Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8887632, 14 pages
https://doi.org/10.1155/2020/8887632

Research Article

Robust and Secure Data Fusion Algorithm Based on Intelligent Sensing in Wireless Sensor Networks

Yong Cheng[1], **Jun Wang**[2], **Shuqiang Ji**[2] **and Ling Yang**[1]

[1]*Jiangsu Key Laboratory of Agricultural Meteorology, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[2]*Department of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China*

Correspondence should be addressed to Yong Cheng; yongch1314@163.com

Received 13 May 2020; Revised 23 September 2020; Accepted 5 November 2020; Published 30 November 2020

Academic Editor: Panagiotis Sarigiannidis

Copyright © 2020 Yong Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Presently, the wireless sensor network (WSN) plays an important role in smart farming. However, due to the limitation of wireless sensor network resources, the time and space correlation of data acquisition is strong. In order to reduce the number of nodes participating in data compression, the robust and secure data fusion algorithm based on intelligent sensing is proposed. The algorithm can divide the whole network into many clusters. In order to maintain energy balance of nodes in the cluster, the probability of each node in each cluster participating in each round of data collection is computed according to the residual energy of the node. On the sink node, the number of sampling rounds of joint reconstruction of collected data is designated according to the application requirements and reconstruction accuracy requirements, and the number of nodes participating in is further reduced. The simulation results show that the number of nodes participating in the data collection of the proposed scheme in this paper is lower than that of the ordinary intelligent sensing LEACH data acquisition scheme. Meanwhile, the proposed scheme can dramatically extend the network lifetime. This paper provides an insight into various needs of WSN used in agriculture and challenges involved in the deployment of WSN.

1. Introduction

The wireless sensor network consists of many low-cost tiny sensor nodes, with the ability to compute, communicate, and acquire data [1–3]. It is used for robust data collection, and it has been widely used in military, environmental monitoring, disaster monitoring, industrial production monitoring, healthcare, smart homes, and other fields in social life. Sensor nodes have limited energy, because its power supply is a battery, and these nodes formed in wireless multihop ad hoc networks through collaborating with each other. Data gathering in the wireless sensor network is a typical form of transmission. In meteorological and environmental monitoring, all nodes periodically send data to the sink node, until the lifetime of the network is over. The real-time requirements for data transmission are often lower, but it is required that each node must periodically send information to the sink node. In meteorological-environmental monitoring, it is required to deploy a large number of meteorological sensor nodes, which will increase network deployment costs. On the other hand, the bandwidth of the network and the energy of the sensor nodes are very limited; thus, how to use the limited resource efficiently to transmit and process meteorological data which are collected by the nodes in the network is one of the most important problems in meteorological sensor network research. In a microclimate sensing network observation system, the spatiotemporal continuity exists between the meteorological factors which leads to the massive node data with high temporal and spatial correlation and high redundancy. Enabling a certain degree of error and time delay of the premise and fusing the meteorological data can improve the data collection rate and reduce the energy consumption, which can prolong the network life [4–10].

Donoho et al. proposed the theory of compressive sensing [11], which provides a new way of thinking data fusion for a meteorological sensor network. The theory states that as long as the signal is sparse or the sparseness can be represented by a sparse basis, you can use a measurement matrix

which is not related to the sparse basis to project this high dimension signal into a low-dimensional space and then by solving a nonlinear optimization problem to reconstruct the original signal from a few projections with high probability. Compressive sensing theory has been applied in wireless sensor networks [12–20], in order to ensure that the original signal can be under the premise of the sparse representation; it applies an encoding algorithm with low complexity in computing and storage capacity of sensor nodes, obtaining vector projection, and then transmits it to the sink node. Due to the idea that the sink nodes' energy is adequate, its computation and storage capacity are strong; hence, it can run a higher complex degree decoding algorithm to recover the original signal. By this way, it can significantly reduce the network data transmission times and energy consumption and also can prolong the lifetime of the network.

Due to the dense deployment of the sensor nodes, there is a lot of spatial correlation between the sensing data [21–26]. Many researches use data spatial correlation to compress the data and use the intelligent sensing theory in the data collection in wireless sensor networks to get a significant compression effect [27–30]. However, the existing research on the data gathering in large-scale wireless sensor networks based on compressive sensing is limited in the simple application of intelligent sensing theory to the process of data collection, and the improvement of network performance is not obvious. Compressive sensing theory is a new information acquisition theory, which is also in the rapid development. If the theory of intelligent sensing is applied in the actual data collection in large-scale wireless sensor networks, there are still many problems to solve and perfect. Therefore, based on the intelligent sensing of wireless sensor network data collection, related research has important practical application value and academic significance. This paper provides an insight into various needs of WSN used in agriculture and challenges involved in the deployment of WSN. Smart farming (SF) has played a major role to enhance more production in the field of agriculture. The solution proposed in this work has been a WSN composed of several units with different sensors, allowing the monitoring of the farming procedure. These sensors send their sensing data to a coordinating unit that processes them and sends them to a database so that any farmer can obtain information about the state of his agricultural exploitation and can act on it adequately.

The remainder of this article is organized as follows. We review related work in Section 2. In Section 3, we describe the robust data fusion method based on intelligent sensing. Section 4 presents the simulation results to illustrate the key features as well as the performance of the designed rating protocol. Finally, conclusions are drawn in Section 5.

2. Related Works

At present, the research on a data collection problem in large-scale wireless sensor networks mainly focuses on three aspects: (1) improving the compression performance of sensing data, so as to reduce the number of the collecting measurements [31, 32]; (2) designing a better projection matrix to reduce the transmission cost of a single measurement [33]; and (3) designing a better routing strategy to match the projection matrix and routing strategy [34–36].

Based on the intelligent sensing data collection process, values for almost all nonzero elements in the measurement matrix are data collected by dense random projection scheme. A dense random projection data collection scheme is used in References [37, 38].

A lot of data collection strategies for sparse random projection are also proposed, such as in Reference [34]; data collection is based on sparse random projection. Sparse random projection can reduce the transmission cost of a single measurement. However, it needs to collect much data. So it cannot improve the network performance. At the same time, the sparse random projection is not suitable for the data collection of a sparse signal. In References [39, 40], the authors also make the measurement matrix become more sparse, which can reduce the communication cost of a single measurement. Due to the randomness of the nonzero elements in the sparse random projection, it is very difficult to match the routing paths in the matrix and real networks.

The design principle of the measurement matrix is to satisfy the RIP (Restricted Isometry Property) condition or low correlation with arbitrary orthogonal representation, whether it is a dense random projection or a sparse random projection. This is also the reason why the measurement matrix cannot be sufficiently sparse, and we can only use the measurement matrix to match the application scene.

References [8, 14, 41–43] proposed an information acquisition scheme, which combines intelligent sensing and LEACH (Low-Energy Adaptive Clustering Hierarchy) for underwater sensor networks. The scheme in [14] is called CS-LEACH. The main idea of this information acquisition scheme is using the LEACH algorithm to divide the sensor network into M clusters at first. In the stable data acquisition phase, the nodes of each cluster use the independent and identically distributed Bernoulli random generator to generate probability p; these nodes of the cluster participate in the data collection with the probability p in each round. The sink node collects data from M cluster head nodes. Each data that the sink node received is a weighted sum of data which are collected by the node participants in collection of the clusters. So the vector of the measured value received by the sink node can be represented as follows:

$$y(n) = \phi(n)u(n) + z(n), \tag{1}$$

where n represents rounds, $z(n)$ represents the noise caused by surroundings, $u(n)$ represents the data collection in the whole sensor network, and $\phi(n)$ represents the measurement matrix; the procedure of establishing the measurement matrix $\phi(n)$ is as follows: if node j belongs to the ith cluster and this node will participate in the data collection, the corresponding $\varphi_{i,j}$ will be set to 1; otherwise, the $\varphi_{i,j}$ will be set to 0. Therefore, the data that the sink node receives from each cluster head node is $y_i = \sum_{j=1}^{N} \varphi_{ij} u_j$.

