IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 25, 2020, accepted March 7, 2020, date of publication March 10, 2020, date of current version March 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2979874

Edge Computing in Internet of Things: A Novel Sensing-Data Reconstruction Algorithm Under Intelligent-Migratoin Stragegy

ZEYU SUN[1,2], XINYAN ZHANG[1], TIAN WANG[3,4], AND ZHIJIAN WANG[5]

[1]School of Computer Science and Information Engineering, Luoyang Institute of Science and Technology, Luoyang 471023, China
[2]School of Information Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China
[3]College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China
[4]Key Laboratories of Computer Vision and Machine Learning, Huaqiao University, Xiamen 361021, China
[5]School of Information Science, Guangdong University of Finance and Economic, Guangzhou 510320, China

Corresponding author: Zhijian Wang (zjian@gdufe.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1604149, in part by the Henan Province Education Department Cultivation Young Key Teachers in University under Grant 2016GGJS-158, in part by the Luoyang Institute of Science and Technology High-Level Research Start Foundation under Grant 2017BZ07, and in part by the Major Project of Basic and Applied Research in Guangdong Universities under Grant 2017WZDXM012.

ABSTRACT Being limited by the node energy, the Internet of Things (IoT) is prone to data redundancy in the process of data acquisition and collection, resulting in a large amount of packet losses in data transmission and making it impossible to guarantee the transmission mechanism with data security. In order to study the credibility of IoT sensing data transmission further, Edge Computing in Internet of Things: A Novel Sensing-data Reconstruction Algorithm under Intelligent-migration strategy (RdS-ImS) is proposed in this paper. From the beginning, this algorithm can build the packet loss model of sensing data on the link into the form which is random and give the data packet loss predictive model based on the compressed sensing theory. Secondly, preventing the random data packet from lossing, the predictive model is applied to recover through the sensing data retransmission mechanism. If it is unavailable to determine random data packet loss, time series prediction algorithm may be applied for recovery. In addition, in case of interruption of the transmission path, such predictive model can upload the sensing data to the cloud computing platform through an alternative path. At the same time, iterative calculation is performed on the path using intelligent algorithms to optimize the path According to simulation results; this predictive model can improve the network runtime quickly but reducing the sensing data packet loss rate effectively, thereby further verifying this method, which is put forward in this paper, is of relatively strong stability and adaptability.

INDEX TERMS Internet of Things, edge computing, intelligent-migration strategy, data reconstruction.

I. INTRODUCTION

As the underlying application system of the Internet of Things (IoT), Wireless Sensor Network (WSN) is a new form of the network enabling the collection, transmission and processing of related sensing data which is in the deployment area, which generally consists of a large amount of sink nodes (Sink) and sensing nodes [1]–[4]. Every node of sensing sends and relays the sensing data to the Sink in a manner which is multi-hop and self-organized. The network which is external can obtain the information of real-time sensing of each node within the area of sensing of the sensor network through the Sink [5]–[7]. The main and crucial function of the WSN is collecting related information which is of sensing in the deployment area. As the network which is monitoring areas is widely and mostly is deployed in the wild, every node which is in the network is equipped with independent power source with the power which is limited. To guarantee the service life of the network, the network nodes energy consumption has become a crucial factor which is restricting the sensor network performance. In the collection process of WSN data, which is traditional, the nodes are supposed to forward other node packets which are of data while transmitting data packets of its own [8]–[11]. Thus, the closerthe Sink is near the node, the more energy it can consume, and the easier it will exhaust the energy of its own ahead of others,

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaofei Wang.

further resulting in "energy hole" in the network. In addition, because of the sensor network nodes dense deployment, there is a large spatial correlation which is spatial for the sensing between neighboring nodes. Meanwhile, there is a large time correlation for the sensing data which is periodic of each node too [12]–[14]. Therefore, there's a great redundancy in terms of time and space of the sensing data of WSN, thereby providing relatively space which is large compressible. To enable the data collection which is efficient and extend the service life of WSN, research on collection strategies of new data are needed urgently in order that not only the network balanced energy consumption can be guaranteed and the "energy hole" can be avoided but also the redundancy of the data within the network can be reduced and the consumption of the energy of the network can be cut down.

In the procedure of WSNs data collection, the main nodes energy consumption is the energy consumption of wireless communication, which can be greatly cut down by reducing the amount of data transmissions [15]–[18]. Therefore, by compressing the data and reducing the volume of data communication in data transmission, the network energy consumption can be lower, and the network service life can be extended in an effective manner [19]–[22]. Traditional data compression techniques, such as wavelet transform, KL transform and Huffman coding, require data collection before compression. Although such methods can reduce the redundancy between data, they cannot reduce the quntity of the data which is transmitted in the network. Though such methods can reduce the redundancy between data, the amount of data transmitted within the network cannot be reduced. Therefore, the proposal of the Compressive Sensing (Compressive Sensing, CS) theory provides a new concept to solve the problem [23], [24].

In recent years, data collection algorithms for wireless sensor networks based on Compressed Sensing (CS) have gradually become a research highlight for researchers. CS data collection can compress the data during data collection to obtain a linear projection that is much shorter than the original signal length. By transmitting the projection data to the Sink, the amount of data transmission within the entire network can be reduced. After the Sink receives the projection data, the optimization algorithm solves an underdetermined equation to achieve high-precision reconstruction of the original data, thereby completing the entire data collection process. In addition, during the process of CS data collection, each node participates in the same number of sampling projections, and the numbers of data packets sent or received are the same, thereby realizing the energy balance of the network and avoiding the "energy hole" problem for traditional data collection processes.

II. RELATED WORK

Featuring excellent compression performance, simple encoder end, complicated decoder end, compressive sensing (CS) technology has been applied in data collection of WSN. The CS-based wireless sensor network data collection algorithm can compress the network data and balance the consumption of the network energy and even extend the service life of the network. Paper [25] firstly realized the combination of CS and WSN data collection, and then proposed the algorithm of compressed sensing data collection. To further decrease the number of the data which is transmitted on the network, researchers have worked on improvement of the rate of data compression and reduction of the consumption of the energy by changing the network routing and observation matrix, and using the temporal and correlation which is spatial of the data in the network. However, all such researches have assumed ideal links of WSNs, that is, there are no bit errors on the links. In practice, as impacted by complex deployment environments and constrained by wireless transceiver power consumption and node hardware cost, package losses and errors are quite common in wireless links on sensor networks and currently there are few studies on compressed sensing data collection algorithms on lossy links. Paper [26] points out when there is a little number of Gaussian independent loss of node data, which is random, within the whole network, the algorithm of the CS reconstruction can utilize the correlation between the data which will reconstruct the lost information. However, the packet loss reasons of the link of WSN in actual practices are complicated. Therefore, it is not impossible to simply describe the reasons by a Gaussian random packet loss model which is independent. Paper [27] analyzes the sensor data collected by the IoT system, indicates that there is a time correlation which is strong between the data of sensor, and proposes a joint optimization algorithm combining CS and spatio-temporal correlation to recover lost node data. However, such algorithm is only applicable to the lost data recovery of the sink end database without considering the packet loss problem recovery during collection of the data. Paper [28] proposes the Distributed and Morphological Operation-based Data Collection Algorithm (DMOA), where it randomly selects some nodes in each round in order to participate in data collection, sparse observation matrix extremely is constructed just based on the node number of the received data in order to reconstruct the raw data of the entire network at the sink end. Though this method aims to solve the CS data collection problem which is under unreliable links, as the extremely sparse assumption is adopted, it is not applicable for environment with weak spatial data correlation across the whole network. In addition, essentially, this algorithm still uses a data collection method which is traditional for data collection of the node, which would result in the problem of imbalanced energy consumption of nodes near the sink that are exhausted prematurely. From the paper [29], the author has proved that distributed compressed sensing can save about 30% of the observed data volume when performing two related signal processing based on experiment. The joint sparse model of the signal group forms the theoretical basis of distributed compressed sensing. In this model, if multiple groups of signals in the signal cluster are sparsely sparse in a certain sparse domain, and such signals are correlated, distributed compressed sensing can be applied to observe such signals with an observation matrix that is not related to the sparse domain. The observation data so obtained can be accurately restored and reconstructed at the receiving end. Based on a comprehensive analysis of the theoretical knowledge and application background of distributed compressed sensing, Paper [30] has proposed a variety of different joint sparse models for different application scenarios and provided corresponding signal compression schemes and reconstruction methods thereof. It has also discussed the selection principle of distributed compressed sensing observation matrix and applied a simple random sparse projection matrix as the observation matrix. Paper [31] makes focused analysis on the problem of restoring and reconstructing the signal group by distributed compressed sensing method and adopts reconstruction error estimation method to improve the existing distributed compressed sensing method. Paper [32] gives the relationship between sparse recovery and the coherence of redundant dictionaries. Presently, the measurement matrices mainly include two types. One is random measurement matrix, which mainly includes the Gaussian random matrix, Bernoulli random sensing matrix, etc. Matrices of such type are usually not related to a vast majority of sparse bases and can meet the high probability Restricted Isometry Property (RIP) characteristics. However, as such matrices are random, the calculation complexity is great. The other is the deterministic matrix. Matrices of such type have superiorities of fast calculation speed and high reconstruction accuracy. However, they're generally designed for a specific type of signals. Paper [33] proposes Mobile Intelligent Computing Based on Compressive Sensing Data Gathering (MIC-CSDG), which applies compressed sensing and processing to reduce the timing signals collected in target tracking, thereby reducing the calculation load. Presently, there have been many research results available on the application of compressed sensing of WSNs. Such results mainly consider approaches to decrease sensor nodes the energy consumption in order to obtain longer network service life mainly in terms of power control, information processing, data collection, routing, etc. Paper [34] proposes A Data Gathering Algorithm Based on Compressive Sensing in Loss Wireless Sensor Networks, (CS-RTSC). In the process of data collection of the wireless sensor network, collected information is observed at the cluster head node. Then the observations are delivered to the sink node. The aggregation node recovers the original data through a reconstruction algorithm. Compared with the leaf nodes far from the cluster head, the leaf nodes near the cluster head are required to transmit more data packets, resulting in the problem of imbalanced energy consumption in the network. Paper [35] proposes a communication mechanism, which compensates for lost data through retransmission. The principle is that if the receiving node cannot receive all the data within the set time, the sending nodes send the data again. If the number of retransmissions exceeds the set range, the sending node does not resend the part of the data. The retransmission mechanism increases the reliability of data transmission. However, this retransmission strategy does not eliminate the possibility of data loss. Paper [36] applies node computing performance to embed specific function algorithms in nodes to complete data analysis. This aims to reduce the number of wireless communications, and thereby reducing the data transmission during the communication process. However, the disadvantages of this method are also obvious. First, the length of time a node works is limited by the energy of the node; second, the analysis capability of a single node is limited during data transmission among multiple nodes.

