Sparse Random Projection Algorithm Based on Minimum Energy Tree in Wireless Sensor Network

Chong Wang, Xia Zhang, and Ou Li
National Digital Switching System Engineering and Technological Research Center, Zhengzhou, China
Email: {wc38022122, zziou}@126.com; zhangxiaatzz@sina.com

Abstract — For the energy-constraint of nodes in wireless sensor networks, a sparse random projection algorithm based on minimum energy consumption tree (SRP-MET) is proposed in this paper. Compressive Sensing is applied to data compression. And it minimizes the number of source node by means of sparse random projection. The relay nodes is chosen on the principle of minimum energy consumption, and the spanning tree routing is created based on the idea of centralized greedy increasing tree to match the projection matrix. Simulation results show that, on the condition of reliable communication links, the proposed algorithm not reconstruct the original data accurately, but also effectively reduce the energy consumption and prolong the network lifetime by balance the network load.

Index Terms—Wireless sensor network, compressive sensing, sparse random projection

I. INTRODUCTION

Wireless Sensor Network (WSN) [1] is a kind of energy-constraint network. It is an important factor to increase the lifetime of WSN by minimizing the energy consumption at sensor networks. In WSN, most of the power of data collection is consumed in transmission and forwarding. To maximize the network lifetime, there are two problems to be solved in a large-scale network: (1) a large number of redundant data to be transported; (2) the unbalanced of energy consumption throughout the networks. Data aggregation, aims at reducing the number of the data to be transported, is one of the methods to maximize the lifetime of WSNs.

In this paper, we propose a method by using the Compressive Sensing (Compressive Sensing, CS) [2] as the data aggregation algorithm. The CS theory was proposed for signal processing for the first time. In CS, if a signal can be transferred into a sparse domain, it can be successfully delivered and recovered with far fewer rates than the Nyquist rate through CS. The CS theory is introduced into WSN for a single-hop network in [3]. The original data vector of sensors can be recovered accurately at the sink with few sample measurements by use of the CS theory. With the technique, it appears that instead of transmitting the original sensors reading data, a few compressed measurements are sent to the sink and the energy consumption is reduced.

In recent literature, data gathering based on CS in wireless sensor networks has been getting more and more attention. In [4], J. Haupt showed that CS could be used in wireless sensor networks and could save the energy consumption by reducing the amount of communication. To extend the CS theory from processing one single signal to multiple signals, Distributed Compressive Sensing (DCS) theory is presented in [5], which rests on a new concept of joint sparsity of a signal ensemble. The authors proposed three kinds of Joint Sparse Models (JSM) by analyzing the intra and inter signal correlation structure, and gave the algorithms for joint recovery of multiple signals from incoherent projections. The application of DCS in WSN reduced the number of measurements per sensor required for accurate reconstruction by modeling the sensor data with the three JSM models.

However, it is necessary to minimize the energy consumption which joint optimize the route protocol and CS measurements matrix in real network [6]. There are two main kind of compressive data gathering method right now: clustering route [7] and spanning tree route [8]. Each of them has its benefits.

The clustering data gathering algorithm, as proposed in [9], each node generates its measurement matrix independently, and transmit its observation to its cluster head and the cluster head get the weight sum of the data in the cluster. At last the cluster heads send their measurements to the sink, and the sink could reconstruct the measurement matrix of the whole network and recover the original sensor readings accurately. The author in [9] presented a method to keep all cluster heads uniformly distributed. Based on the clustering route, a sparsest random scheduling proposed in [10] to minimize the sensor number for each projection. It redesigns the measurement matrix and the sparse domain matrix, so that it could reconstruct the data with high accuracy while only use one sensor node for each projection.

Although the clustering route is easy to be distributed built, as proposed in [11]-[13], the communication energy consumption is not optimal. The spanning tree methods aim at minimizing the transmission distance to save the energy cost. The Compressive Data Gathering (CDG) algorithm is proposed in [14] in multi-hop networks to reduce global scale communication cost for the first time. In CDG, the nodes in the network were organized in a Minimum Spanning Tree (MST) to minimize the transmission distance. And each node in the network
receives and sends the same number of data. In this way, an energy-consumption balance is achieved in the network.