After a data collection round, the sink node can use an orthogonal matching pursuit algorithm (OMP) to reconstruct the information of the whole environment monitoring area

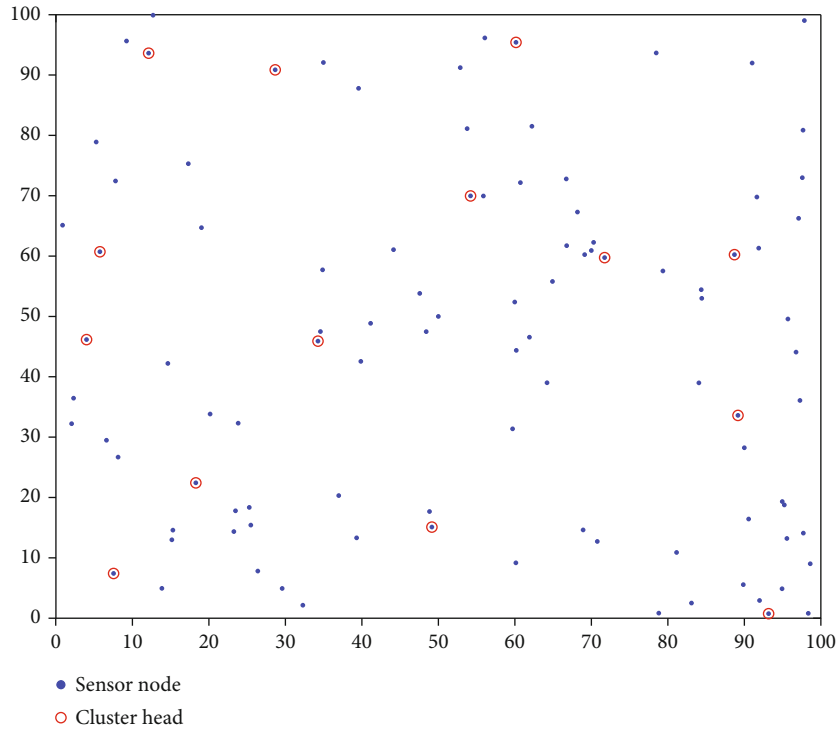

Figure 1: Wireless sensor network developed on a rectangular area.

according to the measurement vector $y(n)$ that is received by the sink node, measurement matrix $\phi(n)$, and basis ψ.

The intelligent sensing theory is introduced into the data fusion of the meteorological sensor network. The temporal and spatial correlation of the cluster head nodes is intelligent and fused, the data transmission of the cluster head is reduced, and the network lifetime is prolonged by constructing the diagonal Gauss matrix. In this model, each node in the model needs to be involved in the process of data acquisition, and how to reduce the amount of data transmission and reduce the number of nodes participating in data acquisition is a problem that needs to be considered.

However, the CS-LEACH does not give the process of constructing the measurement matrix $\phi(n)$ in detail, and also, it does not take the redundant energy of nodes into consideration during the process of constructing the measurement matrix $\phi(n)$. There is no solution to solve the possible abnormal data during the process of collecting data. On the sink node side, the CS-LEACH just reconstructs the information alone for each round of sampling data and does not consider the temporal correlation of data.

So this paper proposes the energy balance data fusion algorithm based on intelligent sensing (EBDFACS), in order to reduce the number of nodes that participates in data compression.

3. Robust Data Fusion Method Based on Intelligent Sensing

3.1. Problem Description. In an enviromental monitoring area with the length L, the width W, as shown in Figure 1, we can collect the meteorological data of the environment monitoring area such as temperature and humidity by deploying wireless sensor nodes regularly on the monitoring area. These deployed sensor nodes can communicate with sink nodes which are located in a point outside of the monitoring area. By selecting part of all nodes to participate in data collection and fusion processing, the problem we need to solve is how to decrease the number of the nodes that participate in data collection in each round and get the information of all nodes in the meteorological environment monitoring at the sink node finally.

3.2. Method Model. It divides the entire monitoring area into many cells of $u \times u$; if we need the meteorological element information of every cell, we need to deploy a sensor node in every cell to collect the data and transmit the data to the sink node. The number of cells in the entire monitoring area is $N = (L \times W)/u^2$.

In temperature monitoring, for example, $x(l, w)$ represents the temperature value collected by the node in location (l, w), where l and w are the coordinates of the node. So the data of the entire area in a sampling round can be represented as follows:

$$F = \begin{bmatrix} x_{1,1} & x_{1,2} \cdots & x_{1,j} \\ x_{2,1} & x_{2,2} \cdots & x_{2,j} \\ \vdots & \cdots & \vdots \\ x_{i,1} & x_{i,2} \cdots & x_{i,j} \end{bmatrix}. \quad (2)$$

It is convenient to express data collected from the network in the form of a vector $f = \mathrm{vec}(F)$; in the formula,

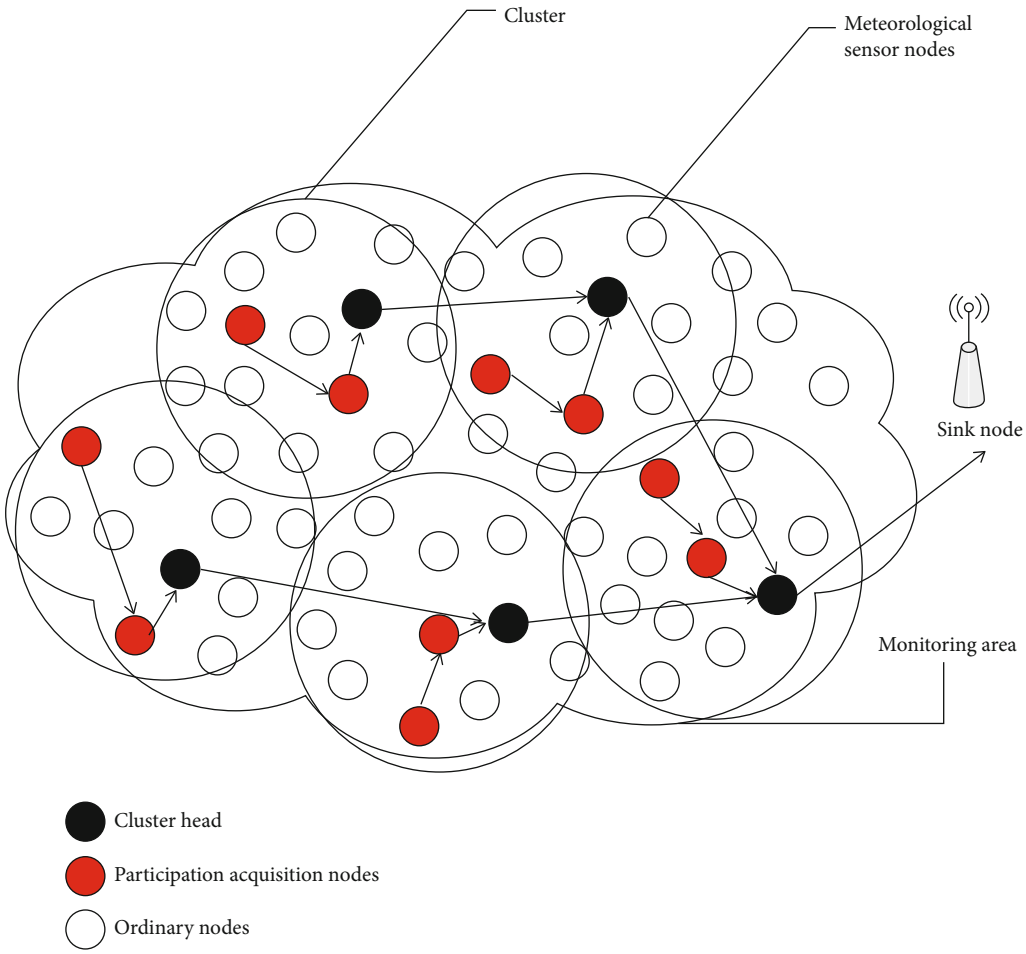

Figure 2: Energy balance data fusion model based on intelligent sensing.

$\mathrm{vec}(F)$ represents the conversion of the matrix F as column vectors to a one-dimensional column vector. The expression is shown as follows:

$$\mathrm{vec}(F) = \left(x_{1,1}, x_{2,1}, \cdots x_{i,1}, x_{1,2}, \cdots x_{i,2}, \cdots x_{i,j}\right)^T. \quad (3)$$

The expression $(\cdots)^T$ represents the vector transpose. As shown in Figure 2, in a sampling round, the nodes collect data with spatial continuity.