Based on the analysis mentioned above, this paper mainly analyzes and summarizes the characteristics of the loss of the link packet in the procedure of CS data collection, then proposes a smart data reconstruction method based on such characteristics. This method applies a hybrid compression sensing data collection strategy to divide the whole network nodes into forwarding nodes which are traditional and CS nodes and designs a sparse observation matrix which is based on the identification of the packet loss in order to observe and sample node data across the network. For traditional forwarding nodes, packet loss is not relevant, and only a sparse observation matrix which is based on identification of packet loss is applied to observe the projection can overcome the impact of packet loss. It is required to apply sparse observation matrix observation which is based on identification of packet loss. In addition, a multi-path transmission mechanism is designed to avoid the occurrence of associated packet loss and ensure the transmission of the reliable link. Finally, corresponding analysis is proposed for three factors affecting the performance of the algorithm.

III. DATA PACKET LOSS MODEL

Compressed sensing is a signal acquisition which is effective and compression method applicable to synchronous sampling and compression of data [37], [38]. Suppose X is a dimensional signal N, Φ is the measurement matrix of $M \times N$ ($M < N$) and the signal $Y = \Phi \times X$, and while X is a linear combination of k basis vectors ($k \ll N$), the signal can be recovered with great probability with Y and $\mathbf{\Phi}$, to improve the probability of accurately recovering the original signal, the number of lines of M of the matrix Φ should satisfy $M \geq c{\cdot}k$, of which k is the sparseness of the signal but c is the function of the sampling rate. If CS technology is applied, sparse basis of the signal X is $\mathbf{\Psi}$, when $X = \Psi \cdot \Theta$, where $||\Theta||_0 = k$, it is expressed as the l_0 normal form, and such signal is called k-spare signal.

Suppose there is a wireless sensor network containing N nodes denoted as $s_1, s_2, s_3, \ldots, s_n$, and their sample values at a certain time are denoted as $x_1, x_2, x_3, \ldots, x_n$. In the process of data collection based on compressed sensing, each node multiplies its own value of sampling by M weights and then sends the result to its next hop node. The last pooling point (Sink) receives M measurements (a linear combination of the perceptual data). The data collection process based on compressive sensing can be expressed in the mathematical form as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1n} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1} & \phi_{m2} & \cdots & \phi_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1}$$

$\boldsymbol{y} = \boldsymbol{\Phi} \times x$, where $y = [y_1, y_2, \ldots, y_m]^T$ is measurement vector collected by the sink, $x = [x_1, x_2, \ldots, x_n]^T$ is the sampled values of the N nodes, $\boldsymbol{\Phi}$ generally is Gaussian random matrix or equal probability Bernoulli matrix.

To explain the process of data collection based on compressed sensing under the topology of multi-hop transmission in sensor network further, a schematic diagram of data collection in a chain structure is shown in Figure 1.

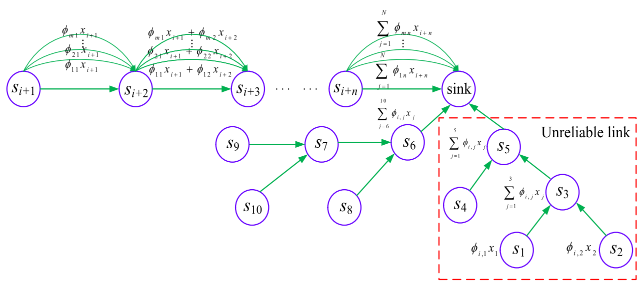


FIGURE 1. CS data collection process based on tree routing.

The set of sensor nodes is $s_1, s_2, s_3, \ldots, s_n$, and the data sampling value so generated is $x_1, x_2, \ldots, x_n$, the data information is, in multi-hop form, transmitted to the sink node. Providing that the traditional mechanism of the tranmission is applied, each node is required to transmit all sample values generated by itself and its downstream nodes, N uncoded data packets can be gotten by the sink node, a total of $N(N+1)/2$ data packets are used for data transmission in the sensor network. In addition, the node is to the convergent node closer, the greater the consumption of the energy is. After applying compressed sensing theory, the most popular N packets can be represented by $M(M \ll N)$ encoded packets. The convergence node will then reconstruct the original data with a higher probability. $\phi_{NI} = \{\phi_{1i}, \phi_{2i}, \ldots, \phi_{mi}\}$ performs a weighting operation on each sample value x_i, where ϕ_i is an element in the observation matrix $\Phi_{M\times N}$. Thus, M data packets are received by the sink node. Then, it requires $M \times N$ data packets to be transmitted by the network in the process of transmission.

Suppose the N nodes are randomly deployed by the sensor network, and the data collection is noted as $d = (d_1, d_2, \ldots d_N)^T$, of which d is sparse under sparse basis $\Psi_{N\times N}$, the observation matrix is $\Phi = (\phi_{ij})_{M\times N}$, in wich the vector of the observation of $M \times N$ is $Y = (y_i)_{M\times 1} = \Phi \cdot \Psi^T \cdot d$. Then the Sink node can reconstruct the original data under certain accuracy constraints by solving the optimization problems shown in Formulas (2) and (3).

$$Y = \Phi \times S = \Phi \times \Psi^{\mathrm{T}} \times d = \Theta \times d \tag{2}$$

$$\hat{d} = argmin \, \|d\|_1 \quad s.t. \; Y = \Phi \cdot S \tag{3}$$

Figure 1 shows the CS data collection process under the tree route. Assuming at the number i observation of CS data collection on a lossy link ($1 \le i \le M$), data $\sum_{j=1}^{3} \phi_{i,j} d_j$ sent by node S_3 to S_5 is lost, the sampling data of child nodes S_1, S_2, S_3 and S_4 of S_5 are all lost. The observation value of the number i observation y_i is changed as:

$$y_i = \left[\phi_{i,1}\phi_{i,2} \; \phi_{i,3} \; \cdots \; \phi_{i,N}\right] \begin{bmatrix} d_5 \\ \vdots \\ d_N \end{bmatrix} \tag{4}$$

We can see that one packet losses during the CS data collection process will cause the collected data of multiple nodes to be lost. Such packet loss "Association Effect" is given rise to by the superimposing of data collected at each not of the multi-hop link during the CS compression sampling process. The closer the node S_k where the packet loss occurs to the Sink, the greater the "Association Effect" of the loss of the packet, especially, of the Sink's one-hop neighbor node packet is lost, the association influence will cause the collected data of all nodes in the network to be lost [39], [40]. After that, according to Formula (4), every round of CS data collection is divided into M observations and performed separately. The intra-network links are lossy links in each observation. If y'_i is used to represent the observation with errors obtained from the number i observation on the lossy link, and $i = 1, 2, \ldots, M$, the vector of the observation obtained by every round is $Y'_{M\times 1} = (y'_1, y'_2, \cdots, y'_M)^{\mathrm{T}}$, and M observations have different degrees of error. Therefore, the accuracy of the data obtained by Sink after optimal reconstruction according to $Y'_{M\times 1}$ will be greatly reduced.

