

Research Article

CS-PLM: Compressive Sensing Data Gathering Algorithm Based on Packet Loss Matching in Sensor Networks

Zeyu Sun ,^{1,2} Rong Tao ,¹ Naixue Xiong ,³ and Xiaoyan Pan¹

¹School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, China

²Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China

³Department of Mathematics and Computer Science, Northeastern State University, USA

Correspondence should be addressed to Rong Tao; taorong_123456@163.com

Received 9 April 2018; Accepted 4 July 2018; Published 5 August 2018

Academic Editor: Dajana Cassioli

Copyright © 2018 Zeyu Sun 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.

The data transmission process in Wireless Sensor Networks (WSNs) often experiences errors and packet losses due to the environmental interference. In order to address this problem, we propose a Compressive Sensing data gathering algorithm based on Packet Loss Matching (CS-PLM). It is proven that, under tree routing, the packet loss on communication links would severely undermine the data reconstruction accuracy in Compressive Sensing (CS) based data gathering process. It is further pointed out that the packet loss in CS based data gathering exhibits the correlation effect. Meanwhile, we design a sparse observation matrix based on packet loss matching and verify that the designed matrix satisfies the Restricted Isometry Property (RIP) with a probability arbitrarily close to 1. Therefore, reliable transmission of the compressed data can be guaranteed by adopting the multipath backup routing among CS nodes. It is shown in the simulation results that, with a 60% packet loss ratio of the link, the CS-PLM algorithm can still ensure the effective reconstruction of the data gathered by the CS algorithm and the relative reconstruction error is lower than 5%. Therefore, it is verified that the proposed algorithm could effectively alleviate the sensitivity to packet losses for the CS based data gathering algorithm on unreliable links.

1. Introduction

The nodes in the wireless sensor networks (WSNs) are usually densely deployed and a lot of redundancy exists in the data gathered, which leads to the waste of the energy of the nodes. The compressive sensing (CS) algorithm is a new technique which could largely reduce the sampling frequency and execute the sampling process in parallel with the compression process. As a result, this technique has drawn much attention by researchers.

In order to balance and reduce the energy consumption of the nodes as well as prolong the network lifetime, researchers have proposed data gathering algorithms based on compressed sensing. At present, most of these algorithms are focused on how to effectively reduce network energy consumption and extend the network lifetime [1–3]. For example, it was proposed in paper [4] to employ sparse measurement matrices to reduce the communication cost of each measurement. The spatial and temporal correlation of the sensing data

was exploited in [5] to improve the compression ratio and further reduce the number of measurements. A multilevel hierarchical clustering topology was employed in [6] to gather the data in the network as well as reduce the number of sent and received packets at each layer of nodes. As a result, the total number of transmitted packets is reduced in the entire network. It was pointed out in [7] that the block diagonal measurement matrix could guarantee the reconstruction accuracy with a smaller number of measurements and a longer network lifetime. In recent years, with the progress in the theory of CS, researchers have started to investigate CS based data gathering algorithms for practical applications. The fact was considered in [8] that the data sparsity in realistic data sets would vary with time and space. Therefore, it was proposed to employ the autoregressive AR model to predict data changes and adaptively adjust the number of measurements to achieve the optimal reconstruction performance. It was pointed out in [9] that the environmental noise of the wireless links imposes prominent influences on the transmission of

the undersampled CS data in the network. An approximate gradient descending algorithm was therefore proposed to reconstruct the compressed data under the influence of noise. For the TF-packet loss problem of wireless links in practical application scenarios, there are relatively few works studying CS based data gathering algorithms. Due to the dynamic and asymmetry of wireless links, channel interference, improper antenna direction and height, etc., unreliable links are often the key issue faced by data gathering algorithms in practical sensor networks [10].

For the CS based data gathering problem on unreliable links, we propose a compressive sensing data gathering algorithm based on packet loss matching (CS-PLM). In this algorithm, the nodes in the network are divided into two types, i.e., the traditional forwarding (TF) nodes and the compressive sensing (CS) nodes. The packet loss of TF nodes does not exhibit the correlation while the lost packets of CS nodes are strongly correlated. In the process of the Compressive Sensing data gathering, a packet loss will lead to the loss of the data gathered from multiple nodes; the packet loss correlation effect is caused by the superimposed transmission of the collected data from each node of the multihop link in the CS compression sampling process. The closer the packet loss node S_k is to Sink, the greater the effect of packet loss is. In particular, if the packet of the next-hop neighbor node of Sink is lost, the correlation effect will result in the loss of data collected by some nodes of the whole network. While TF nodes only relay data in the traditional way of data gathering, packet loss has no correlation. As a result, we design a sparse measurement matrix based on the packet loss matching to recover the gathered data if packet losses occur at TF nodes. Therefore, the recovery problem for lost packets is transformed into the sparse matching sampling process in CS. However, for the lost packets at CS nodes, we design the multipath backup transmission scheme to guarantee the reliable data transmission and avoid the correlation effect for packet loss. Therefore, the impacts of unreliable links are alleviated for the CS based data gathering process and the reconstruction accuracy is guaranteed.

The main contributions of this paper are as follows.

①By analyzing the routing with tree structure, it is pointed out that the packet loss would seriously undermine the reconstruction accuracy of the CS based data gathering process and the packet loss in the CS based data gathering process exhibits the correlation effect.

②We design a SPLM measurement matrix and further prove that this matrix satisfies the Restricted Isometry Property (RIP) with a probability arbitrarily close to 1.

③We propose a multipath backup routing transmission scheme based on Hybrid CS to guarantee the reliable handover of the CS projection data.

2. Related Works

Due to its feature of simple encoding and complex decoding, compressed sensing theory has been widely applied to the area of data collection in WSNs. At present, the research of CS based data collection algorithms in wireless sensor networks mainly focuses on how to use CS technology to reduce the

network energy consumption of the data gathering process in WSNs. Most of these works assume that the network link is an ideal link where the impact of packet loss on the CS based data gathering process is ignored. The number of transmitted data packets was presented in paper [11] under the tree topology with and without CS based data gathering algorithm. Furthermore, a hybrid CS based data gathering method was proposed which combines the conventional relay based data gathering method with the CS based data gathering method. It was shown that the network energy consumption could be further reduced by less transmitted data packets in the proposed protocol.

The application of CS was investigated in [2, 12] under the clustering routing structure. Since single-hop transmission is employed in the clustering topology, the packet loss on the links does not exhibit the correlation. As a result, the CS based data gathering algorithm is insensitive to the packet loss on the links. A CS based data gathering algorithm was proposed in [13] for unreliable links under the cluster topology where the column vectors of the measurement matrix are adjusted according to the packet loss nodes in the cluster. Therefore, the influence of packet loss on CS based data reconstruction can be alleviated. The tree-like multihop network routing topology is often used for large-scale WSNs. The CS based data gathering algorithm with multihop routing is studied in paper [14, 15], where the unreliability of the wireless link was ignored while special attention was paid to the optimal matching between the measurement matrix in CS and the structure of the tree-like routing topology. However, the transmission of CS data packets under the multihop routing requires the weighted superposition of the data from multiple nodes. Multiple original data can be lost once one packet loss occurs. Therefore, the transmission of CS data packets is highly sensitive to the packet loss under the multiple-hop routing.

It was shown via simulations in [16–18] that the data reconstruction accuracy under the tree topology can be seriously affected by the packet losses on the link and the Sparsest Random Scheduling (SRS) was further proposed for CS based data gathering in lossy WSNs. In this protocol, a sparest measurement matrix is constructed according to the reception condition at the Sink end, which is further employed to reconstruct the original sensing data for all the nodes in the network and alleviate the influence of packet losses on CS data reconstruction [19–22]. However, this algorithm is only limited to the application scenarios where the spatial correlation of the sensing data in the network is relatively strong.

