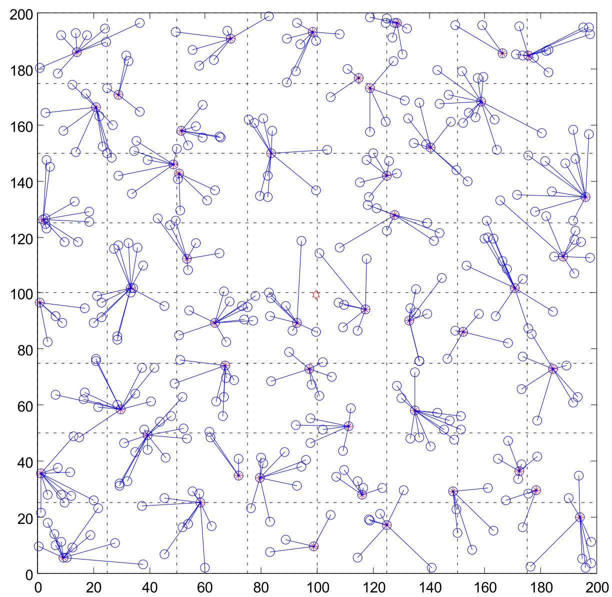


FIGURE 14. Cluster heads distribution and clustering of WSN based on the node distribution density (scene 2).

method in scenario 2. The number of remaining nodes based on the random clustering method still decreases rapidly. Similarly, when the number of remaining nodes is 20, the network is abrupt termination. Compared with the method based on random clustering, in the normal operation of the whole network, the method in this paper has more remaining nodes, the network running time is longer, the running state is better, the energy consumption more balanced.

From the curve of the total energy of the network with the changes of the number of rounds, the same conclusion can be obtained as shown in Fig.16. In different scenarios, the method of density even clustering proposed in this paper has higher total energy per round in the network running than that of the random clustering method. After the energy of the random clustering method runs out, the network using the even clustering method can still continue to run. This shows that the proposed method can reduce energy consumption and extend network lifetime.

C. ANALYSIS OF NETWORK ENERGY CONSUMPTION PARAMETERS

The energy consumption of the network is related to the size of the monitoring area, the number of configured nodes, the number of clusters, the threshold of density of

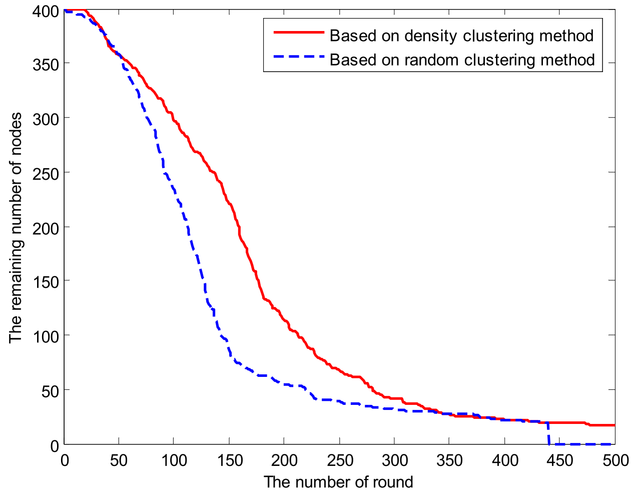


FIGURE 15. Comparison of the number of remaining nodes in the two clustering methods based on node density and random projection nodes (scene 2).

density-based even clustering, and the compression ratio of compressed data collection.The above parameters are separately studied using the density-based even clustering method (DEC) proposed in this paper, and compared with the random clustering method (RC) through simulation.

When the compression ratio of the compressed data acquisition is fixed, the total network energy of the DEC method changes with the number of rounds under the different monitoring area side lengths. As shown in Fig.17, the solid line represents the DEC method proposed in this paper, and the dashed line represents the RC method. The same color represents the same parameter.

It can be seen from the Fig.17 that when the length of the area decreases, the total energy is increased, that is, the energy consumption is reduced. Because the same number of nodes are deployed in a small area, the transmission distance is short and energy consumption is also small. This means that the more nodes deployed in the same area, the less energy will be consumed, but the cost of devices and deployment will increase, and the utilization of nodes will decrease. Therefore, within the allowable cost range, more nodes can be arranged to reduce energy consumption and extend the life of the network. Moreover, it can be seen from the figure that the total energy of the RC method is much lower than that of the DEC method of the same parameter.When the RC method runs to a critical number of nodes, the energy of the network will be exhausted instantaneously.

