IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 12, 2018, accepted June 10, 2018, date of publication June 13, 2018, date of current version July 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2846815

Sparsest Random Sampling for Cluster-Based Compressive Data Gathering in Wireless Sensor Networks

PENG SUN[1], LIANTAO WU[2], ZHIBO WANG[3,4], (Member, IEEE), MING XIAO[5], (Senior Member, IEEE), AND ZHI WANG[1], (Member, IEEE)

[1]State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China
[2]Shanghai Huawei Technologies Corporate Campus, Shanghai 200040, China
[3]School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China
[4]State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
[5]School of Electrical Engineering and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden

Corresponding author: Zhi Wang (wangzhi@iipc.zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 6177020565, Grant U1509215, Grant 6151101271, Grant 61273079, Grant 61773344, and Grant 61502352, in part by the Key Projects of Intergovernmental International Scientific and Technological Innovation Cooperation under Grant 2017YFE0101300, in part by the State Key Laboratory of Robotics under Grant 2017O-05, in part by the Natural Science Foundation of Hubei Province under Grant 2017CFB503, in part by the State Key Laboratory for Novel Software Technology under Grant KFKT2018B09, and in part by the Fundamental Research Funds for the Central Universities under Grant 2042018gf0043.

ABSTRACT Compressive data gathering (CDG) has been recognized as a promising technique to collect sensory data in wireless sensor networks (WSNs) with reduced energy cost and better traffic load balancing. Besides, clustering is often integrated into CDG to further facilitate the network performance. However, existing cluster-based CDG methods generally require a large number of sensor nodes to participate in each compressive sensing (CS) measurement gathering and rarely consider possible node failures due to power depletion or malicious attacks, leading to insufficient energy efficiency and poor system robustness. In this paper, we propose a sparsest random sampling scheme for cluster-based CDG (SRS-CCDG) in WSNs to achieve energy efficient and robust data collection. Specifically, sensor nodes are organized into clusters. In each round of data gathering, a random subset of sensor nodes sense the monitored field and transmit their measurements to the corresponding cluster heads (CHs). Then, each CH transmits the data gathered within its cluster to the sink. In SRS-CCDG, each sensor reading is regarded as one CS measurement, and both intra-cluster and inter-cluster data transmissions can be realized by two methods, i.e., relaying or direct transmission. Furthermore, we propose analytical models that study the relationship between the size of clusters and the energy cost when using different intra-cluster and inter-cluster transmission schemes, aimed at finding the optimal size of clusters and transmission schemes that could lead to minimum energy cost. Then, we present a centralized clustering algorithm based on the theoretical analysis. Finally, we investigate the robustness of signal recovery performance of SRS-CCDG when node failures happen. Extensive simulations demonstrate that SRS-CCDG can significantly reduce the energy cost and improve the system robustness to node failures.

INDEX TERMS Compressive data gathering, cluster, node failures, wireless sensor networks.

I. INTRODUCTION

Advances in computing and communication technologies have led to intensive research effort on wireless sensor networks (WSNs) [1], [2]. WSNs have found extensive applications in urban traffic monitoring and environmental surveillance [3], [4]. Typically, a WSN consists of a number of sensor nodes, which are randomly distributed in the field under surveillance, and a sink node. Generally, sensor nodes are required to collect data periodically and transmit them to the sink through multi-hop routing, and then the information aggregation and extraction tasks are performed at the sink. Considering that sensor nodes usually have limited energy supply and that replacing or recharging the batteries of sensor nodes is difficult in practical WSN deployments, a primary

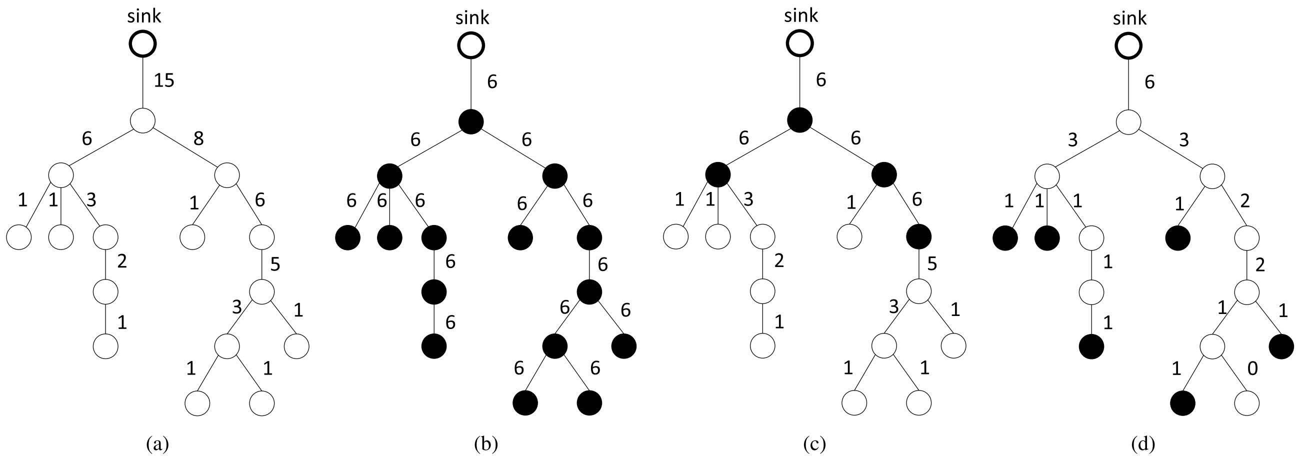


FIGURE 1. Comparison of different data gathering schemes under multi-hop tree-type topology. (a) Non-CS. (b) CDG. (c) H-CDG. (d) SRS-CDG.

objective of data gathering in WSNs is to obtain an accurate approximation of the signal field with as little energy expenditure as possible.

The spatial correlation of sensor readings in WSNs results in an inherent sparsity of data under a proper transform basis, which facilitates the broad uses of compressive sensing (CS) technology [5], [6] in data gathering in WSNs. CS provides a new avenue for energy efficient data gathering in WSNs, as it promises perfect recovery of sparse signals using only a small number of random measurements [7], [8]. In the past few years, there have been considerable research interests in integrating CS into data gathering in WSNs. The scheme of compressive data gathering (CDG) is first proposed in [9] to reduce global scale communication cost and balance the traffic load throughout the entire network. However, in CDG, each sensor node needs to transmit M (M is the required number of measurements for accurate signal recovery) data packets in each round of data gathering. That is, the total number of data transmissions for a network of N sensor nodes is MN, which still incurs high communication cost. To address this problem, hybrid CDG (H-CDG) approaches are proposed in [10]–[12]. In the hybrid methods, the nodes close to the leaf nodes transmit the original data without using the CS technique, while the nodes close to the sink transmit data to sink using the CS method. In [13], a sparsest random scheduling scheme is proposed for CDG (SRS-CDG) in WSNs to further reduce the transmission cost by treating each sensor reading as one CS measurement, where the measurement matrix is a sparsest one. The non-CS, CDG, H-CDG and SRS-CDG schemes are shown in Fig. 1, where the black sensor nodes transmit data by using the CS technique while the white sensor nodes directly transmit data without any compression. The link labels represent the number of transmitted data packets in one round of data gathering. Note that the above mentioned CDG schemes are based on routing trees. Considering the vast advantages that the clustering method has over the tree topology [14]–[16], e.g. fault tolerance and traffic load balancing, a number of cluster-based CDG approaches have been proposed [17]–[20]. In [17], hybrid CS-based data gathering scheme is first applied in clustered WSNs. Based on traditional intra-cluster transmission and CS based inter-cluster transmission schemes, an analytical model is proposed in order to find the optimal size of clusters that could lead to the minimum number of transmissions. In [18], the combination of CDG and clustering in WSNs is realized by utilizing block diagonal matrices as the measurement matrices, which results in significant reduction in transmission power consumption. Nevertheless, these cluster-based CDG schemes still need a large number of intra-cluster data transmissions. Besides, none of the existing works have investigated the robustness of signal recovery performance when node failures happen in the network.