Compared with traditional data collection (non-CS), the nodes far from the sink transmit more packets in CDG. So, a hybrid-CS algorithm was proposed in [15], which applies CS only to the nodes that are overloaded. Hence, the hybrid-CS cost less energy than CDG. To further reduce the global network communication cost, some algorithm was proposed, such as in [16] and [17]. But still, all the nodes must take part in data gathering for each measurement, which is a waste of energy.

To minimize the amount of nodes in each projection, the authors in [18] proposed the sparse random projection compressive sensing algorithm (SRP) which made it possible to reconstruct the original data with only a few nodes in each projection. The SRP selects only a few of nodes randomly for each projection and each selected node gathers a weighed sum to get a measurement. So that, the number of the nodes participated in data gathering is reduced to a very low quantity. However, it requires a lot of relay nodes to transmit the measurement to the sink. So the algorithm is difficult to minimize the network communication energy cost. To match the route with the measurement matrix, there are some methods of data gathering, such as in [8] and [19]. But for the nodes are selected at random for each projection, it is still hard to get a best route to minimize the energy cost because the nodes are selected at random for each projection. To solve this problem, Ebrahimi D in [20] presented a Minimum Spanning Tree Projection (MSTP) for data gathering. It constructed independent forwarding tree for each projection to ensure fewer transmission. In [21], the author considered routing strategy into the designing of the projection matrix, and proposed a compressed sensing algorithm based on data fusion tree, which reduced the number of the measurement.

However, all of the researches above select the relay nodes by the principle of minimum the transmission distance, which leads to too many nodes that are used to achieve the projection. The minimum of the transmission distance doesn’t equal to the minimum of the energy consumption. The forward energy cost of the relay nodes cannot be ignored, especially in SRP data collection.

In this paper, we propose a sparse random projection algorithm based on minimum energy tree (SRP-MET), which aims at minimizing the energy consumption by using SRP for data gathering. We take the analysis of the energy consumption in [10] as lessons and propose a method of relay nodes selection to get a trade-off between the transmission distance and the number of relay nodes. Then we use the method of the greed increasing tree to construct the data transmission route. The SRP is used to reduce further the number of nodes participated into the projection and distribute the energy consumption load more evenly throughout the whole network.

The rest of the paper is organized as follows. In Section II, we introduce the theory of CS and sparse random projection data gathering briefly. Our methodology is described in Section III, including the network model, the energy consumption model, the relay nodes selection algorithm and the minimum energy trees data gathering algorithm. The simulation results and analysis are shown in Section IV. We give our concluding remarks in Section V.

II. BACKGROUND

In this section, we introduce the basic CS theory and the application of sparse random projection for data gathering in wireless sensor networks.

A. Compressive Sensing

According to the traditional theory of compressive sensing, the signal processing of CS can be divided into three parts: sparse representation, compression projection and signal reconstruction.

In WSN, the sensor readings \( x = [x_1, x_2, \ldots, x_n]^T \), where \( x \in \mathbb{R}^n \), are turned as one dimension vector. If \( x \) is sparse when transferred to a sparse domain (such as DCT, DWT), then \( x \) is compressible.

\[
x = \Psi \cdot \theta
\]

(1)

where \( \theta \) is a \( k \)-sparse vector representation of the signal \( x (k \ll n) \), and \( \Psi \) is the sparse domain.