For the density even clustering method, the choice of the threshold (Th) also has an impact on the energy consumption. Fig.18 shows the total energy at different thresholds of the network with 100 and 400 nodes. As can be seen from Fig.18(a), when the thresholds are 1, 2 and 3, the total energy consumption is relatively small, and as the number of rounds increases, the decrease in the energy with a large threshold becomes more significant. This is because when

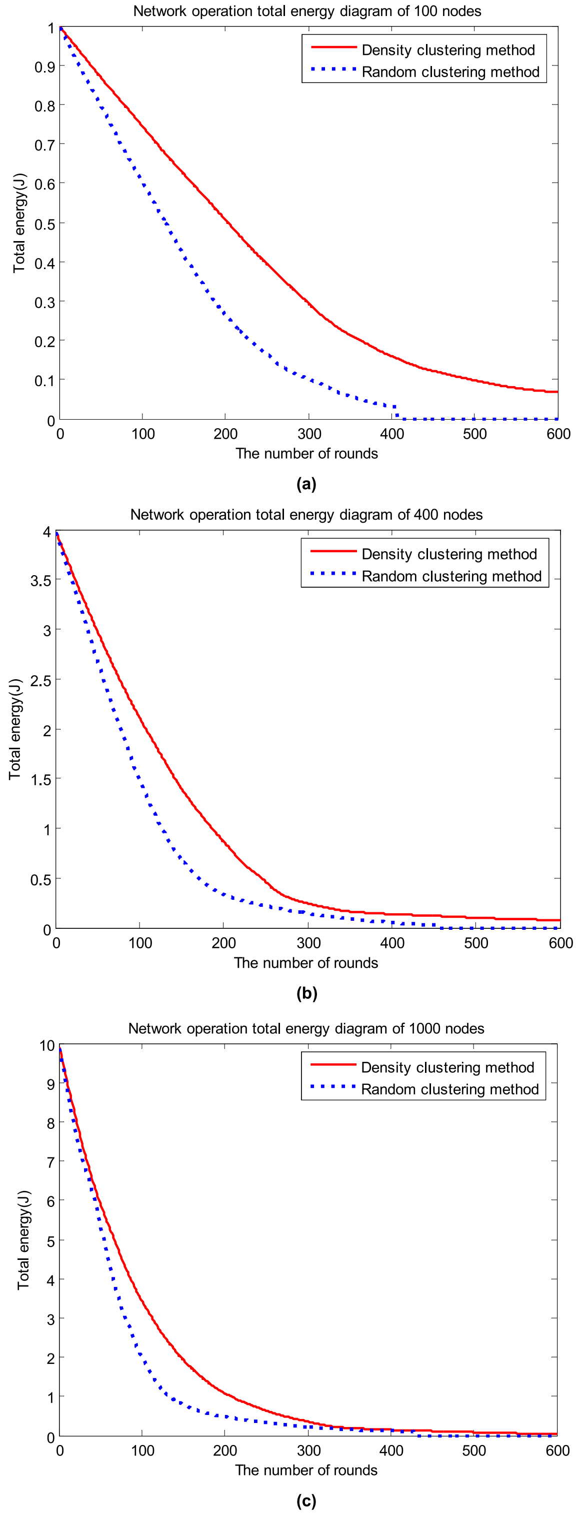


FIGURE 16. The total energy of the network varies with the number of rounds in the two methods (N=100,400,1000). (a) The total energy diagram of the running of a network with 100 nodes. (b) The total energy diagram of the running of a network with 400 nodes. (c) The total energy diagram of the running of a network with 1000 nodes.

the number of rounds increases, the number of remaining nodes continuously decreases, and the number of grids with nodes below the threshold gradually increases, the range of

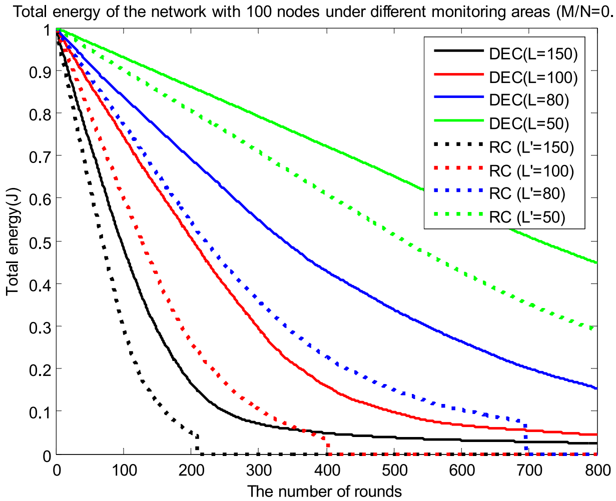
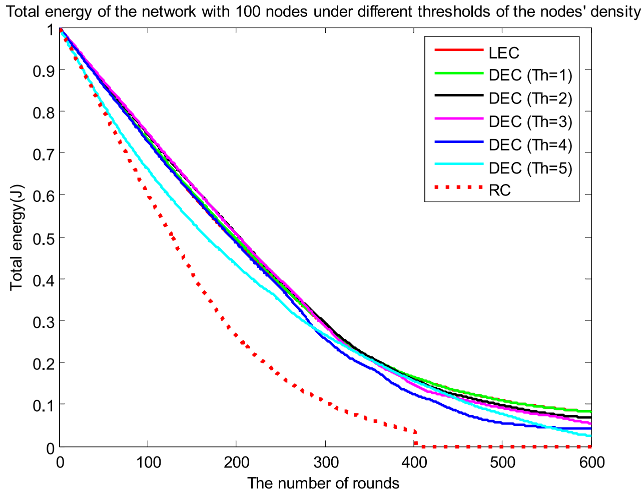
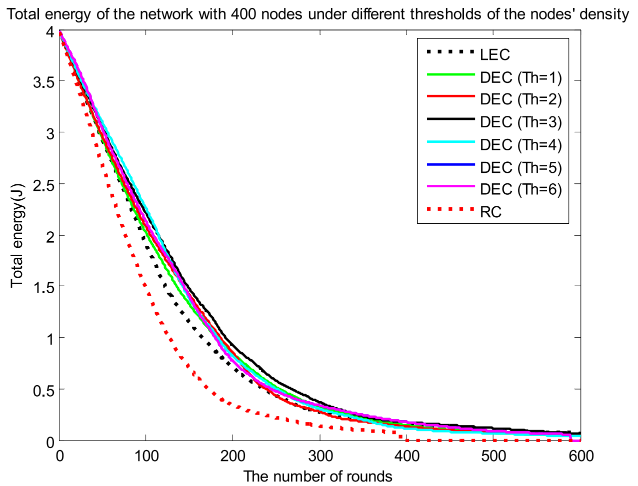