Definition 1: Define the random variable z as the ratio of the numbert of packets which are lost in the sliding window to the window length. It is as this.

$$z = \frac{\sum_{i=1}^{L} (1 - X_i)}{L} \tag{5}$$

where L is the length of the window, which is sliding, the size of L will affect the timeliness of the judgment on one hand and the distribution of the random variable z on the other. The sliding window length $L = 20$ is selected in this paper not only satisfies the requirements of timeliness. In addition, the distribution of the variable z which is random has relatively excellent regularity. Generate 10000 packets of ERL and BRL packets respectively with the simulation tools; because of two types of sequence samples, the Jarque-Bera test is performed on a sample z which is random at a significant level of 5%, respectively, to find their distributions obey the normal distribution law. Note the probability density distribution functions as $f_1(z)$ and $f_2(z)$, as shown in

Formulas (6) and (7).

$$f_1(z) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(z-\mu_1)^2}{2\sigma_1^2}\right) \tag{6}$$

$$f_2(z) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(z-\mu_2)^2}{2\sigma_2^2}\right) \tag{7}$$

Definition 2: External noise, packet conflicts, etc. usually causes random packet loss, and happens randomly and independently. If the bit error wireless channel rate is P_b and length of the the data packet is B bytes, the Element Random Loss (ERL) model packet loss probability of the node is:

$$P_{\mathrm{ERL}} = 1 - (1 - P_{\mathrm{b}})^B \tag{8}$$

Definition 3: Block packet loss is usually caused by network congestion, link-related bursts, etc., and there is correlation between packet losses. To be specific, if the loss of the packet occurs on the link, there will be "continuous packet loss" and the probability of the packet loss decays with time. The packet loss probability of block packet loss is shown in Formulas (9) and (10), where $P_B(i)$ is the number i packet loss probability, and $c \in [0, 1]$ is the attenuation constant. The greater the value, the greater the packet loss probability decay rate corresponding to the blocky packet loss. P_{BRL} is the packet loss probability of the Block Random Loss (BRL) model.

$$P_{\mathrm{B}}(i) = \exp(-\frac{i^2}{2\mathrm{c}^2}) \tag{9}$$

$$P_{\mathrm{BRL}} = \begin{cases} P_{\mathrm{B}}(i), & P_{\mathrm{B}}(i) \geq P_{\mathrm{ERL}} \\ P_{\mathrm{ERL}}, & P_{\mathrm{B}}(i) < P_{\mathrm{ERL}} \end{cases} \tag{10}$$

where, μ_1, σ_1 and μ_2, σ_2 are the mean and variance of the variable z which is random under the BRL model and the ERL model. When the bit error rate P_b and the packet length B of the wireless channel are determined, the values of distribution parameters μ_1, σ_1 and μ_2, σ_2 will be determined accordingly. For example, suppose the bit error rate of the current wireless channel is $P_b = 10^{-3}$, and the length of the the data packet in the network is 50 bytes, the parameter density can be applied to obtain the probability density distribution function of random packet loss and block packet loss under the current link conditions, of which the distribution parameter of z under random packet loss is $\mu_1 = 0.3$, $\sigma_1 = 0.2$ respectively, the distribution parameter of z under blocky packet loss is $\mu_2 = 0.6$, $\sigma_1 = 0.25$. Therefore, before the operation of the network runs, $f_1(z)$ and $f_2(z)$ can be transmitted to the nodes within the network in advance as a prior information for determining the type of packet loss.

As the sliding window observation z is a random variable, which has different probability density distribution functions under different packet loss types, the problem of judging the current link packet loss type could be transformed into a hypothesis test problem, that is, there are two assumptions recorded respectively as: the current packet loss state is random packet loss H_0 and the current packet loss state is blocky packet loss H_1. Under these two assumptions, the probability density distribution function of the random variable z is:

$$f(z|H_0) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(z-\mu_1)^2}{2\sigma_1^2}\right) \tag{11}$$

$$f(z|H_1) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(z-\mu_2)^2}{2\sigma_2^2}\right) \tag{12}$$

Definition 4: Matrix A satisfies the k-order RIP condition. If a constant $\delta_k \in (0, 1)$ exists so that:

$$(1-\delta_k)\,\|k\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\,\|x\|_2^2 \tag{13}$$

If a matrix A satisfies the RIP condition of order $2k$, Formula (13) can be understood as A maintaining a distance between any k-order vectors.

Definition 5: Set A: $R^n \rightarrow R^m$ as a sensing matrix and B: $R^m \rightarrow R^n$ is expressed as a reconstruction algorithm, and (A, B) is called the stable state of C. If for random $x \in \Sigma_k$ and random $e \in R^m$:

$$\|B\,(Ax+e) - x\|_2 \leq C\,\|e\|_2 \tag{14}$$

From the above analysis, the retransmission can overcome the impact of random packet loss at a small cost. If a very small $\varepsilon \leq 10^{-5}$ interference constant is added to the observation matrix, the reconstruction algorithm will not be greatly affected. We consider the fact that the collected data of the sensor network has strong time correlation, in this paper, the temporal data correlation method is applied to improve the the compressed sensing data collection algorithm performance during blocky packet loss. In the process of network operation, the node with lost packet can determine the type of the current packet loss on the link by means of trusted calculation.

IV. RELIABILITY DATA RECONSRUCTION

A. DATA RETRANSMISSION METHOD

The direct idea of addressing link packet loss is to retransmit when packet loss occurs. For a CS data collection algorithm that is highly sensitive to packet loss, can a limited quantiy of retransmissions improve the data reconstruction accuracy? In order to answer this question, this paper models the packet loss of a lossy link and analyzes the impact of retransmission on the performance of CS data collection algorithms which are under models of different packet loss based on experiments. With reference to the Papers [28], [33], [34] and in combination with the actual situation of WSNs, the link packet losses can be divided into three types of element random loss (ERL), block random loss (BRL) and combinational loss (CL).

If D_i is applied to indicate "Decision H_i is true", $(i = 0, 1)$, there are two kinds of errors in the result of this hypothesis test: Virtual Alarm Probability (VAP), $P_f = P(D_1|H_0)$, where ERL is judged as BRL; Missing Alarm Probability (MAP) $P_m = P(D_0|H_1)$, where BRL is judged as ERL. For the application scenario of the paper, the MAP P_m cost is greater than VAP P_f. This is because that in the decision process, if random packet loss is judged as block packet loss, the corresponding cost is only a small loss in the accuracy of packet loss recovery, while block loss is judged as random packet loss, it may result in packet loss under the "association effect". Therefore, the cost function C_{ij} is introduced. When H_i is assumed to be true, the judgment H_j is true. The values are: $C_{10} = \alpha_1$, $C_{01} = \alpha_2$, where $\alpha_1 > \alpha_2 > 0$, $C_{00} = C_{11} = 0$; The minimum risk Bayes decision is selected as the likelihood ratio decision criterion for hypothesis testing.

According to the minimum risk Bayes judgment criterion, we can conclude this.

$$l(z) = \frac{f(z|H_1)}{f(z|H_0)} > \frac{C_{10} - C_{00}}{C_{01} - C_{11}} \cdot \frac{P(H_0)}{P(H_1)} \tag{15}$$

H_1 is judged as true, or H_0 is judged as true. Assume that the probability ratio of block loss to random packet loss on the link is R, that is, $P(H_1)/P(H_0) = R$. By substituting Formulas (6) and (7) into (15), the following result can be obtained:

$$z > \frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2 + \sigma_1\sigma_2\sqrt{(\mu_1 - \mu_2)^2 + 2(\sigma_2^2 - \sigma_1^2)\ln\frac{\alpha_1\sigma_2}{\alpha_2\sigma_1 R}}}{\sigma_2^2 - \sigma_1^2} \tag{16}$$

In which the inequality right side is a constant, which is recorded as the decision threshold ξ. Thus, during the operation of the network, the random value of the variable z is calculated in the sliding window, and the current packet loss status of the link could be determined by judging the z value and the threshold ξ. If $z > \xi$, it is judged that the current packet loss state is block packet loss, otherwise it is random packet loss. If the packet loss status of the current link is determined, the type of packet loss of the next packet loss can be predicted based on the current status, as shown in Fig. 2.

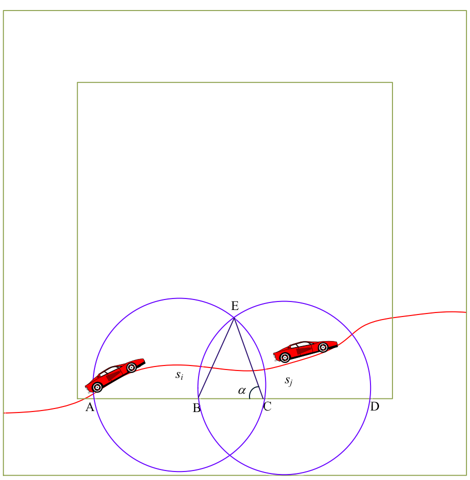


FIGURE 2. Continuous coverage model for mobile nodes.

Take Figure 2 as an example for calculation and analysis. If $\Delta t \to 0$, the node energy of s_j is attenuated by $\Delta E \to 0$, at this time node s_j is to fall into $\mathrm{d}A = \mathrm{d}\gamma \mathrm{d}\alpha$, and the Euclidean distance between node s_i and node s_j could be expressed as:

$$d(i, j) = d(D, \gamma, \alpha) = \sqrt{\gamma^2 + D^2 - 2\gamma D\cos\alpha} \tag{17}$$

In Formulation (17), D is the distance between the aggregation node and node s_i, γ is the distance between tbe aggregation node and node s_j, and angle α is the angle between the the aggregation node and node s_j.