In this paper, we propose a sparsest random sampling scheme for cluster-based compressive data gathering (SRS-CCDG) in WSNs. Specifically, sensor nodes are organized into clusters. In each round of data gathering, a random subset of sensor nodes sense the signal field and each generated sensor reading is treated as one CS measurement. Within each cluster, the nodes that have participated in sensing transmit their measurements to the CH, and then each CH transmits the data gathered within its cluster to the sink through relaying by other CHs. The comparison between traditional cluster-based CDG approaches and SRS-CCDG is shown in Fig. 2 (the required number of measurements for accurate signal recovery is assumed to be 10). Note that in SRS-CCDG both intra-cluster and inter-cluster data transmissions can be implemented through direct transmission or relaying by other intermediate nodes. Furthermore, we propose analytical models that study the relationship between the size of clusters and the energy cost when using different intra-cluster and inter-cluster transmission schemes, aimed at finding the optimal size of clusters and transmission schemes that could lead to minimum energy cost. Then, we present a centralized clustering algorithm based on the theoretical analysis.

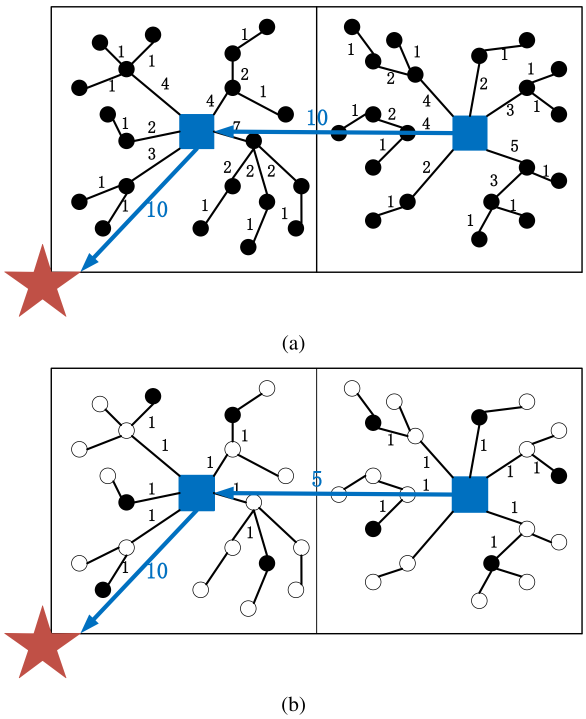


FIGURE 2. Comparison between traditional cluster-based CDG approaches and SRS-CCDG. (a) Traditional cluster-based CDG scheme [17]. (b) SRS-CCDG.

Finally, we investigate the robustness of signal recovery performance of SRS-CCDG when node failures happen in the network. Extensive simulation results demonstrate that SRS-CCDG can significantly reduce the energy cost and improve the system robustness to unavoidable node failures.

The main contributions of this paper can be summarized as follows:

- We propose a sparsest random sampling scheme for cluster-based compressive data gathering in WSNs, where the sparest random sampling scheme (e.g. treating each sensor reading as one CS measurement) is first integrated into clustered WSNs to achieve energy efficient and robust data collection.
- We propose analytical models that study the relationship between the size of clusters and the energy cost when using different intra-cluster and inter-cluster transmission schemes. The optimal size of clusters and transmission schemes are obtained to achieve minimum energy cost.
- We are the first to investigate the robustness of signal recovery performance to node failures for different cluster-based CDG approaches, and simulation results confirm that the performance of SRS-CCDG remains relatively stable when node failures happen, outperforming existing methods.

The remainder of this paper is organized as follows: Section II presents an overview of the SRS-CCDG scheme. Section III presents the analytical model of intra-cluster data transmissions, where the two transmission schemes (e.g. direct transmission and relaying) are compared in terms of the energy cost. Then in Section IV, the relationship between the size of clusters and the total energy cost when using different intra-cluster and inter-cluster transmission schemes is investigated, and the optimal size of clusters and transmission schemes are obtained. Section V provides a centralized implementation of the SRS-CCDG scheme. The simulations and performance evaluations are presented in Section VI. Finally, Section VII concludes the paper.

II. OVERVIEW OF SRS-CCDG

In the network model considered here, we assume that N sensor nodes have been deployed in a square sensing area, to measure the field of some physical phenomena (e.g. temperature, humidity, pressure, etc.). The sensor nodes need to collect data periodically and transmit them to the sink node. First, we make the following reasonable assumptions:

- The sensor nodes are uniformly and independently distributed in the sensing area [17], [21], [22].
- All sensor nodes are aware of their own geographic locations through in-built GPS modules or utilizing other sensor node localization techniques [23], [24].
- The sensor nodes are able to adjust the transmission power according to the required communication distance [25], [26].

In each round of data gathering, the sensor readings of all the N nodes in the network can be represented as a vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$, where x_i is the measurement of the i-th sensor node. Thanks to the spatial correlation of the monitored signal field, $\boldsymbol{x}$ usually has a sparse (compressible) representation under some transform basis [11], [27], [28], e.g. Fourier, DCT, wavelets etc. Formally, if we denote by $\boldsymbol{\Psi}$ the sparsifying basis, we have $\boldsymbol{x} = \boldsymbol{\Psi s}$, and $\boldsymbol{s}$ is a sparse vector, which contains only a small number of nonzero entries, i.e. $\|\boldsymbol{s}\|_0 = K, K \ll N$.

Instead of transmitting all N original sensor readings to the sink as in data gathering without using CS, only M ($M \ll N$) linear projections of $\boldsymbol{x}$ need to be transmitted to the sink in CDG approaches. Then at the sink, we have

$$\boldsymbol{y} = \boldsymbol{\Phi x}, \tag{1}$$

where $\boldsymbol{\Phi}$ is the measurement matrix of size $M \times N$, $\boldsymbol{y}$ is the measurement vector consisting of the M projections received at the sink. Substituting $\boldsymbol{x} = \boldsymbol{\Psi s}$ into (1), we then have

$$\boldsymbol{y} = \boldsymbol{\Phi x} = \boldsymbol{\Phi \Psi s} = \boldsymbol{As}, \tag{2}$$

where $\boldsymbol{A}$ is the equivalent sensing matrix.

Taking the measurement noise into consideration, we have

$$\boldsymbol{y} = \boldsymbol{As} + \boldsymbol{e}, \tag{3}$$

where $\|\boldsymbol{e}\|_2 \leq \xi$ represents the sensing noise induced by the limitations in the sensing device.

Therefore, at the sink node, given the measurement vector $\boldsymbol{y}$, the measurement matrix $\boldsymbol{\Phi}$, and the sparsifying basis $\boldsymbol{\Psi}$, reconstruction of the signal field $\boldsymbol{x}$ can be realized by solving the following constrained l_1-norm based minimization problem

$$\min \|\boldsymbol{s}\|_1 \quad s.t.\ \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{s}\|_2 \le \xi, \tag{4}$$

which can be solved using an efficient solver such as basis pursuit de-noising (BPDN) or various greedy algorithms (e.g. orthogonal matching pursuit, OMP) [29]–[31].

In our method, sensor nodes are organized into clusters, and each cluster has a cluster head (CH). In each round of data gathering, a random subset of sensor nodes sense the signal field, and the generated sensor measurement, together with the sensor node's ID, is organized into a data packet, which is then transmitted to the corresponding CH. After receiving all the data packets generated within its cluster, each CH transmits them to the sink node through relaying by other CHs.