There are other data gathering methods in conventional WSNs such as ARQ, multipath transmission, network coding, etc. However, there are relatively few works studying the reliable CS based data gathering algorithm in WSNs. Furthermore, the CS based data gathering algorithm is much more sensitive to packet losses than conventional methods. Therefore, the study of CS based data gathering algorithm on unreliable links is quite meaningful to the application of CS theory in practical scenarios.

In wireless sensor network, serious packet loss will undermine the communication performance, service quality, and

application effect of sensor network. In recent years, the research premise of the CS based data gathering theory is the ideal link, and, because of the dynamic characteristics of the wireless link, channel interference and asymmetry of conflict, the wrong direction, and height of antenna, the unreliable link issues are commonly encountered in practical applications. There are many methods to ensure the reliable transmission of links in the traditional data gathering methods of wireless sensor networks, but to the best of our knowledge, there is little work for reliable sensor network data gathering method based on compression perception. In addition, the sensitivity of the CS data gathering method to link packet loss is much higher than that of traditional data gathering, so the research on compressed sensing data gathering algorithm under unreliable link is of great significance to the application of compressed sensing technology in real sensor network.

3. Network Model and Problem Description

The CS is a new technique which samples the sparse signal with a frequency below the Nyquist sampling frequency and achieves the projective transformation of the target signal from a high-dimension space to a low-dimension one. The accurate reconstruction of the compressed signal is achieved via the optimal reconstruction algorithm which is widely studied and applied in many areas due to its excellent compression performance.

Assume that N nodes are randomly deployed in the WSN and the gathered data is denoted as $d = (d_1, d_2, \dots, d_N)^T$. Assuming that d is sparse with respect to base $\Psi_{N \times N}$ and the measurement matrix is $\Phi = (\phi)_{M \times N}$, The measured vector $Y_{M \times 1}$ can be expressed as $Y = (y_i)_{M \times 1} = \Phi \cdot \Psi^T \cdot d$. The Sink node can reconstruct the original data with certain accuracy by solving the optimization problem in the following:

$$Y = \Phi \cdot S = \Phi \cdot \Psi^T \cdot d = \Theta \cdot d \quad (1)$$

$$\vec{d} = \operatorname{argmin} \|d\|_p \quad (2)$$

where $\Theta = \Phi \cdot \Psi^T$ is the sensing matrix and $\|d\|_p$ is the l_p norm of the sensing data vector \vec{d} which is defined as

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^N |x_i|^p \right)^{1/p} & 0 < p < +\infty \\ \max_{i=1,2,\dots,N} |x_i| & \end{cases} \quad (3)$$

In the data gathering process of WSNs, each round of CS based data gathering is performed with M times of independent measurements, which is expressed as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{M1} & \phi_{M2} & \cdots & \phi_{MN} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \quad (4)$$

Assume that N ordinary sensor nodes and one immobile Sink node are deployed in the WSN. All the sensor nodes

are uniformly and randomly deployed with fixed locations in a monitoring area of size $a \times a$. The Sink node is at the center of the monitoring area while the sensor nodes periodically gather and transmit the sensing data to the Sink node. Furthermore, the transmission power of the sensor nodes can be adjusted dynamically and adaptively. The Sink node is assumed with strong computation capability so that it can periodically gather and reconstruct the sensing data and acquire the location information for all the nodes in the network. The Minimum Spanning Tree (MST) routing is established by all the nodes in the network to perform data gathering, i.e., a connected undirected graph $G = (V, E(q))$ is constructed where $V = \{v_1, v_2, \dots, v_N\}$ is the set of sensor nodes, $E(q) = \{e_1(q), e_2(q), \dots, e_N(q)\}$ is the set of links in the MST, and $e_i(q)$ indicates that the link is connected with probability q . If we set $p=1-q$, then p indicates the packet loss ratio of the link. In addition, the CS technique is employed for data gathering in the WSN, which exhibits the following features: ① Discrete Fourier Transform (DFT) is employed for the sparse transformation base $\Psi_{N \times N}$ of the sensing data vector. The sparse transformation and the orthogonal sparse base are presented in (5) and (6), respectively. When M measurements are received at the Sink end, we employ the Orthogonal Matching Pursuit (OMP) algorithm to reconstruct the original sensing data. ② The relative error η is adopted in (7) as the metric to indicate the CS based reconstruction accuracy and a lower η means more accurate reconstruction. If the relative error is higher than 5%, the reconstruction is considered as a failure.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)kn} = \sum_{n=0}^{N-1} x(n) W_N^{kn} \quad (5)$$

$$\Psi_j(t) = \frac{1}{\sqrt{N}} e^{i2\pi jt/N} \quad (6)$$

$$\eta = \frac{\|\hat{x} - x\|_2}{\|x\|_2} = \frac{\sqrt{\sum_{n=0}^{N-1} (\hat{x}_n - x_n)^2}}{\sqrt{\sum_{n=0}^{N-1} x_n^2}} \quad (7)$$

The CS based data gathering process on unreliable links under the tree-like topology is illustrated in Figure 1.

If packet loss occurs on the link between S_5 and the Sink end, all the packets corresponding to S_5 as well as the child nodes of S_5 will be lost, which is shown in the frame in Figure 1. Therefore, the loss is more serious if the packet loss occurs on a link closer to the Sink end. Furthermore, since the weighted superposition data packets are transmitted between nodes in the CS based data gathering process, the Sink end also receives a weighted superposition data after each measurement. As a result, the Sink node cannot acquire the information on whether packet loss occurs or the number of lost packets. Assume that the Sink end receives the weighted superposition data of all the nodes in the network as the current measurement, based on which the reconstruction is performed to recover the original sensing data X .

Therefore, the CS based data gathering on unreliable links exhibits the following features: ① one packet loss on the link will result in the data loss of multiple nodes; i.e., the packet

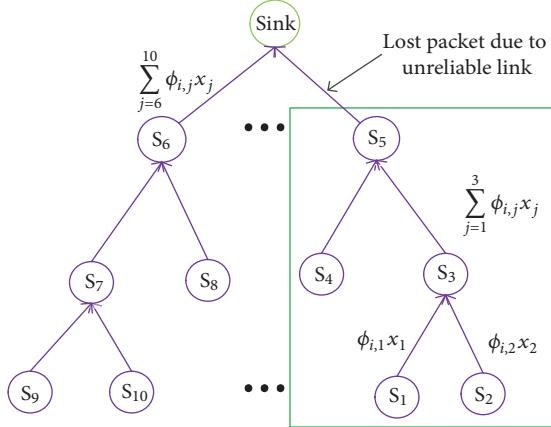


FIGURE 1: CS based data gathering on unreliable links.

loss exhibits the correlation effect. ② The Sink node has no information on the packet loss situation for the nodes in the network and regards the measurement data of the nodes in the network as the data projection to perform reconstruction. That is, the sensing data for compression does not match the sampling of the measurement matrix.

4. Algorithm Design and Realization

4.1. Design for SPLM Measurement Matrix. In order to solve the mismatch problem between the sensing data and the sampling of the measurement matrix, we design a sparse measurement matrix based on Packet-loss Matching (SPLM). In each measurement, the information of the packet loss node is omitted by the measurement matrix. As a result, the packet loss problem for CS based data gathering under tree topology is transformed into the measurement matrix projection problem based on sparse matching. Hence, the large-scale measurement and sampling are accomplished for the data in the network; meanwhile erroneous judgement for the data gathering situation can be avoided at the Sink end. The detailed realization of this process is as follows.