FIGURE 17. **The relation between the total energy and the number of rounds in the monitoring area of different side length of the DEC method and the RC method.**



(a)



(b)

FIGURE 18. **Changes of the total energy with the number of rounds for the three methods (Location-based even clustering (LEC), density-based even clustering (DEC) with different thresholds (Th), random clustering (RC)). (a) Comparison of the total energy of a network with 100 nodes. (b) Comparison of the total energy of a network with 400 nodes.**

forming a cluster increases, and the increase in the distance of transmitted data makes the energy consumption increase significantly. When the threshold is greater than 3, the energy consumption increases significantly and should not be set as a threshold. Moreover, as can be seen from the figure, the effect of location-based even clustering (LEC) is also similar to the DEC method with the thresholds of 1 and 2, but the total energy of the RC method is significantly lower than the proposed method in this paper.

Fig.18(b) is a comparison of the total energy of a network with 400 nodes. The networks with thresholds of 1 and 2 have less energy, and the thresholds of 3 and 4 have more energy. When the threshold is 6, the energy starts to decrease again. Moreover, the energy of the LEC method is also lower than that of DEC method ($1 \leq$threshold≤ 6), and the energy of the RC method is much lower than that of this paper. This shows that the choice of the threshold is related to the total number of nodes, and it is not a monotonous change. But above a certain threshold value, energy will be significantly reduced. Therefore, in actual application, it needs to be tested first and then selected as appropriate.

When the compression ratio of compressed data gathering changes, the energy of the network will also change. As shown in Fig.19, the energy of a network with 100 nodes changes with the number of rounds at different compression ratios. The solid line is the DEC method proposed in this paper. The dashed line is the RC method. As can be seen from the figure, as the compression ratio M/N increases, the energy of the network gradually decreases. Because the increase in the compression ratio means that the number of clusters increases, the amount of data to be transmitted increases, and thus the energy consumption also increases. Moreover, in the case of the same parameters, the RC method consumes much more energy than the DEC method. However, although the compression ratio is small, and the energy consumption is small, the amount of data acquired by the sink is also small, which results in an increase in the reconstruction error. Therefore, it is necessary to select a compression ratio as small as possible to meet the reconstruction error requirement, reduce energy consumption, and extend the network lifetime.

Under different compression ratios, how to select the number of nodes in the monitoring area can be obtained by (18).

$$h^3 = \frac{N^2R^2}{D^2} \tag{21}$$

h is the optimal number of clusters and can be considered as the number of CS required measurements M. R is the radius of the monitoring area, which is 1/2 of the side length L of the monitoring area. Let r be the compression ratio,

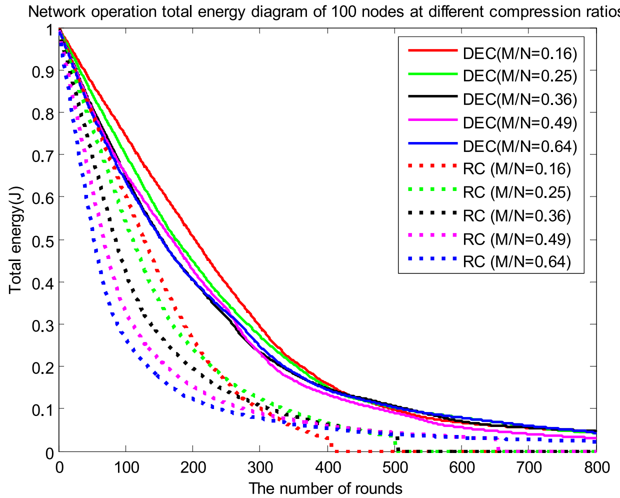


FIGURE 19. Comparison of changes of the total energy of the network with the number of rounds at different compression ratios of the DEC method and RC method.

