Article

Low-Energy Data Collection in Wireless Sensor Networks Based on Matrix Completion

Yi Xu [1] [ID], Guiling Sun [1,*], Tianyu Geng [1] and Jingfei He [2]

[1] College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300071, China; xuyi@mail.nankai.edu.cn (Y.X.); 1010337@mail.nankai.edu.cn (T.G.)

[2] Department of Key Laboratory of Electronic Materials and Devices of Tianjin, School of Electronics and Information Engineering, Hebei University of Technology, Tianjin 300401, China; hejingfei@hebut.edu.cn

* Correspondence: sungl@nankai.edu.cn; Tel.: +86-135-1245-3718

Abstract: Sparse sensing schemes based on matrix completion for data collection have been proposed to reduce the power consumption of data-sensing and transmission in wireless sensor networks (WSNs). While extensive efforts have been made to improve the recovery accuracy from the sparse samples, it is usually at the cost of running time. Moreover, most data-collection methods are difficult to implement with low sampling ratio because of the communication limit. In this paper, we design a novel data-collection method including a Rotating Random Sparse Sampling method and a Fast Singular Value Thresholding algorithm. With the proposed method, nodes are in the sleep mode most of the time, and the sampling ratio varies over time slots during the sampling process. From the samples, a corresponding algorithm with Nesterov technique is given to recover the original data accurately and fast. With two real-world data sets in WSNs, simulations verify that our scheme outperforms other schemes in terms of energy consumption, reconstruction accuracy, and rate. Moreover, the proposed sampling method enhances the recovery algorithm and prolongs the lifetime of WSNs.

Keywords: wireless sensor networks; data collection; sparse sampling; matrix completion

1. Introduction

Wireless Sensor Networks (WSNs) have attracted many researchers' attention and have been deeply studied [1]. With the growing demand for big data, the size of WSNs is increasing, which raises greater challenges to data-collection technologies of WSNs. Moreover, the control of power consumption at the sensor nodes is always a main problem faced by WSNs. To solve the problem, sparse sensing schemes as efficient methods for data collection have been proposed, optimizing the overall energy consumption. It reduces the power consumption of data sensing and data transmission, prolonging the network life.

Sparse sensing schemes concentrate on the reduction of the number of samples and the reconstruction of the missing data using algorithms exploiting the data structure available in the meantime. Based on the sparsity of sensor observations, Compressed Sensing (CS) theory [2] has been used to compress and reconstruct the data. Although CS theory has been widely used in WSNs [3–5], the number of samples is not actually reduced because sensor nodes need to sample all data and then compress. Later according to the research progress of the Matrix Completion (MC) theory [6], researchers exploited the low-rank characteristic of sensor data and applied the MC theory to WSNs [7]. In MC, the low-rank data matrix can be recovered accurately with some samples uniformly at random [6], which means only a portion of data are needed to be sensed and transmitted to the sink, realizing a real sense of sparse sensing.

The data-collection methods based on MC theory mainly include the sampling method and the recovery algorithm. It can reduce the energy consumption from two aspects. One is from reducing the amount of data being sensed and transmitted, and the other one is that many sensor nodes are in the sleep mode when they are not selected to sense data. However, this has a strict requirement for the topology of WSNs. When some nodes are in the sleep mode, the sink may not be able to receive the sensed data due to the communication limits, especially in large-scale WSNs and many realistic environments. Even though the topology can be set up adaptively, it will be energy-consuming and time-consuming as the active nodes are varying. The above problems are ignored by many sampling methods and corresponding algorithms based on MC theory.

Since the application of MC theory in WSNs, many algorithms for solving the MC problem are adapted to WSNs. In the beginning, people use the alternative least squares (ALS) method to solve the data recovery problem exploiting the spatial correlation of WSNs data, which is called the Efficient Data-Collection Approach (EDCA) [7]. Later the Spatio-Temporal Compressive Data Collection (STCDG) [8] was proposed adding the short-term stability features to improve the accuracy. These methods assume the rank of the data is fixed and known, which is unlikely to hold in the practical system. He et al. combined the MC with sparsity constraints, proposing the data recovery method with joint MC and Sparsity Constraints (DRMCSC) [9], and used the alternating minimization to solve the problem.

The recovery methods of previous works make the reconstruction accuracy of WSNs data high; however, the models and the alternative optimization algorithms used in previous works are usually complex and cost a large amount of running time. With the number of data increasing, the recovery efficiency will be a problem. Singular Value Thresholding (SVT) algorithm is a classic algorithm for solving the MC problem quickly [10]. However, the recovery accuracy of SVT algorithm cannot be guaranteed for all kinds of original data and samples. The samples of most previous data-collection methods are obtained randomly with a deterministic sampling ratio. Practical and efficient sampling methods have been proposed such as the adaptive sampling with MC for specific real WSNs data [11], determining the effective sampling set with three sample learning principles and dealing with the corrupted problem [12]. However, the sampling method is complex and energy-consuming.