Definition 1 (link state matrix, LSM). The LSM is defined as the matrix which records the link state information with size $M \times N$, where M is the number of measurements and N is the number of nodes in the network. The entries in LSM are defined as follows:

$$L = (l(i, j))_{M \times N} = \begin{cases} 0, & x_j \notin X \\ 1, & x_j \in X \end{cases} \quad (8)$$

Definition 2 (dense random projections, DRP). Each row of the DRP matrix contains $O(N)$ nonzero elements and the DRP matrix is usually constructed with

$$\Phi_d = (\phi(x, y))_{M \times N} \quad (9)$$

The sparse measurement matrix based on packet loss matching employs the randomness of the packet losses on

realistic links to construct the random sparse measurement matrix. The construction process can be achieved by multiplying the LSM with the DRP element-wise, as shown in

$$\Phi_s = L\Phi_d \quad (10)$$

If the packet loss ratio of the unreliable link is p , then each entry of the SPLM matrix is defined as follows:

$$\phi_s(i, j) = \begin{cases} +1, & \text{with prob. } \frac{(1-p)}{2} \\ -1, & \text{with prob. } \frac{(1-p)}{2} \\ 0, & \text{with prob. } p \end{cases} \quad (11)$$

The design of the measurement matrix should guarantee that most of the orthogonal base satisfies the RIP constraint. However, the proof of the RIP condition is a NP-hard problem. It was pointed in [11] that if the measurement matrix is full-rank, then, after the projection of this matrix, the data can be accurately reconstructed with a probability arbitrarily close to 1. Due to the fact that each element in the SPLM matrix follows the discrete random distribution in (11), each row of Φ_s can be regarded as a random sequence generated by the random variable ξ_n , which can further expressed by a discrete random process $\{\xi(n), n = 1, 2, \dots, N\}$.

Theorem 3. For a matrix $\Phi_s = (\xi_1, \xi_2, \dots, \xi_M)^T$ with independently and identically distributed (i, i, d) random and discrete sequences ξ_i , if the random variable ξ_n which constitutes the sequence follows the distribution in (11), the matrix Φ_s will be full-rank with a probability arbitrarily close to 1.

Proof. Assuming that matrix Φ_s satisfying the conditions above is not full-rank, i.e., a set of coefficients exists for the i th row of the matrix so that

$$\xi_i = a_1\xi_1 + a_2\xi_2 + \dots + a_{i+1}\xi_{i+1} + \dots + a_M\xi_M \quad (12)$$

And not all of the coefficients a_1, a_2, \dots, a_M are zeroes. Define the random process $\{X(n), n = 0, 1, \dots, N\}$ as the row vector ξ_i , the mean, and the variance functions are as follows:

$$\begin{aligned} EX(n) &= \left[\frac{(+) (1-p)}{2} \right] + \left[\frac{(-) (1-p)}{2} \right] + 0 \times p \\ &= 0 \end{aligned} \quad (13)$$

$$DX(n) = E[X(n) - EX(n)]^2 = E[X(n)]^2 = 1 - p \quad (14)$$

□

Define the random process $\{Y(n), n = 0, 1, \dots, N\}$ as $a_1\xi_1 + a_2\xi_2 + \dots + a_{i+1}\xi_{i+1} + \dots + a_M\xi_M$; then the mean and the variance functions are

$$\begin{aligned} EY(n) &= E \left[\sum_{j \in [1, M], j \neq i} a_j \xi_j(n) \right] = \sum_{j \in [1, M], j \neq i} a_j E \xi_j(n) \\ &= 0 \end{aligned} \quad (15)$$

$$\begin{aligned} DY(n) &= E[Y(n) - EY(n)]^2 = E[Y(n)]^2 \\ &= \sum_{j \in [1, M], j \neq i} a_j^2 D\xi_j(n) = \sum_{j \in [1, M], j \neq i} a_j^2 (1-p) \end{aligned} \quad (16)$$

Therefore, $X(n)$ and $Y(n)$ denote different random processes, respectively. For the discrete random process $X(n)$, the possible values of the random variable $X(i)$ are $x(i) \in \{+1, -1, 0\}$. Then the length of the state space I_X is 3^N . For the discrete random process $Y(n)$, the possible values of the random variable $Y(i)$ are $-M+1 \leq y(i) \leq M-1$ with $y(i) \in Z$. Therefore, the length of the state space I_Y is $(2M-1)^N$.

Define the case that (12) holds as event A, the case that not all of the coefficients $a_1, a_2 \dots a_M$ are zeroes as event B, and the case that only one of the coefficients $a_1, a_2 \dots a_M$ is nonzero as event C. Then

$$P(A | B) < P(A | C) \quad (17)$$

Solving the probability $P(A | C)$ can be transformed into solving the probability that two i.i.d random processes $X_1(n)$ and $X_2(n)$ have the same state simultaneously. According to the distribution in (11), different states have different probabilities in the state space for the random process $X(n)$. For the simplicity of analysis and without loss of any generality, we set the parameter in (11) as $p=1/3$. Then

$$P(A | B) < P(A | C) = \left(\frac{1}{3}\right)^N \ll 10^{-3} \quad (18)$$

Therefore, the probability of event A is extremely small; i.e., the original assumption does not hold and the matrix Φ_s is full-rank with a probability close to 1.

To evaluate the performance of the SPLI matrix, the classical CS data gathering algorithm, i.e., the CDG algorithm [18], is selected for comparison. Three different methods were used for the comparison of packet loss processing, as shown in Figure 2, where the CDG-DRP means applying the CDG to an unreliable tree topology with correlation effect of packet losses. A dense measurement matrix is used to measure the entire network data; CDG-SPLI indicates that the CDG algorithm is applied to the unreliable tree topology where packet losses exhibit correlation effects. The sparse measurement matrix based on packet loss tags is used to measure the data of the entire network; the CDG-SPLI-NC indicates that the CDG algorithm employs sparse measurement matrix based on packet loss tags to measure the network data without considering the correlation effect of packet losses. It is shown in Figure 2 that the reconstruction accuracy of the CDG-SPLI algorithm and the CDG-SPLI-NC algorithm is significantly higher than that of the CDG-DRP algorithm which utilizes a dense measurement matrix. This is due to the fact that the Sink end of the CDG-DRP algorithm suffers from the incorrect judgment of data packet reception condition. However, this problem can be avoided by the CDG-SPLI and CDG-SPLI-NC algorithms which employ the SPLI matrix for measurement. It can be seen that the misjudgment of the data packet reception condition at the Sink end can seriously deteriorate the reconstruction accuracy of the CS based data gathering algorithm. As for the two algorithms

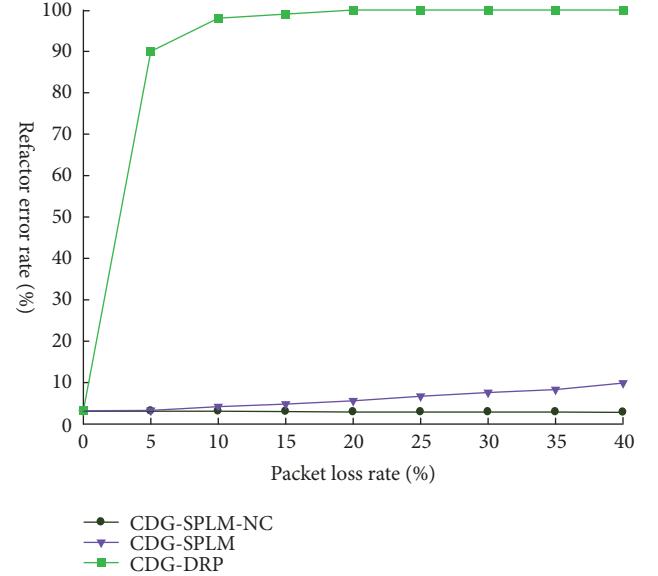


FIGURE 2: Performance comparisons for sparse measurement matrices based on packet loss tags.