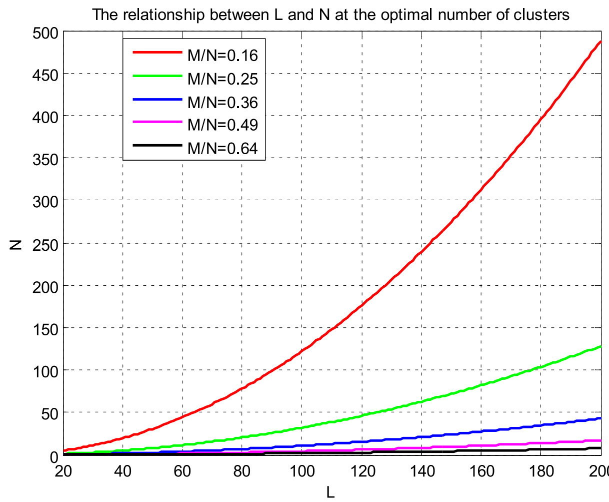


FIGURE 20. The relationship between L and N under different compression ratio at the optimal number of clusters.

then r=M/N=h/N, substituting (17), got

$$r^3N^3 = \frac{N^2L^2}{4D^2} \tag{22}$$

then

$$N = \frac{L^2}{4D^2r^3} \tag{23}$$

When the communication radius D of the node is determined, assuming $D = 50\sqrt{2}$, the relationship between L and N at different compression ratios can be obtained, as shown in Fig.20. According to this figure, we can conveniently find the approximate number of configured nodes in different applications.

Furthermore, according to (17), the total energy consumption of the network of the clustering method can be obtained.

$$E_{clustertotal} = k(2NE_{elect} + NE_{DA} + hE_{fs}D_{toBS}^2 + N\frac{NR^2}{2h^2}E_{fs}) \tag{24}$$

We can compare the proposed method with random walk (RW) data collection method. On the same condition, supposed h is the number of RWs and τ is the RW length. The energy consumption by sending one frame data to sink for each path includes the energy of transmitting and receiving data of (τ-1) nodes, the fusion energy of τ data, and the energy that the last node of RW sends data to the sink.

$$\begin{aligned} E_{path} &= (\tau-1)kE_{Tx} + (\tau-1)kE_{Rx} + \tau kE_{DA} + kE_{toBS} \\ &= (\tau-1)kE_{elect} + kE_{fs}\sum_{i=1}^{\tau-1} d_i^2 + (\tau-1)kE_{elect} \\ &\quad + \tau kE_{DA} + kE_{elect} + kE_{fs} \\ &= (2\tau-1)kE_{elect} + \tau kE_{DA} + kE_{fs}\sum_{i=1}^{\tau-1} d_i^2 + kE_{fs}D_{toBS}^2 \end{aligned} \tag{25}$$

According to [23] and setting the sink at the center of the monitoring area, the expected square distance as

$$E[d^2] = \frac{D^2}{2}, \quad E[D_{toBS}^2] = \frac{L^2}{6} \tag{26}$$

Substituting (26) into (25) and setting the length of RW be $\tau = Nss/h$, the total power consumption of RW is

$$\begin{aligned} E_{RWtotal} &= hE_{path} = h[(2\frac{N}{h} - 1)kE_{elect} + \frac{N}{h}kE_{DA} \\ &\quad + kE_{fs}(\frac{N}{h} - 1)\frac{D^2}{2} + kE_{fs}\frac{L^2}{6}] \\ &= k\{(2N-h)E_{elect} + NE_{DA} \\ &\quad + E_{fs}[(N-h)\frac{D^2}{2} + \frac{hL^2}{6}]\} \end{aligned} \tag{27}$$

Substituting (26) into (24) and R=L/2, then

$$E_{Clustertotal} = k(2NE_{elect} + NE_{DA} + hE_{fs}\frac{L^2}{6} + \frac{N^2L^2}{8h^2}E_{fs}) \tag{28}$$

The energy consumption comparison curves of the clustering method and the RW method given by (28) and (27) are shown in Fig.21 and Fig.22. The xlabel 'Number of groups' refers to the number of clusters for the clustering method, and the number of RWs for the RW method. They are consistent in principle and quantity.

As can be seen from Fig.21, the energy consumption of the clustering method decreases rapidly and then increases slowly with the increase of the number of clusters, so the clustering can reduce energy consumption, and the optimal energy consumption can be achieved by selecting an appropriate number of clusters. The energy consumption of the RW method varies monotonically with the number of RWs and is related to the area range. However, the energy consumption of the clustering method is lower than that of the RW method from a certain number of groups to the maximum number of groups. This is because the randomness of RW cannot select the optimal next node and the optimal path. It can also be seen that when the maximum number of groups, the energy