During the data gathering, instead of sending the the original data \( x \) of size \( n \), \( y = [y_1, y_2, \ldots, y_m]^T \) of size \( m \) where \( m \ll n \) is sent to the sink. The observation value \( y \) is obtained by projection of \( x \) which multiplied by a measurement matrix \( \Phi = [\phi_{ij}]_{mn} \). The compression projection of \( x \) can be expressed as

\[
y = \Phi \cdot x
\]

(2)

In CS theory, if \( \Phi \) and \( \Psi \) satisfy the Restricted Isometry Property (RIP), as proposed in [22], and the number of measurement \( m \) satisfy the condition

\[
m \geq Ck \log n
\]

(3)

Then the original data can be recovered perfectly. However, it is a kind of \( l_0 \)-norm minimization process to reconstruct the originally sparse signal \( \| \theta \|_0 \). And it is proved an N-P hard problem which is difficult to solve in [23]. If the sensing matrix \( \Theta = \Phi \cdot \Psi \) satisfied 2K-order RIP condition, the signal reconstruction process can be transferred into a problem of \( l_1 \)-norm minimization as

\[
\hat{\theta} = \text{argmin} \| \theta \|_1 \text{ s.t. } y = \Theta \theta \quad \Theta = \Phi \cdot \Psi
\]

(4)

where \( \hat{\theta} \) is the reconstructed \( k \)-sparse vector \( \theta \), and the reconstruction data of \( \hat{x} \) is obtained by

\[
\hat{x} = \Psi \cdot \hat{\theta}
\]

(5)

There are already many existing algorithm to recover the originally data such as OMP in [24].
B. Sparse Random Projection Data Gathering

To reduce the amount of sensor for each projection, the author in [18] proposed the Sparse Random Projection (SRP) as the measurement matrix. The elements in SRP matrix are generated as follows:

$$\phi_{ij} = \sqrt{s} \begin{cases} 
1, & p = \frac{1}{2s} \\
0, & p = 1 - \frac{1}{s} \\
-1, & p = \frac{1}{2s} 
\end{cases}$$  \hspace{1cm} (6)$$

where, $\phi_{ij}$ is the element in the matrix, and the parameter $s$ characterizes the sparse degree of the matrix.

If $s = 1$, all of the elements in $\Phi$ are non-zero and the matrix is a dense matrix. While in case $1/s = \log(N)/N$, then there are only $n/s = \log(N)$ non-zero elements on average in each row. Thus, only $n/s = \log(N)$ nodes take part in the data gathering for each projection. It is proved in [18] that the data can be recovered with high accuracy and the number of measurements does not increase too much on the condition of using SRP matrix. Our simulations in Section V also show that the SRP matrix can be used as the measurement matrix in WSN data gathering.

III. METHODOLOGY

In this section, we first give the network model and energy consumption model. By analyzing the energy model, we propose a method of selecting the relay nodes during data transmission. At last, we compare our algorithm of sparse random projection data gathering based on minimum energy tree, and give the analysis of the performance of the algorithm.

A. Network Model

A WSN which aims at environment monitoring such as measuring temperature is assumed in this paper. Assuming that the network contains $n$ sensor nodes and a sink node, CS theory is used to gathering sensor data. The parameters of the network are set as follows:

- All of the sensor nodes are randomly distributed, the sink node is located in the center of the monitoring area;
- The monitoring area is square with the boundary of $a = 200m$ ;
- The sensor nodes are of the same type, and can modify the transmit power dynamically;
- The sink node can get the information and location of the whole network;
- An energy-efficient MAC protocol is used, so retransmission caused by the MAC is not considered in this work.

B. Energy Consumption Model

During CS data gathering, the sensors only need to operate simple linear-calculation, which consume little energy compared with the data communication. So we ignore the energy consumption of data processing. The energy cost of communication is mainly considered. We take the energy consumption model as [25]:

$$E_{TX}(L,d) = (E_{elec} + \epsilon_{amp} \times d^s) \times L$$  \hspace{1cm} (7)$$

$$E_{Rx}(L) = E_{elec} \times L$$  \hspace{1cm} (8)$$

where $E_{TX}(L,d)$ represents the energy consumption for sending a $L$-bit length data over the distance $d$, and $E_{Rx}(L)$ represents the energy consumption for receiving a $L$-bit length data. The amplifier energy is $\epsilon_{amp}$ and the electronics energy consumption is $E_{elec}$. The element $n$ presents the path loss exponent, generally $n = 2$ which means the free space path loss.