Theorem 1: Under the above conditions, the expected hop count is:

$$E(d_h) = \rho\delta \int_{\gamma_{\min}}^{\gamma} \int_{-\alpha_\gamma}^{\alpha_\gamma} \gamma d_{(i,j)} Q\left(\frac{\beta}{\sigma}\right) \mathrm{e}^{-M(\gamma)(1-p_k(\gamma))} \mathrm{d}\alpha \mathrm{d}\gamma \tag{18}$$

where ρ is the density of nodes of the sensor network; δ is the dimension which is proportional of the parameter; $\gamma_{\min}$ is the distance which is the minimum between the the aggregation node and node s_i.

Proof: From the analysis above, if $\Delta t \to 0$, the node s_j energy is attenuated by $\Delta E \to 0$, at this time node s_j is to fall into $\mathrm{d}A = \mathrm{d}\gamma \mathrm{d}\alpha$, and $\mathrm{d}A \to 0$. Because of the node s_j energy consumption, the signal-to-noise ratio of node s_j is greater than the signal-to-noise ratio threshold, that is, $\Psi_j > \Psi_{th}$. The signal-to-noise ratio of any node s_k that is closer to the sink node than the node s_j is less than Ψ_{th}, that is, $\Psi_k < \Psi_{th}$. Therefore, the probability of node s_j to be selected as the next hop is:

$$dP\{N_i = j\} = P\{N_A(d\gamma) = 1\} P\left\{\Psi_j > \Psi_{th}\right\} P\left\{d_{(j,s)} \geq \gamma\right\} \tag{19}$$

In Formulation (10), $N_A(\mathrm{d}\gamma)$ is the quantity of range nodes whose its own area is $\mathrm{d}A\gamma$ from the convergence node. $P\{N_A(\mathrm{d}\gamma) = 1\}$ is expressed as the probability of having a node. $P\{\Psi_j > \Psi_{th}\}$ indicates that the signal-to-noise ratio of the receiving node s_j is greater than the Ψ_{th} threshold. $P\{d_{(j,s)} \geq \gamma\}$ indicates the probability of the minimum distance of γ of the next hop s_j to the aggregation node. If $\mathrm{d}\gamma \to 0$, $N_A(\mathrm{d}\gamma) = 1$ can be approximated as:

$$P\{N_A(\mathrm{d}\gamma) = 1\} \cong 1 - \mathrm{e}^{-\rho\delta\gamma \mathrm{d}\gamma \mathrm{d}\alpha} \tag{20}$$

where, ρ is the sensor network node density; δ is the proportional dimension of the parameter. As $\mathrm{d}\gamma \to 0$, $\mathrm{d}\alpha \to 0$, that is, $\rho\delta\gamma \mathrm{d}\gamma \mathrm{d}\alpha \to 0$, the above formula may be simplified as:

$$P\{N_A(\mathrm{d}\gamma) = 1\} \cong \rho\delta\gamma \mathrm{d}\gamma \mathrm{d}\alpha \tag{21}$$

According to formula (21), if the distance between any node and the sink node is set as d, the received power is:

$$P_r(d) = P_t - PL(d_0) - 10\eta lg\left(d/d_0\right) + X_\sigma \tag{22}$$

where, P_t is the transmit power; $PL(d_0)$ is the loss on the path with the reference distance d_0; η is the path loss index; X_σ is the energy attenuation parameter and satisfies $X_\sigma \sim N(0, \sigma)$ distribution. P_n is the noise power, and the signal-to-noise ratio of the sink node is:

$$\Psi(d) = P_r(d) - P_n \tag{23}$$

$P\{\Psi_j > \Psi_{th}\}$ indicates that the probability of signal-to-noise ratio of the receiving node s_j is greater than the Ψ_{th} threshold is:

$$P\left\{\Psi_j > \Psi_{th}\right\} = P\left\{X_\sigma > \beta\left(d_{(i,j),\Psi_{th}}\right)\right\} = Q\left(\beta\left(d_{(i,j),\Psi_{th}}\right)/\sigma\right) \quad (24)$$

$$\beta\left(d_{(i,j),\Psi_{th}}\right) = \Psi_{th} + P_n = P_t + PL\left(d_0\right) + 10\eta lg\left(d\left(i,j\right)/d_0\right) \quad (25)$$

$$Q\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{-\left(t^2/2\right)}\mathrm{d}t \quad (26)$$

$Q(x)$ indicates the probability of the distribution function. As the energy decay is a log-normal random variable, it is based on the probability of the function of the cumulative distribution of the normal random variable.

$P\{d_{(j,s)} \geq \gamma\}$ indicates that the signal-to-noise ratio of any node s_k that is closer to the sink node than the node s_j is less than the Ψ_{th} probability value. $A(\gamma)$ is applied to indicate that the node s_k which is arbitrary is closer to the area of the area composed of the convergent nodes than the node s_j, that is:

$$\begin{aligned} P\left\{d_{(i,j)} \geq \gamma\right\} &= \sum_{i=0}^{\infty} P\left\{N_{A(\gamma)} = i\right\} p_k\left(\gamma\right)^i \\ &= \sum_{i=0}^{\infty} \frac{\mathrm{e}^{-M(\gamma)}M\left(\gamma\right)^i}{i!} p_k\left(\gamma\right)^i = \mathrm{e}^{-M(\gamma)(1-p_k(\gamma))} \end{aligned} \quad (27)$$

where, $A(\gamma)$ is the overlapping areas of two circles of D with the radius of R_{int} and γ.Thus, $N_{A(\gamma)}$ is the quantity of nodes in $A(\gamma)$, $M(\gamma) = \rho\delta A(\gamma)$ is the nodes average number within the area. In addition, $p_k(\gamma) = P\{\Psi_k \leq \Psi_{th}, k \in A(\gamma)\}$ is the probability of node s_k in $A(\gamma)$, and the received signal-noise ratio is lower than the threshold signal-noise ratio, that is $\Psi_k \leq \Psi_{th}$, therefore:

$$p_k\left(\gamma\right) = \int_{\gamma_{\min}}^{\gamma}\int_{-\alpha_\gamma}^{\alpha_\gamma}\left[1 - Q\left(\frac{\beta}{\sigma}\right)\right]\frac{1}{A\left(\gamma\right)}\mathrm{d}\alpha\mathrm{d}\gamma \quad (28)$$

where $\gamma_{min} = D\text{-}R_{int}$, by substituting Formulas (20), (22)-(28) into Formula (19) to obtain the expected number of hops:

$$E\left(d_h\right) = \rho\delta\int_{\gamma_{\min}}^{\gamma}\int_{-\alpha_\gamma}^{\alpha_\gamma}\gamma d_{(i,j)}Q\left(\frac{\beta}{\sigma}\right)\mathrm{e}^{-M(\gamma)(1-p_k(\gamma))}\mathrm{d}\alpha\mathrm{d}\gamma \quad (29)$$

Therefore, the proof is comleted.

B. COLLECTION AND COMPRESSION OF SPATIAL CORRELATION DATA

For the compressed sensing theory, a special method of signal reconstruction is applied to enable a signal Φ to be accurately recovered from data with fewer sample points than required by the Nyquist criterion. The signal here must be sparse, to be specific, a signal X of length N is k-sparse on an orthogonal basis (that is, it contains k non-zero values, $k \ll N$). Every data element in Φ independently meets the normal distribution of (0,1/N). The function p of the probability density is calculated by Formula (27). Apply an observation matrix that is not related to the transformation basis to project the transformed high-dimensional signal onto a low-dimensional one. Then, by solving an optimization problem, the original signal can be reconstructed with a high probability from these few projections.

Data collected by the member nodes in the cluster generally have correlations in time domain and space domain, and related processing is required to make such data distributed in a sparse manner. First remove the common parts and only keep the private part, then compress the part which is private and add the public part after reconstruction. This can reduce errors of the reconstruction, improve the data processing speed by nodes, and reduce communication costs.

Then, discrete cosine transformation (DCT) is applied to sparsely transform the data. DCT is selected as it can concentrate most of the key signal information in the low-frequency part after signal transformation. The DCT transformation function is as follows:

$$f_k = \sum_{i=1}^{n} x[i]\cos\left[\frac{\pi}{n}k\left(i + \frac{1}{2}\right)\right] \quad m = 1, 2, \ldots, n. \quad (30)$$

The cluster head node delivers the required observation vector obtained by the DCT function transformation to the base station node. In order to reconstruct the data which is original quickly, the control vector size of shall be no greater than 1,000. This chapter applies the mechanism of network clustering, and the quantity of nodes in each cluster is controlled within 20. Therefore, the scale of the observation vector is relatively small, and the base station nodes can reconstruct the data which is original relatively easily.

Theorem 2: For matrix $\Phi_s = (\xi_1, \xi_2, \ldots \xi_M)^T$, ξ_i is a sequence which is discrete random with the same and independent distribution, and the variables ξ_n which is random constituting the sequence all obey the distribution law shown in Formula (30), after that, the matrix Φ_s is in full rank tending to "1".