employing the SPLI measurement matrix, the CDG-SPLI-NC algorithm does not consider the correlation effect of the packet losses and the CS based data reconstruction exhibit a high accuracy. The relative reconstruction error can be as low as 1.8% when the packet loss ratio is 40%. However, for the CDG-SPLI algorithm where the correlation effect exists, the relative reconstruction error exceeds 5% when the packet loss ratio is beyond 10%; i.e., the reconstruction fails. This is due to the fact that the correlation effect would cause the loss of the entire network data when the packet loss ratio is high. Therefore, the severe lack of CS measurements further deteriorates the reconstruction accuracy. Henceforth, in the data gathering process with correlated packet losses, it is not sufficient if we simply solve the misjudgment of the packet reception situation at the Sink end.

4.2. CS-PLM Algorithm. For the unreliable link under the tree topology, the CS based data gathering not only exhibits the correlation effect for packet losses but also suffers from the problem of misjudgment on the data packet reception situation at the Sink end. However, the SPLI matrix can only solve the misjudgment at the Sink end. Therefore relevant mechanisms still need to be studied to solve the problem of correlation effect in the process of CS based data gathering. Essentially, the correlation effect of the packet loss is caused by the weighted superposition processing of data packets during the CS based data gathering process, which is also the advantage of CS based data gathering. Therefore, the most effective method for solving the correlation effect is to guarantee the reliability of link transmission and avoid the appearance of correlated loss.

The cost of the guarantee for the reliable transmission in the whole network link is huge. To reduce the maintenance cost of the network, according to the performance analysis of the SPLM measurement matrix, this paper designs a hybrid

CS method for the data gathering in the network and divides the nodes of entire network node into traditional forwarding (TF) nodes and CS node, where the TF node only forwards data in a traditional data gathering manner, and the packet loss does not exhibit the correlation effect. The CS node transmits and receives data in a CS based data gathering manner and the packet losses are correlated. Therefore, for the data gathering between TF nodes, simply adopting the SPLM measurement matrix can overcome the impacts of packet losses on CS based data reconstruction. However, for CS nodes, in addition to using the SPLM measurement matrix, a corresponding mechanism must be designed to ensure the reliability of data transmission between CS nodes.

This paper designs a transmission mechanism based on multipath backup routing to ensure the reliability of data transmission between CS nodes. Under normal conditions, the CS node uses MST routing to transmit and receive data packets. If a packet loss occurs on the CS link, the transmitting node S_i of the data packet will choose another the transmission path and use the backup path to send data packets to the Sink. The node S_i can be seen as the source while the Sink is the destination node. There are many ways to construct the route from the source to the destination node. To reduce the energy consumption of the backup path transmission, the minimum energy consumption spanning tree is employed for constructing the routing. The energy consumption model of the network is as follows:

$$E(d) = 2\alpha_1 + \alpha_2 d^n \quad (19)$$

where α_1 is the energy consumption coefficient of the circuit and α_2 is the power amplifying coefficient, d is the transmission distance, and n is the path loss factor with $2 \leq n \leq 5$. The value of n usually takes 2 in the free space. Therefore, the minimization problem of the energy consumption for the routing can be modelled as the optimization problem as in (20), where d_0 is the distance from the source node to the destination node, K is the number of hops, and d_i is the link distance between nodes.

$$\begin{aligned} \min & \left(\sum_{i=1}^K 2\alpha_1 + \alpha_2 d_i^n \right) \\ \text{s.t. } & \sum_{i=1}^K d_i = d_0 \end{aligned} \quad (20)$$

$$d_{char} = \left(\frac{2\alpha_1}{\alpha_2(n-1)} \right)^{1/n} \quad (21)$$

The above optimization problem is solved using the Lagrangian multiplier method. The energy consumption of the network is the minimum if and only if the distance between the destination node and the source node is the same for each hop. We further present the value of the characteristic distance d_{char} as in (21). Finally, the optimal number of hop K_{opt} takes the maximum between d and d_{char} . We propose a centralized construction method for the backup path, which can be divided into the following four steps: ① according to the location information of each node, the Sink node

calculates the distance d from itself to the nodes; ② the Sink end compares the distance d to each node with the characteristic distance d_{char} . If $d \leq d_{char}$, a single-hop backup path had constructed. Otherwise, the optimal number of hops K_{opt} is calculated first. Then according to K_{opt} and the equal distance principle for each hop on the link from the CS nodes to the Sink end, the ideal locations of the relay nodes are further derived. ③ According to the ideal locations of the relay nodes, the Sink end chooses the CS nodes nearest to the ideal locations as the relay nodes. ④ The constructed routing with the minimal energy consumption is broadcast from the Sink node to the CS nodes and related relay nodes.

The CS-PLM algorithm first divides the nodes in the entire network into TF nodes and CS nodes. The MST routing tree is built for nodes in the entire network where the number of child nodes for node i in the network is ω_i . Specifically, $\omega_{sink} = N$ for the Sink end and $\omega_j = 0$ for the nodes at the end of the links. Define $\varepsilon = M - 1$ as the threshold for discriminating the node types and the decision is made by each individual nodes. If $\omega_i > \varepsilon$, the node i participates in the data gathering in a CS method and is defined as the CS node. If $\omega_i \leq \varepsilon$, the node participates in the data gathering in a traditional relay method and is defined as the TF node. Therefore, in the data gathering process, the number of transmitted packets $PN(i)$ for node i is

$$PN(i) = \begin{cases} \omega_i + 1, & \omega_i \leq \varepsilon \\ M, & \omega_i > \varepsilon \end{cases} \quad (22)$$

During the data gathering process of the CS-PLM algorithm, the TF nodes gather data along the MST routing in the traditional relay method. As shown by the white nodes in Figure 3, the packet loss does not exhibit the correlation effect. However, the CS nodes gather data along the MST routing in the CS method. As shown by the black nodes in Figure 3, the packet losses are correlated. In the MST routing, for the CS nodes directly connected to the TF nodes, all the gathered node data needs to be weighted and superimposed after completing the data gathering for all child nodes. The superimposed data is further combined into one data packet which is suitable for transmission between CS nodes. Related flag bits are included in the node ID part of the data packets from the child nodes to identify the data reception status of each node.

In the case of an unreliable link, no processing will be performed if packet losses occur in the link between TF nodes. However, if a packet loss occurs in the link between CS nodes, the data packet is directly sent to the Sink end using the minimum energy consumption backup path. After each round of data gathering, the Sink node builds an SPLM measurement matrix according to the packet loss during each measurement process and further employs the SPLM measurement matrix and the M measurements received at the sink end to reconstruct the original sensing data.