In order to lower the transmission energy cost, we need to reduce the number of sensor nodes involved in each CS measurement gathering. Therefore, in SRS-CCDG, we refer to each sensor reading as one CS measurement, and the resulting measurement matrix is a sparsest one containing only one nonzero entry in each row, which can be formulated as follows:

$$\boldsymbol{\Phi}(i,j) = \begin{cases} 0, & if\, j = J(i) \\ 1, & otherwise \end{cases}, \tag{5}$$

where $i = 1, 2, \ldots, M$ is the row index and also the sequence number of the received data packets, $j = 1, 2, \ldots, N$ is the column index, and $J(i)$ is a random index from the interval $[1, N]$, representing the ID of the sensor node whose data packet has been received at the sink. Based on the definition of $\boldsymbol{\Phi}$, one round of data gathering can be modeled as a CS process, which can be expressed as

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \\ \vdots \\ \boldsymbol{\Phi}_M \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \end{bmatrix}, \tag{6}$$

where $\boldsymbol{\Phi}_i$ represents the i-th row of the measurement matrix $\boldsymbol{\Phi}$. The vector $\boldsymbol{y} = [y_1, y_2, \ldots, y_M]^T$ is the CS measurement vector consisting of the M sensor measurements received at the sink. It has been shown in [13] that this kind of measurement matrix satisfies the restricted isometry property (RIP) condition given enough number of measurements. Therefore, recovery of the original signal field $\boldsymbol{x}$ can be realized by solving a sparse signal reconstruction problem as shown in (4).

Note that both intra-cluster and inter-cluster data transmissions can be realized through relaying by intermediate sensor nodes or using direct transmission via transmission power control. The total energy cost of SRS-CCDG is directly related to the size of clusters and the adopted intra-cluster and inter-cluster transmission schemes. In this regard, our goal is to determine the optimal size of clusters and transmission schemes, such that the total energy cost is minimized.

III. ANALYSIS ON INTRA-CLUSTER TRANSMISSION

In this section, we conduct a thorough analysis on the energy cost of intra-cluster data transmissions and make a comparison between the two different transmission schemes, e.g. direct transmission or data relaying. Without loss of generality, we assume that N sensor nodes are uniformly and independently distributed in a square sensing area. The sensing area is divided into small grids of size $a \times a$, and sensor nodes located in a square region of size $Da \times Da$ form a cluster [17], as shown in Fig. 3. We assume that the default transmission range of sensor nodes is r. That is, any two nodes can communicate with each other, without the need for relaying or transmission power control, if their euclidian distance is within r. In this paper, we let $a = \frac{r}{\sqrt{2}}$ so that any two nodes in a grid are within the transmission range of each other. The energy cost of data transmission can be calculated as follows [32]:

$$E_{Tx}(L, d) = E_{elec} \times L + \varepsilon_{amp} \times L \times d^2, \tag{7}$$
$$E_{Rx}(L) = E_{elec} \times L, \tag{8}$$

where $E_{Tx}(L, d)$ represents the energy cost for tranmitting L bits of data over a distance of d, $E_{Rx}(L)$ represents the energy cost for receiving L bits of data, E_{elec} is the energy consumed in the transceiver circuitry at the transmitter or the receiver, and ε_{amp} is the energy consumed at the output transmitter antenna for transmitting one meter. Therefore, the total energy cost for transmitting L bits of data over a distance of d is $E_{Tx}(L, d) + E_{Rx}(L) = 2E_{elec}L + \varepsilon_{amp}Ld^2$, and we define $c_1 = 2E_{elec}L$ and $c_2 = \varepsilon_{amp}L$ in the following analysis.

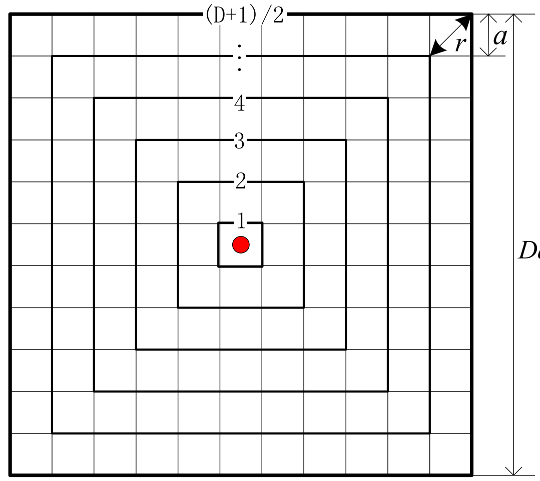


FIGURE 3. The sensing area is divided into small grids of size $a \times a$. Sensor nodes located in a square region of size $Da \times Da$ form a cluster, and the CH is located at the center of the cluster.

A. DIRECT TRANSMISSION

We assume that the CH is located at the center of the cluster, which is the case that produces minimum intra-cluster transmission energy cost [17], and the coordinates of the CH are assumed to be $(0, 0)$. Since the sensor nodes are uniformly distributed in the cluster, the density of the node distribution in the center grid that contains the CH is

$$f_1(x,y)=\frac{1}{a^2},\quad -\frac{a}{2}\le x\le\frac{a}{2},\ -\frac{a}{2}\le y\le\frac{a}{2}. \tag{9}$$

Therefore, the expected squared distance between the sensor nodes in the center grid and the CH, $E(d_1^2)$, can be calculated as

$$\begin{aligned}E(d_1^2)&=\iint_{\Omega_1}\frac{1}{a^2}(x^2+y^2)dxdy\\&=\int_{-\frac{a}{2}}^{\frac{a}{2}}\int_{-\frac{a}{2}}^{\frac{a}{2}}\frac{1}{a^2}(x^2+y^2)dxdy\\&=\frac{4}{a^2}\int_0^{\frac{a}{2}}\int_0^{\frac{a}{2}}(x^2+y^2)dxdy=\frac{a^2}{6},\end{aligned} \tag{10}$$

where $\Omega_1=\left\{(x,y)|-\frac{a}{2}\leqslant x\leqslant\frac{a}{2},-\frac{a}{2}\leqslant y\leqslant\frac{a}{2}\right\}$. Note that if the CH is not located at the center of the cluster, i.e. x_0 and y_0 are not simultaneously equal to 0, we have $E(d_1^2)=\frac{a^2}{6}+x_0^2+y_0^2>\frac{a^2}{6}$, which verifies that the optimal location for the CH is indeed at the center of the cluster.

In a similar way, the expected squared distance from the sensor nodes in the second layer of grids to the CH, $E(d_2^2)$, can be expressed as

$$\begin{aligned}E(d_2^2)&=\iint_{\Omega_2}\frac{1}{(3a)^2-a^2}(x^2+y^2)dxdy\\&=\int_{-\frac{3a}{2}}^{\frac{3a}{2}}\int_{-\frac{3a}{2}}^{\frac{3a}{2}}\frac{1}{8a^2}(x^2+y^2)dxdy\\&\quad-\int_{-\frac{a}{2}}^{\frac{a}{2}}\int_{-\frac{a}{2}}^{\frac{a}{2}}\frac{1}{8a^2}(x^2+y^2)dxdy\\&=\frac{5a^2}{3},\end{aligned} \tag{11}$$

where $\Omega_2=\left\{(x,y)|-\frac{3a}{2}\leqslant x\leqslant\frac{3a}{2},-\frac{3a}{2}\leqslant y\leqslant\frac{3a}{2}\right\}-\Omega_1$.