Proof: Assume that the matrix Φ_s which is satisfying the above conditions is not full rank, that is, there is a set of coefficients for the number i row in the matrix, which makes the following formulation.

$$\xi_i = a_1\xi_1 + a_2\xi_2 + \cdots + a_{i-1}\xi_{i-1} + a_{i+1}\xi_{i+1} + \cdots + a_M\xi_M \quad (31)$$

The formulation is established, and coefficients a_1, $a_2 \ldots, a_M$ are not all zero.

Let the random process $\{X(n), n = 0, 1, \ldots, N\}$ represent the row vector ξ_i, the mean function and variance function are:

$$EX\left(n\right) = (+1)\frac{1-p}{2} + (-1)\frac{1-p}{2} + 0 \times p = 0 \quad (32)$$

$$\begin{aligned} DX\left(n\right) &= E\left[X(n) - EX(n)\right]^2 = E\left[X(n)\right]^2 \\ &= \frac{1-p}{2} + \frac{1-p}{2} = 1 - p \end{aligned} \quad (33)$$

Let the random process $\{Y(n), n = 0, 1, \ldots, N\}$ be $a_1\xi_1 + a_2\xi_2 + \ldots + a_{i-1}\xi_{i-1} + a_i\xi_i + \ldots + a_M\xi_M$, the mean function and variance function are:

$$EY(n) = E\left[\sum_{j\in[1,M],j\neq i} a_j\xi_j(n)\right] = \sum_{j\in[1,M],j\neq i} a_jE\xi_j(n) = 0 \tag{34}$$

Thus, $Y(n)$ and $X(n)$ respectively describe random processes which are different. For the discrete random process $X(n)$, with the value of $X(i)$ random variable $x(i) \in \{-1, 0, 1\}$, then the length of the state space I_x is 3^N; For the process $Y(n)$ which is discrete random, the value of the variable $Y(i)$ which is random is $1 - M \leq y(i) \leq M - 1$, and the state space I_y length is $(2M - 1)^N$.

If event A is defined as the establishment of Formula (31), the coefficients of Event B, $a_1, a_2 \ldots a_{i-1}, a_i, a_{i+1}, \ldots a_M$, which are all non-zero values; The coefficients of event C, $a_1, a_2 \ldots a_{i-1}, a_i, a_{i+1}, \ldots a_M$ have only one non-zero value; Then we get the following conclusion.

$$P(A|B) < P(A|C) \tag{35}$$

Solution of probability $P(A|C)$ could be transformed into solution of probability of independent and identically distributed stochastic processes $X_2(n)$ and $X_1(n)$ with the state which is the same at the same time. In the state space of the random process $X(n)$, different states have different probability values. To facilitate analysis without loss of generality, parameter $p = 1/3$ is taken in Formula (35), and then we got the conclusion as the following.

$$P(A|B) < P(A|C) = \frac{1}{3^N} \ll 10^{-3} \tag{36}$$

Therefore, the proof is comleted.

If weighted vector $A = (a_1, a_2, \ldots a_k)$ is introduced, of which $a_n = n \Big/ \sum_{j=1}^{k} j$, $n \in [1, k]$ indicates the normalized weight of data h^i_{T-k+n} in the sequence, then the received prediction data $\hat{h}^i_{T+1}$ of the number i observation of the node in the $T + 1$ round is:

$$\hat{h}^i_{T+1} = A \times H_k = \sum_{n=1}^{k} \frac{n}{1+2+\cdots+k} h^i_{T-k+n} \tag{37}$$

Assume node S experiences blocky packet loss in the i observation of $T + 1$ round, the node receives data based on the history stored in its own memory, construct a k-order time-dependent sequence H_k; then, the prediction data $\hat{h}^i_{T+1}$ of packet loss can be obtained according to the prediction Formula (36). Among them, the value of the prediction order k is required, in advance, to be determined, and different values of the prediction order k will affect the prediction accuracy of node packet loss. This paper dynamically adjusts the k value according to the prediction error feedback. Define the maximum allowable prediction error as e_{max}, set the initial value as $k_0 = 3$. After the time-series correlation supplementary packet is applied for link packet loss, wait for the real data to be received most recently in the future. After the real transmission data h_{real} is received correctly, the current k value Combination (37) is applied to predict the data prediction value $\hat{h}_{real}$ of the corresponding round of real data hreal again, and the prediction error $e = \left|h_{real} - \hat{h}_{real}\right|$ is calculated. If the predicted error $e \leq e_{max}$, the k value is not required to be adjusted; if the predicted error $e > e_{max}$, respectively adjust the step size as $+\Delta k$ and -Δk, adjust the k value, and calculate the current predicted error e_1, e_2, respectively, if $|e_1| < |e_2|$, select step size as $+\Delta k$, update the k value; otherwise, select step size $-\Delta k$, and update k value.

V. SIMULATION AND ANALYSIS

Assume that the scale of the deployment of sensor networks is medium-sized and small networks, and the network topologies will not change easily. Each node in the network collects data of sensing from the environment periodically, and the amount of sensing data is relatively small. To verify the RdS-ImS algorithm effectiveness, this paper selects 1000 cycles of data collected by 600 sensor nodes, and the data collection interval of each period is 10 minutes. The environment of the simulation is as follows: the sensor nodes are randomly and uniformly arranged in the area of monitoring of $200 \times 200\text{m}^2$ and $300 \times 300\text{m}^2$.The nodes applies MST to construct the entire routing of network. The aggregation node sink is in the position of the center of the network with coordinates of (100,100) and (150,150). The sensor nodes collect the environmental data in the monitoring and deployment area in a periodic manner and transmit the data to the sink end by the CS data collection method of multi-hop routing. DTC basis is applied in the compression sampling process, of which the sparseness of the observation matrix can be controlled by parameters, in this paper, $s \in [0, 1]$, that is, the dense observation matrix is applied; and the OMP algorithm is applied for the end of the sink reconstruction algorithm.

$$\phi_{ij} = \sqrt{s}\begin{cases} +1, & with\ prob. \frac{1}{2s} \\ 0, & with\ prob. 1 - \frac{1}{s} \\ -1, & with\ prob. \frac{1}{2s} \end{cases} \tag{38}$$

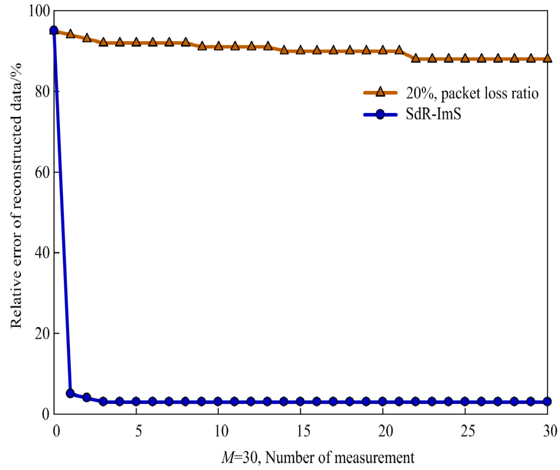
The consumption of the energy of wireless sensor networks adopts the energy consumption model described in paper [20], as shown in Formula (39) and (40), where $E_T(d, l)$ indicates the energy consumption for sending l-bit data, and d indicates the transmission distance α_1 indicates the power consumption of transmitting and receiving circuits, α_2 indicates the distance attenuation coefficient, n is the path loss factor ($2 < n < 5$, generally $n = 2$ in free space); $E_R(l)$ indicates the energy consumption of receiving l-bit data.

$$E_T(d, l) = (\alpha_1 + \alpha_2 \times d^n) \times l \tag{39}$$

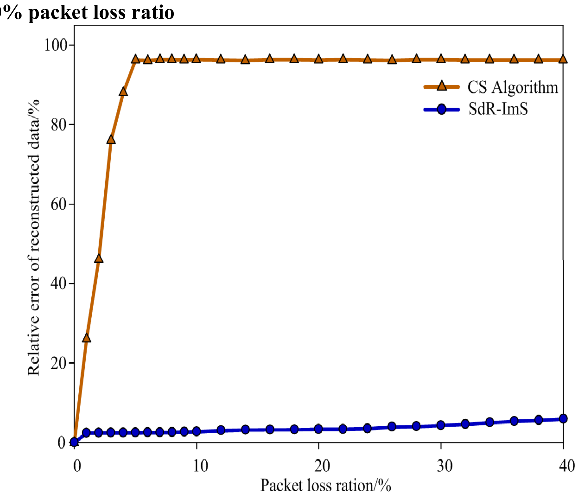
$$E_R(l) = \alpha_1 \times l \tag{40}$$

In this paper, DMOA algorithm [28], MIC-CSDC algorithm [33] and CS-RTSC algorithm [34] are selected as comparison algorithms, of which, DMOA and MIC-CSDC algorithms are classic CS data collection algorithms under tree routing, with the difference lying in that the DMOA algorithm of the entire network nodes participate in the collection of CS data, while the MIC-CSDC algorithm only collects data in the CS mode by nodes that satisfy the number of forwarded data packets greater than the number of observations M. The CS-RTSC algorithm collects and reconstructs the data of the entire work in a quite sparse manner. The data collection process still uses the traditional sensor network data collection method. Performance evaluation indicators refer to relative reconstruction errors and network lifetime. The smaller the relative reconstruction error and the relative reconstruction error, the higher the accuracy of the reconstruction algorithm will be. The simulation data selected in this paper is sourced from the real temperature data of the GreenOrbs system. Therefore, there is a strong spatial correlation between the data and the threshold is set as $\delta_k \in [0.1, 0.5]$. When the relative reconstruction error is less than the threshold, the reconstruction is considered successful, otherwise the reconstruction is considered failed. The life of the network is defined as the time from when the network starts until the node which is the first dies. The longer the network lifetime, the better the algorithm can resist energy consumption. Set the simulation parameters as follows: CS observations $M \in [0, 50]$, maximum number of retransmissions for packet loss max_num=5.The packet loss applies combinational loss composed of 70% random packet loss and 30% block packet loss. Time series correlation prediction supplement order is $k \in [1, 5]$.