4.3. Data Gathering and Analyses. The data gathering process of the CS-PLM algorithm can be divided into three stages. The first stage is initialization of the sensor network, in

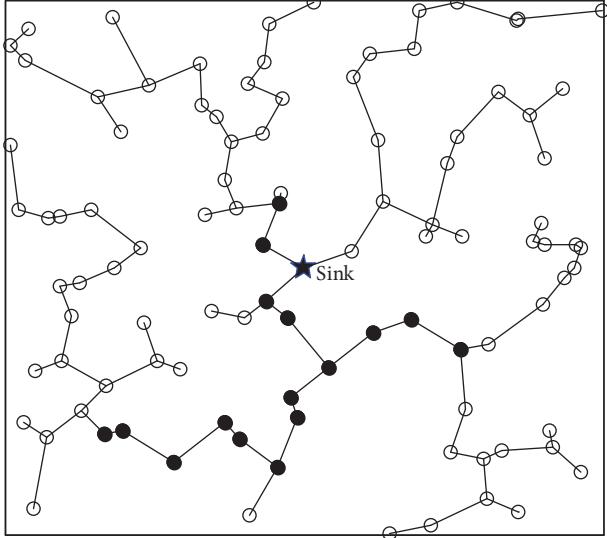


FIGURE 3: Distributions of different node types in the CS-PLM algorithm.

which the node networking, the accumulation of the a priori information for the reception state of each link, and the configuration for the node measurement vector are accomplished. At the second stage, the CS based data gathering is performed on the lossy link and the effective CS based sampling and gathering is achieved for all the nodes in network. At the third stage, the CS based reconstruction is achieved for the sampled data and the original data is further acquired for the nodes in the network. The operation of the algorithm is detailed as follows:

(1) Initialization of the sensor network: initiated by the Sink end, the minimum spanning tree (MST) routing is established to achieve the node networking. Then the Sink end broadcasts the heartbeat data packet to all the nodes in the network. At the reception of the heartbeat data packet, the nodes in the network turn on its own timer. Within the time period T_1 , the nodes transmit and relay the data packet along the routing to the Sink end and the data packet contains the heartbeat information. In this process, each node tracks the real-time packet reception state on its own reception link and stores the result in its memory. This result in the memory will be further employed as the a priori information for the packet loss type prediction which is based on the sliding window scheme. That is, the reception sequence $\{X_i\}$ for each node is initiated till the end of T_1 . Then the Sink end first broadcasts the random seed ξ . At the reception of the random seed, node i combines the seed with its own ID and generates the message (ξ, ID_i) . A column of the measurement $(\varphi_{i1}, \varphi_{i2}, \dots, \varphi_{iM})^T$ is further generated for each node and stored in its own memory.

(2) CS based data gathering: in the CS based data gathering process, each node multiplies the gathered data d_i with the corresponding measurement coefficient φ_{ij} according to the routing. Then the messages are added up sequentially and relayed to the Sink end. When a node recognizes its packet loss in this process, the probability density functions $f_1(z)$ and $f_2(z)$ of the random variable z under two different

packet loss types are determined according to the current bit error rate P_b of the link and the data packet length B . According to (15), the decision threshold ξ for the two types of packet loss is calculated. In the sliding window, the value of the random variable z is calculated and we further compare the value of z with the threshold η . When $z < \eta$, the retransmission scheme is employed for packet recovery. Assuming that the maximal number of retransmission is max_num , the sliding window is updated in a real-time method during the retransmission. If max_num is reached and the received data of the node still remains unrecovered, then the value of z will be compared with the threshold ξ according to the latest update in the sliding window. The current packet loss type is further predicted again according to the comparison result and the corresponding recovery schemes will be adopted. Specifically, the prediction based on the correlation of the temporal sequence is adopted when $z > \eta$ and a k -order temporal correlated sequence H_k is constructed for the node with lost packets. According to (17), the prediction value $h^T T_i + 1$ is calculated for the lost packet and further transmitted to the next one-hop node, assuming that the predicted packet is identical to the lost packet. Therefore, the CS based measurement and sampling is finished and the prediction order k is updated.

(3) Data reconstruction: when the Sink end receives M measurements in one round of data gathering, the measurement vector $Y = (y_1, y_2, \dots, y_M)^T$ is constructed and the Sink end reconstructs the measurement matrix $(\varphi_{ij})_{M \times N}$ according to the random seed ξ and the IDs in the whole network. According to the sparse base $\Psi_{N \times N}$, the CS reconstruction algorithm is employed to reconstruct the sparse signal S and we can recover the original signal vector d by the calculation of $d = \Psi \cdot S$. According to the process stated above, the prediction of the packet loss type, lost packet retransmission, and the temporal correlation based prediction can be simply executed serially and the execution complexity is $O(1)$. As a result, for performance-limited single nodes, this algorithm could guarantee the real-time online operation. For the whole network, at least $N \times M$ data packets have to be received or transmitted for one round of data gathering operation of the CS-PLM algorithm where N is the number of nodes in the network while M is the number of measurements. During the operation of the CS based data gathering algorithm, receiving or transmitting one data packet involves one multiplication, some additions, and no complex regressive operations. Furthermore, according to the CS theory, the number of measurements $M \geq O(K \log N)$ where K is the sparsity degree of the original data, which could be regarded as a constant. By considering the whole network, the complexity of the CS-PLM algorithm is $O(N \log N)$, which indicates that the CS-PLM algorithm causes no additional computations in comparison with conventional CS densely projective data gathering algorithms.

5. Performance Evaluation

In order to evaluate the performance of the CS-PLM algorithm, we employ the tool MatlabR2014a to perform the

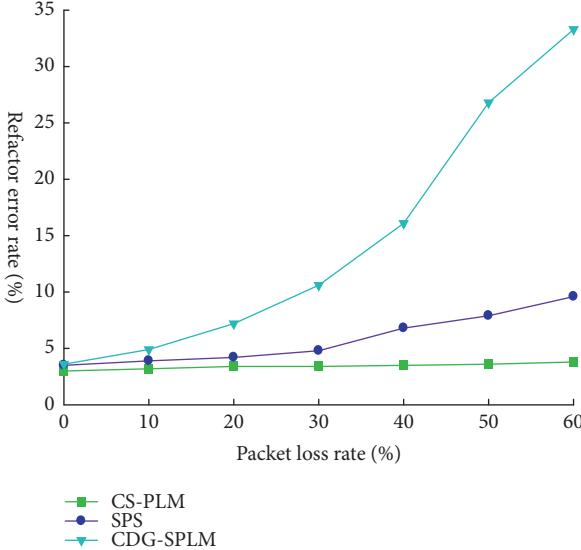


FIGURE 4: Reconstruction error with different packet loss ratio.

simulations. Based on the network model in this paper, we assume that the network is deployed in a 200×200 monitoring area and 400 sensor nodes are placed in the network. The original data sources are assumed to abide by the two-dimensional Gaussian distribution.

The reconstruction accuracy of the CS-PLM algorithm is illustrated with different network packet loss ratio in Figure 4. For comparison, we also investigate the performances of the Sparsest Random Scheduling (SRS) based CS algorithm [2] and the CDG-SPLM algorithm. It is shown in Figure 4 that the reconstruction error performances of these three algorithms increase with the packet loss ratio while the CS-PLM algorithm outperforms the others. The performance gap in terms of reconstruction error between the CS-PLM algorithm and the others also widens with the packet loss ratio. The reason for the performance gap lies in the fact that the SRS algorithm employs the sparsest measurement matrix to reconstruct the original data. On the one hand, the amount of sampled data is far from enough for the purpose of measurement. On the other hand, packet loss will reduce the number of measurements and the reconstruction accuracy is further undermined due to the lack of sampling. Although the CDG-SPLM algorithm could overcome the misjudgment problem for the packet reception state at the Sink end, the correlation of the packet loss in the CS based data gathering process would seriously affect the reconstruction accuracy. However, the proposed CS-PLM algorithm could accomplish as many measurements as possible and overcome the misjudgment problem at the Sink end as well as the packet loss correlation. These improvements could guarantee the reliability of the CS based data gathering process on unreliability links and the highly accurate CS based reconstruction. Furthermore, according to the simulation results, the CS-PLM algorithm could still guarantee the effective reconstruction for the gathered data with a 60% packet loss ratio, which proves the robustness of the proposed algorithm.

