# Compressive Sensing meets Unreliable Link: Sparsest Random Scheduling for Compressive Data Gathering in Lossy WSNs


Xuangou Wu
Anhui University of Technology
Maanshan, China
University of Science and Technology of China
Hefei, China
wxgou@mail.ustc.edu.cn

Panlong Yang
Institute of Communication Engineering, PLAUST
Nanjing, China
panlongyang@gmail.com

Taeho Jung
Illinois Institute of Technology
Chicago,USA
tjung@hawk.iit.edu

Yan Xiong
University of Science and Technology of China
Hefei, China
yxiong@ustc.edu.cn

Xiao Zheng
Anhui University of Technology
Maanshan,China
zhx@ahut.edu.cn



## ABSTRACT

Compressive Sensing (CS) has been recognized as a promising technique to reduce and balance the transmission cost in wireless sensor networks (WSNs). Existing efforts mainly focus on applying CS to reliable WSNs, namely, each wireless link is 100% reliable. However, our experimental results show that traditional compressive data gathering (CDG) could result in arbitrarily bad recovery performance, when the wireless links are lossy. In this paper, we study the impact of packet loss on compressive data gathering and ways to improve its robustness using sparsest random scheduling (SRS). The key idea of our scheme is to treat each sampling value as one CS measurement, which helps us to reduce the impact of packet loss on the recovery accuracy. Our scheme also outperforms the tradition CDG in reliable WSNs in that our scheme has significantly lowered transmission cost. To achieve this, we present a sparsest measurement matrix where each row has only one nonzero element. More importantly, we propose a representation basis to sparsify the gathering data, and prove that our measurement matrix satisfies the restricted isometric property (RIP) with high probability. Extensive experimental results show our scheme can recover the data accurately with packet loss ratio up to 15%, while traditional CDG can hardly recover the data under similar or even better conditions.




*MobiHoc'14,* August 11–14, 2014, Philadelphia, PA, USA.



http://dx.doi.org/10.1145/2632951.2632969 .



## Categories and Subject Descriptors:

C.2.1 [Computer-Communication Networks]: Network Operations, Network monitoring

## Keywords

Data gathering, compressive sensing, sparsest random scheduling, restricted isometry property.


## 1. INTRODUCTION

Wireless sensor networks (WSNs) have been widely used for collectively monitoring and disseminating information about various phenomena of interest [7, 31]. For example, ExScal [1] is an intrusion detection network with more than 1000 sensor nodes. And typically, GreenOrbs [11] and CitySee [15] systems have been built for continuously collecting environmental data including temperature, humidity, illumination, and carbon dioxide etc. Leveraging the spatial-temporal properties in sensory data from real deployments, many Compressive Sensing (CS) based data gathering techniques have been proposed to reduce and balance the in-network data transmission cost (e.g. [13, 16, 17, 25, 26, 29]). As far as we are concerned, all existing works on compressive data gathering (CDG) focus on reliable WSNs, without considering the impact of packet loss on the recovery performance. Unfortunately, our experimental results show that the recovery performance of traditional CDG could be seriously degraded by packet loss, e.g., the recovery accuracy could be arbitrarily bad even with 2% packet loss rate. This problem is especially pronounced in many large scale sensor networks whose packet loss ratio could reach 20% [14]. The prevalence of unreliable links in real-world WSNs [11] has posed a fundamental challenge for maintaining the recovery accuracy under unsatisfactory and lossy conditions.

In general, there are two reasons preventing the traditional CDG from working well in lossy network. (1) There are too many sensors involved for gathering single measurement, when adopting tree-based routing, one packet loss may cause severe decrease on the quality of that measurement; (2) CS theory is originally developed for minimizing the number of measurements, rather than the cost of each measurement, unfortunately, so far we can not find any representation basis for making the measurement matrix sufficiently sparse. To address such deficiencies, we develop a novel CDG scheme that achieves good resilience to packet loss. In our scheme, unlike the traditional CDG, each row of the measurement matrix contains only one nonzero element and each sampling value is considered as one CS measurement. In spite of some existing research in this direction [26][27], our work differs from theirs in that: (1) in sparse random projections [26], each row of measurement matrix requires $O(\log N)$ nonzero elements where $N$ is the network size; (2) single sensor temporal random scheduling data gathering [27] suffers from a lack of a decent theoretical analysis, and this scheme can not work in distributed WSNs. In contrast, we find a suitable representation basis based on Gaussian joint distribution model. More importantly, we are the first to prove that the proposed measurement matrix and representation basis satisfy restricted isometric property (RIP) with high probability, this guarantees that the gathered data can be accurately recovered at the sink. The contributions of this paper are three folds.

(1) We are the first to investigate the impact of packet loss on the performance of traditional CDG in terms of recovery accuracy. We show that the recovery accuracy could be very bad even with 2% packet loss ratio. This motivates us to revisit and enhance existing CDG for achieving better resilience to packet loss.

(2) We propose a novel sparsest random scheduling based CDG scheme (SRS-DG). We present a detailed characterization of its performance through both analytical and numerical results. Our scheme can reduce the impact of packet loss on the recovery accuracy. We also prove that the proposed measurement matrix and representation basis satisfy RIP with high probability.

(3) We empirically demonstrate the effectiveness of our scheme on real data set from CitySee [15]. In particular, we show that the recovery error of our scheme is only 0.1 with packet loss ratio up to 15%. In contrast, traditional CDG can hardly recover the data under similar or even better conditions.

The rest of this paper is organized as follows. In section 2 presents the preliminaries of CS. The motivation and challenges are given in section 3. The detailed design of SRS-DG is presented in Section 4. In section 5, we provide the transmission cost analysis of SRS-DG and its algorithm implementation. Section 6 reports our experimental results. We present a literature review of existing work in section 7 and make a conclusion in Section 8.

## 2. PRELIMINARIES

CS is a new compression and sampling paradigm compared with traditional Shannon's sampling theorem [2, 4, 10]. CS theory asserts that a relatively small number linear combination of a compressible or sparse signal can contain most of its salient information. Assuming that $\mathbf{s} \in \mathbb{R}^N$ is a $k$-sparse signal, which is only $k$ nonzero components or $(N-k)$ smallest components can be ignored. Thus, the information can be extracted from $\mathbf{s}$ by $\mathbf{y} = \Phi\mathbf{s}$, where $\Phi$ is an $M \times N$ measurement matrix, $\mathbf{y} \in \mathbb{R}^M$ is measurement vector and $M \ll N$. To recover the signal $\mathbf{s}$, two problems need to be answered: (1) How to design $\Phi$ such that the salient information can be extracted from any $k$-sparse signal? (2) How to design recovery algorithm to reconstruct $\mathbf{s}$ from $\mathbf{y}$? To answer the first problem, $\Phi$ should satisfy the restricted isometric property (RIP) [6]:

Definition 1 ([6]). *A matrix $\Phi$ satisfies the restricted isometry property (RIP) of order $k$ if there exists a $\delta_k \in (0,1)$ such that*

$$(1-\delta_k)\|\mathbf{s}\|_2^2 \le \|\Phi\mathbf{s}\|_2^2 \le (1+\delta_k)\|\mathbf{s}\|_2^2 \qquad (1)$$

*for all $k$-sparse vectors $\mathbf{s} \in \mathbb{R}^N$.*

Candès, Romberg, and Tao [5] and Donoho [10] have shown many random matrices that satisfy the RIP such as Gaussian identity distribution matrix, $\pm 1$ Bernoulli matrix and so on.