Fig 3(a) compares the performance of the CS data collection algorithm on an ideal lossless link and a lossy link with an average packet loss rate of 20%. With the increase of the quantity of observations, the reconstruction error of the SdR-ImS algorithm decreases sharply. After M is bigger than 15, the reconstruction error tends to be stable and less than 3%; for links with an average packet loss rate of 20%, though the relative error decreases with the increase of M, the value is still greater than 85%. Therefore, the quality of the data reconstruction algorithm of the CS data collection algorithm on the actual lossy link is much lower than the evaluation result on the ideal link. In case of a lossy link, due to the "association effect" of packet loss, simply by increasing the number of observations M cannot effectively improve the reconstruction accuracy. Fig 3(b) is the change of the error of the data reconstruction with the loss rate of the link packet with the fixed number of observations of $M = 30$. Under the CS algorithm, when the average loss rate of the packet exceeds 5%, the error of the data reconstruction has been up to 95%; while in this paper, when the average loss rate of the packet of SdR-ImS algorithm is 5%, the data reconstruction rate is 2.3%. Therefore, with the increase of the loss rate of the link packet, the relative reconstruction error of CS data increases rapidly. When the link packet loss rate is 20%, the data reconstruction error is greater than 96.1%. This indicates that even a slight packet loss on the link would seriously degrade the performance of the CS collection algorithm.



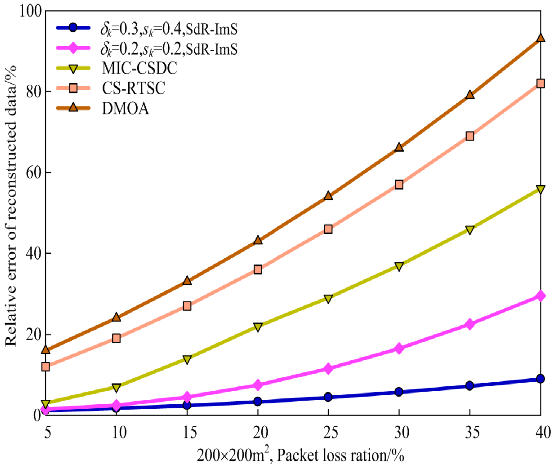
(a) Comparison of relative error between SdR-ImS algorithm and 20% packet loss ratio



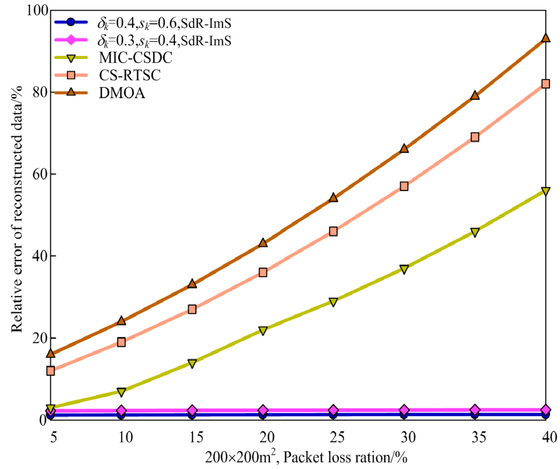
(b) Comparison of relative error between SdR-ImS algorithm and CS algorithm

FIGURE 3. The impact of packet loss on the performance of data gathing based on CS in lossy link.

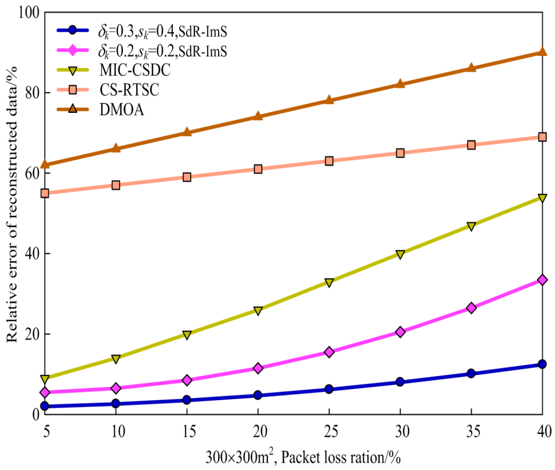
Figure 4 shows the comparison of data loss rate and reconstruction error between the algorithm in this paper and the three algorithms under different parameters (δ_k, s_k) under different monitoring areas. According to Fig 4(a), the parameters applied in the SdR-ImS algorithm in this paper $\{(\delta_k = 0.3, s_k = 0.4), (\delta_k = 0.2, s_k = 0.2)\}$. With the increase of time, the data loss rates and reconstruction errors of the four algorithms increase correspondingly. However, for the SdR-ImS algorithm used in this paper, with the increase of the data packet loss rate, the increase in reconstruction error is relatively slow. Compared with the SdR-ImS algorithm, the other three algorithms have larger reconstruction errors. The main reason lies in is that this paper applies compressed sensing technology to reconstruct the WSN data in order to



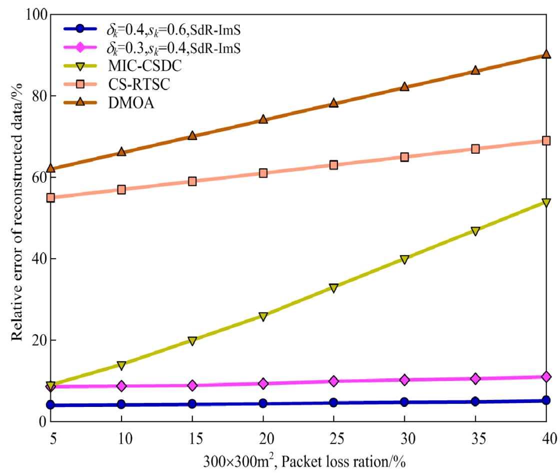
(a){(δ_k=0.3,s_k=0.4), (δ_k=0.2,s_k=0.2)}, Data packet loss rate and reconstruction error



(b){(δ_k=0.4,s_k=0.6), (δ_k=0.3,s_k=0.4)}, Data packet loss rate and reconstruction error



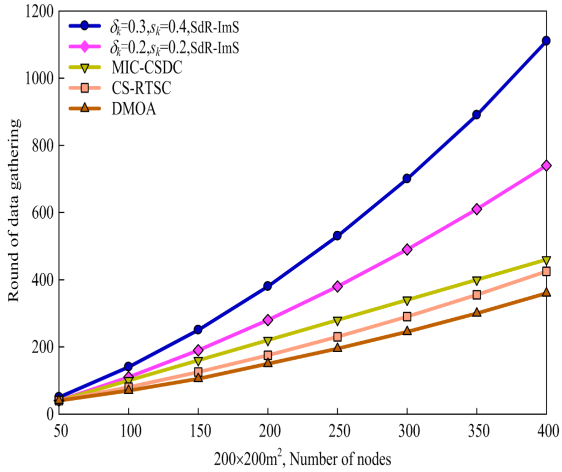
(c){(δ_k=0.3,s_k=0.4), (δ_k=0.2,s_k=0.2)} Data packet loss rate and reconstruction error



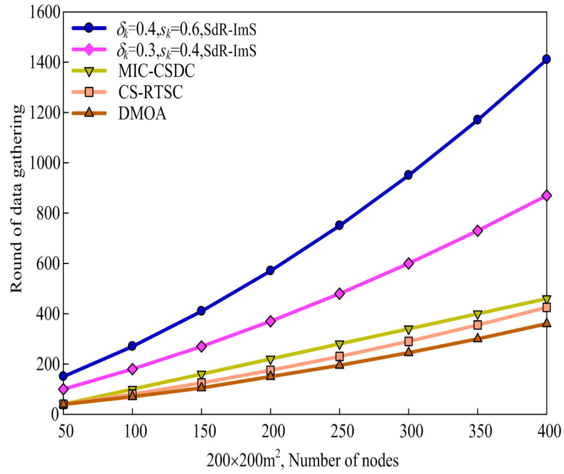
(d) {(δ_k=0.4,s_k=0.6), (δ_k=0.3,s_k=0.4)} Data packet loss rate and reconstruction error

FIGURE 4. Comparison of Effects of Different Parameters {(δ_k = 0.4, s_k = 0.6), (δ_k = 0.3, s_k = 0.4)} on data packet loss rates and reconstruction errors of four Algorithms.