Following this pattern, the distribution density of the sensor nodes in the h-th ($h\ge 2$) layer of grids is $\frac{1}{((2h-1)a)^2-((2h-3)a)^2}$. Thus, the expected squared distance between the sensor nodes in the h-th layer of grids and the CH, $E(d_h^2)$, is calculated as

$$\begin{aligned}E(d_h^2)&=\iint_{\Omega_h}\frac{1}{((2h-1)\cdot a)^2-((2h-3)\cdot a)^2}(x^2+y^2)dxdy\\&=\int_{-\frac{(2h-1)a}{2}}^{\frac{(2h-1)a}{2}}\int_{-\frac{(2h-1)a}{2}}^{\frac{(2h-1)a}{2}}\frac{1}{8(h-1)a^2}(x^2+y^2)dxdy\\&\quad-\int_{-\frac{(2h-3)a}{2}}^{\frac{(2h-3)a}{2}}\int_{-\frac{(2h-3)a}{2}}^{\frac{(2h-3)a}{2}}\frac{1}{8(h-1)a^2}(x^2+y^2)dxdy\\&=\frac{[(2h-1)^4-(2h-3)^4]a^2}{48(h-1)}=\frac{[4(h-1)^2+1]a^2}{3}.\end{aligned} \tag{12}$$

Since the number of grids in the h-th ($h\ge 2$) layer is $8(h-1)$, and the number of sensor nodes in each grid is $\frac{1}{(Da)^2}\cdot a^2=\frac{1}{D^2}$, the expected intra-cluster transmission energy cost when using direct transmission scheme is given as follows

$$\begin{aligned}E_{intra}^d&=\frac{c_1+c_2\frac{a^2}{6}}{D^2}+\sum_{h=2}^{\frac{D+1}{2}}\frac{8(h-1)}{D^2}\cdot\left\{c_1+c_2\frac{[4(h-1)^2+1]a^2}{3}\right\}\\&=\frac{1}{D^2}\cdot(c_1+c_2\frac{a^2}{6})+c_1\sum_{h=2}^{\frac{D+1}{2}}\frac{8(h-1)}{D^2}\\&\quad+c_2\sum_{h=2}^{\frac{D+1}{2}}\frac{8(h-1)}{D^2}\cdot\frac{[4(h-1)^2+1]a^2}{3}\\&=\frac{1}{D^2}\cdot(c_1+c_2\frac{a^2}{6})+\frac{8c_1}{D^2}\sum_{h=1}^{\frac{D-1}{2}}h+\frac{8c_2a^2}{3D^2}[4\sum_{h=1}^{\frac{D-1}{2}}h^3+\sum_{h=1}^{\frac{D-1}{2}}h]\\&=\frac{1}{D^2}\cdot(c_1+c_2\frac{a^2}{6})+\frac{c_1(D^2-1)}{D^2}+\frac{c_2a^2(D^4-1)}{6D^2}\\&=c_1+\frac{c_2a^2D^2}{6}.\end{aligned} \tag{13}$$

Note that when the direct transmission scheme is adopted for intra-cluster transmission, the sensor nodes close to the CH will consume less energy, which makes it convenient for CH replacement when necessary.

B. DATA RELAYING

When intra-cluster data transmissions are realized through relaying by other sensor nodes in the cluster, the sensor nodes in the center grid take only one hop to transmit their data to the CH. Obviously, the nodes in the h-th layer take h hops to transmit data to the CH. Therefore, the expected number of intra-cluster data transmissions is given as

$$E_T=\frac{1}{D^2}\cdot 1+\sum_{h=2}^{\frac{D+1}{2}}\frac{8(h-1)}{D^2}\cdot h, \tag{14}$$

given that one hop represents a communication distance of a, the expected intra-cluster transmission energy cost when using data relaying is

$$\begin{aligned}E_{intra}^r&=\frac{1}{D^2}\cdot(c_1+c_2a^2)+\sum_{h=2}^{\frac{D+1}{2}}\frac{8(h-1)}{D^2}\cdot h\cdot(c_1+c_2a^2)\\&=\frac{1}{D^2}\cdot(c_1+c_2a^2)+\frac{8(c_1+c_2a^2)}{D^2}\sum_{h=2}^{\frac{D+1}{2}}(h-1)h\\&=\frac{1}{D^2}\cdot(c_1+c_2a^2)+\frac{8(c_1+c_2a^2)}{D^2}\left[\sum_{h=1}^{\frac{D-1}{2}}h^2+\sum_{h=1}^{\frac{D-1}{2}}h\right]\\&=\frac{1}{D^2}\cdot(c_1+c_2a^2)+(c_1+c_2a^2)\left[\frac{(D^2-1)}{3D}+1-\frac{1}{D^2}\right]\\&=(c_1+c_2a^2)\left[\frac{D}{3}-\frac{1}{3D}+1\right].\end{aligned} \tag{15}$$

Intuitively, when the cluster size (i.e. the value of D) is relatively small, direct transmission will consume less energy, while using data relaying will be more energy efficient when the cluster size becomes larger. Now, we make a comparison between the two intra-cluster transmission schemes. The difference of energy cost between the two intra-cluster transmission schemes is defined as $delta = E^r_{intro} - E^d_{intro}$. Assume that there are 1000 sensor nodes uniformly distributed in a 1000m $\times$ 1000m two-dimensional sensing area. We show the relationship between the energy cost difference and the size of clusters for the following example set of system parameters: $E_{elec} = 50$ nJ/bit, $\varepsilon_{amp} = 100$ pJ/bit/m^2, each transmitted packet contains $L = 1024$ bits of data, and a is set to 7 when the transmission power is -10dBm [33]. As we can observe from Fig. 4, when the cluster size $D \leq 42$, direct transmission scheme produces less energy cost than data relaying in intra-cluster data transmissions. The reason is that when the cluster size is relatively small, the majority of sensor nodes in the cluster have a small communication distance to the CH, thus the energy cost of direct transmission is small. In contrast, the relaying scheme needs more than one transmission to send data to the CH, leading to larger energy consumption. However, when the cluster size is relatively large, the result is reversed.

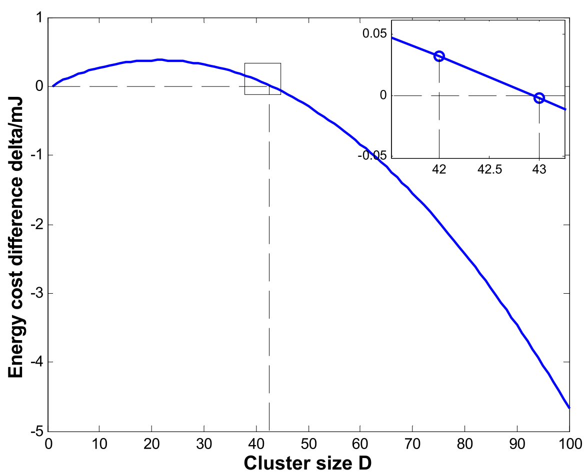


FIGURE 4. Difference of intra-cluster transmission energy cost between using direct transmission and data relaying as the transmission scheme.

IV. ANALYSIS ON INTER-CLUSTER TRANSMISSION AND THE OPTIMAL CLUSTER SIZE

In this section, we first investigate the energy cost of inter-cluster data transmissions under two different inter-cluster transmission schemes. Assume that there is a sink node located at the corner of the square sensing area, and the coordinates of the sink are (0, 0). This assumption is typical in WSNs [17] as the sink is usually placed outside the sensing area for easy deployment. The results in this paper can be easily extended to other cases when the sink is not located at the corner of the sensing area.

In each round of data gathering of SRS-CCDG, after collecting all the data packets generated within their clusters, the CHs then transmit them to the sink node along a backbone tree. In our analytical model, the backbone tree is constructed as shown in Fig. 5 [17]: 1) all CHs transmit data to their left-neighbor CH until reaching the leftmost cluster; 2) for clusters at the leftmost column, CHs transmit data to down-neighbor CHs until reaching the left-bottom cluster; and 3) the CH of the left-bottom cluster transmits data to the sink. Since the cluster size is D, it takes D hops to transmit a data packet from one CH to its neighboring CH. For the cluster at the left-bottom corner, it takes approximately $\frac{D}{2}$ hops to transmit a data packet from the CH to the sink. Note that in SRS-CCDG, inter-cluster data transmissions can be realized in two ways, i.e. through relaying by some intermediate sensor nodes or direct transmission from one CH to its neighboring CH via transmission power control. Next, we propose analytical models that study the inter-cluster transmission energy cost under the two transmission schemes.