To answer the second problem, the signal $\mathbf{s}$ can be recovered via $\ell_1$ optimization as

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \| \mathbf{s} \|_1 \quad s.t. \ \mathbf{y} = \Phi\mathbf{s} \qquad (2)$$

If $\Phi$ satisfies RIP and $M \ge O(k \cdot log(N/k))$ , then $\mathbf{s}$ can be recovered successfully with high probability. If the measurement vector $\mathbf{y}$ contains noise, then the signal $\mathbf{s}$ can be recovered via

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \| \mathbf{s} \|_1 \quad s.t. \ \|\Phi\mathbf{s} - \mathbf{y}\|_2^2 \le \epsilon \qquad (3)$$

where $\epsilon$ bounds the noise. There already exist many efficient algorithms to solve the above problems such as basis pursuit [6], orthogonal matching pursuit (OMP) algorithm [24], CoSaMP [20] and so on.

However, the real sensory signals are almost compressible signals instead of sparse signal. Compressible signal can usually be transformed into sparse signal via representation basis transformation. For example, a smooth signal $\mathbf{x} \in \mathbb{R}^N$ can usually be transformed into a sparse signal $\mathbf{s}$ under discrete cosine transformation (DCT) basis or discrete wavelet transformation (DWT) basis. The measurement vector $\mathbf{y}$ can be expressed as

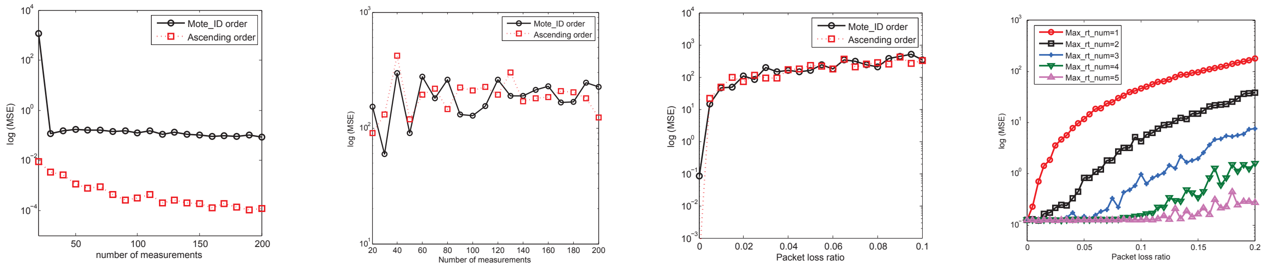
$$\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\mathbf{s} \qquad (4)$$

where $\Psi$ is a $N \times N$ representation basis. If $\Phi\Psi$ satisfies RIP, the sparse signal $\mathbf{s}$ can be recovered accurately with high probability. Then $\mathbf{x}$ can be recovered via $\mathbf{x} = \Psi\mathbf{s}$.

## 3. MOTIVATION AND CHALLENGES

In this section, we will analyze the recovery performance of traditional CDG in lossy WSNs, and demonstrate that the recovery accuracy is seriously impacted by data packet loss. Based on the experimental results from real deployments, we further investigate the reasons why traditional CDG is not robust to packet loss. Correspondingly, we present the idea of SRS-DG that guarantees good recovery performance even under severe packet loss.

**Observation**: Our experiments were performed on a real-world data set from CitySee system [19]. CitySee system was deployed in an urban area of Wuxi City, China, which contained thousands of wireless sensor nodes for environmental monitoring. In CitySee system, each sensor sampled once every 10 minutes and sent its sampling value to the sink. We evaluate sensory data recovery performance of traditional CDG under different data packet loss ratios and different number of CS measurements. In our experiment, Gaussian

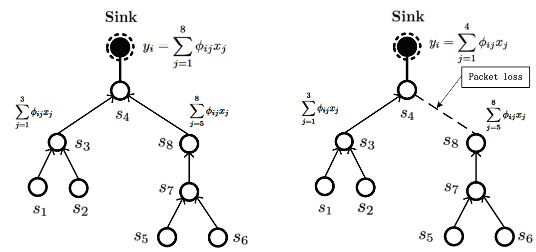



(a) $pl\_ratio = 0$ (b) $pl\_ratio = 0.05$ (c) $M = 120$ (d) $M = 120$

**Figure 1: Recovery error comparisons. (a) No data packet loss. (b) Different number of measurements without retransmission. (c) Different data packet loss ratio without retransmission. (d) Different data packet loss ratio with retransmission.**

random matrix and discrete cosine transformation (DCT) are considered as measurement matrix and sparse representation basis, which are denoted by $\Phi_G$ and $\Psi_{dct}$ respectively. The experimental data set is 256 temperature sampling values in the same time interval from 256 sensors. First, we sort the sensory data by Mote_ID order. For better justification, we also evaluate their recovery performance under different orders. Notice that in practice, it is impossible to reorder the sensory data according to their value which is unknown at the beginning. The sparse level under Mote_ID order is around 25, and the sparse level under ascending order is around 5.

Next, we evaluate the recovery performance of the traditional CDG under three different network configurations. (1) Sensor network is reliable without data packet loss. (2) Sensor network is unreliable without allowing lost packet retransmission. (3) Sensor network is unreliable with allowing lost packet retransmission. We build a routing tree containing 256 sensor nodes. Each packet is dropped at random with uniform probability (packet loss ratio) on each link. To simplify the expression, the packet loss ratio and the number of measurements are denoted by *pl_ratio* and $M$ respectively. In our experiment, we use mean square error (MSE) as the recovery error metric. Fig 1 (a) shows the recovery error with no data packet loss. Two types orders sensory data can be recovered accurately, the recovery error of Mote_ID order is decreased within 0.1 (MSE) with at least 30 measurements. And the recovery error of ascending order is even less than 0.01 (MSE) with at least 20 measurements. Fig 1 (b) shows the recovery error under $pl\_ratio = 5\%$ without lost packet retransmission. The recovery error is higher than 60 in the majority of cases. It demonstrates that the sensory data can hardly be recovered under both ascending order and Mote_ID order. Moreover, even we increase the number of measurements, the recovery performance is still poor. Fig 1 (c) displays the recovery error under different packet loss ratios without retransmission. The result shows that the recovery performance could be very bad even with low packet loss ratio ($pl\_ratio < 2\%$). We next evaluate the recovery performance with retransmissions. Fig 1 (d) depicts the mean recovery error under different *maximum number of retransmissions*, which is denoted by $Max\_rt\_num$. Unfortunately, limited number of retransmissions does not help much in improving the recovery accuracy. Moreover, retransmission may cause extra communication overhead which is clearly undesirable. This issue is especially pronounced in many large scale sensor networks, the average packet loss ratio could still reach 20% with $Max\_rt\_num = 30$ [14].




(a) Reliable Network (b) Lossy Network

**Figure 2: CS based data gathering in tree routing network.**

**Insight**: We identify at least two reasons why traditional CDG can not achieve satisfactory recovery performance under lossy WSNs. The first reason is that there are too many sensors involved for gathering single measurement, therefore one packet loss may cause severe drop on the quality of that measurement; The second reason is that CS theory is originally developed for minimizing the number of measurements, rather than the cost of gathering each measurement, unfortunately, so far we can not find any presentation basis for making the measurement matrix sufficient sparse. Although CS theory asserts that it has a strong robustness to noise and data loss [5], distributed sensor networks has many unique features which make traditional CS theory not capable of handling the packet loss. For example, Fig 2 displays the data packet loss under tree based routing. Let $s_j$ denote the $j^{th}$ sensor, the corresponding sampling value and projection element are $x_j$ and $\phi_{ij}$ respectively ($j = 1, 2, \cdots, 8$). Fig 2 (b) shows that the data packet of $s_8$ is lost when it is sent to $s_4$ during the $i^{th}$ measurement gathering. Because the $i^{th}$ measurement is calculated as $y_i = \sum_{i=1}^{8} \phi_{ij} x_i$, all involved sensory data including $x_5$, $x_6$, $x_7$, and $x_8$ would be lost. In this case, one packet loss could cause severe drop on the quality of gathered measurements. Because this case could happen to every measurement, simply increasing the number of measurements can not improve the recovery performance.