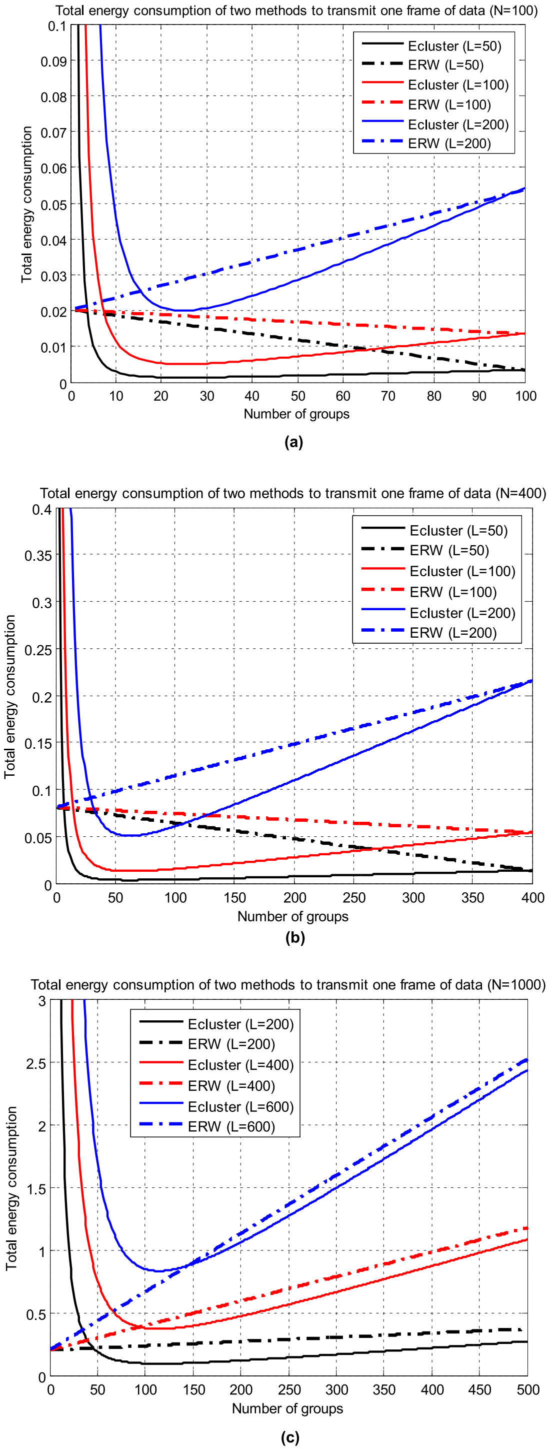


FIGURE 21. Relation of the total energy consumption with the number of groups under different numbers of the nodes in the clustering method and the RW method. (a) Comparison of the total energy consumption of the two methods for a network with 100 nodes (When the area side length L=50,100,200). (b) Comparison of the total energy consumption of the two methods for a network with 400 nodes (When the area side length L=50,100,200). (c) Comparison of the total energy consumption of the two methods for a network with 1000 nodes (When the area side length L=200,400,600).

consumption of the two methods is the same. Because if 100 nodes are divided into 100 groups, the two methods are actually the same, of course, the same energy consumption.

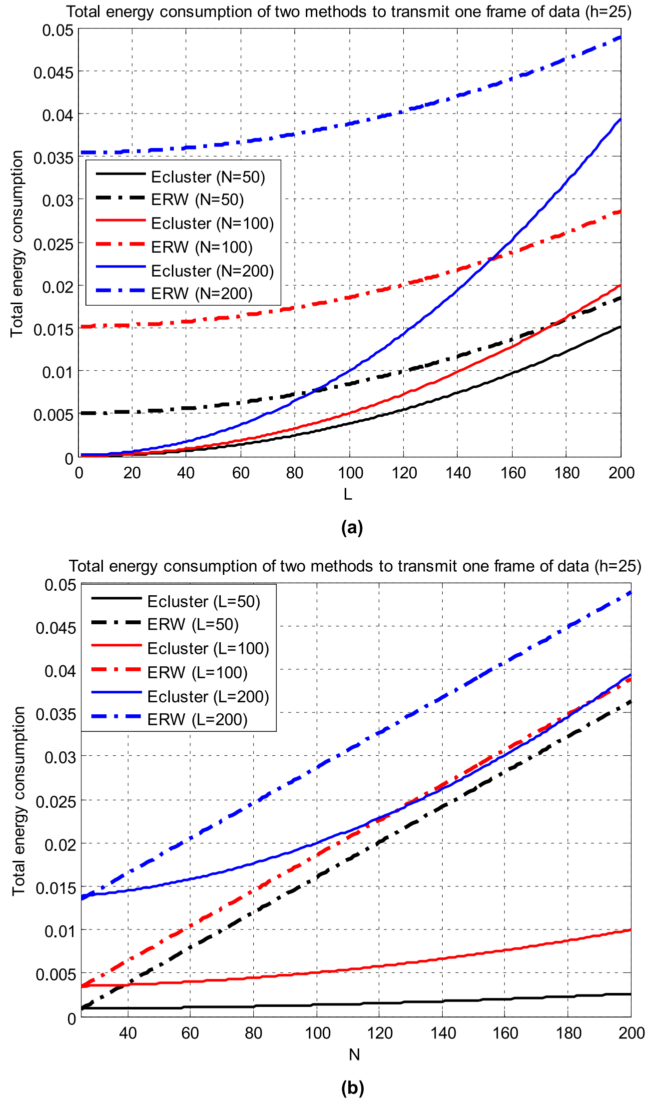


FIGURE 22. Comparison of the total network energy consumption of the DEC method and the RW method at a fixed number of groups. (a) The comparison of the total network energy consumption of the two methods with the change of the side length of the area (When the number of the nodes N=50,100,200). (b) Comparison of the total network energy consumption of two methods with the change of the number of nodes (when the side length of the area L=50,100,200).