We assume the WSN is large-scale and monitors slowly changing environment parameters over time such as the temperature, the humidity, and the light. Moreover, the sensor nodes sense the data at set intervals and the interval is short, usually several to dozens of seconds. This scenario usually exists in agriculture and environmental protection field. With these assumptions, to recover the data accurately and fast and to reduce the energy consumption in WSNs, we propose a low-energy data-collection method based on sparse sensing with MC theory in this paper. Our method includes a Rotating Random Sparse Sampling (RRSS) method and a Fast Singular Value Thresholding (FSVT) algorithm. The proposed method can be used not only for WSNs but also in the context of more complex networks as Wireless Sensor and Actuator Networks (WSANs) [13,14] or the Wireless Sensor, Actuator and Robot Network (WSARN) [15]. The main contributions of this paper are summarized as follows:

- We propose a low-energy data-collection method for large-scale WSNs including a load-balancing random sampling method and a corresponding algorithm to recover the original data from the samples accurately and fast. The data-collection method can be combined with any topologies of WSNs easily without the limit to communication change in sparse sensing schemes.
- We design a rotating varying sampling ratio based on sparse sampling and set a time schedule, on which sensor nodes are in the sleep mode most of the time to greatly reduce the energy consumption.
- We construct the model with Tikhonov regularization for recovering the original signal from the samples and use the SVT algorithm with the Nesterov technique to solve it. Theoretical explanations of the effectiveness of the algorithm have been given.

- With two real-world data sets, we evaluate the proposed sampling method with the corresponding FSVT algorithm compared the traditional sampling method and other data-collection methods. The simulation results show that the proposed sampling method can enhance the FSVT algorithm and the recovery performance outperforms other methods in terms of the reconstruction accuracy and rate.

The rest of this paper is organized as follows. Section 2 presents the basic collection method with matrix completion for WSNs including the basic MC theory and the random sampling method. The proposed sampling method is described in Section 3. According to the proposed sampling method, the problem formulation and the solution are given in Section 4. Section 5 presents the experiments results of the proposed methods compared with the state-of-the-art methods in the term of reconstruction accuracy and running time. Finally, Section 6 concludes the paper.

2. Data Collection Based on MC

In a WSN system, considering N sensor nodes and one sink, sensor nodes sense and transmit the signal to the sink. We define a matrix $X_{N \times T}$ to represent the original data from N sensor nodes over T time slots without loss of generality. In the data matrix X, the row index corresponds to the node index and the column index corresponds to the time slot so that an element X_{ij} represents the data for a specific sensor node i in the jth time slot.

MC is to solve the problem of recovering a data matrix from a few samplings of its entries. Candes et al. [6] proved that if the number m of sampled entries obeys

$$m >> Cn^{1.2}r \log n \tag{1}$$

for some positive numerical constant C, then with very high probability, most $n \times n$ matrices of rank r can be perfectly recovered by solving a simple convex optimization problem.

Many data-collection methods for WSNs are developed based on the MC theory [7–9]. For these methods, at each time slot, some sensor nodes are selected to sense data randomly with a deterministic sampling ratio and then transmit the data to the sink. After T time slots, the sink gets the measurements matrix, which is generally incomplete. Only the positions corresponding to the selected nodes offer valid sensing data. The values in the other positions are usually set as zeros. Here we adopt a linear operator $\mathcal{A}(\cdot)$ to represent the sampling procedure:

$$\mathcal{A}(X) = M \tag{2}$$

where X is the original data without the loss of generality and M is the incomplete measurements matrix. It can be also seen as a Hadamard product of X and a sampling matrix, which is linear [16]. As this operator is linear, it is easy to recover X from M with the constraint of Equation (2). Previous studies have shown that data collected from WSNs are highly spatial and temporary correlated [17], which brings the low-rank characteristic of the matrix X. According to the work of the MC theory [6], it is highly possible to recover a low-rank matrix from a subset of its entries. Thus, the data recovery problem can be formulated as follows:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \text{rank}(X) \\ \text{subject to} \quad & \mathcal{A}(X) = M. \end{aligned} \tag{3}$$

The brief graphical representation for the data-collection method based on MC is shown in Figure 1. Circles represent the nodes and the $\times$ in the middle represents the sink. Through the selected sensing nodes, which are shown as black circles, the sink gets the measurements matrix and then recovers the original data according to the Equation (3). Because only a portion of sensor nodes need

to sense and transmit data, the energy consumption has been reduced and the lifetime of WSNs has been prolonged.
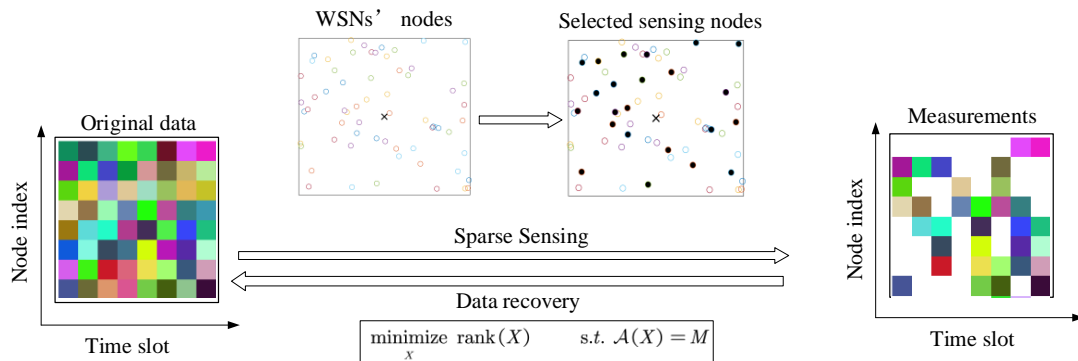

Figure 1. The data-collection method based on matrix completion.

3. The Rotating Random Sparse Sampling Method

From the description of the data-collection method based on MC theory, it can be known that the original data can be recovered from the samples. However, the positions of the selected sensing nodes have great impacts on the recovery performance and the energy consumption of WSNs. Although the previous random sampling method ensures the above two aspects to an extent, it has some difficulties in implementing it considering the communication limit when random nodes are in the sleep mode. Moreover, it may generate all-zeros column in the measurements matrix when the sampling ratio is low, which poses a challenge to the recovery algorithm.