**Challenges**: We now investigate how to reduce (or possibly avoid) the consequences of data packet loss on the quality of gathered measurements. Intuitively, if each measurement is represented by only one sampling value, then the received measurements only contains sampling noise. This is because the sink would either gather the entire measurement successfully or lose the entire measurement. Traditional CS theory focuses on minimizing the number of measurements, rather than the number of participation nodes for each measurement. It requires the measurement matrix to satisfy any orthogonal representation basis of compression signal. Although Donoho *et al.* proposed that the partial fourier coefficient can also recover original signal ([5, 10]), it is not suitable for distributed WSNs.

In this paper, we aim to let each sampling value represent one CS measurement. If this goal can be realized, we can reduce both the data transmission cost, and, more importantly, the consequences of packet loss on the data recovery. Then it is possible to develop a novel CDG which can be used even in lossy sensor networks. However, achieving this goal is not trivial, there are two primary challenges: The first challenge is to design appropriate measurement matrix and representation basis, this is critical sparsifying the sensory data and letting each sensory value represent one measurement; The second challenge is to ensure that the sensory data can be accurately recovered under the previous design. To overcome these challenges, we develop a novel CDG scheme SRS-DG in the rest of this paper.

# 4. SRS-DG DESIGN

In this section, we provide the detailed design and analysis of each sampling values as one CS measurement, namely, the measurement matrix and representation basis design and analyze in SRS-DG. To let each sampling value represent one CS measurement, each row of the measurement matrix should contain only one nonzero element. In order to facilitate the expression, we call this type of measurement matrix as *sparsest measurement matrix*, and the data gathering process as *sparsest random scheduling* for data gathering (SRS-DG). In Section 4.1, we present our sparsest measurement matrix design and its rationale. Accordingly, in Section 4.2, we design a suitable representation basis. We verify that our representation basis can sparsify the spatial correlation sensory data in Section 4.3. We also prove that the sparsest measurement matrix and representation basis satisfy RIP with high probability in Section 4.4, this ensures high recovery accuracy.

Assume that there are $N$ sampling values that need to be gathered, which are denoted by $\mathbf{x} = [x_1, x_2, \cdots, x_N]^T$. During the data gathering, the measurement matrix is denoted by $\Phi = [\phi_1, \phi_2, \cdots, \phi_N]^T$, the $i^{th}$ CS measurement is calculated as $y_i = \sum_{j=1}^{N} \phi_{ij} x_j$. If $\phi_{ij}$ is nonzero, the $j^{th}$ sensor node requires to participate in the $i^{th}$ CS measurement gathering.

## 4.1 Measurement matrix design and analysis

In this subsection, we present the design of sparsest measurement matrix, where each sampling value represents one CS measurement. We also present the design rationale behind our sparsest measurement matrix.

**Measurement matrix design:** To balance the resource consumption in WSNs, it is reasonable to define our sparsest measurement matrix $\Phi_e$ as

$$\Phi_e(i,j) = \begin{cases} 1 & j = r_i \\ 0 & otherwise \end{cases} \tag{5}$$

where $i = 1, 2, \cdots M$, $j = 1, 2, \cdots N$, $r_i$ represent the independent and identically distributed random indices, and $r_i < r_{i+1}$, $r_i \in [1, N]$. Based on the definition of $\Phi_e$, each row of $\Phi_e$ has only one nonzero element. Then, one round of data gathering can be expressed as

$$\mathbf{x_r} = \begin{bmatrix} x_{r_1} \\ x_{r_2} \\ \vdots \\ x_{r_M} \end{bmatrix} = \begin{bmatrix} \phi_{e_1} \\ \phi_{e_2} \\ \vdots \\ \phi_{e_M} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \tag{6}$$

where $\mathbf{x_r}$ is CS measurement vector, $\phi_{e_i}$ represents the $i^{th}$ row of $\Phi_e$. If $\mathbf{x_r}$ is reconstructed by $\mathbf{x}$ based on CS theory, then the decoder can be expressed as

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_1 \ \ s.t. \ \ \|\Phi_e \Psi \mathbf{s} - \mathbf{x_r}\|_2^2 \le \epsilon \tag{7}$$

where $\mathbf{s} = \Psi^{-1}\mathbf{x}$, $\Psi$ is orthogonal basis of $\mathbf{x}$, $\epsilon$ bounds the noise in $\mathbf{x_r}$, $\hat{\mathbf{s}}$ is the recovery signals of $\mathbf{s}$.

**Feasibility analysis:** The design of sparsest measurement matrix ensures that each sampling value represents one CS measurement, and reduces the impact of packet loss on the recovery performance. The next challenge is to design an appropriate representation basis, by taking into account both the data transmission cost and data recovery quality. In particular, an appropriate representation basis, $\Psi$, should satisfy the following two conditions: (1) $\Psi$ can sparsify the gathering sensory data; and (2) $\Phi_e\Psi$ satisfies RIP with high probability.

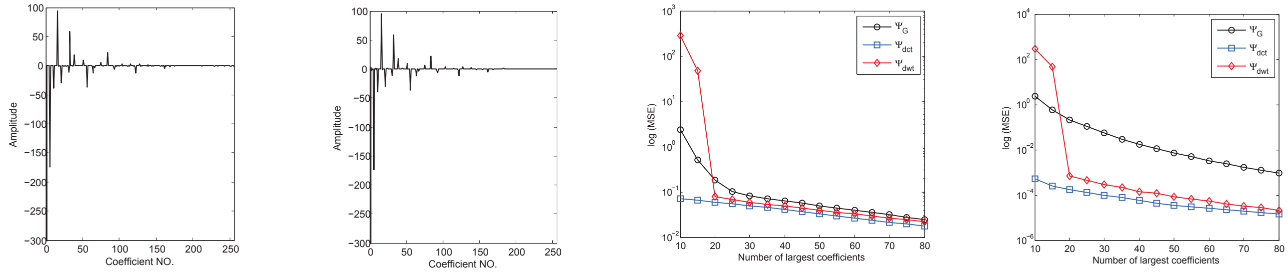
We next demonstrate, from the information extraction aspect, why each row of our measurement matrix contains only one nonzero element. If sensory data $\mathbf{x}$ is sparse under representation basis $\Psi$, it can be recovered from a small number of CS measurements (i.e., $\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\mathbf{s}$). The decoding process is to recover the sparse signal $\mathbf{s}$ instead of directly recovering the sensory data $\mathbf{x}$. The sparse signal $\mathbf{s}$ can be recovered because each component of $\mathbf{y}$ contains a part of information of $\mathbf{s}$, namely, each component of $\mathbf{y}$ is a linear combination of $\mathbf{s}$. However, each component of $\mathbf{x}$ is also a linear combination of $\mathbf{s}$. Each component of $\mathbf{x}$ and $\mathbf{y}$ contains the information of the sparse signal $\mathbf{s}$. So, if we design or select a suitable representation basis and make it satisfy RIP with sparsest measurement matrix, the sparse signal $\mathbf{s}$ can be recovered from a part of $\mathbf{x}$.

In what follows, we give the detailed design of our representation basis and illustrate why it satisfies the above two conditions.

## 4.2 Representation basis design

In [5, 10], Donoho *et al.* have shown that many random matrices satisfy the RIP with any orthonormal representation basis. In addition, they show that if one can find a sparsity representation basis which is similar to a certain random matrix, it is possible to satisfy RIP. Our representation basis is developed based on the correlation of sensory data.