In the energy model, the element $\epsilon_{amp}$ and $E_{elec}$ are constant, so if the length of data $L$ is fixed, the smaller of the transmission distance $d$, the less the node energy consumption. But for the energy consumption of the whole network, the more relay nodes are involved, the more energy is expended as electronics consumption.

C. Selection of Relay Nodes

By analyzing the energy model, the author in [26] gave the method to select the relay node in ideal network which we can always find the relay node at the ideal position.

**Definition 1:** Given the distance $d$ and the optimal number of the relays $K - 1$, then the energy consumption is minimized when all the hop distances are equal to $d/K$.

**Definition 2:** The optimal number of hops $K_{opt}$ is always one of the two followings:

$$K_{opt} = \left\lfloor \frac{d}{d_{char}} \right\rfloor$$  \hspace{1cm} (9)$$

where $d_{char}$ is the characteristic distance, which is independent from $d$. The element $d_{char}$ depends on the energy model parameter, and is given by

$$d_{char} = \sqrt[1-n]{\frac{E_{elec}}{\epsilon_{amp}}}$$  \hspace{1cm} (10)$$

However, it is impossible to achieve the ideal path in real network. A minimum energy consumption routing algorithm based on geographical location information (GLB-DMECR) is proposed in [27]. But it needs the node to get the information of neighbor node, which involved too much control energy consumption.

In this paper, we adopt centralized routing generation method to build the route at the sink node. For sink get the information of the whole network, the relay nodes are
D. The Minimum Energy Consumption Tree Method

Based on the algorithm above, we present the minimum energy consumption tree method. In SRP data gathering, we adopt the centralized strategy to generate the measurement matrix at the sink node, so that the route can match the measurement matrix.

Define one measurement as \( m \) projections, by which the sink could recover the original. Our goal is to minimizing the energy consumption for one single measurement. For the measurement matrix is generated randomly in SRP, each projection is independent of each other. So our goal can be turned to minimize the energy consumption of the each projection. We construct independent tree for each projection.

The method of Greedy Incremental Tree (GIT) is used to construct multicast tree broadcast path. First, we get the initial tree as the minimum energy consumption path between the sink and the nearest sources node; then find the next nearest source node from the existing tree; after that add the node and its path to the tree into the existing minimum energy consumption tree; do this until the last source node so that the minimum energy consumption tree is accomplished. The measurement matrix is broadcasted to the network while constructing the route. The detail algorithm is as shown in Table II.

**Table I: Minimum Energy Consumption Relay Nodes Selection**

<table>
<thead>
<tr>
<th>Algorithm 1: Minimum Energy Consumption Relay Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> The source node ( S ) and the destination node ( D )</td>
</tr>
<tr>
<td>The feature distance ( K_{opt} )</td>
</tr>
<tr>
<td><strong>Output:</strong> The path from ( S ) to ( D )</td>
</tr>
<tr>
<td><strong>Begin</strong></td>
</tr>
<tr>
<td>1: get the ideal relay position ( C_i ), ( i=1,\ldots,K_{opt}-1 )</td>
</tr>
<tr>
<td>2: initialize the path set ( P=[S] )</td>
</tr>
<tr>
<td>3: initialize the candidate node set ( T=[i</td>
</tr>
<tr>
<td>4: for ( i=1 ) to ( i=K_{opt}-1 )</td>
</tr>
<tr>
<td>5: find the nearest node ( n_i ) from the ideal position ( C_i )</td>
</tr>
<tr>
<td>6: if ( (n_i\in T &amp; &amp; d(n_i,D)&lt;d(n_i,D)) )</td>
</tr>
<tr>
<td>7: add ( n_i ) to the path set ( P )</td>
</tr>
<tr>
<td>8: delete ( n_i ) from the candidate set ( T )</td>
</tr>
<tr>
<td>9: end if</td>
</tr>
<tr>
<td>10: end for</td>
</tr>
<tr>
<td><strong>End</strong></td>
</tr>
</tbody>
</table>