We design the RRSS method based on MC. The random characteristic of the sampling method remains to ensure the security of WSNs. In contrast to the previous random sampling method, the sampling ratio is varying over time slots and it may be different among sensor nodes.

We define a round include a transmission period and c time slots, during which sensor nodes sense. Each round every sensor node only samples once and transmits once so that the overall sampling ratio is $p = 1/c$ for the whole WSN. With these definitions, the proposed RRSS method is explained as follows in chronological order.

At the beginning of each round, which is also called the transmission period, all sensor nodes will wake up independently, set up the topology and transmit the data collected last round. N data will be collected at the sink as each sensor node sense once in a round. The data package of each sensor node includes the data value x, the number of sensing time slot j and the node ID i so that data value x can be formed as X_{ij} at the sink. It is noteworthy that our method can be easily combined with any WSN topology and data compression method because all sensor nodes are active when transmitting and the number of data each sensor node transmits is deterministic.

After the transmission period, during the first time slot of each round, each sensor node i chooses a uniform random number a_{iu} between 0 and 1 at the uth round and then compares a_{iu} with a threshold. The sensor node is selected to sense if its random number a_{iu} is less than the threshold. The threshold is related to the sparse sampling ratio p and defined as follows:

$$P_i(j) = \frac{p}{1 - p \times ((j-1) \bmod c)} , \quad i \in G_u, \tag{4}$$

where $j = 1, 2, ...T$. $u = \text{ceil}(j/c)$ is the nearest integer greater than or equal to (j/c) and represents the number of current round. We define the set G_u in which sensor nodes have not been selected to sense in the uth round. The explanation of this definition is given later in this section after the chronological description.
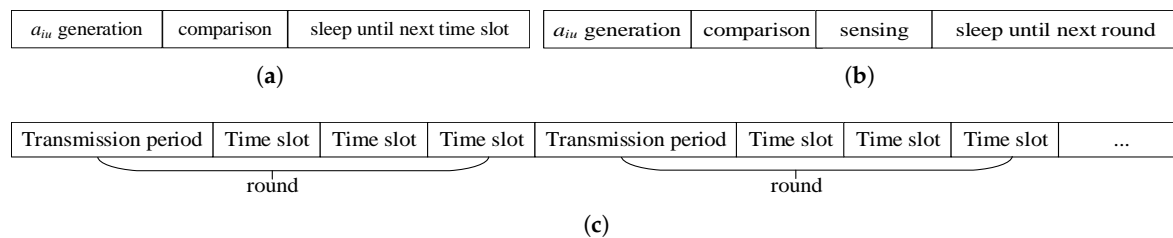
The selected nodes sense and send their sensed data to the sink during the jth time slot, then it will not belong to the set G_u anymore. After that, these selected nodes will fall into sleep mode during the remaining time slots of this round to reduce energy consumption.

In the following time slots of this round, the nodes $i \in G_u$ repeat the process of random number generation and comparison with the threshold as well as falling into sleeping mode after being selected to sense.

When re-entering a new round, all nodes will wake up and start the data transmission and the random number generation process, repeating the same sampling process as the previous round. The time scheduling of the process is shown in Figure 2.

Because a round includes c time slots, the mean number of selected nodes during each time slot is $N/c = N \times p$, which we define as r so that Equation (4) can be transformed into

$$P_i(j) = \frac{r}{N - r \times ((j-1) \bmod c)}, \quad i \in G_u. \tag{5}$$

a_{iu} generation	comparison	sleep until next time slot

(a)

a_{iu} generation	comparison	sensing	sleep until next round

(b)

Transmission period	Time slot	Time slot	Time slot	Transmission period	Time slot	Time slot	Time slot	...

round　　　　　　　　　　　round

(c)

Figure 2. Time scheduling. (**a**) A time slot of non-sensing nodes. (**b**) A time slot of sensing nodes. (**c**) Time scheduling of the network when $c = 3$.

It can be seen that the probability $P_i(j)$ changes with the time slot varying, which makes the sampling ratio keep p and the number of selected nodes keep r. Because each node only senses once in a round, the number of nodes in the set G_u reduces and can be formulated as $N(j) = N - r \times ((j-1) \bmod c)$ so that the expected number of selected sensor nodes during each time slot is:

$$\begin{aligned} E &= N(j) \times P_j \\ &= (N - r \times ((j-1) \bmod c)) \times \frac{r}{N - r \times ((j-1) \bmod c)} \\ &= r. \end{aligned} \tag{6}$$

Because the method for selecting sensor nodes to sense and to transmit is rotating and random, and nodes will fall into the sleep mode after sensing once in a round, the RRSS method can reduce and balance the energy consumption of the network as well as sampling randomly and sufficiently to make the system secure and the recovery performance more accurate and fast.

It is notable that the proposed method is suitable for monitoring slowly changing environment parameters over time. The time length of a round is usually several minutes. The time length of a transmission period t_r is according to the size of WSNs as it needs to transmit all data to the sink. The time length of a time slot t_s can be several to dozens of seconds as described above. So, the time length of a round can be expressed as $(t_r + t_s \times c)$.