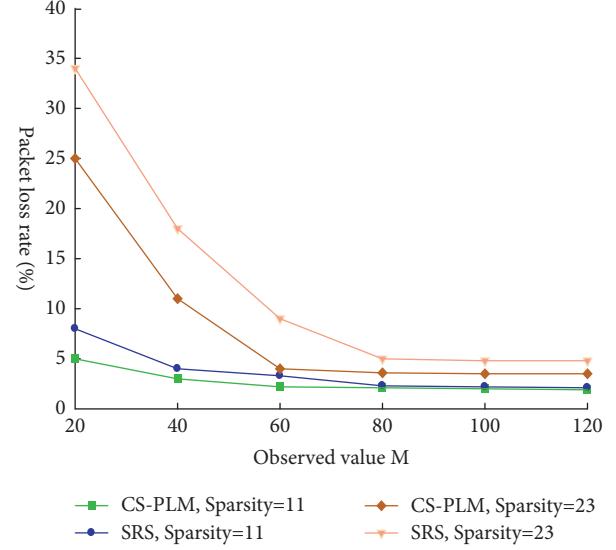


FIGURE 5: Network packet loss ratio with different sparsity Degree.

In order to evaluate the performance of the algorithm under different degree of correlation, we choose two data sets with different degree of sparsity; i.e., the results are illustrated in Figure 5 for the reconstruction accuracy of the CS-PLM algorithm and the SRS algorithm, where the link packet loss ratio is set to 20%. It is shown in Figure 5 that, for the same algorithm, a lower sparsity degree would improve the reconstruction accuracy because the data correlation degree increases with the sparsity degree and, according to the CS theories, the reconstruction accuracy can be further improved. When the sparsity degree is set to 11, the proposed algorithm only slightly outperforms the SRS algorithm but when the sparsity degree is 23, the performance of the CS-PLM algorithm is far better than that of the SRS algorithm. This is because when the data is weakly correlated, the SRS algorithm could not recover the data precisely since it is only based on a small amount of compressed samples. However, the CS-PLM algorithm accomplishes as many measurements as possible and the effective reconstruction can thus be guaranteed for data sets with common degree of correlation. Generally, the CS-PLM algorithm could effectively reduce the strong dependency of the CS based data gathering algorithms on the correlation of the data set. The reliable CS based data gathering can be further guaranteed on unreliable links for data sets with ordinary correlation degree.

The network lifetime performances for different algorithms are compared in Figure 6 where the CDG-Retransmission algorithm further adopts the retransmission scheme based on the CDG algorithm to tackle the packet loss on unreliable links. The network lifetime is defined as the number of rounds from the beginning to the first node failure. It is shown in Figure 6 that the proposed CS-PLM algorithm increases the network lifetime by 400% and 140% in comparison with the SRS algorithm and the CDG-Retransmission scheme. This is due to the energy imbalance of the SRS algorithm which causes the capability vacancy and the early failure of the network. The retransmission scheme is adopted

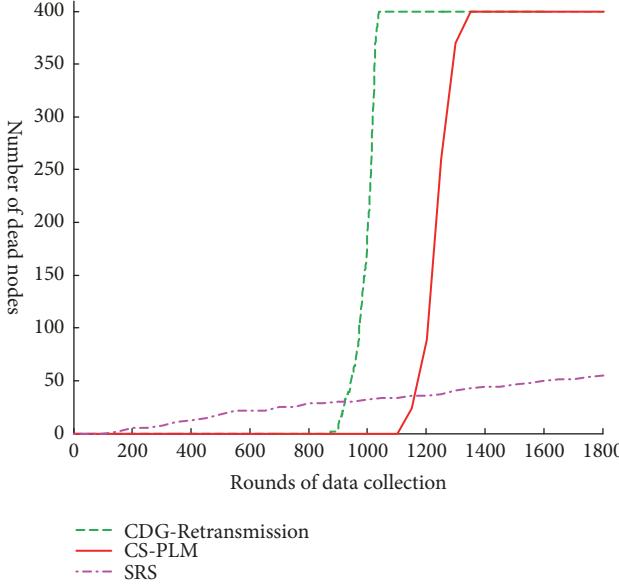


FIGURE 6: Network lifetime for the CS based algorithms.

for the CDG-Retransmission algorithm for packet loss recovery while more packets are transmitted and the network lifetime is therefore shortened.

According to the theory of CS, the correlation between the measurement matrix and the sparse base of the data will affect the reconstruction performance of the algorithm. To verify the performance of the CS-PLM algorithm under different sparse basis, we choose two sparse bases, i.e., the Discrete Cosine Transform (DCT) and Discrete Fourier Transform (DFT) as a comparison. Related results are shown in Figure 7. It is shown in Figure 7 that, with different packet loss ratio, the proposed algorithm performs better in the DCT base; i.e., the DCT base exhibits weaker correlation with the SPLM matrix in comparison with the DFT base and we can obtain a higher reconstruction accuracy by choosing the DCT base to perceive the changing sparsity degree.

The impacts of packet loss ratio on relative reconstruction error are illustrated in Figure 8 with different number of measurements. The number of measurements for the CS-PLM algorithm is set to 60, 80, 120, and 140, respectively. It is shown in Figure 8 that when the number of measurements is 60, the relative reconstruction error changes significantly with the packet loss ratio and the increasing packet loss ratio would undermine the recovery accuracy. However, when the number of measurements increases, the impacts of packet loss ratio on the reconstruction accuracy are further alleviated. For example, when the number of measurement is 140, the relative reconstruction errors are almost the same for the packet loss ratio of 10%, 30%, and 50%. The reason behind this phenomenon is that the proposed CS-PLM algorithm adopts the measurement matrix based on packet loss matching to sample the data in the network. The increase of the packet loss ratio would lead to a higher sparsity degree in the measurement matrix. In order to maintain the reconstruction accuracy, the number of measurements

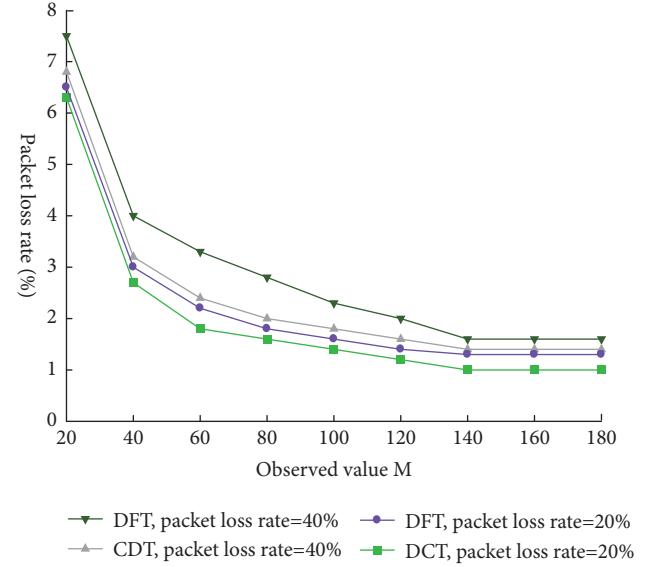


FIGURE 7: Relative reconstruction errors under different sparse bases.