As real-world environmental data usually exhibit strong spatial correlation, Gaussian joint distribution is an effective and reasonable model for further theoretical analysis [9]. Without loss of generality, we define Gaussian kernel function, $\mathcal{K}(x_i, x_j)$, as $\mathcal{K}(x_i, x_j) = exp\{\frac{-d_{ij}^2}{2\sigma^2}\}$, where $\mathcal{K}(x_i, x_j)$ represents the correlation between $x_i$ and $x_j$, $d_{ij}$ is the distance between the $i^{th}$ sensor and the $j^{th}$ sensor. How much effect between $x_i$ and $x_j$ depends on the parameter $\sigma$, $\mathcal{K}(x_i, x_i) = 1$. The parameter $\sigma$ can be estimated from




(a) (b) (c) (d)

**Figure 3: The transformed signal and recovery error comparisons. (a) Transformed coefficients with Mote_ID order. (b) Transformed coefficients with ascending order. (c) Recovery error comparison with Mote_ID order. (d) Recovery error comparison with Ascending order.**

training sensory data by maximum likelihood or Bayesian framework [22]. Based on the above assumptions, the correlation matrix $G$ with $N$ sensor points can be expressed as

$$G = \begin{bmatrix} e^{\frac{-d_{11}^2}{2\sigma^2}} & e^{\frac{-d_{12}^2}{2\sigma^2}} & \cdots & e^{\frac{-d_{1N}^2}{2\sigma^2}} \\ e^{\frac{-d_{21}^2}{2\sigma^2}} & e^{\frac{-d_{22}^2}{2\sigma^2}} & \cdots & e^{\frac{-d_{2N}^2}{2\sigma^2}} \\ \vdots & \vdots & \vdots & \vdots \\ e^{\frac{-d_{N1}^2}{2\sigma^2}} & e^{\frac{-d_{N2}^2}{2\sigma^2}} & \cdots & e^{\frac{-d_{NN}^2}{2\sigma^2}} \end{bmatrix} \quad (8)$$

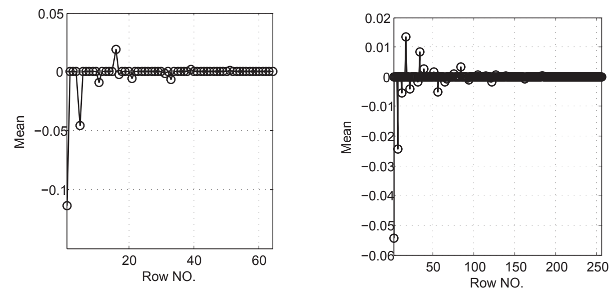
We map $N$ sensors into 2-dimensional array based on real locations and the distance between neighboring sensors is unitary. Then, $G$ is a Toeplitz matrix which can be diagonalized, namely, $G$ can be expressed as $G = \Psi_G \Lambda \Psi_G^{-1}$, where $\Psi_G$ is orthonormal eigenvector basis, $\Lambda$ is the diagonal matrix whose diagonal entries are the corresponding eigenvalues of $G$. We use $\Psi_G$ as an orthonormal representation basis, then $\mathbf{x}$ can be represented by $\mathbf{x} = \Psi_G \mathbf{s}$, where $\mathbf{s}$ is transformed signal ($\mathbf{s} = \Psi_G^{-1}\mathbf{x}$). If $\mathbf{s}$ is sparse and $\Phi_e \Psi_G$ satisfy RIP, $\Psi_G$ would be considered as a representation basis of sparsest random projections.

## 4.3 Does $\Psi_G$ sparsify sensory data?

In this section, we experimentally show that $\Psi_G$ indeed sparsifyies the spatial correlation signal. We also give statistical analysis of $\Psi_G$ to explain why it can sparsify the spatial correlated sensory data. To evaluate the sparse performance of $\Psi_G$, we compare it with DCT basis and DWT basis, which are denoted by $\Psi_{dct}$ and $\Psi_{dwt}$ respectively.

**Experimental verification:** We use the same data set as in Section 3. We still examine two orders: Mote_ID order and ascending order. Ascending order exhibits stronger spatial correlation, and Mote_ID order has weaker spatial correlation.

Fig 3 (a) and (b) show the transformed coefficients under $\Psi_G$ corresponding to two orders. The transformed coefficients under $\Psi_G$ are mainly concentrated in a few components, both orders have a good sparse performance. To recover a compressible signal based on CS theory, the number of CS measurement is proportional to the sparsity of the sensory data. For the same spatial correlated signal, different representation basis obtain different sparse level signal. Fig 3 (c) and (d) show the recovery quality under different numbers of largest transformed coefficients corresponding to different orders. The number of largest transformed coefficients can be considered as sparse level of transformed signal. Fig 3 (c) shows that $\Psi_G$ can sparsify our sensory data, its performance is very close to the one of $\Psi_{dct}$ and




(a) N=64 (b) N=256

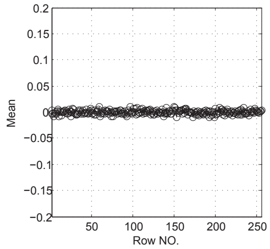
**Figure 4: The mean of every row of $\Psi_G^{-1}$ with $\sigma^2 = 1$ and different $N$.**

4-layer 'haar' $\Psi_{dwt}$. Fig 3 (d) illustrates that the recovery performance of $\Psi_G$ is not as good as $\Psi_{dct}$ and $\Psi_{dwt}$, this is because ascending order is a theoretical ideal case.

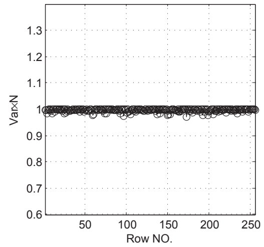
**Statistical analysis:** We also illustrate that $\Psi_G$ can sparsify the spatial correlation data through analyzing the mean and variance of its row elements. Fig 4 shows the mean of every row of $\Psi_G^{-1}$ with $\sigma^2 = 1$ and $N = 64$, 256, there are only a few rows whose mean values are nonzero. This also explains that why the transformed signal $\mathbf{s}$, $\mathbf{s} = \Psi_G^{-1}\mathbf{x}$, is sparse when the spatial correlation signal $\mathbf{x}$ changes smoothly. For the real sensory data, we can learn the optimal value of $\sigma$ based on the historic sensory data. Based on the above analysis, we show that $\Psi_G$ can efficiently sparsify the spatial correlated signal.

## 4.4 Does $\Phi_e \Psi_G$ Obey RIP?

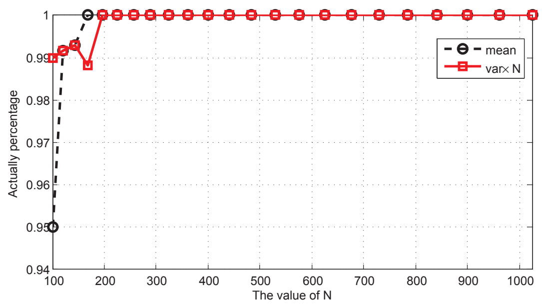
In this subsection, we prove that $\Phi_e \Psi_G$ satisfy RIP with high probability, this ensures high recovery accuracy. We first present the statistic properties of $\Psi_G$. The mean and variance of each row in $\Psi_G$ with $N = 256$ are shown in Fig 5 (a) and (b) respectively. Both the mean and variance of each row are very stable. Fig 5 (c) illustrates the actual percentage of the mean of each row in the range of $(-0.02, 0.02)$ and $N$ times variance of each row in the range of $(-0.05, 0.05)$. Fig 5 (c) illustrates that both the mean and variance of each row exhibit similar trends with the increasing of $N$. Based on the law of large numbers, each row of $\Psi_G$ can be considered as a random sequence generated by a random variable. $\Psi_G$ is generated by $N$ random variables denoted by $\xi_1, \xi_2, \cdots \xi_N$. These random variables have the same numerical characteristics, namely, $\mathbb{E}(\xi_i) = 0$, $\mathrm{Var}(\xi_i) = \mathbb{E}(\xi_i^2) = 1/N$, $(i = 1, 2, \cdots, N)$.