**Table II: Minimum Energy Consumption Tree Construction**

<table>
<thead>
<tr>
<th>Algorithm 2: Minimum Energy Consumption Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> The ( i )th row of the measurement matrix ( \Phi_i=(\phi_{i1},\phi_{i2},\ldots,\phi_{iv}) )</td>
</tr>
<tr>
<td><strong>Output:</strong> The tree node set of the ( i )th projection</td>
</tr>
<tr>
<td><strong>Begin</strong></td>
</tr>
<tr>
<td>1: initialize the source node set ( S_i=[j</td>
</tr>
<tr>
<td>2: initialize the tree node set ( T_i={\text{sink}} )</td>
</tr>
<tr>
<td>3: while ( S_i\neq \Phi )</td>
</tr>
<tr>
<td>4: find the nearest node ( n_i ) from ( S_i ) to ( T_i )</td>
</tr>
<tr>
<td>5: get the path ( P_i ) from ( n_i ) to ( T_i )</td>
</tr>
<tr>
<td>6: add ( P_i ) to ( T_i )</td>
</tr>
<tr>
<td>7: delete ( n_i ) from ( S_i )</td>
</tr>
<tr>
<td>8: end while</td>
</tr>
<tr>
<td><strong>End</strong></td>
</tr>
</tbody>
</table>

During data gathering, sink broadcast the measurement matrix to each node through the route path. Then the nodes which correspond to non-zero elements in the matrix send the weight sum of its reading and received data to their father node. By this method, the sink get one measurement finally.

![Fig. 1. Schematic of one projection with SPT algorithm.](image1)

![Fig. 2. Schematic of one projection with MET algorithm.](image2)
algorithm and the latter is a schematic of the MET algorithm. We can see that on the same condition, the MET algorithm needs fewer relay nodes. According to the calculation of the energy model parameters in [25] \( E_{\text{elec}} = 50nJ/\text{bit} \) and \( E_{\text{amp}} = 10pJ/\text{bit/m}^2 \), we can see that the electronics energy consumption \( E_{\text{elec}} \) is much larger than the amplifier energy \( E_{\text{amp}} \). Although the SPT algorithm gets the shortest path, it wastes too much energy on electronics energy consumption and its energy consumption is not minimal.

Of course, there are still some limitations. For we generate our route and measurement matrix centrally, if the communication environment was very poor, the performance of the algorithm decreased very quickly.

IV. SAMPLING PERFORMANCE

In this section, we give our simulation and analysis of the algorithm which implements with MATLAB R2012b. The monitoring area is square with the boundary of 200*200.

A. The Accuracy of Reconstruction Data

As is shown in Fig. 3 is the comparison of normalized reconstruction error of SRP and dense matrix for different data. The data is generated by two-dimensional Gaussian distribution with different sparsity under DCT domain. The sparsity of the SRP matrix is \( n/\log(n) \), which only needs \( n/\log(n) \) node for each projection on average. Compared with the dense measurement matrix, the sparse random matrix needs a little more measurements to recover the original data. But with the use of SRP matrix, we can save a lot of nodes for each projection.

B. Comparison of Participated Node Number

To compared, Fig. 4 shows the comparison of the number of transmit packages in one measurement between non-CS, Hybrid-CS and SRP-CS on the condition of minimum spanning tree route. It is obviously that with the use of CS theory, the number of transmit packages decrease significantly. But compared with hybrid-CS, the packages in SPT-CS could not decrease too much. This is because too many relay nodes are involved to transmit data. It is necessary to optimize the route for SRP-CS data gathering.

C. Comparison of Energy Consumption with Different CS Data Gathering Algorithm

To evaluate energy consumption performance of the SRP-MET algorithm, we compared it with the existing data gathering algorithm: Hybrid-CS proposed in [15]
and SRP-SPT algorithm which uses shortest path spanning tree as the route path for sparse random projection data gathering.