4. Problem Formulation and the Solution

Using the RRSS method, the sink obtains the observed data defined as $B_{N \times T}$. We use a sampling matrix $Q_{N \times T}$ of which the element is 1 when the data is observed or 0 otherwise. According to Equation (4), Q is defined as follows in detail:

$$
Q = \begin{pmatrix}
0 & 1 & 0 & \cdots & 0 & \cdots & 0 & 1 & \cdots \\
0 & \cdots & 1 & 0 & \cdots & 1 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 1 & 0 & \cdots & 0 & \cdots & 1 & 0 & \cdots \\
1 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 & \cdots
\end{pmatrix}.
\tag{7}
$$

$$\underbrace{\qquad\qquad}_{\text{a round}} \qquad \underbrace{\qquad\qquad}_{\text{a round}}$$

In a round, there is only one 1 in a row while the others in the row are zeros and the sampling process can be formulated as:

$$
B = Q \circ X,
\tag{8}
$$

where $\circ$ represents the Hadamard product of two matrices. Namely, $B_{ij} = Q_{ij} \times X_{ij}$. To recover the original data X from the observed data matrix B, based on its low-rank property, intuitively, we consider the rank minimization problem as

$$
\min_{X} \{ \mathrm{rank}(X), \ \text{subject to } B = Q \circ X \}
\tag{9}
$$

with $\mathrm{rank}(X)$ being the rank of X. However, this problem is NP-hard [18]. Thus, people turn to solve its convex approximation [10] which reads as

$$
\min_{X} \{ \|X\|_*, \ \text{subject to } B = Q \circ X \},
\tag{10}
$$

where $\|X\|_*$ is the nuclear norm which is defined as sum of all singular values of X. Rather than directly solving Equation (10), we focus on its penalty problem, by employing positive parameters λ, μ as

$$
\min_{X} \left\{ \Phi(X) := \frac{1}{2} \|Q \circ X - B\|^2 + \lambda \|X\|_* + \frac{\mu}{2} \|X\|_F^2 \right\}.
\tag{11}
$$

The term $\mu \|X\|_F^2$ is actually the Tikhonov regularization [19] which is developed based on the idea that Q is ill-posed. In fact, various nonconvex low-rank regularizations are also developed [20–23]. However, here, we just consider the convex regularizations and then we can use theory-proved acceleration techniques. The convex composite optimization structure of this problem allows us to apply the forward-backward splitting method [24]. The proximal map [25] of $\lambda \|X\|_*$ is SVT operator [10]

$$
\mathcal{S}_\tau(X) = U \mathrm{Diag} \{ (\sigma_i - \tau)_+ \} V
\tag{12}
$$

where U, V and σ_i comes from the singular value decomposition of $X = U \mathrm{Diag}\{(\sigma_i)\} V$. Please note that the gradient of $\frac{1}{2} \|Q \circ X - B\|^2 + \frac{\mu}{2} \|X\|^2$ can be easily calculated as

$$
\nabla \left(\frac{1}{2} \|Q \circ X - B\|^2 + \frac{\mu}{2} \|X\|^2 \right) = Q \circ X - Q \circ B + \mu X.
$$

Obviously, the gradient is Lipschitz with $1 + \mu$. Direct use of forward-backward splitting method to (11) yields the following scheme

$$
X^{k+1} = \mathcal{S}_{\gamma \lambda} (X^k - \gamma (Q \circ X^k - Q \circ B + \mu X^k)),
\tag{13}
$$

where γ is the stepsize. It has been well-known that the forward-backward splitting can be accelerated by the Nesterov technique which is frequently used in sparse or low-rank signal processing [26–32]. By introducing an auxiliary positive sequence $(t_k)_{k\geq 0}$ with $t_0 = 1$ and $(Y^k)_{k\geq 0}$, the accelerated scheme can be described as: for any given X^0, set $Y^0 = X^0$, $t_0 = 1$,

$$X^{k+1} = \mathcal{S}_{\gamma\lambda}(Y^k - \gamma(Q \circ Y^k - Q \circ B + \mu X^k)), \tag{14a}$$

$$t_{k+1} = \frac{1 + \sqrt{t_k^2 + 1}}{2}, \tag{14b}$$

$$Y^{k+1} = X^k + \frac{t_k - 1}{t_{k+1}}(X^k - X^{k-1}). \tag{14c}$$

The iterations stop when k exceeds the maximum number of iterations, $\|X^{k+1} - X^k\|_F / \|X^k\|_F$ is smaller than a predefined tolerance parameter ϵ, or $\|B - Q \circ X^{k+1}\|_F / \|B\|_F$ is smaller than ϵ.

Both Equations (14b) and (14c) just involve basic linear algebra computations. The only costly part for computation is Equation (14a) which needs the SVD computing. Compared with the forward-backward splitting method convergence with speed $O(\frac{1}{k})$, the fast SVT Equation (14) can improve the speed to $O(\frac{1}{k^2})$ [27], where k denotes the iterations. According to (Theorem 4.4, [27]), if setting $\gamma = \frac{1}{1+\mu}$, we immediately get the following result.

Proposition 1. *Assume that $(X^k)_{k\geq 0}$ generated by the FSVT Equation (14) minimizing Φ, and X^* is any solution to Equation (11). If $\gamma = \frac{1}{1+\mu}$, then it holds that*

$$\Phi(X^k) - \Phi(X^*) \leq \frac{2(1+\mu)\|X^0 - X^*\|^2}{(k+1)^2} \tag{15}$$

with X^0 being any starting point.

5. Performance Evaluation

The experiments are conducted on MATLAB platform with real-world data sets to evaluate the performance of the proposed method and other data-collection methods. We firstly introduce the experiment environments and parameters setting, and then show the results and analysis of the experiments.