(a) Mean ($N = 256$).




(b) variance ($N = 256$).




(c) Actual percentage of mean and variance.

**Figure 5: The mean and variance of every row of $\Psi_G$ with $\sigma^2 = 1$ and different $N$ values.**

Let $\Theta = \Phi_e \Psi_G$, since the nonzero elements of each row of $\Phi_e$ are independent of each other, we can assume that each row of $\Theta$ is chosen independently at random from $\Psi_G$. Meanwhile, $\Theta$ is generated by ($i.i.d$) random variables $\xi_{r_1}, \xi_{r_2}, \cdots, \xi_{r_M}$. In the following, we give the definition of *sub-Gaussian* and the related corollary.

Definition 2 (**sub-Gaussian** [3]). *A random variable $\xi$ is called sub-Gaussian if there exists a constant $c > 0$ such that $\forall \lambda \in \mathbb{R}$*

$$\mathbb{E}\left(e^{\lambda\xi}\right) \leq e^{c^2\lambda^2/2} \tag{9}$$

*We say $\xi \sim Sub(c^2)$ iff $\xi$ satisfies the above inequality.*

Corollary 1. *If the $i^{th}$ row of $\Theta$ is considered as a sequence generated by random variables $\xi_{r_i}$ $(i = 1, 2, ..., M)$, then $\xi_{r_i} \sim Sub(2)$.*

Proof. Because $\mathbb{E}(\xi_{r_i}) = 0$ and $\mathrm{Var}(\xi_{r_i}) = \mathbb{E}(\xi_{r_i}^2) = 1/N$, then for all $\lambda \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}\left(e^{\lambda\xi_{r_i}}\right) &= \mathbb{E}\left(\sum_{n=0}^{\infty}\frac{\lambda^n\xi_{r_i}^n}{n!}\right) = 1 + \sum_{n=2}^{\infty}\frac{\lambda^n\mathbb{E}\left(\xi_{r_i}^n\right)}{n!} \\ &\leq 1 + \sum_{n=2}^{\infty}\frac{\lambda^n}{n!} \leq e^{|\lambda|} - |\lambda| \leq e^{2\lambda^2/2}\end{aligned} \tag{10}$$

The proof of $e^{|\lambda|} - |\lambda| \leq e^{2\lambda^2/2}$ is given in Appendix. So, $\xi_{r_i} \sim Sub(2)$, $(i = 1, 2, \cdots, N)$. □

Theorem 1 ([8]). *Suppose $\xi_{\mathbf{r}} = [\xi_{r_1}, \xi_{r_2}, \cdots, \xi_{r_M}]^T$, where each $\xi_{r_i}$ is i.i.d. $\xi_{r_i} \sim Sub(c^2)$ and $\mathbb{E}\left(\xi_{r_i}^2\right) = \sigma^2$. Then*

$$\mathbb{E}\left(||\xi_{r_i}||_2^2\right) = M\sigma^2 \tag{11}$$

*Moreover,$\forall\alpha \in (0, 1), \forall\beta \in [c^2/\sigma^2, \beta_{max}]$, there exists a constant $c^*$ such that $\mathbb{P}\left(||\xi_{\mathbf{r}}||_2^2 \leq \alpha M\sigma^2\right) \leq e^{(-M(1-\alpha)^2/c^*)}$ and $\mathbb{P}\left(||\xi_{\mathbf{r}}||_2^2 \geq \beta M\sigma^2\right) \leq e^{(-M(\beta-1)^2/c^*)}$.*

In the following theorem, we prove that $\Theta$ satisfies RIP with probability tending to 1.

Theorem 2. *Fix $\delta \in (0, 1)$ and each row of $\Theta$ satisfies $Sub(2)$, if $M = O(k\log(N/k))$, then with high probability, $\Theta$ satisfies $(1-\delta) \leq \frac{||\Theta\mathbf{v}||_2^2}{||\mathbf{v}||_2^2} \leq (1+\delta)$ for all $N$-dimensional $k$-sparse signal $\mathbf{v}$.*

Proof. Suppose that the $i^{th}$ row of $\Theta$ is generated by $\xi_{r_i}$. Because each row of $\Theta$ is randomly selected based on scheduling sensor, $\xi_{r_1}, \xi_{r_2}, \cdots, \xi_{r_M}$ is *i.i.d.* with $\xi_{r_i} \sim Sub(2)$. To simplify the proof, we normalize $\Theta$ and $\Theta = \sqrt{N/M}[\theta_1, \theta_2, \cdots, \theta_M]^T$, $\mathbb{E}(\theta_{ij}) = 0$ and $\mathrm{Var}(\theta_{ij}) = \mathbb{E}(\theta_{ij}^2) = 1/N$. We have

$$\begin{aligned}&\mathbb{E}(< \sqrt{\frac{N}{M}}\theta_i, \mathbf{v} >) = \mathbb{E}(\sqrt{\frac{N}{M}}\sum_{j=1}^{N}\theta_{ij}v_j) \\ &= \sqrt{\frac{N}{M}}\sum_{j=1}^{N}\mathbb{E}(\theta_{ij})v_j = 0\end{aligned} \tag{12}$$

$$\begin{aligned}&\mathrm{Var}(< \sqrt{\frac{N}{M}}\theta_i, \mathbf{v} >) = \mathrm{Var}(\sqrt{\frac{N}{M}}\sum_{j=1}^{N}\theta_{ij}v_j) \\ &= \frac{N}{M}\sum_{j=1}^{N}\mathrm{Var}(\theta_{ij})v_j^2 = \frac{1}{M}\sum_{j=1}^{N}v_j^2 = \frac{||\mathbf{v}||_2^2}{M}\end{aligned} \tag{13}$$

$$\begin{aligned}&\mathbb{E}(||\Theta\mathbf{v}||_2^2) = \mathbb{E}(\sum_{i=1}^{M}(< \sqrt{\frac{N}{M}}\theta_i, \mathbf{v} >)^2) \\ &= \sum_{i=1}^{M}\mathbb{E}(< \sqrt{\frac{N}{M}}\theta_i, \mathbf{v} >)^2 = \sum_{i=1}^{M}\mathrm{Var}(< \sqrt{\frac{N}{M}}\theta_i, \mathbf{v} >) \\ &= \sum_{i=1}^{M}\frac{||\mathbf{v}||_2^2}{M} = ||\mathbf{v}||_2^2\end{aligned} \tag{14}$$

Based on Theorem 1, we set $\alpha = 1-\delta$ and $\beta = 1+\delta$, and the following inequality can be obtained $\mathbb{P}\left(\frac{||\Theta\mathbf{v}||_2^2}{||\mathbf{v}||_2^2} \leq 1-\delta\right) \leq e^{-M\delta^2/c^*}$ and $\mathbb{P}\left(\frac{||\Theta\mathbf{v}||_2^2}{||\mathbf{v}||_2^2} \geq 1+\delta\right) \leq e^{-M\delta^2/c^*}$, then we have

$$\mathbb{P}\left(\left|\frac{||\Theta\mathbf{v}||_2^2}{||\mathbf{v}||_2^2} - 1\right| \geq \delta\right) \leq 2e^{-M\delta^2/c^*} \tag{15}$$

Since there are $(N,\ k)$ possible $k$-dimensional subspaces of $\Theta$, based on Sterling's approximation, we have $(N,\ k) \leq (eN/k)^k$, and the probability of $k$-sparse signal $\mathbf{v}$ which satisfies $\left|\frac{||\Theta\mathbf{v}||_2^2}{||\mathbf{v}||_2^2} - 1\right| \geq \delta$ is

$$(eN/k)^k \cdot 2e^{-\frac{M\delta^2}{c^*}} = 2e^{-\frac{M\delta^2}{c^*} + k\log(\frac{N}{k}) + 1} \tag{16}$$

Therefore, when $M = O(k\log(N/k))$ , the probability of $\Theta$ satisfies $(1-\delta) \leq \frac{||\Theta\mathbf{v}||_2^2}{||\mathbf{v}||_2^2} \leq (1+\delta)$ for all $k$-sparse signal $\mathbf{v}$, it trends to 1. □

According to the above analysis, we know that $\Psi_G$ can sparsify spatial correlation signal, and satisfy the RIP with $\Phi_e$ as well.