When the nodes are deployed evenly in the network, first of all, the clustering algorithm is used to select the cluster head, and the M cluster heads are selected in the cluster head election. Namely, M cluster structures are established in the whole network.

After the establishment of the cluster structure, the stable data collection and transmission stage is started. In this stage, we can combine the intelligent sensing random measurement process with the random selection of nodes in the cluster, in which not every node in clusters needs to be involved in the data collection. Each cluster node generates a random number, and compared with the probability threshold, it decides whether this node anticipates the data acquisition process in the sampling round. A weight coefficient is set on every node. The node multiplies the weight coefficients of the node by its own collecting value after collecting the data value and then sends the results to the next node. The next node multiplies the weight coefficients of the node by its own collecting value, and the result includes the value of the parent nodes. The results continue forwarding. In the process of forwarding, the data is processed until transmitted to the cluster head. The cluster head node receives the fusion result and sends the result value to the sink node by the shortest path algorithm. This process corresponds to the random measurement process of intelligent sensing, equivalent to building a measurement matrix ϕ to observe the data collected from the nodes for the entire network, that is, $y = \phi f$, $y \in R^M$ $M \prec \prec N$. Each node involved in the data acquisition in a cluster fuses into a set of data and sends this set to the sink node, and the sink node in a sampling turn can receive M data in all.

Due to the data sequence existing in temporal correlation, this temporal correlation can be used to reconstruct the sampled data, repeat data acquisition, and transfer operation by T sampling rounds totally. The sink node received these fusion results of the T times sampling round; then, it ran a decoding algorithm of compression sensing. Moreover, it can recover data of T sampling rounds by joint

reconstruction in the whole network. The sampling round number T can change according to application requirements. If it needs to reconstruct the original data sequence effectively, it is required to meet the prerequisites for intelligent sensing; the dimension of measured values of M is the number of clusters that need to satisfy inequality as follows:

$$M > cK \log^{N/K}, \tag{4}$$

where c is a constant and it is more than 1, N is the length of the original data sequence, and K is the data sparse degree. An energy balance data fusion model based on intelligent sensing is shown in Figure 2.

When the nodes in the clusters collected abnormal data, it will destroy the sparse of sampling data of the entire network. The abnormal data are transmitted to the sink node, the reconstruction algorithm on the sink node cannot effectively recover the abnormal data, and the data error between the recovered data and the original collected data is too large; it cannot meet the requirements of the application; at this time, it needs to deal with the abnormal data, and the abnormal value will be discarded or sent to the sink node to make early warning.

It sets an exception threshold e in the entire network, and the nodes compare the collected data to the threshold e; when the acquisition data x_n is larger than the threshold e, then x_n is identified as abnormal data and needs to be processed. Then, the node does not need to participate in the data overlay fusion. The fusion data is transmitted directly to the next node which is involved in the data acquisition process, and the node immediately discards the abnormal data or directly transmits it to the sink node for processing.

3.3. Algorithm Description

3.3.1. Observation Measure. In the data stable transmission stage, the cluster head knows the location information of each member node, and the member nodes send the information to the cluster head node in a certain sequence. Each sensor node has a random number generator to generate random number w; if the random number w is less than the probability threshold p, the node is involved in the data acquisition; the calculation of the probability threshold p can be expressed as follows:

$$p = \max\left(\frac{E_{\text{resident}}}{E_{\text{o}}}, C_{\text{prob}}\right), \tag{5}$$

where E_{resident} is the residual energy of the node, E_{o} is the initial energy of the node, and C_{prob} is the probability of the nodes participating in the data collection. $E_{\text{resident}}/E_{\text{o}}$ is the percentage of the remaining energy and the initial energy of the nodes. Each of these sampling round nodes with higher residual energy values has higher probability to participate in data acquisition, and the counter is involved in the control of the number of data acquisition nodes, achieving balanced energy consumption. Since the process of clustering is random, w is also randomly generated, so the sensor node selection based on residual energy cannot destroy this randomness. The number of nodes participating in the data acquisition is consistent in each cluster; each node has a weight coefficient φ_n, and the coefficient φ_n will be set as 1 when the cluster node will participate in the data acquisition round; the cluster head specifies a starting node; the path from the starting node to a cluster head is according to the shortest route principle; the node-collected data x multiplied by the weight coefficient φ_n is sent to the next participating node; the next node multiplies the weight coefficients of the node by its own collected value; and the result is added to the data forwarded by the former node. The results continue forwarding; the final results reach the cluster head node. The cluster head node will send the result of the fusion and the node number of the data acquisition nodes to the sink node.

In a sampling round, the weight coefficients of each node in the whole network can be abstracted into matrix ϕ, and the matrix ϕ can be expressed as follows:

$$\phi = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots \varphi_{1,j} & \cdots & \varphi_{1,N} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots \varphi_{2,j} & \cdots & \varphi_{2,N} \\ \vdots & & \cdots \varphi_{i,j} & \cdots & \vdots \\ \varphi_{M,1} & \varphi_{M,2} & \cdots \varphi_{M,j} & \cdots & \varphi_{M,N} \end{bmatrix}, \tag{6}$$

where M is the number of clusters within the whole sensor network and N is the number of nodes in the network sensor. Each row in the matrix corresponds to each cluster in the network, and each column corresponds to each sensor node. In the matrix ϕ, if the node j belongs to the cluster i and the node is involved in the sampling data acquisition, the weight coefficient $\varphi_{i,j}$ is 1 and the other position is 0, and if one node weight coefficient is 0, the node does not participate in data acquisition.

Matrix ϕ is corresponding to the observation matrix of the compressive sensing, and the data fusion process is corresponding to the random measurement of the intelligent sensing. The received data which was sent by M clusters in the sink node is as follows:

$$y = \phi f = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots \varphi_{1,j} & \cdots & \varphi_{1,N} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots \varphi_{2,j} & \cdots & \varphi_{2,N} \\ \vdots & & \cdots \varphi_{i,j} & \cdots & \vdots \\ \varphi_{M,1} & \varphi_{M,2} & \cdots \varphi_{M,j} & \cdots & \varphi_{M,N} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ \vdots \\ x_N \end{bmatrix}. \tag{7}$$

The fusion result of the cluster head node transmitted to the sink node is $y_i = \sum_{j=1}^{N} \varphi_{i,j} x_j$. x_j is the data that is collected from the node of the ith cluster, and $\varphi_{i,j}$ is the weight coefficient of the corresponding node in the ith cluster. y is the measurement result. y is also the data sequence of the fusion value collected from each cluster. y is a $M \times 1$ column vector, where $y \in R^M$. After the completion of a sampling round,

Table 1: Experimental parameters.

Parameter name	Parameter definition	Value (unit)
K	Sparseness	2
E	Anomaly threshold	40 (°C)
E_{elec}	RF energy coefficient	50 (nJ/bit)
ε_{fs}	Power amplifier circuit energy coefficient under free-space model	10 (pJ/bit/m^2)
ε_{mp}	Power double fading model amplifying circuit energy factor	0.0013 (pJ/bit/m^4)
d_0	Distance threshold	$\sqrt{\varepsilon_{\text{fs}}/\varepsilon_{\text{mp}}}m = 87$ (m)
E_0	Initial energy	0.5 (J)
N	The total number of nodes	100

each cluster head node sends the fusion data to the sink node. After receiving the data from the sample round, the node id can be used to rebuild the observation measurement matrix ϕ, and the data from each cluster makes the measurement value vector y.

$$y = \phi f = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots 0 & \cdots & \varphi_{1,N} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots \varphi_{2,j} & \cdots & \varphi_{2,N} \\ \vdots & & \cdots \varphi_{i,j} & \cdots & \vdots \\ \varphi_{M,1} & \varphi_{M,2} & \cdots \varphi_{M,j} & \cdots & \varphi_{M,N} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_u \\ \vdots \\ x_N \end{bmatrix}. \tag{8}$$

When the collected data compared with the abnormal data threshold e of a node is confirmed as abnormal data x_u, the node needs to deal with the abnormal data to avoid abnormal data impact on the measurement and reconstruction. Processes are as follows: the node detects the abnormal data, the nonabnormal data, and the received data fusion, and the node detects the abnormal data, the nonabnormal data, and the received data fusion. We discard the abnormal data directly and forward the normal data to the next nodes till to the sink node. In the first cluster, for example, an abnormal value is collected and the processing procedure is shown in formula (8).