5.1. Experiment Environments and Parameters Setting

5.1.1. Real-world Data Sets

Two sets of data from two institutes are used for simulations to evaluate the performance of data-collection methods, one of which is often used in previous WSNs data-collection works [7,9,33], the other one is generated recently [34].

The first dataset is a real-world trace from the Intel Berkeley Research Lab and the temperature data collected by 54 sensor nodes on 1 March 2004 [35] are selected. The sensors sense the temperature once every 31 s and the real-world trace forms the $X \in \mathbb{R}^{54 \times 2880}$ which represent the temperature values sampled by 54 nodes in one day. The sensors are arranged in the Intel Berkeley Research Lab and the venue of sensors deployment is shown in Figure 3.
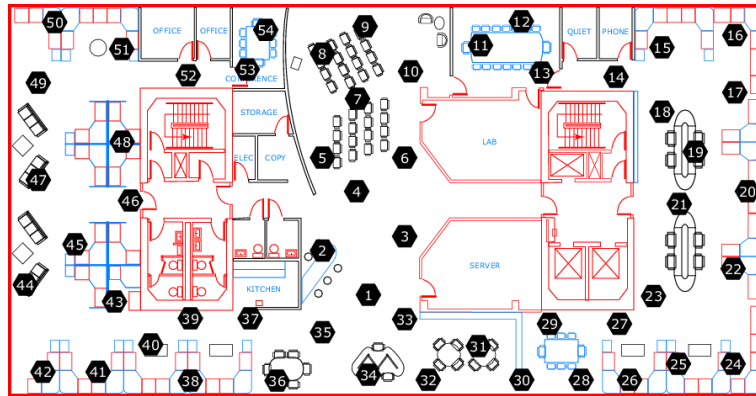
Figure 3. The venue of sensors deployment from Intel Berkeley Research Lab.

The second dataset is the sensing mote data from the Data-Sensing Lab [34]. The dataset is named as the Strata New York 2012 held at the New York Hilton Midtown generated in October 2012 in New York, NY and the temperature humidity data are selected. There are 40 sensors and 1724 time slots forming the matrix $X \in \mathbb{R}^{40 \times 1724}$. The brief map indicating the position of each sensor mote is shown in Figure 4.

The hexagon with number inside in Figures 3 and 4 represents different sensor nodes with number ID. It is easily seen from the two brief maps that the environment of sensors deployment is usually complex. So, using the basic sparse sampling method will cause some unnecessary energy consumption or demand strict topology as some sensor nodes do not need to transmit data. However, with the proposed RRSS method, all sensor nodes are active at the end of a round and only transmit one value to the sink so that any protocols and routings can be used.
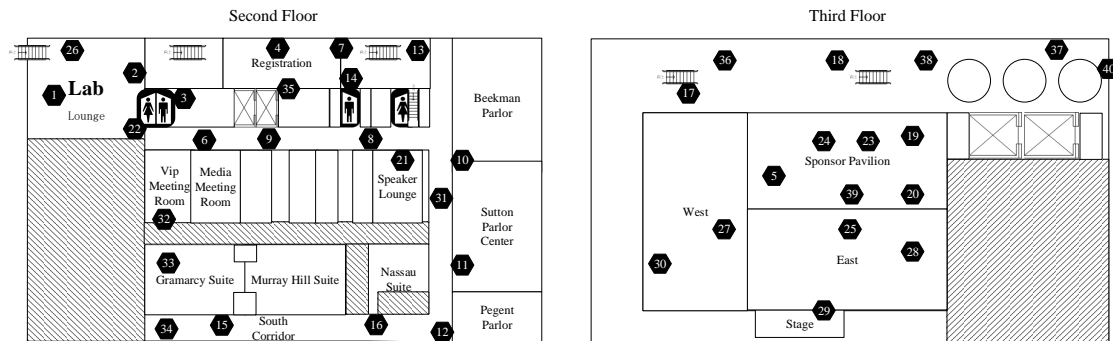
Figure 4. The venue of sensors deployment from Data-Sensing Lab.

From the sensor nodes of the above two labs, the sensing temperature data are obtained and are formed as two matrices, $X \in \mathbb{R}^{54 \times 2880}$ form Intel Berkeley Research Lab and $X \in \mathbb{R}^{40 \times 1724}$ from Data-Sensing Lab. However, with the disturbance when sensing and transmitting, there are many missing data in the two matrices. Actually, the real data curve should be smooth and continuous. To evaluate the performance better and according to the characteristic of WSNs data, the missing measurements are preprocess through the 2D Median filter with the 3×3 box. The sensed measurements are remained the same to preserve the characteristic of the original data to the greatest extent.

5.1.2. Parameters Setup

To evaluate the accuracy of the reconstructed data, the Normalized Mean Absolute Error (NMAE) is used to measure the recovery performance which is defined as:

$$\text{NMAE} = \frac{\sum_{i,j:(i,j)\notin\Omega} |\hat{X}_{ij} - X_{ij}|}{\sum_{i,j:(i,j)\notin\Omega} |X_{ij}|}, \tag{16}$$

where Ω is the set of sampling values in X and the $\hat{X}$ is the recovered data. Using NMAE, only the error of data which are not sampled is considered as the samples are obtained directly. NMAE is used in many previous data-collection works for WSNs [8,9] and is suitable to evaluate the performance of data-collection methods.