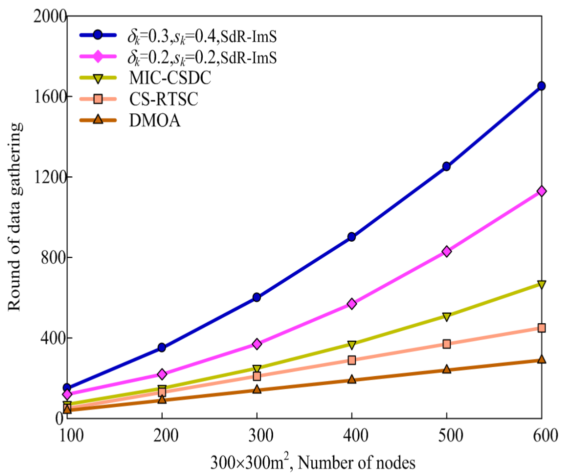
reduce the impact of repeated data on the transmission of the entire network. By comparing the four algorithms, the reconstruction error of the SdR-ImS algorithm in this paper is smaller than the other three algorithms under the same data packet loss rate. And the average reconstruction error thereof is 16. 78%. The parameter in Fig 4(b) is {(δ_k = 0.4, s_k = 0.6), (δ_k = 0.3, s_k = 0.4)}. Compared with Fig 4(a), the parameter values are all increased, and the reconstruction error are less than that of Fig 4(a). Compared with the other three algorithms, the reconstruction error is also smaller. Meanwhile, according to Fig 4(b), in this paper, the SdR-ImS algorithm stabilizes as the data packet loss rate increases. This verifies that the SdR-ImS algorithm can maintain a relatively stable reconstruction error under the condition of high data packet loss rate in this paper. Therefore, it is demonstrated that the SdR-ImS algorithm has strong stability and effectiveness, and its average reconstruction error is 25. 02%. With the increase of the monitoring area, the data loss rate and reconstruction error of the four algorithms all represent an increasing trend, as shown in Fig 4(c) and Fig 4(d). Taking Fig 4(c) as an example, while the data packet loss is 20%, the reconstruction error of the SdR-ImS algorithms in this paper is 4.7% and 11.5%. And the values for the other three algorithms are 26%, 61%, 74%, respectively. The average reconstruction error of the SdR-ImS algorithm in this paper is much smaller than those of the other three algorithms, and the average reconstruction error is 45.56%. The main reason lies in that, in this paper, based on the application of compressed sensing, the time series correlation prediction supplement order is introduced to effectively control the reconstruction accuracy, reduce the reconstruction error, and improve the data fusion efficiency thereof. The other three algorithms have not introduced related optimization algorithms. Instead, they use compressed sensing technology



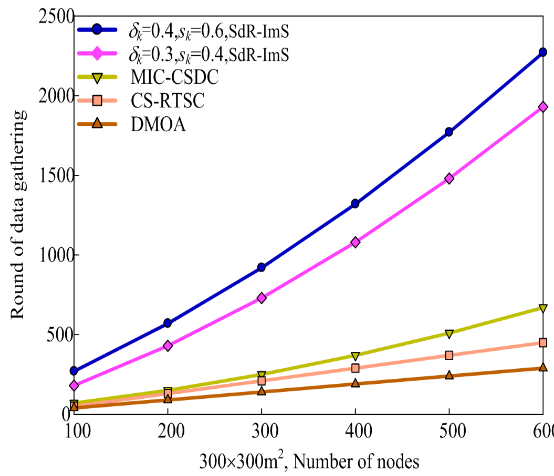
(a) {(δ_k=0.3,s_k=0.4), (δ_k=0.2,s_k=0.2)} comparison of the number of nodes and the number of data collection rounds



(b) {(δ_k=0.4,s_k=0.6), (δ_k=0.3,s_k=0.4)} comparison of the number of nodes and the number of data collection rounds



(c) {(δ_k=0.3,s_k=0.4), (δ_k=0.2,s_k=0.2)} comparison of the number of nodes and the number of data collection rounds



(d) {(δ_k=0.4,s_k=0.6), (δ_k=0.3,s_k=0.4)} comparison of the number of nodes and the number of data collection rounds

FIGURE 5. Comparison of effects of different parameters {(δ_k = 0.4, s_k = 0.6), (δ_k = 0.3, s_k = 0.4)} on the number of nodes and the number of data collection rounds.

directly to reconstruct the data, and the data reconstruction effect is relatively poor.

Figure 5 gives the comparison of the number of nodes and the number of data collection rounds under different parameters. According to Fig 5(a), with the increase number of sensor nodes in these four algorithms, the numbers of data collection rounds increase accordingly. When the quantity of the sensor nodes is 300, the numbers of data collection rounds corresponding to the SdR-ImS algorithm in this paper is 700, 490. And the values for other algorithms are 340, 290 and 245 respectively. In this paper, under the effect of the same number of sensor nodes, the average number of data collection rounds of the SdR-ImS algorithm is much higher than the other three algorithms. The average value is 304, which is much higher than the values of the other three algorithms. The main reason lies in that, with the increase of the number of sensor nodes, the sensing ability between nodes is strengthened accordingly. Meanwhile, in this paper, the sensor nodes as independent and randomly distributed discrete random sequences by SdR-ImS algorithm, and data collection of the entire network is done by random variables composed of sensor nodes. The MIC-CSDC algorithm and CS-RTSC algorithm apply a tree structure to complete the data collection. DMOA applies energy comparison to determine the sink node, and then the sink node selects the member nodes to complete the tree-type data fusion. The above three algorithms fail to consider the impact of network runtime on data collection. The analysis in Fig 5(b) is much similar to that in Fig 5(a). Fig 5(c) shows a comparison diagram of the number of nodes and the amount of data collection rounds in the monitored area of 300 × 300m^2. According to Fig 5(c), with the increasing of the amount of the sensor nodes, the number of data collection rounds of the four algorithms increase accordingly. However, the extent of the SdR-ImS algorithm in this paper is far greater than the other three algorithms. When the quantity of sensor nodes is 400, the numbers of data collection rounds corresponding to the SdR-ImS algorithm in this paper is 900, 571. And the values for other algorithms are 370, 290 and 190 respectively. In this paper, under the effect of the same number of sensor nodes, the average number of data collection rounds of the SdR-ImS algorithm is much higher than the other three algorithms. The average value is 453, which is much higher than the values of the other three algorithms. According to the analysis above, the data collection efficiency of the SdR-ImS algorithm in this paper is much higher than other three algorithms, further illustrating that the SdR-ImS algorithm in this paper is provided with relatively strong data fusion ability.

VI. CONCLUSION

The paper proposes Edge Computing in Internet of Things: A Novel Sensing-data Reconstruction Algorithm under Intelligent-migration strategy (RdS-ImS) with compressed sensing as research background. During the data collection process of the algorithm, in case of data packet loss on the link, the dynamic real-time system design is completed by means of dynamic parameter fitting. Random type packet loss is recovered by retransmission mechanism. Meanwhile, a time correlation sequence prediction recovery model is designed to reduce the impact of data packet loss on the entire wireless sensor network and ensure the normal operation of the wireless sensor network. According to simulation results, that the SdR-ImS algorithm in this paper can effectively suppress reconstruction errors under different data packet loss rates, and effectively improve the number of data collection rounds under the action of different sensor nodes, which further verifies the effectiveness and stability of the SdR-ImS algorithm in this paper. In the next phase, focus should be made on approaches to apply the compressed sensing technology to complete the calculation of characteristics such as wireless sensor network coverage, tracking and deployment.

REFERENCES

[1] T. Wang, Y. Mei, W. Jia, X. Zheng, G. Wang, and M. Xie, "Edge-based differential privacy computing for sensor–cloud systems," *J. Parallel Distrib. Comput.*, vol. 136, pp. 75–85, Feb. 2020.

[2] Z. Sun, C. Li, L. Wei, Z. Li, Z. Min, and G. Zhao, "Intelligent sensor-cloud in fog computer: A novel hierarchical data job scheduling strategy," *Sensors*, vol. 19, no. 23, p. 5083, Jan. 2019.

[3] 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," *J. Parallel Distrib. Comput.*, vol. 131, pp. 189–199, Sep. 2019.

[4] T. Wang, P. Wang, S. Cai, Y. Ma, A. Liu, and M. Xie, "A unified trustworthy environment establishment based on edge computing in industrial IoT," *IEEE Trans Ind. Informat.*, to be published, doi: 10.1109/TII.2019.2955152.

[5] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of edge computing and deep learning: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, to be published, doi: 10.1109/COMST.2020.2970550.

[6] X. Liu, T. Qiu, B. Dai, L. Yang, A. Liu, and J. Wang, "Swarm intelligence-based rendezvous selection via edge computing for mobile sensor networks," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2020.2973401.

[7] Z. Sun, J. Liu, Z. Li, T. Wang, Z. Wang, F. Jia, and C. Lai, "CSR-IM: Compressed sensing routing-control- method with intelligent migration-mechanism based on sensing cloud-computing," *IEEE Access*, vol. 8, pp. 28437–28449, 2020.

[8] W. Tian, C. Zhihan, W. Shuo, W. Jianhuang, Q. Lianyong, L. Anfeng, X. Mande, and L. Xiaolong, "Privacy-enhanced data collection based on deep learning for Internet of vehicles," *IEEE Trans. Ind. Informat.*, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8945334, doi: 10.1109/TII.2019.2962844.