3.3.2. Joint Reconstruction. When the sink node receives the data from each cluster all over the network in a sampling round, the data is temporarily stored, and the sink node does not run the intelligent sensing reconstruction algorithm immediately. When performing the data acquisition process in the network, it runs T times and T sampling rounds in all. Each cluster head node transmits the data to the sink node in each round.

After the sink node receives the fusion result of the T sampling rounds, the measurement value y of each sampling round is reformed to the new measurement value Y. $Y = [y(1), y(2), \cdots y(i) \cdots y(T)]^T$, where $y(i)$ represents the measurement sequence of the ith sampling round of the entire network. That is, a diagonal matrix is recomposed of the measurement matrix of each sampling round.

$$\phi = \begin{bmatrix} \phi(1) & & & & & \\ & \phi(1) & & & & \\ & & \ddots & & & \\ & & & \phi(i) & & \\ & & & & \ddots & \\ & & & & & \phi(T) \end{bmatrix}, \tag{9}$$

where $\phi(i)$ represents the measurement matrix of the ith sampling round of the entire network. At this point, the calculation process of the fusion result Y of the T sampling rounds can be expressed as follows:

$$Y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(i) \\ \vdots \\ y(T) \end{bmatrix} = \phi F = \begin{bmatrix} \phi(1) & & & & & \\ & \phi(1) & & & & \\ & & \ddots & & & \\ & & & \phi(i) & & \\ & & & & \ddots & \\ & & & & & \phi(T) \end{bmatrix} \begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(i) \\ \vdots \\ f(T) \end{bmatrix}, \tag{10}$$

because F can be sparsely represented as follows:

$$F = \psi\theta. \tag{11}$$

Formula (11) is a transformation of formula (10). The problem can be transformed to the convex optimization problem of the L1 norm. $\min \|\psi^T F\|_1$ is subjected to $Y = \phi\psi\theta = \Theta\theta$. The sink node can run common reconstruction algorithms to solve the equation $\bar{\theta} = \text{argmin}\|\psi^T F\|_1$, subject to $Y = \phi\psi\theta$. There are many reconstruction algorithms, such as the orthogonal matching pursuit algorithm. It can get the approximate vector $\bar{\theta}$ of the sparse coefficient $\theta\theta = \psi^T F$, it runs inverse transformation $\bar{F} = \psi\bar{\theta}$ again, and it can get the approximate value $\bar{F}$ of the original signal. $\bar{F}$ is a sampling

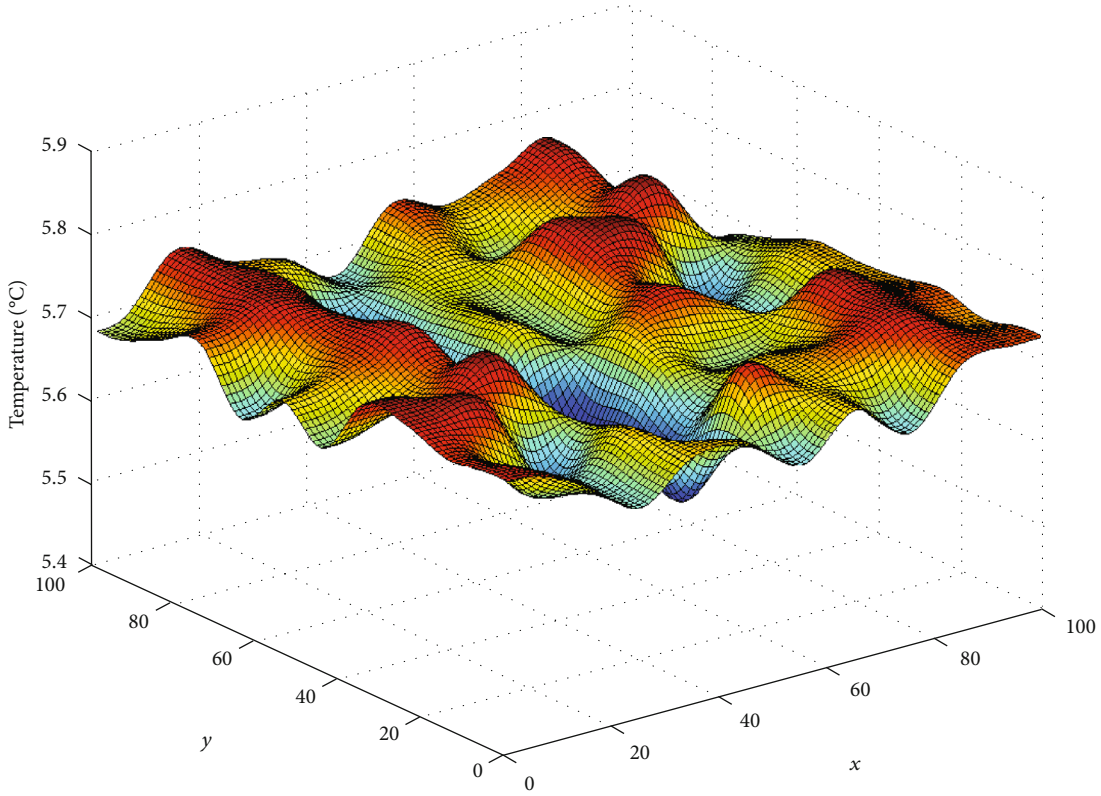

Figure 3: The three-dimensional graph of collected data for a sampling round.

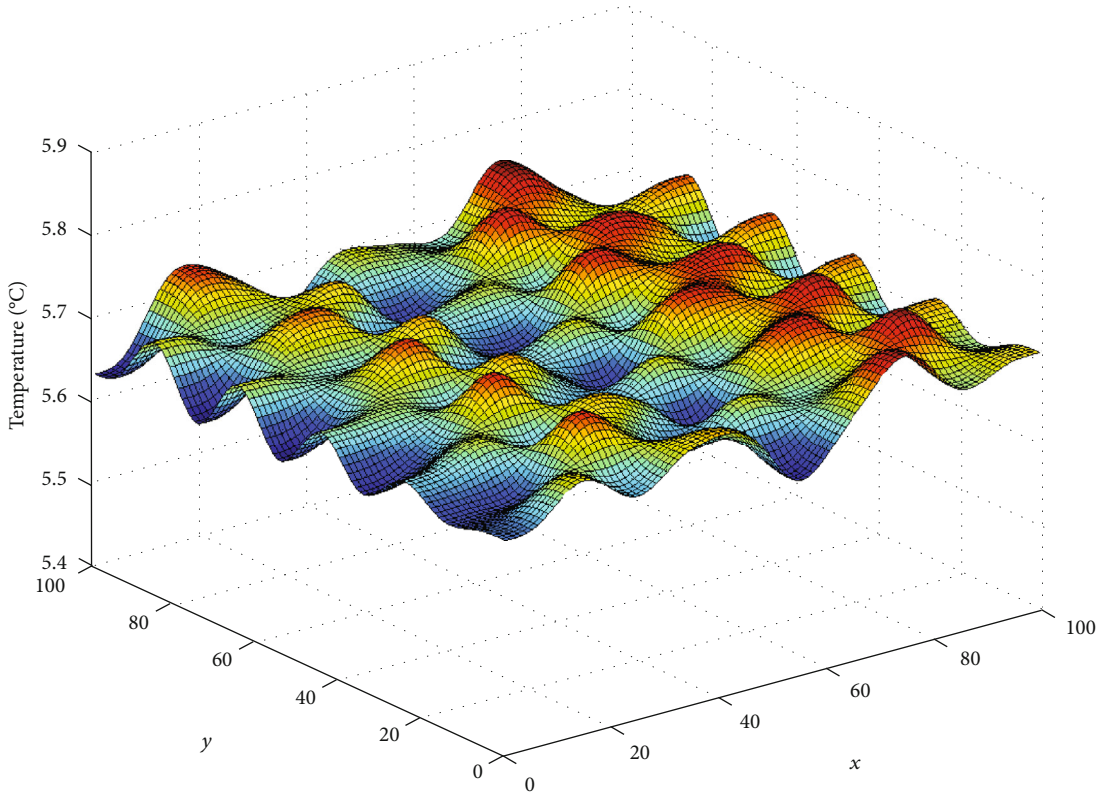

Figure 4: The three-dimensional graph of recovered data for a sample round on the sink node.