The methods are tested over different overall sampling ratios p. p determines the number of effective value of the measurements matrix. The data recovery experiments are conducted with the sampling ratio changing from 5% to 50%. The traditional random sampling method based on MC is used. With a fixed p, through a random selection process of 1-value elements in Q, the sampling matrix Q can be obtained so that the sampling process is done according to the Equation (2). The proposed RRSS method generates the Q with the varying p according to the Equation (4).

For the reconstruction methods, the iteration number threshold is set as 1000. The parameters in the Equation (13) are set as $\lambda = 0, 1, \mu = 0.1$. The stepsize γ is influenced by the matrix Q and will be adjust for different data sets. The initial matrix of X, X^0 is a $N \times T$ matrix containing pseudorandom values drawn from the standard uniform distribution on the open interval (0 1). The predefined tolerance parameter ϵ for the algorithm to be terminated is set as $\epsilon = e^{-5}$.

5.2. Results and Analysis of Experiments

The experiments on the recovery performance of the proposed data-collection method compared with the EDCA, STCDG and DRMCSC methods are implemented. Moreover, the traditional sampling method with the FSVT algorithm and the proposed RRSS method with the FSVT algorithm are also conducted, which are labeled as FSVT-Random and FSVT-RRSS. The recovery performance includes the recovery accuracy using the metric NMAE and the running time of the algorithm. In each experiment, the process of randomly sampling data and recovering data was repeated 10 times. The experimental results presented were the mean NMAE and mean running time, which are shown in Figures 5–8.

Figures 5 and 6 show the experimental results of the recovery performance for the temperature data from Intel Berkeley Research Lab. It can be seen that the proposed method, the FSVT reconstruction algorithm with the RRSS method, outperforms other methods in both reconstruction accuracy and time especially when the sampling ratio is low. Not only the RRSS sampling method enhances the FSVT method, improving the reconstruction accuracy, but also it decreases the running time as shown in Figure 6.

Figures 7 and 8 show the experimental results of the recovery performance for the temperature data from Data-Sensing Lab. The results are similar to that of the data from Intel Berkeley Research Lab, but the recovery accuracy is not as good as that, overall. Actually, the stepsize γ of the FSVT algorithm in the Equation (13) decreases as it is more difficult to find the correct solution compared with that for the data from Intel Berkeley Research Lab. The FSVT algorithm is more suitable for large-scale WSNs although the proposed RRSS method can improve its ability in solving the recovery problem with small size data.
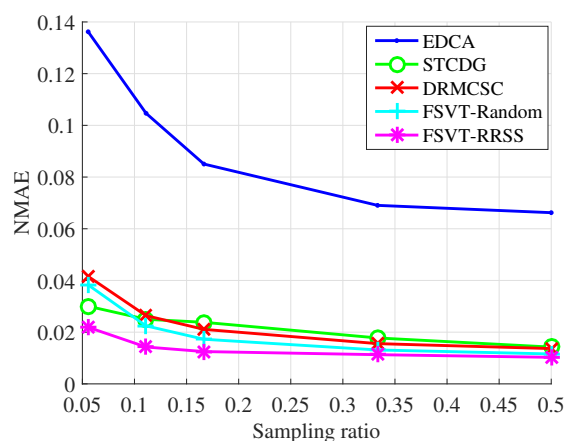
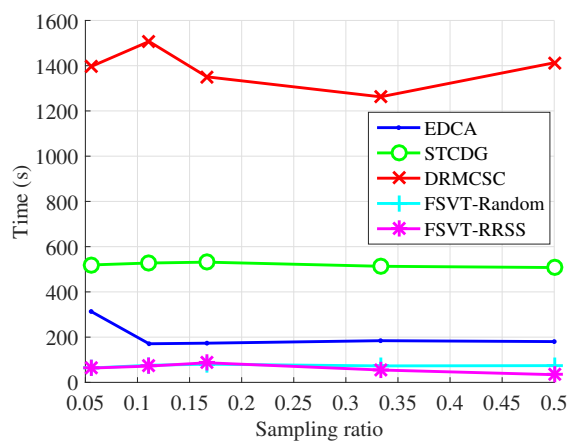**Figure 5.** NMAE over sampling ratio for the temperature from Intel Berkeley Research Lab.

Figure 6. Running time over sampling ratio for the temperature from Intel Berkeley Research Lab.

It is seen in Figures 6 and 8 that there is little change in running time of a method with increasing sampling ratios. We think this is because the iteration number of each algorithm reaches the maximum, which is set as 1000 for all methods. The stopping criterion has been described above. Because the time length of each iteration is close, the total running time of an algorithm to recover the original data is close with different sampling ratios.
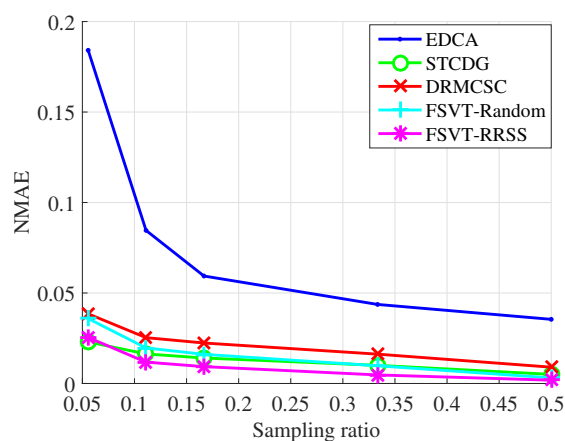
Figure 7. NMAE over sampling ratio for the temperature from Data-Sensing Lab.
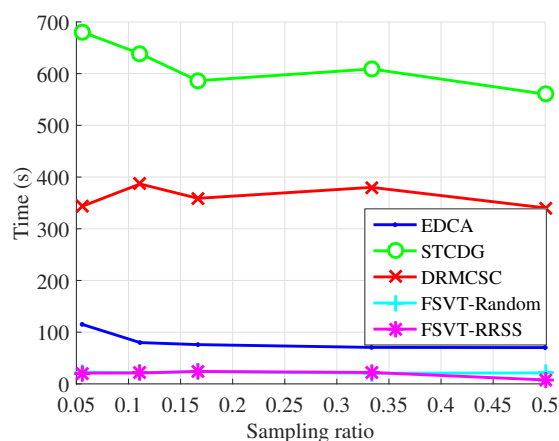
Figure 8. Running time over sampling ratio for the temperature from Data-Sensing Lab.