## 5. IMPLEMENTATION AND ANALYSIS

In this section, we first give the algorithm implementation of sparsest random scheduling for data gathering (SRS-DG), and then analyze the transmission cost of our scheme.

### 5.1 Algorithm Implementation

To implement SRS-DG scheme, we first investigate how to select $M$ random scheduling sensors from $N$ sensors to meet $\Phi_e$. Then we focus on how to adaptively adjust the number of CS measurements based on the recovery error from the previous rounds. For the first task, we develop a probabilistic scheduling strategy which can satisfy the proposed measurement matrix and ensure balanced sensor participation. For the second task, we utilize the recovery error of currently received sampling values to adjust our probabilistic scheduling. We next detailize the implementation of SRS-DG, which contains two components of algorithm implementation. (1) The sink component is responsible for the sensory data recovery and random scheduling probability assignment/adjustment as shown in Algorithm 1. (2) Each sensor component is responsible for sampling and transmitting data to the sink as shown in Algorithm 2.

**Sink Component:** In Algorithm 1, the inputs are representation basis $\Psi_G$, upper bound $\epsilon_{ub}$, lower bound $\epsilon_{lb}$ on recovery error, and the scheduling probability step length $\triangle p$. In the initializing stage, the scheduling probability is set to 1. The outputs are the recovery sensory data $\hat{\mathbf{x}}$ and sensor scheduling probability $p_s$. $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}_{\mathbf{r}}$ are denoted as recovery sparse signal, the whole recovery sensory data, and recovery received sensory data, respectively. During the data gathering, the indices of received sensory data need to be recorded in $\Omega_{\mathbf{r}}$ which has been used for generating $\Phi_e$. If the recovery error $\epsilon$ is greater than $\epsilon_{ub}$, which means there is a lack of CS measurements for recovering the sensory data, we will increase the number of CS measurements. Otherwise, we will decrease the number of CS measurements, namely, the scheduling probability $p_s$ needs to be adjusted when the recovery error is out of the range of $(\epsilon_{lb}, \epsilon_{ub})$. Notice that, the scheduling probability should not change frequently since the sparse level of sensory data is relatively stable.

**Sensor Component:** In Algorithm 2, each sensor obtains a sensing value periodically and decides whether or not to participate in the data gathering. In particular, each sensor will generate a random number $p_r \in (0,1)$, if the random number is greater than its scheduling probability, the sensor will send its value to the sink along the shortest routing path. Otherwise, the sensor will not sample and send its value.

### 5.2 Transmission Cost Analysis

In this subsection, we analyze transmission cost of SRS-DG and compare it with traditional CDG. We compare SRS-DG with three traditional CDG schemes, *dense random projections* for data gathering (DRP-DG) [17], *sparse random projection* for data gathering (SRP-DG) [26] and *hybrid CS* for data gathering (hybrid CS-DG) [18]. For simplicity of analysis, we assume that sensor network contains $N$ sensors, the average hop distance from any sensor to the sink is $H$. In DRP-DG, each CS measurement gathering require each sensor participation once. In SRP-DG, each row of measurement matrix can only contain $O(\log N)$ non-zero elements, namely, each measurement gathering needs to transmit at least $O(\log N)$ data packets. In hybrid CS-DG

**Algorithm 1:** Sensory data recovery and scheduling probability assignment.

```
Input  : Ψ_G, ε_ub ε_lb, △p.
Output: x̂, p_s
Φ_e ← 0;                       /* Initializing measurement matrix */
Ω_r ← {i| if x_i received };   /* Record received index */
j ← 1;
foreach i ∈ Ω_r do
    Φ_e(j, i) ← 1;             /* Assign measurement matrix */
    j ← j + 1;
Θ ← Φ_e ∗ Ψ_G;
x̂ = CS_Recovery (x̂_r, Θ);     /* Recovery sensory data */
x̂_r ← x̂_{Ω_r};                 /* Extract received recovery sensory data */
ε = MSE(x_r, x̂_r);             /* Recovery error */
p_s ← |Ω_r|/N;                 /* Current scheduling probability */
if ε > ε_ub then
    p_s ← p_s + △p;            /* Increase scheduling probability */
    broadcast p_s to all sensors;
if ε < ε_lb then
    p_s ← p_s − △p;            /* Decrease scheduling probability */
    Broadcast p_s  to  all  sensors;
```

**Algorithm 2:** Random scheduling data gathering for the $i^{th}$ sensor

```
Input  : p_s
Output: x_i
p_r ← rand();                  /* Generate a random probability */
if p_r ≤ p_s then
    Sampling x_i;
    Send x_i to the sink;
```

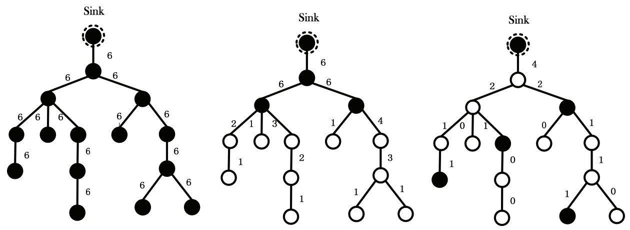
, if the number of transmission data packets is larger than CS measurements, the sensor carries out CS compression. Fig 6 shows three types of CS based data gathering schemes under multi-hop tree topology network. In Fig 6, the black sensors represent the participation compression sensor nodes during CS based data gathering and the link labels represent the number of transmission data packets during a round data gathering. Fig 6 illustrates that transmission cost of hybrid CS-DG less than DRP-DG, and transmission cost of our scheme outperforms than hybrid CS-DG and DRP-DG schemes. Considering hybrid CS-DG comes from DRP-DG, we only analyze and compare DRP-DG and SRP-DG with our scheme.

For DRP-DG, each row has $O(N)$ non-zero elements, and each CS measurement gathering requires at least $O(N)$ data packets. Assuming that the sparse level of gathering data is $k$, the sink needs to gather $O(k \log N)$ measurements to recovery sensory data, then a round data transmission cost, $TC_{drp}$, is

$$TC_{drp} = O(N \cdot k \log N) = O(kN \log N) \tag{17}$$

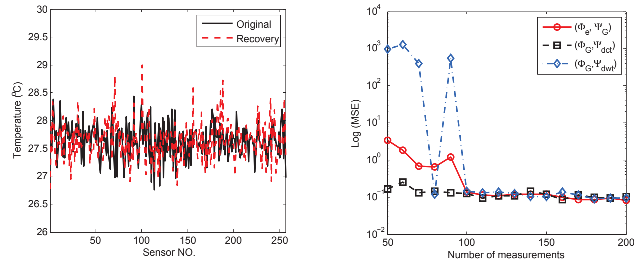
For SRP-DG, the number of nonzero elements is $O(\log N)$ and $O(k^2 \log N)$ measurements are required to recover sensory data. But each nonzero element is random, each CS measurement transmission cost is $O(H \log N)$. The transmission cost of one round, $TC_{srp}$, is

$$TC_{srp} = O(H \log N \cdot k^2 \log N) = O(k^2 H \log^2 N) \tag{18}$$




(a) DSP-DG. (b) Hybrid CS-DG. (c) SRS-DG.