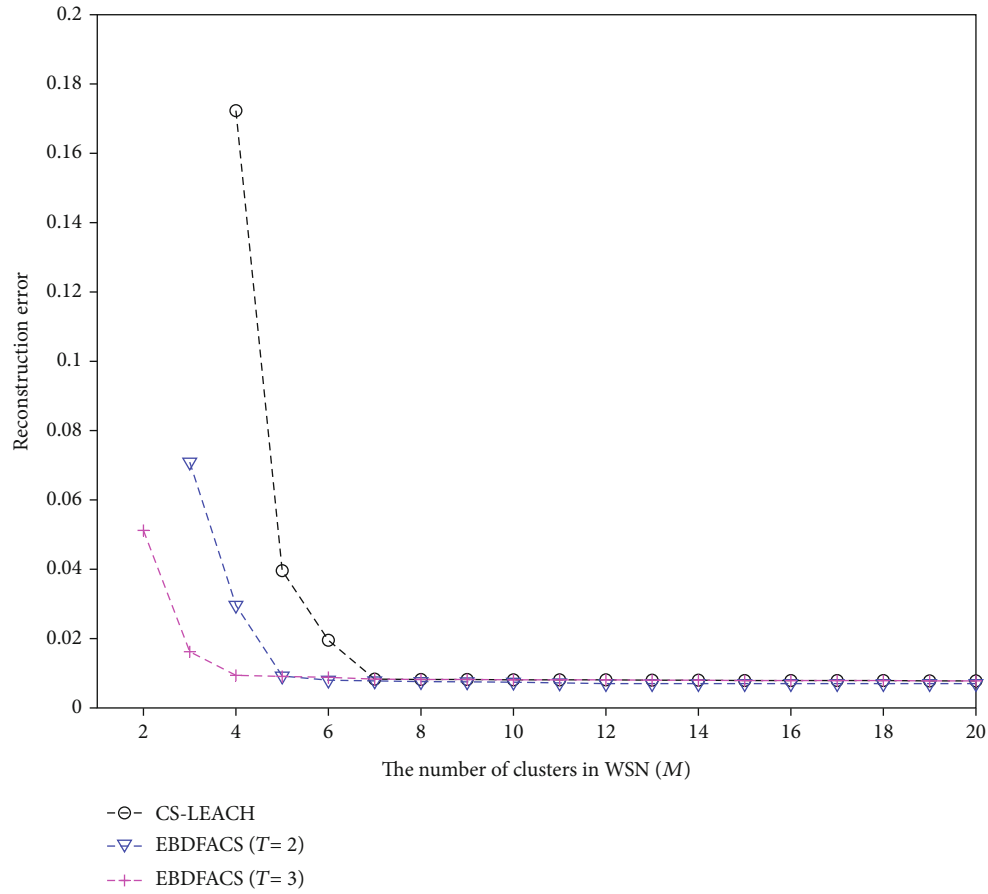

Figure 5: Reconstruction error changes with cluster number M.

data recovery value of the data from all nodes of T sampling rounds in the entire network.

The process of joint reconstruction on the sink node is not a fixed sampling round T process, because encoding and decoding are relatively independent. Sink nodes can be used to change the T value; when T is 1, it can be adapted to a different scale.

The steps of the energy balance data fusion algorithm based on intelligent sensing are as follows:

Step 1. After deploying meteorological sensor nodes in the monitoring area, the node runs the clustering algorithm in the network to establish M clusters, and each cluster head communicates with the sink node by the shortest path routing. The cluster head of each cluster transmits its node information to the sink node; then, the network enters the stable data acquisition stage.

Step 2. In one sampling round, the nodes in each cluster are generated by a random number generator and W is compared with the probability threshold p. If p is less than the threshold value, it is involved in the data collection. At the begining, we build the shortest path and the packets are transmitted along this path. In this transmission, the data is collected and the relay nodes do the multiplication process until the cluster head node. In the process of forwarding and merging, the nodes will collect data and abnormal e. If the abnormal data is directly transmitted to the received data and then the weight is set to 0, the value is processed according to the application requirements.

Step 3. The data is transmitted to the cluster head node of each cluster, and the cluster head node forwards the data $y_i (y_i = \sum_{j=1}^{N} \omega_{i,j} x_j)$ and the node id of the data acquisition to the sink node in a multihop way.

Step 4. According to different application requirements, the sink node constructs the fusion result of T sampling rounds, it constructs a new measurement value vector $YY = Y(1) \cdots Y(T)$, and the measurement matrix $\phi(T)$ of each sampling round is reconstructed by using the node information of data acquisition. The diagonal measurement matrix ϕ is composed of the measurement matrix of each sampling round. The reconstruction algorithm of intelligent sensing runs on the sink node; then, it recovers the sampling data of T sampling rounds of all nodes in the whole network.

4. Performance Analysis

In this paper, the performance of the energy balance data fusion algorithm based on intelligent sensing is evaluated by using the MATLAB tool. The data is the measurement

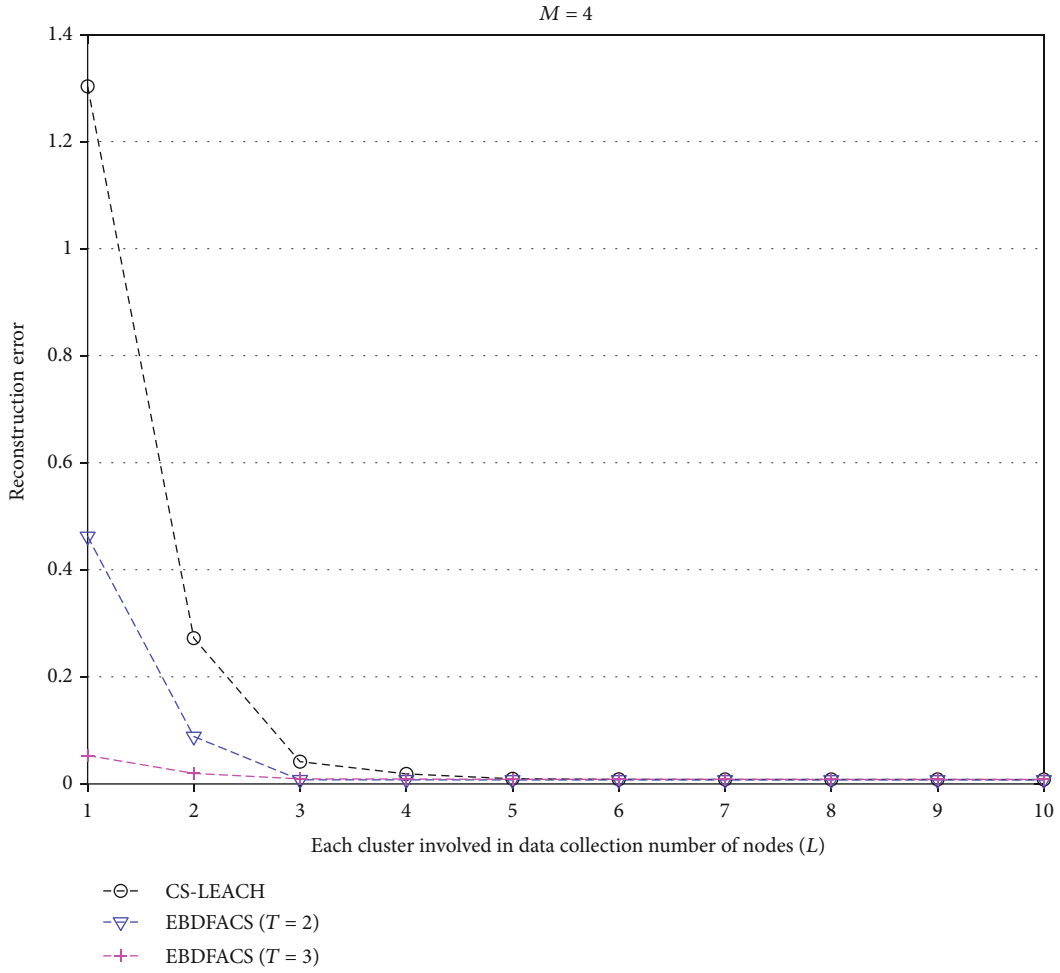

Figure 6: The relationship between the reconstruction error and the number of nodes involved in data acquisition in each cluster, when the cluster number is 4.

of ambient environment temperature from EPFL Sensor Scope WSN [13]. 100 adjacent nodes of the data set in the network are extracted as 100 meteorological sensor nodes. The data acquisition interval is set to 4 minutes. The calculation formula of data reconstruction error adopts the calculation formula of relative error ε, as shown as follows:

$$\varepsilon = \frac{\left\|\bar{f} - f\right\|_2}{\|f\|_2} = \frac{\sqrt{\sum_{i=1}^{n} (\bar{x}_i - x_i)^2}}{\sqrt{\sum_{i=1}^{n} x_i^2}}. \quad (12)$$

The reconstruction algorithm on the sink node uses the orthogonal matching pursuit algorithm. In the simulation, the data recovery algorithm runs 400 times, the average value of the reconstruction error of all times is the same as that of the data reconstruction error, and the related experimental parameters are shown in Table 1. In the experimental network, there are 100 sensor nodes which are deployed in the 100×100 square meter monitoring region, that is, about a sensor node within each 10×10 square meter region.