Fig.22 shows the network energy consumption comparison curves for the DEC method and the RW method when the number of groups is fixed. Fig.22(a) shows the energy consumption of the network with the changes of the length of the area under different number of the nodes. Fig.22(b) shows the energy consumption of the network with the changes of the number of nodes under different area lengths. Since the number of the groups is 25, the number of the node also starts from 25. It can be seen that as the side length increases, the network energy consumption increases; as the number of nodes increases, the network energy consumption also increases. However the energy consumption of the RW method is always much greater than that of the DEC method. It is because that though the RW method adapts to the theory

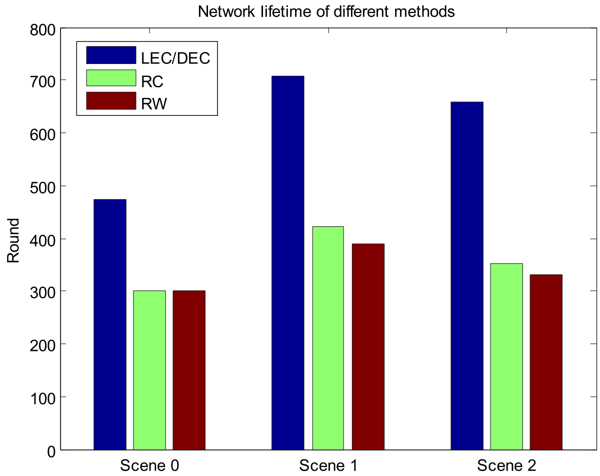


FIGURE 23. **Comparison of network lifetime of different methods.**

of random projection and reduces the power consumption of the sink broadcasting, the paths of RW are full of uncertainty due to the randomness of the RW, resulting in an imbalance of network energy consumption and shortening the network lifetime.

Next these methods are applied to compare network lifetimes. For Scene 0 of Table 1, the LEC method presented in this paper is compared with the RC method and the RW method. For Scene 1 and Scene 2 in Table 2, the DEC method proposed in this paper is used for comparison. As can be seen from the comparison result illustration in Fig.23, the network lifetime of the proposed method is much higher than the other two methods, which further demonstrates the effectiveness of the proposed method.

D. COMPRESSIVE DATA GATHERING BASED ON EVEN CLUSTERING

The specific implementation steps of compressive data gathering method based on even projection include the following:

(1) According to the network parameters, determine the number of grids.

(2) Even selection of projection nodes. For the WSN with uniform distribution of nodes, an even clustering method based on spatial locations is used. The WSN with uneven distribution of nodes adopt even clustering based on distribution density of nodes, thus the projection nodes are evenly distributed with more energy. The projection nodes are equivalent to the cluster head.

(3) Establishment and rotation of the cluster. After a projection node is selected for the first time, a cluster centered on the projection node can be established, and the rotation and rules of the next cluster can be determined.

(4) Construction of projection matrix. The sensor nodes that need to transmit data in each round of transmission are determined by the measurement matrix in terms of CS theory. Therefore, the structure of the measurement matrix determines the number of sensing nodes, the transmission distance and the energy consumption, as well as the performance of reconstruction.

(5) Routing mechanism of transmission node. The sensor node's transmission path, that is, the routing mechanism, is the key to the energy consumption and efficiency of the entire network transmission.

(6) The even clustering-based WSN data gathering. After the above process is completed, the data can be collected. Furthermore, all the information of all the clusters in the area can be collected, or only the data of the areas of interest can be collected. The division of clusters greatly increases the flexibility of network data collection and improves system efficiency.

According to the compressive data gathering process based on even clustering, Fig.24 shows the clustering effect of the WSN with 100 nodes. The nodes of different color represent nodes of different projection rows. The weighted sum of nodes of the same color is transmitted to the sink, the sink obtains a measurement. The corresponding weighted sums of all the nodes of different color are M measurements, and thus the signal reconstruction can be achieved.

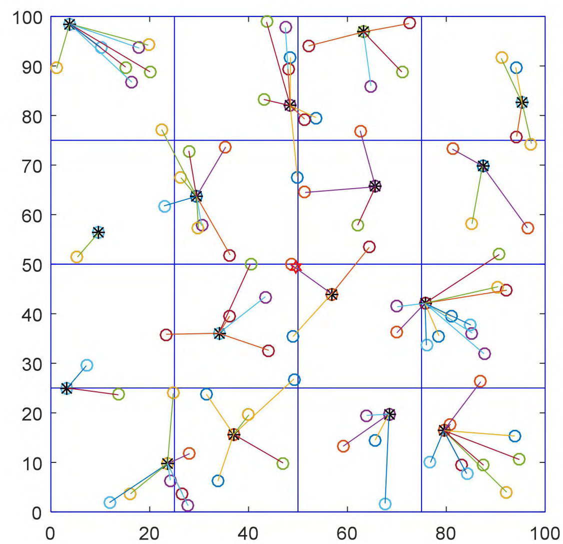


FIGURE 24. **Clustered schematic of compressive data gathering based on even clustering.**

In location-based clustering, the number of grids is the number of projection nodes needed, which ensures the number of projection nodes. However, in density-based clustering, the number of projection nodes is reduced due to disrupted grid constraints. How to guarantee to get M projection values is the first problem to be solved.