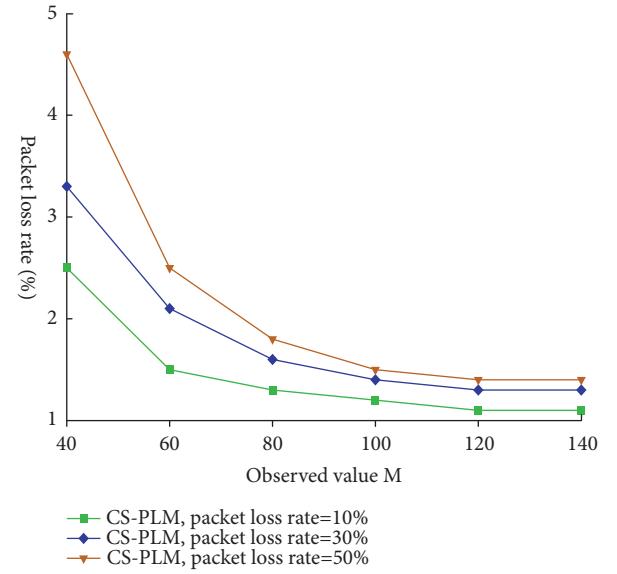


FIGURE 8: Relative reconstruction errors with different number of measurements.

should be increased to deal with the higher sparsity degree. Therefore, more measurements can alleviate the impacts of the packet loss ratio on the reconstruction accuracy to some extent.

When the decay coefficient of the event source is $n=0.01$ and the neighborhood range is $r=2$, the performance of the CS-PLM algorithm is shown in Figure 9. When the bit error rate is relatively low, i.e., $Pb=10^{-5}$, the performances of these three algorithms are almost the same. However, when the bit error rate is as high as $Pb=10^{-3}$, the CS-PLM algorithm outperforms the others. It is shown in Figure 7 that when $Pb=10^{-3}$, the SNR for data reconstruction of the CDG algorithm is 27.33dB and the high bit error rate obviously

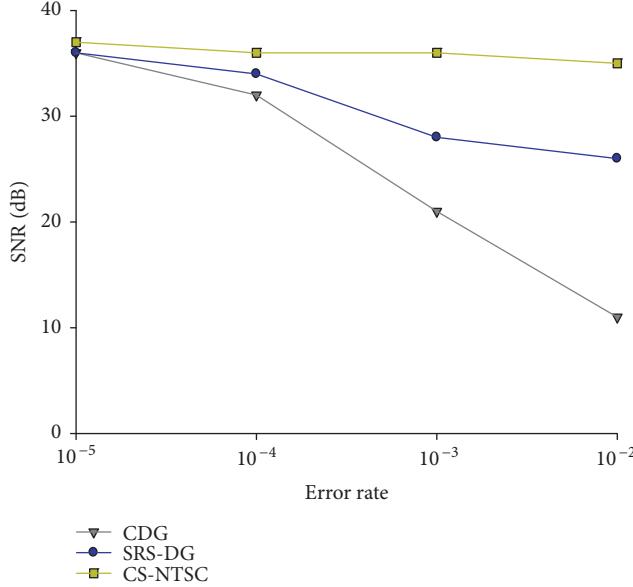


FIGURE 9: Performance analyses of the CS-PLM algorithm.

affects the performance of the CDG algorithm. The SNR for data reconstruction of the SRS-DG algorithm is 29.72dB. Since the SRS-DG algorithm is designed for recovering block loss, the data block is abandoned once transmission error occurs and the error-free nodes are measured with the sparse measurement matrix. Therefore, the measured information is reduced every time. The lost data block is then compensated by increasing the number of measurements through the next round of data gathering and the reconstruction SNR is therefore increased. As a result, the SNR is not high for this round of data reconstruction. The data reconstruction SNR obtained by the CS-PLM algorithm is 35.91dB. The transmission error is predicted by the spatial correlation of the data under certain conditions and the amount of abandoned information is therefore reduced. Henceforth, in the wireless scenarios with high bit error rate, the CS-PLM algorithm does not cause additional communication energy consumption and it can overcome the impacts of erroneous data blocks on data reconstruction. The efficiency of the CS-PLM algorithm is therefore verified.

For the CS-PLM algorithm, under the premise of ensuring the constant data compression rate, different network sizes require different measurement numbers in the data gathering process, which in turn causes the changing threshold for the judgment of node type. Therefore, different network size leads to different proportions of node types, which not only affects the total packet throughput of the network, but also affects the complexity of constructing the backup paths between CS nodes as well as the reliability of their transmission. Therefore, the performance of the algorithm is affected by the network size and related results are illustrated in Figure 10, where the packet loss ratio on the link is set to 20%. According to the curves in Figure 10, it is shown that as the number of nodes increases, the proportion of CS nodes decreases in the network. As for the reason, in large-scale

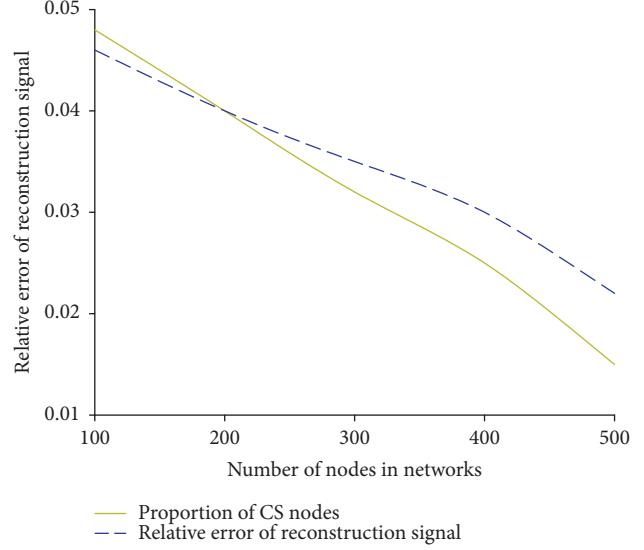


FIGURE 10: Relative reconstruction error with different network scale.

WSNs, the branches in tree topologies will be rather large and there are many smaller branches. However, according to the node classification rules in hybrid CS schemes, usually only the nodes in the main branch will be classified as CS nodes. Therefore, the proportion of CS nodes decreases, so does the complexity of constructing backup paths between CS nodes. According to the curves of the relative reconstruction error in Figure 10, the reconstruction accuracy of the CS-PLM algorithm improves with the increasing number of nodes in the network. This is due to the fact that as the proportion of CS nodes in the network decreases, the packet losses occurring in the network are uncorrelated with a larger probability. As a result, the reconstruction is less affected by the reliability of the backup paths and higher reconstruction accuracy can be guaranteed. According to the simulations above, the CS-PLM algorithm exhibits better performances in large-scale WSNs. Since the SPLI measurement matrix is designed in the CS-PLM algorithm, according to the CS theory, the correlation between the measurement matrix and the sparse base will affect the reconstruction performance of the algorithm. In order to investigate the performances of the CS-PLM algorithm under different sparse bases, we adopt the DCT base and the DFT base for comparison, as shown in Figure 11. It is shown that, with different number of measurements, the algorithm performs better under the DCT base. Therefore, it can be concluded that the SPLI matrix exhibits a lower correlation with the DCT base. That is, a better reconstruction performance of the algorithm can be ensured by choosing the DCT base to sense the changing of the data sparsity.