In the simulation experiment, the relationship between the reconstruction error of the algorithm and the number of clusters M in the network is considered. The relationship between the reconstruction error and the number of nodes L in the data acquisition is also considered. And the performance comparison between EBDFACS and CS-LEACH [14] and the abnormal data processing are verified.

There is a high spatial correlation for sampling values of the dense sensor node deployment; Figure 3 shows the spatial three-dimensional graph of all nodes in a sampling round. It can be seen that there is continuous change of the node data in the space. There is a high spatial correlation.

First of all, this simulation verifies the validity of the EBDFACS algorithm (data in Figure 3 for joint reconstruction) and restores the data series (three-dimensional graph shown in Figure 4); as can be seen from Figure 4, the difference between the collected data in Figure 3 and the recovered

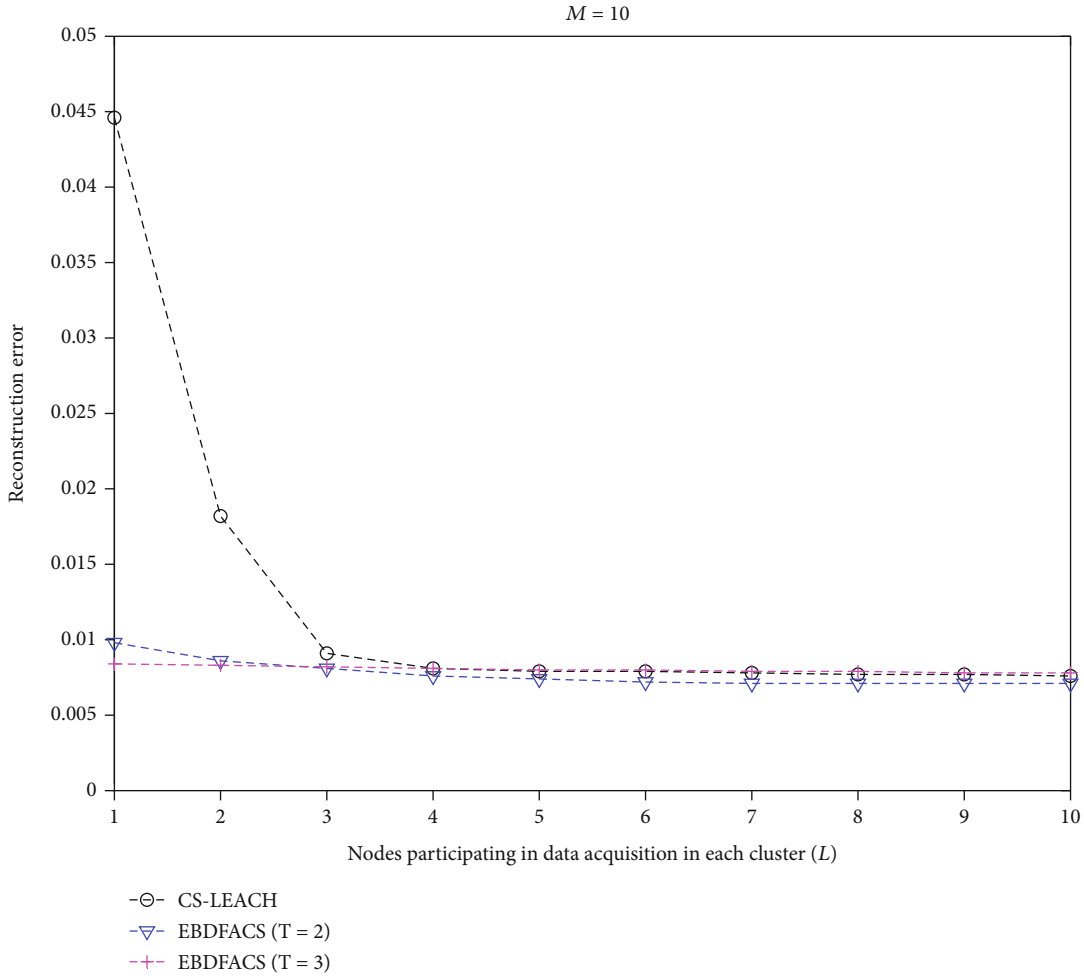

Figure 7: The relationship between the reconstruction error and the number of nodes involved in data acquisition in each cluster, when the cluster number is 10.

data in Figure 4 is not large, and data reconstruction error of only 0.0081 meets the validity of data reconstruction.

In the energy balance data fusion algorithm based on compressive sensing, the number of clusters in the network corresponds to the measured value M of the intelligent sensing. With the increase in measured values, the relative error of data recovery is smaller; that is, with the increase in the number of clusters in the network, the error of data recovery will be less. To randomly select 2 nodes within each cluster to participate in a sample round of data collection, for example, it can be seen from Figure 5 that with the increase in the number of clusters within the network structure, the reconstruction error of CS-LEACH and EBDFACS ($T = 2$, $T = 3$) shows a downward trend. In the CS-LEACH protocol, when the number of clusters was increased from 4 to 5, the reconstruction error showed a rapid decline. When the number of clusters is more than 7, the change in reconstruction error is very small, and when the number of clusters is less than 4, it is unable to reconstruct the original data. And in the EBDFACS algorithm, the decline trend of the reconstruction error is relatively small. When the number of sampling rounds is 2 and the cluster number reaches 3, it can effectively reconstruct the original data. When T equals 3 and the number of clusters in the network is 1, the original data sequence cannot be reconstructed, and the reconstructed error can reach 0.0512; when the number of clusters is 2, the reconstruction effect is better. When the number of clusters is larger than 4, the reconstruction error is stable. When the number of clusters is 4, the reconstruction error of EBDFACS ($T = 3$) is better than that of the data recovery of CS-LEACH (0.1723). From the figure, it can be seen that in the EBDFACS algorithm, the requirements for the cluster number is relatively low for the entire network; as long as the number of clusters is more than two, it can effectively recover the original data; this algorithm is more suitable to the network with fewer clusters.