In Fig.24, the non-zero coefficient nodes in each row is distributed. First, different labels are set for the non-zero nodes in different rows. All projection nodes add up the weights of the same labels to obtain the measurement value. Therefore, for this method, the number of projection nodes has nothing to do with the number of measurements, but acts as a cluster head to aggregate the data.

This problem is related to the measurement matrix. In the non-sparse measurement matrix, the elements of the measurement matrix are non-zero. In the process of projection, there are weight coefficients for each of N nodes, so a weighted sum of N nodes can only get one measurement. To get M measured values, the weight sum is transmitted once every time slot. For sparse measurement matrix, some nodes have a measurement coefficient of 0, so they do not have to transmit data or just act as a route, sending only the weighted sum of the previous node. If a node's measurement coefficients are not zero in both rows, it can only send one weighted sum at a time, and the next weighted sum after a time slot. Obviously, if there is only one non-zero element in each column of the measurement matrix, the non-zero nodes of each row do not interfere with each other, and the transmission of M measurement values can be performed at the same time without the completion of the time slots. Thus this measurement matrix is the most efficient, this transfer is the fastest and more real-time.

VI. CONCLUSION

For the problems of random selection and unbalanced position of projection nodes, this paper proposes a compressed data gathering method based on even projection. For the WSN with uniformly distributed nodes, a location-based even clustering method is proposed. The clustering is implemented with the same size of the grids, which ensures the positional balance of the projected nodes. For the WSN with uneven distributed nodes, a node density-based even clustering method is proposed. The DEC method, taking into account the factors of location and density, reduces the energy consumption at isolated points, equalizes the energy, and extends the network lifetime. Moreover, The analysis and simulation of the relevant parameters affecting the network energy consumption were analyzed. Compared with the random projection node method and the random walk method, the proposed method performs well and the network lifetime is significantly extended. In the next step, we will consider the application of artificial intelligence to further optimize the routing topology of the network and make more in-depth research on signal reconstruction to obtain better compressed data collection results.

REFERENCES

[1] D. Ebrahimi and C. Assi, "Compressive data gathering using random projection for energy efficient wireless sensor networks," *Ad Hoc Netw.*, vol. 16, pp. 105–119, May 2014.

[2] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. IPSN*, Nashville, TN, USA, Apr. 2006, pp. 134–142.

[3] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 92–101, Mar. 2008.

[4] J. Luo, L. Xiang, and C. Rosenberg, "Does compressed sensing improve the throughput of wireless sensor networks?" in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 2010, pp. 1–6.

[5] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. 15th Annu. Int. Conf. (MobiCom)*, Beijing, China, Sep. 2009, pp. 145–156.

[6] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Efficient measurement generation and pervasive sparsity for compressive data gathering," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3728–3738, Dec. 2010.

[7] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[8] J. Haupt and R. Nowak, "Signal reconstruction from noisy randomized projections with applications to wireless sensing," in *Proc. IEEE/SP Workshop Stat. Signal Process.*, Jul. 2005, pp. 1182–1187.

[9] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Proc. 6th Int. Symp. Info. Process. Sensor Netw.*, 2007, pp. 331–339.

[10] T. Wimalajeewa and P. K. Varshney, "Wireless compressive sensing over fading channels with distributed sparse random projections," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 1, no. 1, pp. 33–44, Mar. 2015.

[11] R. Rong and H. Oh, "Adaptive sparse random projections for wireless sensor networks with energy harvesting constraints," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 1, pp. 113–122, Dec. 2015. [Online]. Available: https://link.springer.com/article/10.1186/s13638-015-0324-3

[12] M. T. Nguyen and N. Rahnavard, "Cluster-based energy-efficient data collection in wireless sensor networks utilizing compressive sensing," in *Proc. IEEE Military Commun. Conf.*, Nov. 2013, pp. 1708–1713.

[13] M. T. Nguyen, K. A. Teague, and N. Rahnavard, "CCS: Energy-efficient data collection in clustered wireless sensor networks utilizing block-wise compressive sensing,"*Comput. Netw.*, vol. 106, pp. 171–185, Sep. 2016.

[14] X. G. Wu, Y. Xiong, W. C. Huang, H. Shen, and M. X. Li, "An efficient compressive data gathering routing scheme for large-scale wireless sensor networks," *Comput. Elect. Eng.*, vol. 39, no. 6, pp. 1935–1946, 2013.

[15] S. Abbasi-Daresari and J. Abouei, "Toward cluster-based weighted compressive data aggregation in wireless sensor networks," *Ad Hoc Netw.*, vol. 36, no. 1, pp. 368–385, 2016.

[16] X. Xu, R. Ansari, A. Khokhar, and A. V. Vasilakos, "Hierarchical data aggregation using compressive sensing (HDACS) in WSNs," *ACM Trans. Sensor Netw.*, vol. 11, no. 3, 2015, Art. no. 45.

[17] D. Ebrahimi and C. Assi, "Optimal and efficient algorithms for projection-based compressive data gathering," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1572–1575, Aug. 2013.