6. Conclusions

In order to address the CS based data gathering problem on unreliable links, we have proposed a CS-PLM algorithm. We designed the SPLM measurement matrix by analyzing the influence of the packet loss on CS based data gathering

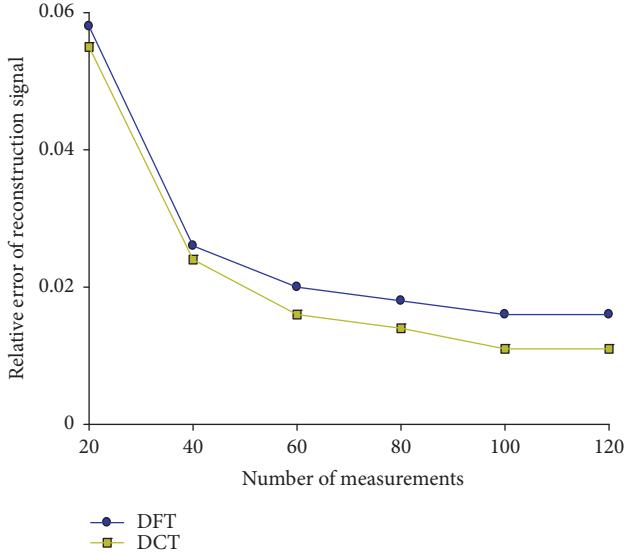


FIGURE 11: Relative reconstruction error under different sparse bases.

and further verified through simulations that the correlation of packet losses would undermine the reconstruction performance of the SPLM measurement matrix. Therefore, the nodes in the network are divided into TF nodes and CS nodes. Packet losses between TF nodes do not exhibit correlation and we simply adopt the SPLM measurement matrix to perform measurement projection. However, besides adopting the SPLM measurement matrix for measurement projection, the CS nodes also guarantee the transmission reliability by the minimum energy consumption backup paths and avoid the occurrence of correlated lost packets. It was shown in the simulation results that when the packet loss ratio on the link is 60%, the CS-PLM algorithm could still guarantee the effective reconstruction of the compressed data. Compared with other algorithms, the proposed algorithm showed great improvements in terms of reconstruction accuracy and the sparsity degree of the data set. It can accurately reflect the influences of the network size, sparse base, and the number of measurements for data gathering on the performance of the CS-PLM algorithm. Future work may focus on the impacts of packet loss ratio for mobile nodes on the reconstruction accuracy when the network flow is sufficiently large or small.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by Henan Province Education Department Cultivation Young Key Teachers in University

under Grant no. 2016GGJS-158; Henan Province Education Department Natural Science Foundation under Grants nos. 19A520006 and 18B520026; Luoyang Institute of Science and Technology High-level Research Start Foundation under Grant no. 2017BZ07; Natural Science and Technology Research of Henan Province Department of Science Foundation under Grants nos. 182102210428 and 162102310474.

References

- [1] W. Fang, X. Yao, X. Zhao, J. Yin, and N. Xiong, "A Stochastic Control Approach to Maximize Profit on Service Provisioning for Mobile Cloudlet Platforms," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 4, pp. 522–534, 2018.
- [2] T. Wang, J. Zhou, M. Huang et al., "Fog-based storage technology to fight with cyber threat," *Future Generation Computer Systems*, vol. 83, pp. 208–218, 2018.
- [3] V. K. Singh and M. Kumar, "A Compressed Sensing Approach to Resolve The Energy Hole Problem in Large Scale WSNs," *Wireless Personal Communications*, pp. 1–17, 2017.
- [4] P.-F. Wu, F. Xiao, C. Sha, H.-P. Huang, R.-C. Wang, and N.-X. Xiong, "Node scheduling strategies for achieving full-view area coverage in camera sensor networks," *Sensors*, vol. 17, no. 6, article no. 1303, 2017.
- [5] R. Zhou, C. Yang, J. Wan, W. Zhang, B. Guan, and N. Xiong, "Measuring complexity and predictability of time series with flexible multiscale entropy for sensor networks," *Sensors*, vol. 17, no. 4, 2017.
- [6] V. Angayarkanni, V. Akshaya, and S. Radha, "Design of a Compressive Sensing Based Fall detection System for Elderly Using WSN," *Wireless Personal Communications*, vol. 98, no. 1, pp. 421–437, 2018.
- [7] F. Pareschi, M. Mangia, D. Bortolotti et al., "Energy Analysis of Decoders for Rakeness-Based Compressed Sensing of ECG Signals," *IEEE Transactions on Biomedical Circuits and Systems*, 2017.
- [8] M. K. Khan, M. Shiraz, K. Z. Ghafoor, S. Khan, A. S. Sadiq, and G. Ahmed, "EE-MRP: Energy-Efficient Multistage Routing Protocol for Wireless Sensor Networks," *Wireless Communications and Mobile Computing*, vol. 2018, no. 1, Article ID 6839671, pp. 1–13, 2018.
- [9] M. Wu, L. Tan, and N. Xiong, "Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications," *Information Sciences*, vol. 329, pp. 800–818, 2016.
- [10] Z. Wu, N. Xiong, Y. Huang, and Q. Gu, "Optimal service distribution in WSN service system subject to data security constraints," *Sensors*, vol. 14, no. 8, pp. 14180–14209, 2014.
- [11] C. Yang, D. Puthal, S. P. Mohanty, and E. Kougnanos, "Big-Sensing-Data Curation for the Cloud is Coming: A Promise of Scalable Cloud-Data-Center Mitigation for Next-Generation IoT and Wireless Sensor Networks," *IEEE Consumer Electronics Magazine*, vol. 6, no. 4, pp. 48–56, 2017.
- [12] S. Khanna and C. R. Murthy, "Communication-efficient decentralized sparse Bayesian learning of joint sparse signals," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 617–630, 2017.
- [13] B. Sun, Y. Guo, N. Li, and D. Fang, "Multiple Target Counting and Localization Using Variational Bayesian em Algorithm in Wireless Sensor Networks," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2985–2998, 2017.

- [14] Z. Sun, X. Xing, C. Li, Y. Nie, and Y. Cao, "A Novel Nonlinear Multitarget k-Degree Coverage Preservation Protocol in Wireless Sensor Networks," *Journal of Sensors*, vol. 2016, 2016.
- [15] T. Wang, J. Zeng, Y. Lai et al., "Data collection from WSNs to the cloud based on mobile Fog elements," *Future Generation Computer Systems*, 2017.
- [16] H. Djelouat, H. Baali, A. Amira, and F. Bensaali, "Joint sparsity recovery for compressive sensing based EEG system," in *Proceedings of the 2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB)*, pp. 1-5, Salamanca, September 2017.
- [17] 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 Networks*, vol. 23, no. 4, pp. 1201-1222, 2017.
- [18] T. Wang, Z. Peng, J. Liang et al., "Following targets for mobile tracking in wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 12, no. 4, 2016.
- [19] Z. Sun, Y. Shu, X. Xing, W. Wei, H. Song, and W. Li, "LPOCS: A novel linear programming optimization coverage scheme in wireless sensor networks," *Ad Hoc & Sensor Wireless Networks*, vol. 33, no. 1-4, pp. 173-197, 2016.
- [20] G. Xing, M. Li, T. Wang, W. Jia, and J. Huang, "Efficient rendezvous algorithms for mobility-enabled wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 47-60, 2012.
- [21] S. Wang, H. Yi, L. Wu, F. Zhou, and N. N. Xiong, "Mining probabilistic representative gathering patterns for mobile sensor data," *Journal of Internet Technology*, vol. 18, no. 2, pp. 321-332, 2017.
- [22] Z. Lu, W. W. Li, and M. Pan, "Maximum lifetime scheduling for target coverage and data collection in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 2, pp. 714-727, 2015.