The number of clusters in the network affects the quality of data reconstruction on the sink node. In each cluster, the number of nodes involved in data collection also affects the quality of reconstruction error. It can be seen from Figures 4–6 that reconstruction error decreases with the increase in the number of sampling nodes in the cluster; when the proportion of the sample node in the cluster tends to 100%, reconstruction error tends to 0. When the number of clusters in the network is small, the number of nodes participating in the data acquisition of the CS-LEACH protocol

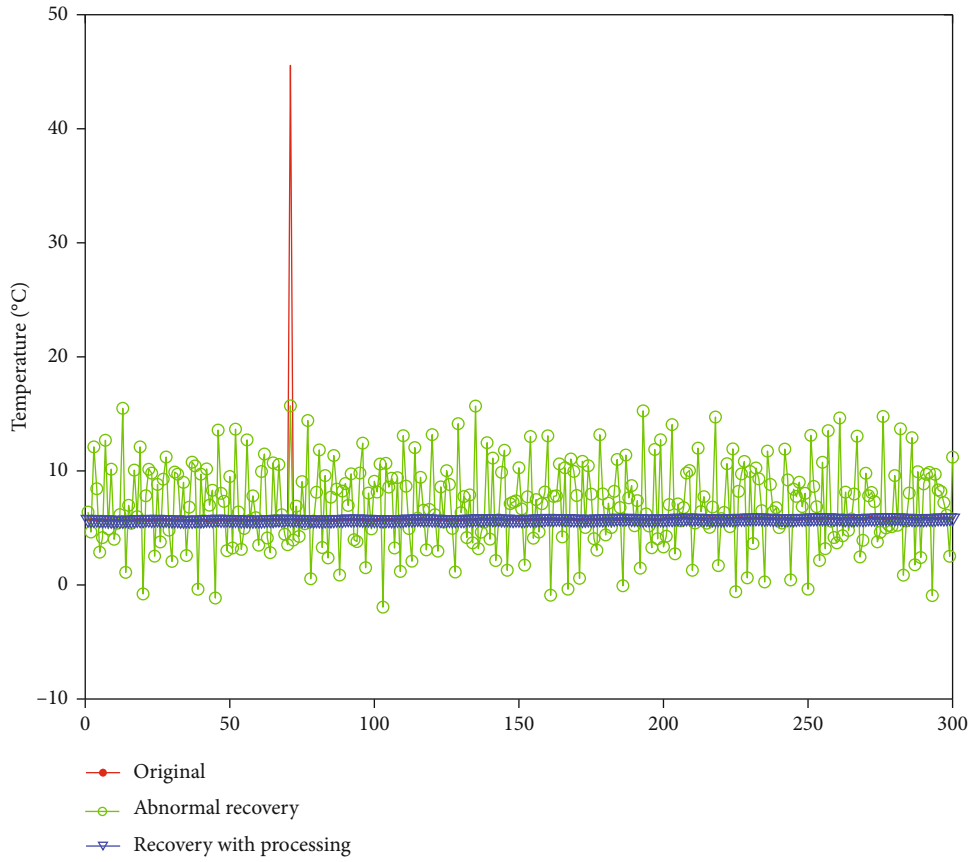

Figure 8: Recovery effect for joint reconstruction after processing sparse anomalies.

has a big influence on the reconstruction error. In the case of four clusters in the network, the reconstruction error is 1.3032; in the same case, the reconstruction error of EBDFACS ($T = 3$) is 0.0529. In order to achieve the same reconstruction error, the CS-LEACH needs three nodes to participate in the data acquisition in each cluster. After the number of nodes participating in the collection reaches four, the reconstruction error of CS-LEACH and EBDFACS ($T = 3$) is close to having the same values.

When the number of clusters is relatively large in the network, as shown in Figure 7, only one node in the cluster participates in data acquisition, the EBDFACS ($T = 3$) reconstruction error is only 0.0086, relative to 0.0446 of CS-LEACH, and the effect is much better. And they can achieve the same error in at least 4 nodes. Thus, in the energy balance data fusion algorithm based on the intelligent sensing algorithm, the requirement for the number of nodes participating in data acquisition is lower.

From the comparison between Figures 6 and 7, it can be seen that relative to the number of nodes participating in data acquisition, the number of clusters in the network has a great influence on the reconstruction error. When the number of clusters is relatively large, even if the number of nodes participating in the collection is relatively small, it can also achieve a better reconstruction effect. From the comparison between EBDFACS ($T = 2$) and EBDFACS ($T = 3$), it can be seen that when the number of nodes involved in the data acquisition is small, the reconstruction error of $T = 2$ is 0.002 higher than that of $T = 3$. However, when the number of nodes involved in the data acquisition is big, the reconstruction error of $T = 2$ is lower than that of $T = 3$. This is due to the increase in the number of sampling rounds; the correlation between the data becomes small but the error increases.

Joint reconstruction of the energy balance data fusion algorithm based on compressive sensing is a kind of flexible joint reconstruction, and the number of sampling rounds can change according to the application requirements. From the above analysis, we can see that EBDFACS is very suitable for smaller numbers of clusters within the network, better suited for a small-scale short sampling period of meteorological sensor networks. Our scheme is also suitable for large-scale networks by adjusting the sample number of rounds. Relative to CS-LEACH, EBDFACS can reach smaller reconstruction error, higher precision for data recovery, and lower demands on clusters within the number of nodes involved in the data collection.

When the network is involved in the data acquisition of the abnormal data, it is necessary to deal with the abnormal data. If the abnormal value is not processed, the original data sequence cannot be recovered. As shown in Figure 8, if the nodes do not deal with the abnormal value of the data sequence, the deviation between the recovered data sequence and original data sequence is very large, the green recovery value and the red value of the original basic do not coincide, and the reconstruction error is 0.7735. When the processing method of this paper can be used to deal with the abnormal

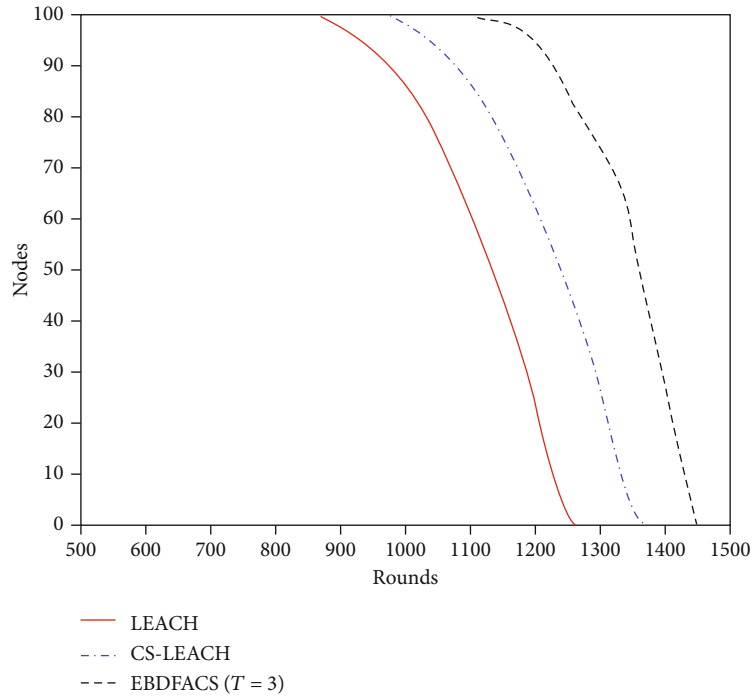

Figure 9: Life cycle of the network.

value, the blue line is the basic and the red value of the recovery value is basically maintained. The reconstruction error of the data is only 0.0086. It can be seen that the reconstruction effect is very good after processing the abnormal value.

Figure 9 is a life cycle comparison for different methods; when we run the experiment in the 853rd round, there appears the first node death for the LEACH protocol, while after the CS-LEACH in the 967th round, the first dead node comes in the 1113th round for EBDFACS ($T = 3$).

5. Conclusion

According to the problems in the meteorological sensor network, such as sensor nodes being deployed in a large scale and network bandwidth resource and node energy constraints in the network, for efficient transmission and processing of data acquisition, this paper proposed the energy balance data fusion algorithm based on intelligent sensing. In the process of random selection of nodes participating in the data acquisition process, in order to promote uniform energy consumption of nodes, the node residual energy is introduced to calculate the probability p. In view of the abnormal data in the network, in order to avoid the effect of outliers on the fusion and reconstruction operations, this paper proposed an algorithm to process the data and which can process the data according to the application requirements of the abnormal value. In order to reduce the number of clusters and the number of nodes involved in data acquisition, measurement values of T sampling rounds are jointly reconstructed using temporal and spatial correlation of perception data on the sink node. And this kind of joint reconstruction is a kind of flexible joint reconstruction; it can be adjusted according to the application requirements. This paper designs experimental simulation of the algorithm to analyze the effectiveness of the proposed algorithm, and the requirements for the number of clusters and the number of nodes are shown in performance comparison. The research results show that EBDFACS can effectively recover the original data sequence, relative to CS-LEACH, it can get better recovery results in the case of the small number of clusters in the network, and the requirements for the number of nodes participating in data collection are lower. Due to the flexibility of the sampling rounds T, the algorithm is suitable for all kinds of network scale, especially for the small number of clusters in the network. Finally, the validity of the algorithm is verified. The energy balance data fusion algorithm based on intelligent sensing has a good recovery effect; moreover, it can reduce the number of nodes that participate in data acquisition and also can effectively extend the network lifetime.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 41975183 and 41875184).