[18] Z. Chen, G. Yang, L. Chen, and J. Xu, "Constructing maximum-lifetime data-gathering tree in WSNs based on compressed sensing," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 5, pp. 1–11, 2016.

[19] F. Xiao, G. Ge, L. Sun, and R. Wang, "An energy-efficient data gathering method based on compressive sensing for pervasive sensor networks," *Pervasive Mobile Comput.*, vol. 41, pp. 343–353, Oct. 2017. [Online]. Available: https://www.researchgate.net/publication.

[20] J.-S. Lee and T.-Y. Kao, "An improved three-layer low-energy adaptive clustering hierarchy for wireless sensor networks," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 951–958, Dec. 2016.

[21] W. Wei, W. Dan, and J. Yu, "Energy efficient distributed compressed data gathering for sensor networks," *Ad Hoc Netw.*, vol. 58, pp. 112–117, Apr. 2017.

[22] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Hawaii Int. Conf. Syst. Sci.*, Jan. 2000, pp. 3005–3014.

[23] R. B. Fletcher, "Energy efficient compressed sensing in wireless sensor networks via random walk," Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. Tennessee Chattanooga, Chattanooga, TN, USA, 2011.

[24] M. Sartipi and R. Fletcher, "Energy-efficient data acquisition in wireless sensor networks using compressed sensing," in *Proc. Data Compress. Conf.*, 2011, pp. 223–232.

[25] M. T. Nguyen, "Minimizing energy consumption in random walk routing for wireless sensor networks utilizing compressed sensing," in *Proc. Int. Conf. Syst. Syst. Eng.*, Jun. 2013, pp. 297–301.

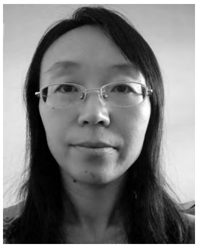
[26] M. T. Nguyen and K. A. Teague, "Compressive sensing based random walk routing in wireless sensor networks," *Ad Hoc Netw.*, vol. 54, pp. 99–110, Jan. 2017.

[27] X. Meng, X. Shi, Z. Wang, S. Wu, and C. Li, "A grid-based reliable routing protocol for wireless sensor networks with randomly distributed clusters," *Ad Hoc Netw.*, vol. 51, pp. 47–61, Nov. 2016.

[28] R. Xie and X. Jia, "Transmission-efficient clustering method for wireless sensor networks using compressive sensing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 806–815, Mar. 2014.

[29] S. P. Tirani and A. Avokh, "On the performance of sink placement in WSNs considering energy-balanced compressive sensing-based data aggregation," *J. Netw. Comput. Appl.*, vol. 107, pp. 38–55, Apr. 2018.

[30] H. Zheng, F. Yang, X. Tian, X. Gan, X. Wang, and S. Xiao, "Data gathering with compressive sensing in wireless sensor networks: A random walk based approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 1, pp. 35–44, Jan. 2015.
[31] P. Zhang, S. Wang, K. Guo, and J. Wang, "A secure data collection scheme based on compressive sensing in wireless sensor networks," *Ad Hoc Netw.*, vol. 70, pp. 73–84, Mar. 2018.
[32] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comp. Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, May 2008.
[33] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[34] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[35] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[36] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.
[37] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Process.*, vol. 86, no. 5, pp. 533–548, 2006.
[38] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Info. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[39] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for Bernoulli and SubGaussian ensembles," *Constructive Approx.*, vol. 28, no. 3, pp. 277–289, 2008.
[40] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
[41] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[42] J.-S. Lee and W.-L. Cheng, "Fuzzy-logic-based clustering approach for wireless sensor networks using energy predication," *IEEE Sensors J.*, vol. 12, no. 9, pp. 2891–2897, Sep. 2012.
[43] D. C. Hoang, R. Kumar, and S. K. Panda, "Realisation of a cluster-based protocol using fuzzy C-means algorithm for wireless sensor networks," *IET Wireless Sensor Syst.*, vol. 3, no. 3, pp. 163–171, 2013.



JIANHUA QIAO received the B.S. degree from the Shenyang Institute of Chemical Technology, Shenyang, China, in 1997, and the M.S. degree from the Taiyuan University of Science and Technology, Taiyuan, China, in 2002. She is currently pursuing the Ph.D. degree with Taiyuan University of Technology, Taiyuan. She is currently an Associate Professor and a Master Supervisor with the Taiyuan University of Science and Technology. Her research interests include wireless sensor networks, compressed sensing theory applications, and embedded systems.

XUEYING ZHANG received the B.S. and M.S. degrees from the Harbin Institute of Shipbuilding Engineering, Harbin, China, in 1985 and 1988, respectively, and the Ph.D. degree from Harbin Engineering University, Harbin, in 1998. She is currently a Professor and a Ph.D. Supervisor with the Taiyuan University of Technology, Taiyuan, China. She has authored and co-authored over 100 conference and journal papers, presided over the completion of more than 20 national and provincial projects, and has been granted over three national invention patents in her research areas. Her current research interests include speech signal processing, big data applications, multimedia communications, and Internet of Things technology.