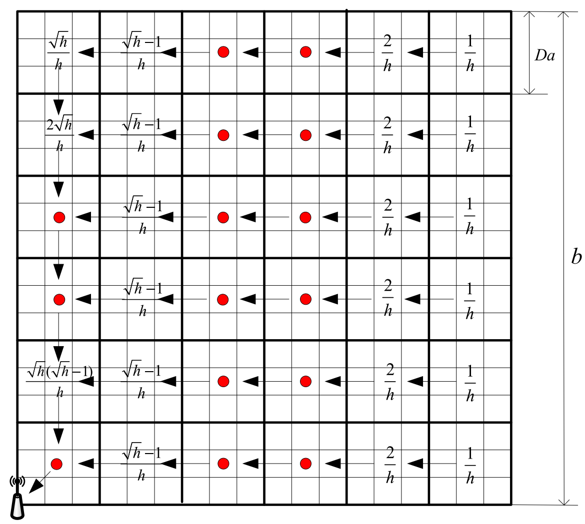


FIGURE 5. The inter-cluster data transmissions.

Assume the edge length of the sensing area is b, so the total number of clusters in the network is $h = \frac{b^2}{D^2 a^2}$. As shown in Fig. 5, the CHs at the rightmost column (denoted by the first column) have a probability of $\frac{1}{h}$ to transmit data, and the CHs at the second column transmit data with a probability of $\frac{2}{h}$. Except the leftmost column, the expected number of transmissions for all CHs to forward one data packet is expressed as

$$[1/h + 2/h + \ldots + (\sqrt{h} - 1)/h] * \sqrt{h} = \frac{\sqrt{h} - 1}{2}. \quad (16)$$

After the data packets are forwarded to the leftmost column, the CHs there then transmit data to their down-neighbor CHs until reaching the CH of the left-bottom cluster. Therefore, except the left-bottom cluster, the expected number of transmissions for all the CHs at the leftmost column to forward one data packet is

$$\begin{aligned}\sqrt{h}/h + 2\sqrt{h}/h + \ldots + (\sqrt{h}-1)\sqrt{h}/h \\ = [1/h + 2/h + \ldots + (\sqrt{h}-1)/h]\sqrt{h} \\ = \frac{\sqrt{h}-1}{2}. \end{aligned} \tag{17}$$

Finally, for the cluster at the left-bottom corner, the expected number of transmissions for the CH to send one data packet to the sink node is $\frac{\sqrt{h}\cdot\sqrt{h}}{h} = 1$.

Therefore, the expected energy cost of inter-cluster transmission when using direct transmission between neighboring CHs for gathering M sensor measurements is given as

$$\begin{aligned} E^{d}_{inter} = (\frac{b}{Da} - 1) \cdot M \cdot (c_1 + c_2 D^2 a^2) \\ + M(c_1 + c_2(\frac{D}{2})^2 a^2), \end{aligned} \tag{18}$$

while when inter-cluster data transmissions (i.e. transmission between neighboring CHs) are implemented through relaying by some intermediate sensor nodes, the expected inter-cluster transmission energy cost is

$$\begin{aligned} E^{r}_{inter} &= (\frac{b}{Da} - 1) \cdot M \cdot D \cdot (c_1 + c_2 a^2) + \frac{D}{2} M(c_1 + c_2 a^2) \\ &= (\frac{Mb}{a} - \frac{MD}{2}) \cdot (c_1 + c_2 a^2). \end{aligned} \tag{19}$$

An important task of our method is to determine the optimal size of clusters and intra-cluster and inter-cluster transmission schemes, such that the total energy cost is minimized. We now present the relationship between the size of clusters and the total energy cost when using different intra-cluster and inter-cluster transmission schemes in Fig. 6. Note that in this figure, the parameters are configured as follows: E_{elec}=50 nJ/bit, ε_{amp}=100 pJ/bit/m^2, the grid length $a = 7$m, the edge length $b = 1000$m, the number of sensor nodes $N = 1000$, and the required measurement number $M = 200$. As shown in Fig. 6, 'intraD' and 'intraR' represent that intra-cluster transmission adopts direct transmission and data relaying, respectively. 'interD' and 'interR' mean using direct transmission or data relaying in inter-cluster data transmissions. We can observe from the figure that when using a proper cluster size, the total energy cost is minimized when direct transmission scheme is adopted in both intra-cluster and inter-cluster transmission. In this case, the total energy cost can be calculated as

$$E^{dd}_{sum} = E^{d}_{inter} + M \cdot E^{d}_{intra}, \tag{20}$$

which is a convex function of D. Therefore, the optimal size of clusters can be determined numerically via a gradient-descent algorithm, as shown in Algorithm 1. It turns out that when $D = 5$, the total energy cost is minimized. In summary, the optimal size of clusters is $D^* = 5$ and the optimal intra-cluster and inter-cluster transmission schemes are both direct transmission. In the following sections, the optimal case is simply referred to as SRS-CCDG for clarity.

Algorithm 1 Gradient-Descent Algorithm for Optimal Cluster Size

Input: E_{elec}, ε_{amp}, Packet length L;
Node number N, Required measurement number M,
Stepsize η, Maximum number of iterations *itermax*
Output: Optimal cluster size D^*

```
Initialize cluster size D;
∇E = dE^{dd}_{sun}/dD = bc2Ma − (bc1M/a)·(1/D^2) − (4c2Ma^2/3)·D;
for i = 1 to itermax do
    D = D - η*∇E;
end
```

We present the data gathering process of SRS-CCDG when using direct transmission as both the intra-cluster and inter-cluster transmission scheme in Fig. 7. Points marked with different colors represent that the sensor nodes are organized into different clusters, the red lines represent direct transmission between the nodes and their corresponding CHs, and the direct data transmissions between neighboring CHs are represented by blue lines.

V. MINIMUM ENERGY COST CLUSTERING ALGORITHM

This section presents a centralized clustering algorithm to achieve energy efficient SRS-CCDG. We assume that the sink node has the full knowledge of the network topology. The sink will divide the sensor nodes into clusters, choose a CH for each cluster, and construct a backbone tree that connects all CHs to the sink.

Based on the theoretical analysis in Section IV, we can determine the optimal number of clusters for a given square sensing area with edge length b as

$$h^* = \left\lceil \frac{b^2}{(D^* a)^2} \right\rceil. \tag{21}$$

Therefore, the objective of SRS-CCDG is to divide the N sensor nodes in the network into h^* clusters and minimize

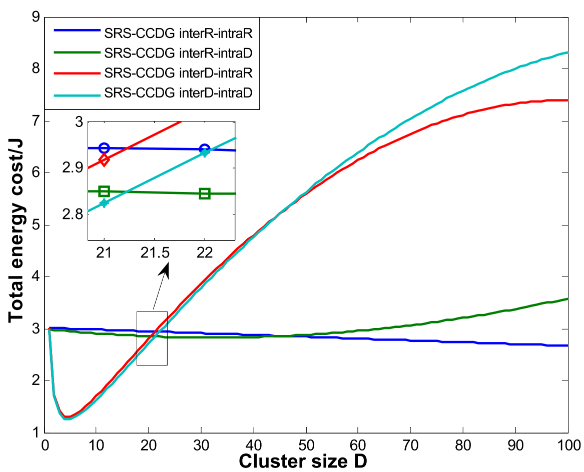


FIGURE 6. The relationship between the size of clusters and the total energy cost when using different intra-cluster and inter-cluster transmission schemes in SRS-CCDG.

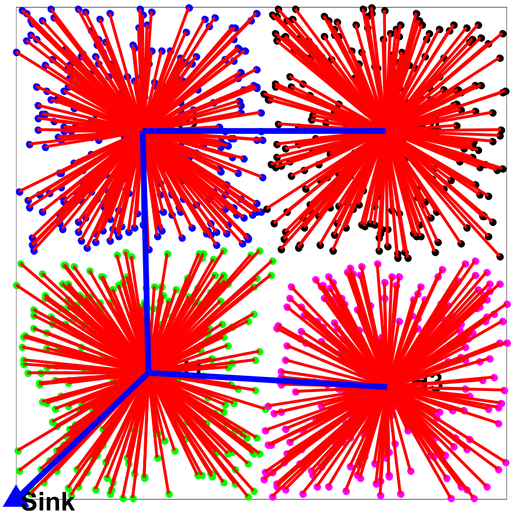


FIGURE 7. Data gathering process of SRS-CCDG when using direct transmission scheme in both intra-cluster and inter-cluster transmission.

the total energy cost, which can be formulated as

$$\begin{aligned} &\min \ \{E_{inter} + M \cdot E_{intra}\} \\ &s.t.\ E_{CH_i} \geq E_t, \quad i = 1, 2, \ldots, h^*, \end{aligned} \tag{22}$$

where E_{CH_i} represents the residual energy of the i-th CH, and E_t is the minimum energy required to be a CH. That is, when a CH's remaining energy is less than E_t, we need to conduct CH replacement.