Moreover, the performance of the RRSS method for the reduction of energy consumption is also evaluated. Here we use the network lifetime as the metric. The power consumption model adopted in our study and the definition of the network lifetime are similar to those in previous works [7,8]. The lifetime of a network is in accordance with the time slot of the first node running out of its energy. The energy consumption is simplified as the packages of transmitting data. To show the performance of methods more intuitively, we regard the number of data package that each sensor node senses in a time slot and transmits in the transmission period as one. In addition, we set the death threshold as $N \times p$. It is the purpose of the sampling methods that the network can live during the whole time slots and the WSN will not break ahead of time. However, when using the traditional random sampling method, some nodes die ahead of time. The proposed sampling method makes every sensor sense and transmit data at the same times in each round, i.e., the network reaches load balance. Because of this advantage, the network can live as expected with the RRSS method as shown in Figures 9 and 10.
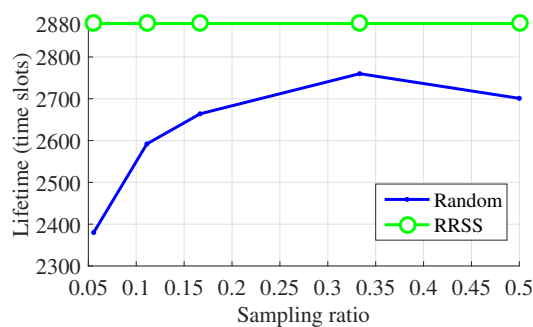

Figure 9. Lifetime over sampling ratio for WSNs from Intel Berkeley Research Lab.
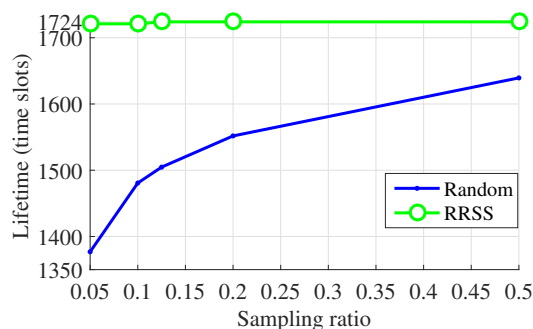

Figure 10. Lifetime over sampling ratio for WSNs from Data-Sensing Lab.

Figure 11 shows the accelerated speed of the FSVT method compared with the original SVT method. The reconstruction error of FSVT method is close to the minimum when the iteration number is 100 while that is more than 200 for the original SVT method, i.e., the FSVT algorithm can be twice faster than the original SVT algorithm, so the FSVT algorithm is used here.
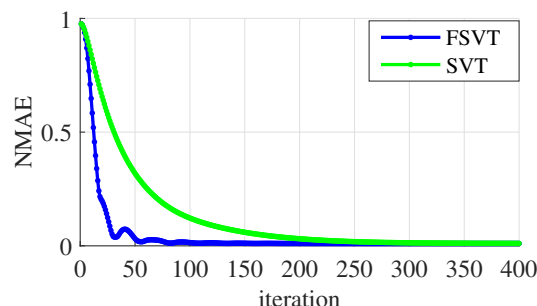

Figure 11. NMAE over iteration time between FSVT and SVT.

6. Conclusions

In this paper, we propose a novel data-collection method based on MC for large-scale WSNs. We design a RRSS method, in which sensor nodes are selected randomly and uniformly and are in the sleep mode most of the time to greatly reduce the energy consumption. From the samples obtained by the proposed sampling method, a corresponding FSVT algorithm is given to recover the data accurately and fast. The experiments are conducted with two real-world data sets and the simulation results show that the proposed sampling method can enhance the corresponding algorithm. Moreover, the proposed data-collection method outperforms other methods in terms of the reconstruction accuracy and rate. Our future work can be conducted form two aspects: combining our sampling method with other data gathering methods to further reduce the energy consumption and dealing with robust MC problem of WSNs with our method.

References

1. Rawat, P.; Singh, K.D.; Chaouchi, H.; Bonnin, J.M. Wireless sensor networks: A survey on recent developments and potential synergies. *J. Supercomput.* **2014**, *68*, 1–48. [CrossRef]
2. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]
3. Luo, C.; Wu, F.; Sun, J.; Chen, C.W. Compressive data gathering for large-scale wireless sensor networks. In Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, Beijing, China, 20–25 September 2009; pp. 145–156.
4. He, J.; Sun, G.; Li, Z.; Zhang, Y. Compressive data gathering with low-rank constraints for wireless sensor networks. *Signal Process.* **2017**, *131*, 73–76. [CrossRef]
5. Chen, J.; Jia, J.; Deng, Y.; Wang, X.; Aghvami, A.H. Adaptive Compressive Sensing and Data Recovery for Periodical Monitoring Wireless Sensor Networks. *Sensors* **2018**, *18*, 3369. [CrossRef] [PubMed]
6. Candès, E.J.; Recht, B. Exact matrix completion via convex optimization. *Found. Comput. Math.* **2009**, *9*, 717. [CrossRef]
7. Cheng, J.; Jiang, H.; Ma, X.; Liu, L.; Qian, L.; Tian, C.; Liu, W. Efficient data collection with sampling in WSNs: Making use of matrix completion techniques. In Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM 2010), Miami, FL, USA, 6–10 December 2010; pp. 1–5.