**Figure 6: Communication cost comparisons with different CDG schemes in multi-hop tree-type topology.**




(a) Original and recovery. (b) Recovery errors comparison.

**Figure 7: Recovery performance without packet loss.**

For SRS-DG, each measurement only has one nonzero entry, and its transmission cost is $O(H)$. Because our proposed sparsest measurement matrix and representation basis satisfy RIP, it suffices to obtain $O(k \log N)$ measurements for achieving accurate data recovery. As a result, the transmission cost of one round SRS-DG, $TC_{srs}$, is

$$TC_{srs} = O(H \cdot k \log N) = O(kH \log N) \tag{19}$$

So, if $H = O(N)$ such as multi-hop chain topology network, $TC_{srs} = TC_{drp} = O(kN \log N) < O(k^2 N \log^2 N) = TC_{srp}$. If $H = O(\log N)$ such as multi-hop tree topology network, $TC_{srs} = O(k \log^2 N) < TC_{drp} = O(kN \log N)$ and $TC_{srs} = O(k \log^2 N) < TC_{srp} = O(k^2 \log^3 N)$. According to the above transmission cost analysis and comparison, it demonstrates SRS-DC also reduce transmission cost in lossless sensor networks compared with the state of the art traditional CDG schemes.

# 6. EVALUATION

We conduct extensive experiments using CitySee data set to evaluate the performance of our scheme in both reliable and lossy sensor network. In reliable WSNs with no packet loss, we compare our scheme with some traditional CDG schemes. In lossy WSNs, we evaluate the robustness of our scheme under different packet loss ratios. In particular, we choose random projections as a baseline scheme, it requires fewer measurements and can obtain better recovery performance compared with sparse random projections [5, 17]. In dense random projections, DCT and 4-layer 'haar' wavelet basis (DWT) are considered as different representation bases with recovery algorithm OMP [24].

## 6.1 Reliable sensor network

As analyzed in Section 5.2, our scheme achieves significantly lower transmission cost than using dense random projections. In our experiment, we evaluate our scheme from the following two aspects:

1. Evaluate recovery quality of gathering data using $\Phi_e$ and $\Psi_G$ as measurement matrix and representation basis respectively, which compares recovery data with the original data.

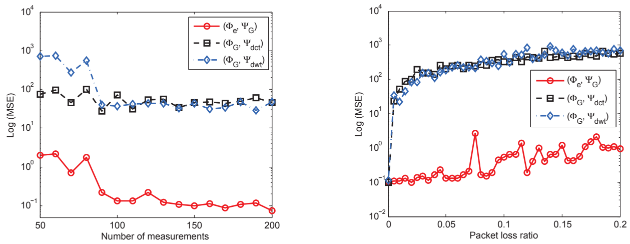
2. Compare the recovery errors using ($\Phi_e$, $\Psi_G$), ($\Phi_G$, $\Psi_{dct}$) and ($\Phi_G$, $\Psi_{dwt}$) as the pairs of measurement matrix and representation basis, which displays the recovery performance of our scheme compared with dense random projections.

The experimental data set contains 256 temperature sampling values of the CitySee system [15], which are collected from the same monitoring area during the same sampling time interval. Fig 7 (a) shows the original sampling data and the recovery data using 100 CS measurements with our scheme. The mean square error (MSE) of our scheme is 0.1642. Fig 7 (b) depicts recovery error comparisons among ($\Phi_e$, $\Psi_G$), ($\Phi_G$, $\Psi_{dct}$) and ($\Phi_G$, $\Psi_{dwt}$). The recovery performance of ($\Phi_e$, $\Psi_G$) can be as good as dense random projections when the number of measurements is greater than 90. It shows that our scheme can recover the data accurately without increasing the number of measurements.

## 6.2 Lossy sensor network

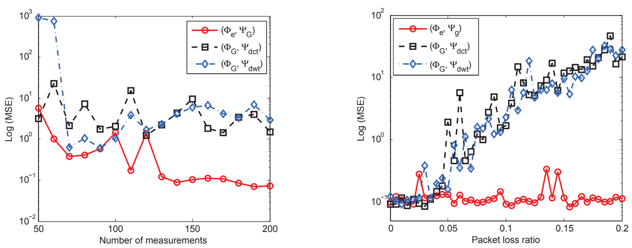
Then we evaluate the robustness of our scheme under different packet loss ratios. In this experiment, we adopt multi-hop tree routing topology. Each packet is dropped at random with uniform probability (packet loss ratio). Firstly, we carry out the experiment without allowing retransmission. Fig 8 (a) shows the recovery errors with different number of measurements under packet loss ratio 10% . We find that the recovery performance of dense random projection could be very bad even when the number of measurements is increased. In contrast, our scheme can always achieve satisfactory recovery accuracy, and the accuracy increases as the number of measurements increases. The MSE of $(\Phi_G, \Psi_{dct})$ and $(\Phi_G, \Psi_{dwt})$ are greater than 20, while the MSE of our scheme is always less than 0.1. Fig 8 (b) shows the recovery errors with different data packet loss ratios and 160 measurements. It demonstrates that the gathering data of dense random projections cannot be recovered even with 2% data packet loss ratio. The experiment results validate that our scheme is robust to packet loss.

Secondly, we carry out the experiment by allowing retransmission. Although the retransmission strategy is usually used for networks packet loss, it would increase transmission cost and resource consumption. Moreover, many real deployed systems shown that the packet loss ratio could still reach 20% with retransmission strategy [14]. Fig 9 displays the recovery error of different schemes when the max number of retransmissions (denoted by $Max\_rt\_num$) for each packet is 2. Fig 9 (a) demonstrates that the recovery errors of $(\Phi_G, \Psi_{dct})$ and $(\Phi_G, \Psi_{dwt})$ are always large with increasing the number of measurement. Fig 9 (a) also demonstrates that when the packet loss ratio is 5%, we can hardly recover the data using dense random projection schemes.




(a) $pl\_ratio = 0.1$ (b) $M = 160$

**Figure 8: Recovery performance without retransmission.**




(a) $pl\_ratio = 0.05$ (b) $M = 160$

**Figure 9: Recovery performance with retransmission ($Max\_rt\_num = 2$).**

# 7. RELATED WORK

The emergence of CS theory has opened up a new research avenue for in-network compression and sampling. For example, D. Baron *et al.* [23] proposed distributed CS to compress multi-signal exploiting both intra- and inter-signal correlation structures. In [12, 21], Haupt, J. *et al.* applied CS theory to single-hop data gathering in WSNs to obtain efficient compression for network data. In [16], Luo *et al.* applied CS to reduce data transmission cost in large-scale WSNs. In [25, 29, 30], J. Wang *et al.* proposed a dual-layer compressed aggregation and adaptive the number of measurements scheme. These techniques exploited *dense measurement matrix* to gather CS measurements, the transmission cost of each CS measurement is $O(N)$ because each row of measurement matrix has $O(N)$ nonzero elements. To recover $k$-sparse sensory data, it requires $O(k \log N)$ measurements. In [18], J. Luo *et al.* proposed that applying CS naively may not bring any improvement for WSN data gathering and proposed a hybrid-CS data gathering scheme. In [17], C. Luo *et al.* discovered that $[I, R]$ measurement matrix has also good RIP, $I$ is $M \times M$ identity matrix and $R$ is $M \times (N - M)$ dense random matrix. But each CS measurement matrix transmission cost is still $O(N)$. In [26], W. Wang *et al.* proposed *sparse measurement matrix* can also obtain the salient information of compressible signal, each row of it has $O(\log N)$ nonzero entries, but it requires $O(k^2 \log N)$ CS measurements to recovery $k$-sparse sensory data. Based on sparse measurement matrix, Lee. S *et al.* [13] proposed low coherence projections for efficient routing data gathering. In [27, 28], X. Wu *et al.* proposed a temporal random sampling data gathering scheme, which only considered one sensor in temporal domain. They didn't consider spatial signal in large-scale WSN data gathering and its representation basis is inflexible. To the best our knowledge, most of existing CS base CDG techniques were discussed in reliable WSNs.