Now, we present the centralized clustering algorithm. This algorithm consists of two steps: 1) after determining the optimal number of clusters h^*, select h^* CHs from the N sensor nodes and divide them into h^* clusters; 2) construct a backbone routing tree that connects all CHs to the sink node. The above problem has been proven to be NP-hard [34], which can be solved by an iterative method introduced in [17].

1) Randomly select h^* CHs from the N sensor nodes, and connect sensor nodes to their closest CHs.
2) In each cluster, choose a new CH, such that the total intra-cluster transmission energy cost is minimized.
3) Repeat the above two steps until the CHs would not change any more.
4) Construct a backbone routing tree that connects all CHs to the sink node using a minimum spanning tree (MST) based method.

VI. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed SRS-CCDG, and make a comparison with two traditional schemes, i.e. the classical compressive data gathering (CDG) [9] and the cluster-based hybrid CDG (HCS-CCDG) [17]. Note that in CDG, both intra-cluster and inter-cluster data transmissions can be implemented through relaying by other nodes or direct transmission. Thus, we first try to find the optimal intra-cluster and inter-cluster transmission schemes in CDG, and then use the best performance of CDG as the baseline.

A. SIMULATION SETTINGS

Since energy efficiency is one of the dominating issues in WSNs, we use the energy cost as the performance metric. In this paper, we only consider the energy cost for data gathering as the energy cost for establishing routing information is negligible [13]. The network parameters are configured as shown in Table 1.

TABLE 1. Parameter configuration.

E_{elec}	50 nJ/bit
ε_{amp}	100 pJ/bit/m^2
Grid length a	1-7m (7 default)
Boundary length b	500-1000m (1000 default)
Sensor nodes number N	500-1000 (1000 default)
Threshold of Measurement number M	100-200 (200 default)

B. ENERGY COST COMPARISON

We first present the relationship between the energy cost and the size of clusters when using different intra-cluster and inter-cluster transmission schemes in CDG. As we can observe from Fig. 8, the most energy efficient intra-cluster and inter-cluster transmission schemes are both direct transmission. Therefore, in the following simulations, CDG using direct transmission scheme both in intra-cluster and inter-cluster transmission is utilized as a baseline. Note that both intra-cluster and inter-cluster transmission in HCS-CCDG are realized through data relaying.

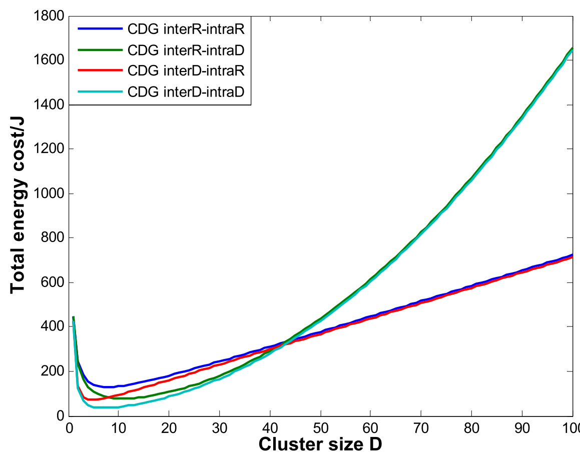
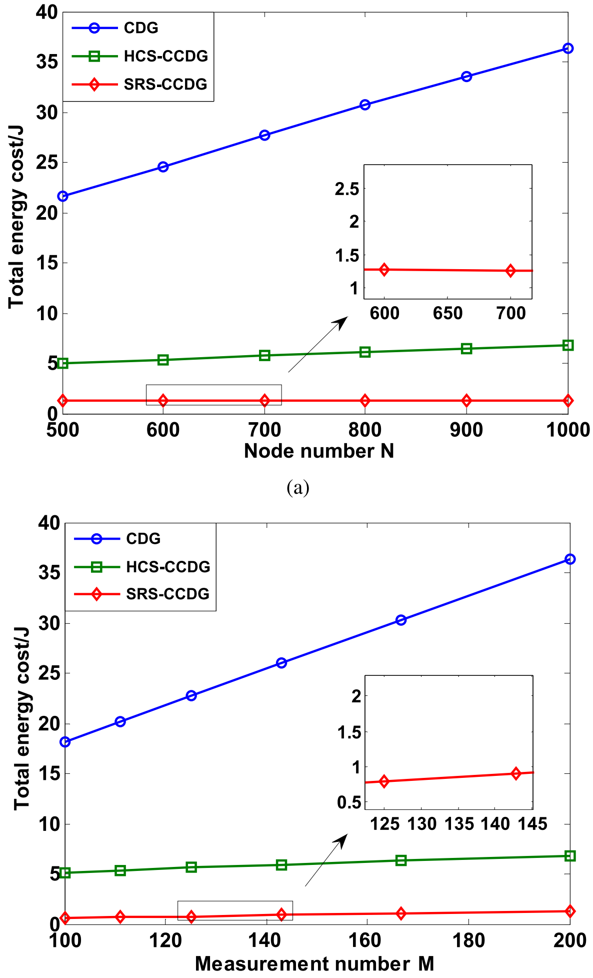


FIGURE 8. The relationship between the energy cost and the size of clusters when using different intra-cluster and inter-cluster transmission schemes in CDG.

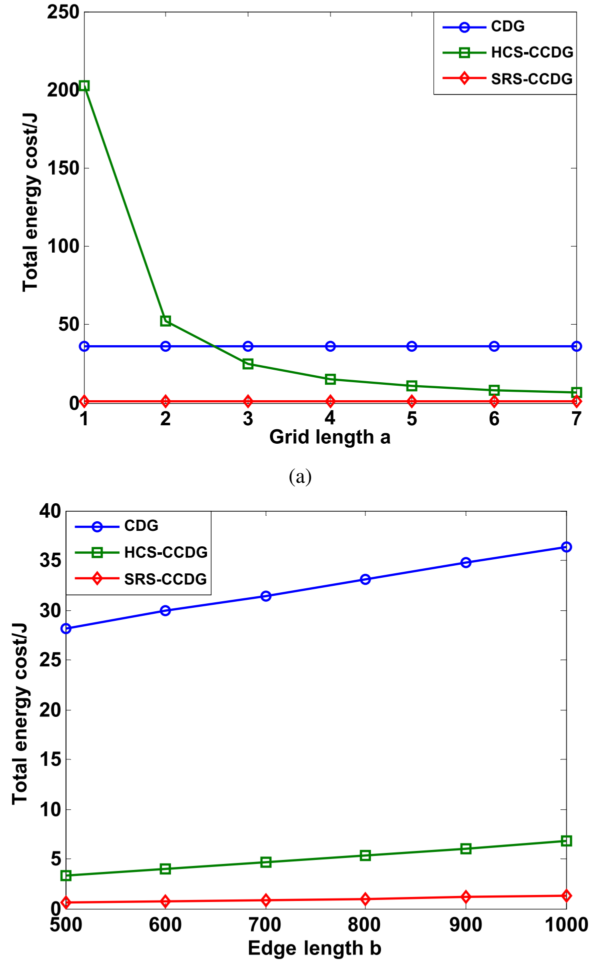
Now, we present the energy cost comparisons among CDG, HCS-CCDG and our SRS-CCDG. Fig. 9(a) shows the result when the number of sensor nodes in the network varies. Since both CDG and HCS-CCDG require all sensor nodes to participate in each CS measurement gathering, their energy cost increases as N rises. In contrast, SRS-CCDG just needs



(a)

(b)

FIGURE 9. Energy cost comparisons among different data gathering schemes under varying node number *N* and measurement number *M*. (a) Varying node number *N*. (b) Varying measurement number *M*.