8. Cheng, J.; Ye, Q.; Jiang, H.; Wang, D.; Wang, C. STCDG: An efficient data gathering algorithm based on matrix completion for wireless sensor networks. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 850–861. [CrossRef]

9. He, J.; Sun, G.; Zhang, Y.; Wang, Z. Data Recovery in Wireless Sensor Networks With Joint Matrix Completion and Sparsity Constraints. *IEEE Commun. Lett.* **2015**, *19*, 2230–2233. [CrossRef]

10. Cai, J.F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [CrossRef]

11. Xie, K.; Wang, L.; Wang, X.; Xie, G.; Wen, J. Low cost and high accuracy data gathering in WSNs with matrix completion. *IEEE Trans. Mob. Comput.* **2018**, *17*, 1595–1608. [CrossRef]

12. Xie, K.; Ning, X.; Wang, X.; Xie, D.; Cao, J.; Xie, G.; Wen, J. Recover corrupted data in sensor networks: A matrix completion solution. *IEEE Trans. Mob. Comput.* **2017**, *16*, 1434–1448. [CrossRef]

13. Akyildiz, I.F.; Kasimoglu, I.H. Wireless sensor and actor networks: Research challenges. *Ad Hoc Netw.* **2004**, *2*, 351–367. [CrossRef]

14. Primeau, N.; Falcon, R.; Abielmona, R.; Petriu, E.M. A Review of Computational Intelligence Techniques in Wireless Sensor and Actuator Networks. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 2822–2854. [CrossRef]

15. Curiac, D.I. Towards wireless sensor, actuator and robot networks: Conceptual framework, challenges and perspectives. *J. Netw. Comput. Appl.* **2016**, *63*, 14–23. [CrossRef]

16. Ando, T. Majorization relations for Hadamard products. *Linear Algebra Its Appl.* **1995**, *223–224*, 57–64. [CrossRef]

17. Vuran, M.C.; Akan, Ö.B.; Akyildiz, I.F. Spatio-temporal correlation: Theory and applications for wireless sensor networks. *Comput. Netw.* **2004**, *45*, 245–259. [CrossRef]

18. Natarajan, B.K. Sparse approximate solutions to linear systems. *SIAM J. Comput.* **1995**, *24*, 227–234. [CrossRef]

19. Golub, G.H.; Hansen, P.C.; O'Leary, D.P. Tikhonov regularization and total least squares. *SIAM J. Matrix Anal. Appl.* **1999**, *21*, 185–194. [CrossRef]

20. Mohan, K.; Fazel, M. Iterative reweighted algorithms for matrix rank minimization. *J. Mach. Learn. Res.* **2012**, *13*, 3441–3473.

21. Hu, Y.; Zhang, D.; Ye, J.; Li, X.; He, X. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2117–2130. [CrossRef] [PubMed]

22. Lu, C.; Tang, J.; Yan, S.; Lin, Z. Generalized nonconvex nonsmooth low-rank minimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 4130–4137.

23. Sun, T.; Jiang, H.; Cheng, L. Convergence of proximal iteratively reweighted nuclear norm algorithm for image processing. *IEEE Trans. Image Process.* **2017**, *26*, 5632–5644. [CrossRef] [PubMed]

24. Combettes, P.L.; Wajs, V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168–1200. [CrossRef]

25. Parikh, N.; Boyd, S. Proximal algorithms. *Found. Trends Optim.* **2014**, *1*, 127–239. [CrossRef]

26. Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 87.

27. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [CrossRef]

28. O'donoghue, B.; Candes, E. Adaptive restart for accelerated gradient schemes. *Found. Comput. Math.* **2015**, *15*, 715–732. [CrossRef]

29. Becker, S.R.; Candès, E.J.; Grant, M.C. Templates for convex cone problems with applications to sparse signal recovery. *Math. Program. Comput.* **2011**, *3*, 165. [CrossRef]

30. Toh, K.C.; Yun, S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pac. J. Optim.* **2010**, *6*, 15.

31. Sun, T.; Cheng, L. Reweighted fast iterative shrinkage thresholding algorithm with restarts for l_1-l_1 minimisation. *IET Signal Process.* **2016**, *10*, 28–36. [CrossRef]

32. Sun, T.; Barrio, R.; Jiang, H.; Cheng, L. Convergence rates of accelerated proximal gradient algorithms under independent noise. *Numer. Algorithms* **2018**, 1–24. [CrossRef]

33. Piao, X.; Hu, Y.; Sun, Y.; Yin, B.; Gao, J. Correlated spatio-temporal data collection in wireless sensor networks based on low rank matrix approximation and optimized node sampling. *Sensors* **2014**, *14*, 23137–23158. [CrossRef] [PubMed]

34. Sensor Mote Data from Data Sensing Lab. Available online: http://datasensinglab.com/data/ (accessed on 10 October 2018).

35. Sensor Data from Intel Berkeley Research Lab. Available online: http://db.csail.mit.edu/labdata/labdata.html (accessed on 15 May 2018).