References

[1] J. Hu, C. Chen, T. Qiu, and Q. Pei, "Regional-centralized content dissemination for eV2X services in 5G mmwave-enabled IoV," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7234–7249, 2020.

[2] L. Liu, C. Chen, T. Qiu, M. Zhang, S. Li, and B. Zhou, "A data dissemination scheme based on clustering and probabilistic broadcasting in VANETs," *Vehicular Communications*, vol. 13, pp. 78–88, 2018.

[3] S. Wan, Z. Gu, and Q. Ni, "Cognitive computing and wireless communications on the edge for healthcare service robots," *Computer Communications*, vol. 149, pp. 99–106, 2020.

[4] T.-T. Goh, Z. Xin, and D. Jin, "Habit formation in social media consumption: a case of political engagement," *Behaviour & Information Technology*, vol. 38, no. 3, pp. 273–288, 2019.

[5] S. Wan, X. Xu, T. Wang, and Z. Gu, "An intelligent video analysis method for abnormal event detection in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2020.

[6] Y. Wu, H. Huang, Q. Wu, A. Liu, and T. Wang, "A risk defense method based on microscopic state prediction with partial information observations in social networks," *Journal of Parallel and Distributed Computing*, vol. 131, pp. 189–199, 2019.

[7] Z. Gao, H.-Z. Xuan, H. Zhang, S. Wan, and K.-K. R. Choo, "Adaptive fusion and category-level dictionary learning model for multi-view human action recognition," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9280–9293, 2019.

[8] S. Wan, Y. Zhao, T. Wang, Z. Gu, Q. H. Abbasi, and K.-K. R. Choo, "Multi-dimensional data indexing and range query processing via Voronoi diagram for internet of things," *Future Generation Computer Systems*, vol. 91, pp. 382–391, 2019.

[9] Y. Zhang, K. Wang, Q. He et al., "Covering-based web service quality prediction via neighborhood-aware matrix factorization," *IEEE Transactions on Services Computing*, p. 1, 2019.

[10] Y. Zhang, G. Cui, S. Deng, F. Chen, Y. Wang, and Q. He, "Efficient query of quality correlation for service composition," *IEEE Transactions on Services Computing*, p. 1, 2018.

[11] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[12] G. Li, X. Gao, L. Guo, J. Lin, Y. Gao, and M. Liao, "A multimodel based range query processing algorithm for information collection in cps," *International Journal of Distributed Sensor Networks*, vol. 11, no. 8, Article ID 403267, 2015.

[13] C. Chen, Y. Zhang, M. R. Khosravi, Q. Pei, and S. Wan, "An intelligent platooning algorithm for sustainable transportation systems in smart cities," *IEEE Sensors Journal*, p. 1, 2020.

[14] H. Zheng, S. Xiao, X. Wang, X. Tian, and M. Guizani, "Capacity and delay analysis for data gathering with compressive sensing in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 917–927, 2013.

[15] L. Xiang, J. Luo, and C. Rosenberg, "Compressed data aggregation: energy-efficient and high-fidelity data collection," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1722–1735, 2013.

[16] X.-Y. Liu, Y. Zhu, L. Kong et al., "<italic>CDC</italic>: compressive data collection for wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2188–2197, 2015.

[17] S. Wan, Y. Zhang, and J. Chen, "On the construction of data aggregation tree with maximizing lifetime in large-scale wireless sensor networks," *IEEE Sensors Journal*, vol. 16, no. 20, pp. 7433–7440, 2016.

[18] S. K. Goudos, "Joint power allocation and user association in non-orthogonal multiple access networks: an evolutionary approach," *Physical Communication*, vol. 37, p. 100841, 2019.

[19] Y. Zhang, C. Yin, Q. Wu, Q. He, and H. Zhu, "Location-aware deep collaborative filtering for service recommendation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–12, 2020.

[20] S. Wan, R. Gu, T. Umer, K. Salah, and X. Xu, "Toward offloading internet of vehicles applications in 5G networks," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2020.

[21] S. Wan, X. Li, Y. Xue, W. Lin, and X. Xu, "Efficient computation offloading for internet of vehicles in edge computing-assisted 5G networks," *The Journal of Supercomputing*, vol. 76, no. 4, pp. 2518–2547, 2020.

[22] H. Gao, Y. Xu, Y. Yin, W. Zhang, R. Li, and X. Wang, "Context-aware QoS prediction with neural collaborative filtering for internet-of-things services," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4532–4542, 2020.

[23] S. Ding, S. Qu, Y. Xi, and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, pp. 520–530, 2020.

[24] C. Wang, G. Liu, H. Huang, W. Feng, K. Peng, and L. Wang, "MIASec: enabling data indistinguishability against membership inference attacks in MLaaS," *IEEE Transactions on Sustainable Computing*, vol. 5, no. 3, pp. 365–376, 2020.

[25] S. Ding, S. Qu, Y. Xi, and S. Wan, "A long video caption generation algorithm for big video data retrieval," *Future Generation Computer Systems*, vol. 93, pp. 583–595, 2019.

[26] C. Chen, T. Xiao, T. Qiu, N. Lv, and Q. Pei, "Smart-contract-based economical platooning in blockchain-enabled urban internet of vehicles," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4122–4133, 2020.

[27] Y. Tang, B. Zhang, T. Jing, D. Wu, and X. Cheng, "Robust compressive data gathering in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2754–2761, 2013.

[28] Y. Zhao, H. Li, S. Wan et al., "Knowledge-aided convolutional neural network for small organ segmentation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1363–1373, 2019.

[29] D. Jin, S. Shi, Y. Zhang, H. Abbas, and T.-T. Goh, "A complex event processing framework for an adaptive language learning system," *Future Generation Computer Systems*, vol. 92, pp. 857–867, 2019.

[30] S. Wan, M. Li, G. Liu, and C. Wang, "Recent advances in consensus protocols for blockchain: a survey," *Wireless Networks*, vol. 26, no. 8, pp. 5579–5593, 2020.

[31] M. Mahmudimanesh, A. Khelil, and N. Suri, "Reordering for better compressibility: efficient spatial sampling in wireless sensor networks," in *in 2010 IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing*, pp. 50–57, IEEE, 2010.

[32] N. Lv, C. Chen, T. Qiu, and A. K. Sangaiah, "Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in SAR images," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5530–5538, 2018.

[33] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *in*

Proceedings of the 6th international conference on Information processing in sensor networks, pp. 331–339, ACM, 2007.

[34] S. Lee and A. Ortega, "Joint optimization of transport cost and reconstruction for spatially-localized compressed sensing in multi-hop sensor networks," in *Asia Pacific Signal and Info. Proc. Assoc. Summit (APSIPA)*, Singapore, 2010.

[35] S. K. Goudos, P. D. Diamantoulakis, and G. K. Karagiannidis, "Multi-objective optimization in 5G wireless networks with massive MIMO," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2346–2349, 2018.

[36] C. Chen, Q. Pei, and X. Li, "A GTS allocation scheme to improve multiple-access performance in vehicular sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1549–1563, 2016.

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

[38] B. Wang, L. Zhang, H. Ma, H. Wang, and S. Wan, "Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction," *Complexity*, vol. 2019, 13 pages, 2019.

[39] S. Lee, S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega, "Spatially-localized compressed sensing and routing in multi-hop sensor networks," in *in International conference on geosensor networks*, pp. 11–20, Springer, 2009.

[40] S. K. Goudos, Z. D. Zaharis, and K. B. Baltzis, "Particle swarm optimization as applied to electromagnetic design problems," *International Journal of Swarm Intelligence Research*, vol. 9, no. 2, pp. 47–82, 2018.

[41] S. Wan and S. Goudos, "Faster R-CNN for multi-class fruit detection using a robotic vision system," *Computer Networks*, vol. 168, article 107036, 2020.

[42] P. Zhao, J. Li, F. Zeng, F. Xiao, C. Wang, and H. Jiang, "ILLIA: Enabling k -Anonymity-Based Privacy Preserving Against Location Injection Attacks in Continuous LBS Queries," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1033–1042, 2018.

[43] L. Li, T.-T. Goh, and D. Jin, "How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis," *Neural Computing and Applications*, vol. 32, no. 9, pp. 4387–4415, 2020.