(a)

(b)

FIGURE 10. Energy cost comparisons among different data gathering schemes under varying network parameters. (a) Varying grid length *a*. (b) Varying edge length *b*.

M (M is the required number of measurements for accurate signal recovery, and $M \ll N$) sensor nodes to participate in sensing and treats each sensor reading as one CS measurement, the energy cost of SRS-CCDG is significantly reduced compared to the two baseline methods and is independent of N. The reason why HCS-CCDG consumes less energy than CDG is that in intra-cluster transmission, each sensor node in the cluster just needs one data transmission, while in CDG, each sensor node in the cluster needs to transmit M projections to the CH. The relationship between the energy cost and the required measurement number M for the three approaches is shown in Fig. 9(b). As noted in the figure, the energy cost of all the three schemes rises when M increases, which is understandable. What is different among them is the growth rate versus M. In CDG, each sensor node needs to conduct M transmissions, and the total number of transmissions is $O(MN)$. Increasing M has no impact on the intra-cluster transmission of HCS-CCDG but does make a difference to the inter-cluster transmission. For SRS-CCDG, the total number of transmissions is $O(M)$. Therefore, the growth rate of the energy cost for SRS-CCDG is the lowest among them, while the energy cost of CDG increases most rapidly.

Next, we present the impact of different network parameters on the energy cost of the three schemes. Fig. 10(a) shows the energy cost comparison result when the grid length a varies. Note that a represents the distance of one-hop transmission. Thus, given the sender node and the destination, a smaller a means a larger number of hops when using relaying as the transmission scheme, resulting in higher energy cost. As shown in the figure, HCS-CCDG consumes more energy when a becomes smaller as data relaying is adopted as both intra-cluster and inter-cluster transmission scheme. However, both the CDG and SRS-CCDG here use direct transmission as intra-cluster and inter-cluster transmission scheme, the energy cost is determined by the communication distance between the sender and the receiver and is independent of the number of hops, thus the energy cost of them remains stable when a changes. The energy cost comparison result under varying edge length b is shown in Fig. 10(b). As the sensing area expands (i.e. b grows), the average communication distance from all sensor nodes to the sink node increases, thus the energy cost of all the three methods rises. What remains unchanged is that SRS-CCDG is always the most energy efficient data gathering method.

C. SYSTEM ROBUSTNESS TO NODE FAILURES

Sensor node failures tend to happen in WSNs due to energy depletion and various malicious attacks. In this part, we investigate the robustness of data gathering of the three approaches when node failures happen. Note that the data set used here is the temperature data trace provided by the Sensorscope project [35], and we select the measurements of 128 sensor nodes as the ground truth. Since both CDG and HCS-CCDG require all sensor nodes to participate in each CS measurement gathering, except for the difference in energy cost, the impact of node failures on signal recovery performance is identical for them. Thus, we integrate CDG and HCS-CCDG as 'CDG/HCS-CCDG' here.

We first present the signal recovery performance comparison when the number of failed nodes varies. Note that the signal recovery performance is measured by the reconstruction error, which is defined as

$$\varepsilon = \frac{||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2}{|||\boldsymbol{x}|_2}, \tag{23}$$

where $\boldsymbol{x}$ is the original signal field, and $\hat{\boldsymbol{x}}$ is the recovered signal field. Note that in this simulation, N is set to 128 and M is 25. As shown in Fig. 11(a), the signal recovery performance of 'CDG/HCS-CCDG' deteriorates when the number of failed nodes increases. In SRS-CCDG, however, each sensor reading is treated as one CS measurement, and if the failed node happens to participate in CS measurement gathering, we can simply discard it and employ a nearby normal sensor node as a substitute. As a result, the signal recovery performance of SRS-CCDG remains relatively stable even though the number of failed nodes increases.

Fig. 11(b) shows the signal recovery performance comparison with fixed number of failed nodes and increasing number of measurements. The number of failed nodes is 8 here, and the number of measurements increases from 25 to 70 with an interval of 5. As we can observe from the figure, in 'CDG/HCS-CCDG', increasing the number of measurements cannot improve the signal recovery performance when the number of failed nodes reaches a certain value. This is because each CS measurement in 'CDG/HCS-CCDG' is a linear combination of all sensor readings, and when the number of failed nodes is fixed, increasing the number of measurements doesn't increase the amount of useful information at the sink. In SRS-CCDG, however, increasing the number of measurements indeed provides more useful information for accurate signal recovery. Thus, the recovery error declines with increased number of measurements.

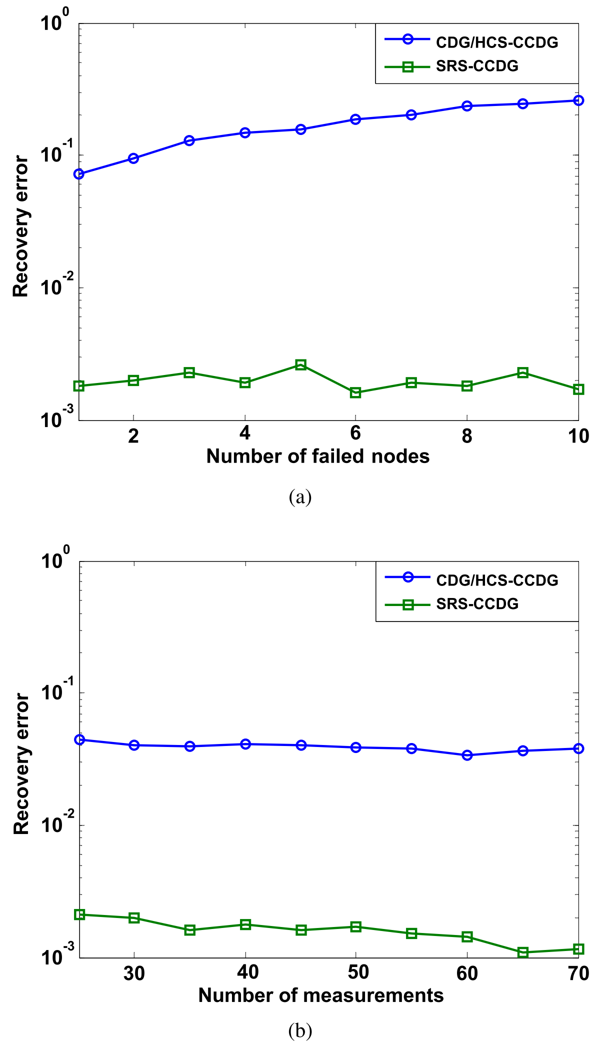


FIGURE 11. **Signal recovery performance comparisons of different data gathering schemes when node failures happen. (a) Fixed number of measurements and varying number of failed nodes. (b) Fixed number of failed nodes and varying number of measurements.**

VII. CONCLUSION

In this paper, we propose a sparsest random sampling scheme for cluster-based compressive data gathering in WSNs. Specifically, sensor nodes are organized into clusters. In each round of data gathering, a random subset of sensor nodes sense the signal field and transmit their measurements to the corresponding CHs. Then, each CH transmits the data gathered within its cluster to the sink node. In order to reduce the number of sensor nodes involved in each CS measurement gathering, we treat each sensor reading as one CS measurement in SRS-CCDG. Both intra-cluster and inter-cluster data transmissions can be realized using two methods, i.e. relaying by other intermediate sensor nodes or direct transmission. Furthermore, we propose analytical models that study the relationship between the size of clusters and the energy cost when using different intra-cluster and inter-cluster transmission schemes, aimed at finding the optimal size of clusters and transmission schemes that could lead to minimum energy cost. Then, we present a centralized clustering algorithm based on the theoretical analysis. Finally, we investigate the robustness of signal recovery performance of SRS-CCDG when node failures happen in the network. Extensive simulations are performed, and results demonstrate that SRS-CCDG can significantly reduce the energy cost of data gathering and improve the system robustness to unavoidable node failures.