The parameters of the simulation are set as follows: the number of the nodes in the network is set as 400, the energy of each node is $1J$, and the length of each packet is 1024bit, the data compression ratio is $\rho = M/N = 0.25$, and the energy model parameters is set as proposed in [25] $E_{\text{elec}} = 50\mu J/\text{bit}$ and $E_{\text{amp}} = 10\mu J/\text{bit}$. We can see that compared with hybrid-CS which used dense measurement matrix, the SRP data gathering method cost less energy. This is because only a few number of nodes are participated in each projection. Compared to the SRP-SPT algorithm, the proposed SRP-MET algorithm could save more energy consumption. This is because the SRP-MET algorithm is a method of joint optimization of the relay node number and the transmission distance to achieve the minimum energy consumption of the whole network.

**Fig. 6.** Energy consumption of the network, compared with hybrid-CS and SRP-SPT algorithm.

**Fig. 7.** Number of dead nodes in the network, compared with hybrid-CS and SRP-SPT algorithm.

**Fig. 8.** Lifetime of the network, compared with hybrid-CS and SRP-SPT algorithm.

In this paper, we proposed a sparse random projection algorithm based on minimum energy consumption tree. We adopted sparse random matrix as the measurement matrix to reduce the number of the source node. By analyzing the energy consumption model and the practicality of the network, centralized strategy was used to get the minimum energy consumption relay nodes. Then, we presented the SPT-MET data gathering algorithm. Compared with the existing SPT algorithm, the proposed algorithm made a trade-off between the number of the participated nodes and the transmit distance to optimize the energy consumption. The simulation showed that it reduced the energy consumption and increase the life time of the whole network dramatically on the condition of a reliable communication link. We can see that for the nodes participated in the data gathering are random in SRP data gathering, it could not balance the energy consumption perfectly. But it is obvious that the SRP-MET algorithm could prolong the lifetime of the network. This is because for each node, although the transmission distance increase, the probability of being chosen as the relay nodes decrease. Which saves a considerable electronics energy consumption, and could live much longer.

V. CONCLUSION

In this paper, we proposed a sparse random projection algorithm based on minimum energy consumption tree. We adopted sparse random matrix as the measurement matrix to reduce the number of the source node. By analyzing the energy consumption model and the practicality of the network, centralized strategy was used to get the minimum energy consumption relay nodes. Then, we presented the SPT-MET data gathering algorithm. Compared with the existing SPT algorithm, the proposed algorithm made a trade-off between the number of the participated nodes and the transmit distance to optimize the energy consumption. The simulation showed that it reduced the energy consumption and increase the life time of the whole network dramatically on the condition of a reliable communication.

For we generate our route and measurement matrix centrally, if the communication environment was very poor, it was necessary to rebuild the whole route to guarantee the accuracy of the reconstruction data. Thus,
the performance of the algorithm decreased very quickly. For further work, we will focus on the unreliable communication link data gathering. And the spanning tree method still needs too many relay nodes, which leads high probability of packet loss. Thus, distributed clustering is more suitable for the unreliable link situation. We will optimize the clustering method according to the characteristics of the CS data gathering to minimize the network energy consumption.

ACKNOWLEDGMENT

This work was supported by National Science and Technology Major Projects of China under grant No. (2014ZX03006003). We are grateful to the anonymous reviewers for their helpful comments.

REFERENCES


Chong Wang was born in Shandong Province, China, in 1989, master candidate. He received the B.S. degree from the Beihang University (BUAA), Beijing, in 2012 in electronic information engineering. He is currently pursuing the M.S. degree with the Department of Communication and Information System, NDSC. His research interests include wireless sensor networks, routing protocol, and communication technology.

Xia Zhang was born in Shandong Province, China, in 1979, Ph. D. supervisor. Her research interests include computer network, wireless sensor networks, and information operations.

Ou Li was born in Henan Province, China, in 1961, professor, Ph. D. supervisor. His research interests include wireless sensor networks, cognitive radio networks, and Ad-hoc networks.

©2015 Journal of Communications 746