[9] X. Liu, T. Qiu, X. Zhou, T. Wang, L. Yang, and V. Chang, "Latency-aware path planning for disconnected sensor networks with mobile sinks," *IEEE Trans Ind. Informat.*, vol. 16, no. 1, pp. 350–361, Jan. 2020.

[10] Z. Sun, R. Tao, N. Xiong, and X. Pan, "CS-PLM: Compressive sensing data gathering algorithm based on packet loss matching in sensor networks," *Wireless Commun. Mobile Comput.*, vol. 2018, Aug. 2018, Art. no. 5131949.

[11] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," *IEEE Netw.*, vol. 33, no. 5, pp. 156–165, Sep. 2019.

[12] Z. Sun, L. Wei, C. Xu, and Z. Lv, "An event-driven mechanism coverage algorithm based on Sensing-Cloud-Computing in sensor networks," *IEEE Access*, vol. 7, pp. 84668–84679, 2019.

[13] S. Shen, Y. Han, X. Wang, and Y. Wang, "Computation offloading with multiple agents in edge-computing–supported IoT," *ACM Trans. Sensor Netw.*, vol. 16, no. 1, pp. 1–27, Feb. 2020.

[14] X. Liu, A. Liu, T. Qiu, B. Dai, T. Wang, and L. Yang, "Restoring connectivity of damaged sensor networks for long-term survival in hostile environments," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1205–1215, Feb. 2020.

[15] X. Yi, S. Guiling, G. Tianyu, and L. Zhouzhou, "Compressive multi-timeslots data gathering with total variation regularization for wireless sensor networks," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 648–651, Apr. 2019.

[16] Z. Sun, Y. Zhang, Y. Nie, W. Wei, J. Lloret, and H. Song, "CASMOC: A novel complex alliance strategy with multi-objective optimization of coverage in wireless sensor networks," *Wireless Netw.*, vol. 23, no. 4, pp. 1201–1222, May 2017.

[17] C. Lv, Q. Wang, W. Yan, and J. Li, "Compressive sensing-based sequential data gathering in WSNs," *Comput. Netw.*, vol. 154, pp. 47–59, May 2019.

[18] J. Huang and B.-H. Soong, "Cost-aware stochastic compressive data gathering for wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1525–1533, Feb. 2019.

[19] X. Liu, T. Qiu, and T. Wang, "Load-balanced data dissemination for wireless sensor networks: A nature-inspired approach," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9256–9265, Dec. 2019.

[20] Q. Yuan, H. Zhou, Z. Liu, J. Li, F. Yang, and X. Shen, "CESense: Cost-effective urban environment sensing in vehicular sensor networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3235–3246, Sep. 2019.

[21] Z. Sun, L. Wei, C. Xu, T. Wang, Y. Nie, X. Xing, and J. Lu, "An energy-efficient cross-layer-sensing clustering method based on intelligent fog computing in WSNs," *IEEE Access*, vol. 7, pp. 144165–144177, 2019.

[22] V. K. Singh, M. Kumar, and S. Verma, "Accurate detection of important events in WSNs," *IEEE Syst. J.*, vol. 13, no. 1, pp. 248–257, Mar. 2019.

[23] X. Li, X. Tao, and Z. Chen, "Spatio-temporal compressive sensing-based data gathering in wireless sensor networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 198–201, Apr. 2018.

[24] Z. Sun, G. Zhao, and X. Xing, "ENCP: A new energy-efficient nonlinear coverage control protocol in mobile sensor networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 15, Dec. 2018.

[25] T. Wang, H. Luo, X. Zheng, and M. Xie, "Crowdsourcing mechanism for trust evaluation in CPCS based on intelligent mobile edge computing," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 6, Oct. 2019, Art. no. 62.

[26] Y. Wu, H. Huang, N. Wu, Y. Wang, M. Z. Alam Bhuiyan, and T. Wang, "An incentive-based protection and recovery strategy for secure big data in social networks," *Inf. Sci.*, vol. 508, pp. 79–91, Jan. 2020.

[27] T. Wang, J. Zeng, Y. Lai, Y. Cai, H. Tian, Y. Chen, and B. Wang, "Data collection from WSNs to the cloud based on mobile fog elements," *Future Gener. Comput. Syst.*, vol. 105, pp. 864–872, Apr. 2020.

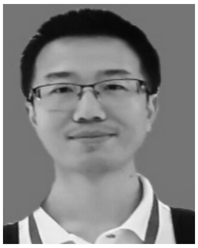
[28] Y. Nie, H. Wang, Y. Qin, and Z. Sun, "Distributed and morphological operation-based data collection algorithm," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 7, Jul. 2017, Art. no. 155014771771759.

[29] Z. Sun, L. Li, X. Xing, Z. Lv, and N. N. Xiong, "A novel nodes deployment assignment scheme with data association attributed in wireless sensor networks," *J. Internet Technol.*, vol. 20, no. 2, pp. 509–520, 2019.

[30] X. Liu, T. Wang, W. Jia, A. Liu, and K. Chi, "Quick convex hull-based rendezvous planning for delay-harsh mobile data gathering in disjoint sensor networks," *IEEE Trans. Syst., Man, Cybern. Syst.*, pp. 1–11, 2019, doi: 10.1109/TSMC.2019.2938790.
[31] W. Tian, B. M. Z. Alam, W. Guojun, Q. Lianyong, W. Jie, and T. Hayajneh, "Preserving balance between privacy and data integrity in edge-assisted Internet of Things," *IEEE Internet Things J.*, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8891776, doi: 10.1109/.JIOT.2019.2951687.
[32] H. Tao, M. Z. A. Bhuiyan, M. A. Rahman, T. Wang, J. Wu, S. Q. Salih, Y. Li, and T. Hayajneh, "TrustData: Trustworthy and secured data collection for event detection in industrial cyber-physical system," *IEEE Trans Ind. Informat.*, vol. 16, no. 5, pp. 3311–3321, May 2020.
[33] Z. Sun, X. Xing, B. Song, Y. Nie, and H. Shao, "Mobile intelligent computing in Internet of Things: An optimized data gathering method based on compressive sensing," *IEEE Access*, vol. 7, pp. 66110–66122, 2019.
[34] H. Zhe, Z. Xia, Z. Dalong, Z. Ce, and D. Siyuan, "A data gathering algorithm based on compressive sensing in lossy wireless sensor networks," in *Proc. 2nd Int. Conf. Frontiers Sensors Technol. (ICFST).*, Shenzhen, China, Apr. 2017, pp. 146–153.
[35] Z. Sun, G. Zhao, and X. Pan, "PM-LPDR: A prediction model for lost packets based on data reconstruction on lossy links in sensor networks," *Int. J. Comput. Sci. Eng.*, vol. 19, no. 2, pp. 177–188, 2019.
[36] W. Tian, Q. Lei, S. A. Kumar, L. Anfeng, B. M. Z. Alam, and M. Ying, "Edge computing based trustworthy data collection model in the Internet of Things," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2020.2966870.
[37] X. Pan, C. Li, X. Xing, Z. Sun, and J. Liu, "A dynamic cluster job scheduling optimisation algorithm based on data irreversibility in sensor cloud," *Int. J. Embedded Syst.*, vol. 11, no. 5, pp. 551–561, 2019.
[38] H. Zheng, W. Guo, and N. Xiong, "A kernel-based compressive sensing approach for mobile data gathering in wireless sensor network systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2315–2327, Dec. 2018.
[39] Z. Liu, Z. Li, M. Li, W. Xing, and D. Lu, "Path reconstruction in dynamic wireless sensor networks using compressive sensing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 4, pp. 1948–1960, Aug. 2016.
[40] W. Zhao, S. Han, W. Meng, D. Sun, and R. Q. Hu, "BSDP: Big sensor data preprocessing in multi-source fusion positioning system using compressive sensing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8866–8880, Sep. 2019.

ZEYU SUN received the B.S. degree in computer science and technology from the Henan University of Science and Technology, in 2003, and the M.Sc. degree from Lanzhou University, in 2010, and the Ph.D. degree from Xi'an Jiaotong University, in 2017. He is currently an Associate Professor with the School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, Henan, China. He is also an External Tutor of master's students with the School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, Henan. His research interests include wireless sensor networks, mobile computing, and the Internet of Things.

XINYAN ZHANG received the B.S. degree in computer application technology from Henan Normal University, in 1998, and the M.S. degree in software engineering from the Huazhong University of Science and Technology, in 2008. Since 2006, she has been a Lecturer with the Luoyang Institute of Science and Technology. Her research interests include intelligent algorithm, data mining, and wireless sensor networks.



TIAN WANG received the B.Sc. and M.Sc. degrees in computer science from Central South University, Changsha, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2011. He is currently a Professor with the College of Computer Science and Technology, Huaqiao University, Xiamen, China. His research interests include wireless sensor networks, cloud computing, and fog computing.

ZHIJIAN WANG received the B.Sc. and M.Sc. degrees in computer science and the Ph.D. degree in control theory and control engineering from Central South University, in 1992, 1995, and 2007, respectively. He is currently a Professor with the Information Science School, Guangdong University of Finance and Economics, China. His research interests include system modeling, software engineering, and supply chain management.