REFERENCES

[1] U. Lee, B. Zhou, M. Gerla, E. Magistretti, P. Bellavista, and A. Corradi, "MobEyes: Smart mobs for urban monitoring with a vehicular sensor network," *IEEE Wireless Commun.*, vol. 13, no. 5, pp. 52–57, Oct. 2006.

[2] C. Zhang, R. Lu, X. Lin, P.-H. Ho, and X. Shen, "An efficient identity-based batch verification scheme for vehicular sensor networks," in *Proc. 27th Conf. Comput. Commun. (INFOCOM)*, Apr. 2008, pp. 246–250.

[3] X. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018.

[4] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, and C. Liu, "Exploring human mobility patterns in urban scenarios: A trajectory data perspective," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 142–149, Mar. 2018.

[5] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[8] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.

[9] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. 15th Annu. Int. Conf. Mobile Comput. Netw.*, 2009, pp. 145–156.

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

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

[12] L. Xiang, J. Luo, and A. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw. (SECON)*, Jun. 2011, pp. 46–54.

[13] X. Wu, Y. Xiong, P. Yang, S. Wan, and W. Huang, "Sparsest random scheduling for compressive data gathering in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5867–5877, Oct. 2014.

[14] M. Youssef, A. Youssef, and M. Younis, "Overlapping multihop clustering for wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 12, pp. 1844–1856, Dec. 2009.

[15] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: Recent developments and deployment challenges," *IEEE Netw.*, vol. 20, no. 3, pp. 20–25, May 2006.

[16] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 366–379, Oct./Dec. 2004.

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

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

[19] C. Zhao, W. Zhang, Y. Yang, and S. Yao, "Treelet-based clustered compressive data aggregation for wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4257–4267, Sep. 2015.

[20] X. Li, X. Tao, and G. Mao, "Unbalanced expander based compressive data gathering in clustered wireless sensor networks," *IEEE Access*, vol. 5, pp. 7553–7566, 2017.

[21] D. Wang, L. Lin, and L. Xu, "A study of subdividing hexagon-clustered wsn for power saving: Analysis and simulation," *Ad Hoc Netw.*, vol. 9, no. 7, pp. 1302–1311, 2011.

[22] S. Ji, R. Beyah, and Z. Cai, "Snapshot and continuous data collection in probabilistic wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 3, pp. 626–637, Mar. 2014.

[23] K. Yedavalli and B. Krishnamachari, "Sequence-based localization in wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 81–94, Jan. 2008.

[24] J.-F. Huang, G.-Y. Chang, and G.-H. Chen, "A historical-beacon-aided localization algorithm for mobile sensor networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 6, pp. 1109–1122, Jun. 2015.

[25] S. Lin *et al.*, "ATPC: Adaptive transmission power control for wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 12, no. 1, p. 6, 2016.

[26] H. U. Yildiz, B. Tavli, and H. Yanikomeroglu, "Transmission power control for link-level handshaking in wireless sensor networks," *IEEE Sensors J.*, vol. 16, no. 2, pp. 561–576, Jan. 2016.

[27] C. Karakus, A. C. Gurbuz, and B. Tavli, "Analysis of energy efficiency of compressive sensing in wireless sensor networks," *IEEE Sensors J.*, vol. 13, no. 5, pp. 1999–2008, May 2013.

[28] L. Wu, P. Sun, M. Xiao, Y. Hu, and Z. Wang, "Sparse signal aloha: A compressive sensing-based method for uncoordinated multiple access," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1301–1304, Jun. 2017.

[29] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[30] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

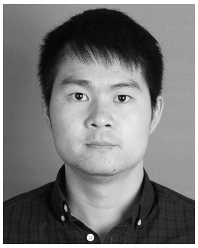
[31] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[32] J.-H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 4, pp. 609–619, Aug. 2004.

[33] L. Wu, K. Yu, D. Cao, Y. Hu, and Z. Wang, "Efficient sparse signal transmission over a lossy link using compressive sensing," *Sensors*, vol. 15, no. 8, pp. 19880–19911, 2015.

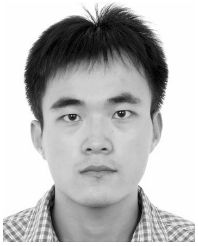
[34] P. K. Agarwal and C. M. Procopiuc, "Exact and approximation algorithms for clustering," *Algorithmica*, vol. 33, no. 2, pp. 201–226, 2002.

[35] F. Ingelrest, G. Barrenetxea, G. Schaefer, M. Vetterli, O. Couach, and M. Parlange, "SensorScope: Application-specific sensor network for environmental monitoring," *ACM Trans. Sensor Netw.*, vol. 6, no. 2, p. 17, 2010.



PENG SUN received the B.E. degree in automation from Tianjin University, China, in 2015.

He is currently pursuing the Ph.D. degree with the College of Control Science and Engineering, Zhejiang University, China. His research interests include wireless sensor networks, array signal processing, and compressive sensing.

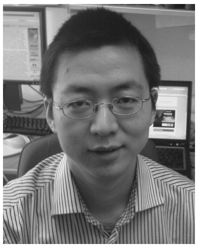


LIANTAO WU received the B.E. degree in automation from Shandong University, China, in 2012, and the Ph.D. degree in control science and engineering from Zhejiang University, China, in 2017.

He is currently with Shanghai Huawei Technologies Corporate Campus. His research interests include sensor networks, wireless communications, and compressive sensing.

ZHIBO WANG (M'10) received the B.E. degree in automation from Zhejiang University, China, in 2007, and the Ph.D. degree in electrical engineering and computer science from the University of Tennessee, Knoxville, in 2014.

He is currently an Associate Professor with the School of Cyber Science and Engineering, Wuhan University, China. His current research interests include mobile crowd sensing systems, cyber-physical systems, recommender systems, and privacy protection. He is a member of ACM.



MING XIAO (S'02–M'07–SM'12) received the bachelor's and master's degrees in engineering from the University of Electronic Science and Technology of China, Chengdu, in 1997 and 2002, respectively, and the Ph.D. degree from the Chalmers University of Technology, Sweden, in 2007.

From 1997 to 1999, he was a Network and Software Engineer with China Telecom. From 2000 to 2002, he also held a position with the Sichuan communications administration. Since 2007, he has been with the Department of Information Science and Engineering, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Sweden, where he is currently an Associate Professor. Since 2012, he has been an Associate Editor for the IEEE Transactions on Communications, has been an Senior Editor for the IEEE Communications Letters since 2015, and the IEEE Wireless Communications Letters from 2012 to 2016. He has been the Lead Guest Editor for the IEEE Journal on Selected Areas in Communications Special issue on Millimeter Wave Communications for future mobile networks since 2017.

ZHI WANG (M'99) received the B.E. degree from Shenyang Jianzhu University in 1991, the M.S. degree from Southeast University in 1997, and the Ph.D. degree from the Shenyang Institute of Automation, Chinese Academy of Sciences, in 2000.

From 2000 to 2002, he was a Post-Doctoral Fellow where he conducted research with the Institut National Polytechique de Lorraine, France, and Zhejiang University, China, respectively. Since 2002, he has been an Associate Professor with the College of Control Science and Engineering, Zhejiang University, and became a Doctoral Advisor in 2006. From 2010 to 2011, he conducted research as an Advanced Scholar at UW–Madison, USA. His main research focus is on the cyber physical system, including sensor network, collaborative signal and information processing, real-time theory and system, integrated schedule and control, secure and privacy protection, and related platforms, such as networked acoustic arrays, networked visual and optical arrays, and industrial communication networks. He is a member of the ACM. He serves as a general co-chair and a TPC member for a number of CPS and sensor networks-related international conferences, and the Committee Member for the China Computer Federation Sensor Network Technical Committee and the China National Technical Committee of Sensor Network Standardization.