# 8. CONCLUSION AND FUTURE WORK

In this paper, we discussed a novel sparsest random scheduling scheme for compressive data gathering in lossy sensor networks. Although traditional compressive sensing based data gathering can recovery gathering data accurately, these techniques were not workable in lossy sensor network according to our real experimental results. Our scheme can efficiently avoid recovery performance degraded by link loss, also can reduce transmission cost in reliable networks. We carried out experiment in real CitySee data set, experiment results demonstrated our scheme can recover the sensory data accurately both in reliable and lossy sensor networks. In our scheme, we only consider spatial sparsest random scheduling for compressive data gathering. Future work will extend to spatial-temporal domain to joint optimal sparsest random scheduling for compressive data gathering.

## Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (No.61232018, No.61272487, No.61170233, No.61003311).

# 9. REFERENCES

[1] A. Arora, R. Ramnath, E. Ertin, P. Sinha, S. Bapat, V. Naik, V. Kulathumani, H. Zhang, H. Cao, M. Sridharan, et al. Exscal: Elements of an extreme scale wireless sensor network. In *Embedded and Real-Time Computing Systems and Applications, 2005. Proceedings. 11th IEEE International Conference on*, pages 102–108. IEEE, 2005.

[2] R. Baraniuk. Compressive sensing. *Signal Processing Magazine, IEEE*, 24(4):118–121, 2007.

[3] V. Buldygin and I. Kozachenko. *Metric characterization of random variables and random processes*, volume 188. Amer Mathematical Society, 2000.

[4] E. Candès. Compressive sampling. In *Proceedings oh the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, pages 1433–1452, 2006.

[5] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[6] E. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[7] C. Chong and S. Kumar. Sensor networks: Evolution, opportunities, and challenges. *Proceedings of the IEEE*, 91(8):1247–1256, 2003.

[8] M. Davenport. *Random observations on random observations: Sparse signal acquisition and processing.* PhD thesis, Texas A&M University, 2010.

[9] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong. Model-driven data acquisition in sensor networks. In *Proceedings of the Thirtieth international conference on Very large data*

*bases-Volume 30*, pages 588–599. VLDB Endowment, 2004.
[10] D. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
[11] GreenOrbs. http://www.greenorbs.org/all/greenorbs.htm.
[12] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak. Compressed sensing for networked data. *Signal Processing Magazine, IEEE*, 25(2):92–101, 2008.
[13] S. Lee, S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega. Compressed sensing and routing in multi-hop networks. Technical report, Technical Report, University of Southern California, 2009.
[14] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, L. Mo, W. Dong, Z. Yang, M. Xi, J. Zhao, et al. Does wireless sensor network scale? a measurement study on greenorbs. In *INFOCOM, 2011 Proceedings IEEE*, pages 873–881. IEEE, 2011.
[15] Y. Liu, X. Mao, Y. He, K. Liu, W. Gong, and J. Wang. Citysee: Not only a wireless sensor network. *IEEE NETWORK*, 27(5):42–47, 2013.
[16] C. Luo, F. Wu, J. Sun, and C. Chen. Compressive data gathering for large-scale wireless sensor networks. In *Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 145–156. ACM, 2009.
[17] C. Luo, F. Wu, J. Sun, and C. Chen. Efficient measurement generation and pervasive sparsity for compressive data gathering. *Wireless Communications, IEEE Transactions on*, 9(12):3728–3738, 2010.
[18] J. Luo, L. Xiang, and C. Rosenberg. Does compressed sensing improve the throughput of wireless sensor networks? In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–6. IEEE, 2010.
[19] X. Mao, X. Miao, Y. He, X.-Y. Li, and Y. Liu. Citysee: Urban co_2 monitoring with sensors. In *INFOCOM, 2012 Proceedings IEEE*, pages 1611–1619. IEEE, 2012.
[20] D. Needell and J. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.
[21] M. Rabbat, J. Haupt, A. Singh, and R. Nowak. Decentralized compression and predistribution via randomized gossiping. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 51–59. ACM, 2006.
[22] C. E. Rasmussen. Gaussian processes for machine learning. 2006.
[23] S. Sarvotham, R. Baraniuk, M. Duarte, D. Baron, and M. Wakin. Distributed compressive sensing, 2009.
[24] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, 2007.
[25] J. Wang, S. Tang, B. Yin, and X. Li. Data gathering in wireless sensor networks through intelligent compressive sensing. In *INFOCOM, 2012 Proceedings IEEE*, pages 603 – 611. Ieee, 2012.
[26] W. Wang, M. Garofalakis, and K. Ramchandran. Distributed sparse random projections for refinable approximation. In *Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on*, pages 331–339. IEEE, 2007.
[27] X. Wu and M. Liu. In-situ soil moisture sensing: measurement scheduling and estimation using compressive sensing. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*, pages 1–12. ACM, 2012.
[28] X. Wu, M. Liu, and Y. Wu. In-situ soil moisture sensing: Optimal sensor placement and field estimation. *TOSN*, 8(4):33, 2012.
[29] L. Xiang, J. Luo, C. Deng, A. Vasilakos, and W. Lin. Dual-level compressed aggregation: Recovering fields of physical quantities from incomplete sensory data. *Arxiv preprint arXiv:1107.4873*, 2011.
[30] L. Xiang, J. Luo, and A. Vasilakos. Compressed data aggregation for energy efficient wireless sensor networks. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*, pages 46–54. IEEE, 2011.
[31] J. Yick, B. Mukherjee, and D. Ghosal. Wireless sensor network survey. *Computer Networks*, 52(12):2292–2330, 2008.

# APPENDIX

In Corollary 1, we need to prove $e^{|\lambda|} - |\lambda| \leq e^{2\lambda^2/2}$ .

PROOF. If $e^{|\lambda|} - |\lambda| \leq e^{2\lambda^2/2}$ is corrected, we only need to proof $\ln(e^{|\lambda|} - |\lambda|) \leq \lambda^2$, which is equivalent to $\frac{\ln(e^{|\lambda|}-|\lambda|)}{\lambda^2} \leq 1$. Since

$$\frac{\ln(e^{|\lambda|} - |\lambda|)}{\lambda^2} = \frac{\ln(e^{|\lambda|} - |\lambda| - 1 + 1)}{\lambda^2} \leq \frac{e^{|\lambda|} - |\lambda| - 1}{\lambda^2}$$
$$= \frac{1}{\lambda^2} \sum_{n=2}^{\infty} \frac{(|\lambda|)^n}{n!} = \sum_{n=2}^{\infty} \frac{(|\lambda|)^{n-2}}{n!}$$

If $|\lambda| < 1$ , then

$$\frac{\ln(e^{|\lambda|} - |\lambda|)}{\lambda^2} \leq \sum_{n=2}^{\infty} \frac{(|\lambda|)^{n-2}}{n!} < \sum_{n=2}^{\infty} \frac{1}{n!} = e - 1 - 1 < 1$$

If $|\lambda| \geq 1$ , then $\frac{\ln(e^{|\lambda|}-|\lambda|)}{\lambda^2} \leq \frac{|\lambda|}{\lambda^2} \leq 1$
So, $e^{|\lambda|} - |\lambda| \leq e^{2\lambda^2/2}$ □