

# Improved Compressed Network Coding Scheme for Energy-Efficient Data Communication in Wireless Sensor Networks

Ying Zhou, Lihua Yang\*, and Longxiang Yang

Jiangsu Key Laboratory of Wireless Communication, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210003, China [e-mail: 2013010202@njupt.edu.cn, yanglh@njupt.edu.cn, yanglx@njupt.edu.cn]

\*Corresponding author: Lihua Yang

*Received October 21, 2016; revised January 21, 2017; accepted February 25, 2017; published June 30, 2017*

---

## Abstract

An improved energy-efficient compressed network coding method is proposed for the data communication in the wireless sensor networks (WSNs). In the method, the compressed sensing and network coding are jointly used to improve the energy efficiency, and the two-hop neighbor information is employed to choose the next hop to further reduce the number of the transmissions. Moreover, a new packet format is designed to facilitate the intermediate node selection. To theoretically verify the efficiency of the proposed method, the expressions for the number of the transmissions and receptions are derived. Simulation results show that, the proposed method has higher energy efficiency compared with the available schemes, and it only requires a few packets to reconstruct measurements with reasonable quality.

---

**Keywords:** Compressed sensing, network coding, compressed network coding, wireless sensor networks, energy-efficient

---

This work was supported by the National Basic Research Program of China (973 Program) (Grant no. 2013CB329104), the National Natural Science Foundations of China (Grants no. 61372124, no. 61427801, and no. 61401232), and Natural Science Foundation of the Jiangsu Higher Education Institutions of China (14KJB510026).

## 1.Introduction

Wireless Sensor Networks (WSNs), typically consisting of a vast number of densely deployed and collaborative battery-powered sensors, have been widely used for the various monitoring and measuring purposes in many applications, such as environmental, industrial and military [1]. The energy-limited sensors of the networks, once deployed, may be inaccessible and hard to be replaced, especially in harsh, hostile and catastrophic environments. Therefore, the energy is the bottleneck problem restricting the development of WSNs. In the WSNs, the energy consumption mainly consists of three stages: sensing, data processing and data delivering, and the energy consumed by the data delivering is dominated. Therefore, delivering data with high energy efficiency in the network is the key to solve the energy constrain for the WSNs.

Since the observation regions of the different sensors are overlapping and the data observed at a sensor over time change slowly, the sensed data from a WSN are either spatially or temporally correlated [2]. Therefore, data compression techniques which exploit the inherent correlation to reduce the amount of data transferred in the WSNs have been extensively studied [3], [4]. However, these techniques may not be efficient due to the deficiency of global correlation knowledge between the source data. Recently, Compressed Sensing (CS) [5], as an advanced sampling theory, provides a new data compression solution without the global knowledge, and it can reduce the number of the data transfer since a few measurements are needed for reconstructing the original data in the sink. In [6], a framework employed CS for the distributed data compression in the WSNs is developed, which improves the energy efficiency by exploiting the intra-signal and inter-signal correlations. However, there is no practical communication scheme to realize the framework in [6]. To solve the problem, some actual multi-hop routing topologies are presented for large-scale WSNs in [7], [8], while the interactions between data compression and routing are decoupled, which causes a low efficiency of the data processing.

Besides the data storage and forwarding capabilities, sensors in WSNs also have the function of calculation. Thus, Network Coding (NC) [9] has also been provided to reduce the node energy consumption and improve the network capacity during data delivering. Based on the theory of NC, the intermediate nodes merge the input packets by the simple algebraic operations [10] instead of just storing and forwarding packets along routing. Meanwhile, the introduction of NC in designing a communication scheme for the WSNs can overcome the dynamic nature of links and exploit the broadcast characteristic of WSNs [11]. However, until the feasibility of random linear network coding (RLNC) is verified in [12], NC has not been applied to the actual communication protocol design for the WSNs [13]. RLNC is that the intermediate nodes linearly combine the input data with random coefficients to generate the output packets, and the network codes can be constructed and efficiently transmitted to receivers in the distributed manner. Nevertheless, these schemes in [11-13] suffer from a 'all-or-nothing' constrained in the process of decoding, which will cause serious packet losses as the sink cannot get any information before receiving as many coded packets as the original packets.

Recently, to effectively reduce the energy consumption by the data transferred, the data communication schemes joint the CS and RLNC are discussed in [14-18]. Besides the less communication consumption, two additional advantages can be obtained for the coupled schemes: (a) CS removes the "all or nothing" constrained in NC and (b) NC provides a method

to perform in-network compression for CS. In [14], a RNC method coupled CS and RLNC in a natural fashion is firstly given, but it never gives a practical data communication scheme to realize the RNC. Literature [15] fills this gap and presents a method coupled CS and NC based on the similarity between the random projections in CS and the operations to generate outputs by RLNC in the intermediate nodes, but the header explosion problem has not been considered in the method. To solve the problem, some practical methods [16], [17] are presented. In [16], a Compressed Network Coding method in finite fields is provided where the RLNC is performed in a Galois Field (GF) at the intermediate nodes to prevent the size of payload from being increased. The authors in [17] provide a method called NetCompress to reduce the size of header in the packet by designing a novel encoding scheme using a sparse measurement matrix [17]. But these schemes in [14-16] only focus on reducing the number of data transmissions and ignore the number of data receptions, which have almost the same energy cost as data transmissions for the short-range communications in the WSNs [1]. Literature [18] gives a distributed data storage scheme based on the compressed network coding, where the reduction of the data receptions is considered by designing the reception condition of the packets in the intermediate nodes. However, all of the above schemes [14-18] have never considered the selection of the intermediate nodes, which can maximize the effect of one transmission and greatly reduce the communication consumption.

To solve the above problem and further improve the energy efficiency, we propose a distributed data storage scheme based on compressed network coding with high energy efficiency for the WSNs. By coupling the compressed sensing and network coding, and utilizing the two-hop neighbor information and a new designed packet format to choose the intermediate node in the coupled scheme, one can greatly improve the energy efficiency of the data communication in the WSNs.

The rest of this paper is organized as follows. The background and system model are presented in Section 2. In Section 3, we introduce our proposed protocol in detail and derive the expressions of the numbers of communications. Section 4 discusses the simulation results and conclusions are given in Section 5.

*Notation:*  $\Omega(i)$  denotes the set of one-hop neighbors of the node  $i$ , and  $\mathbf{A} \cap \mathbf{B}$  is the intersection of two sets.  $\setminus$  denotes the subtraction of two sets, i.e.  $\mathbf{A} \setminus \mathbf{B} = \mathbf{A} - \mathbf{A} \cap \mathbf{B}$ .  $|\mathbf{A}|$  means the number of the elements in set  $\mathbf{A}$ .  $d(a, b)$  represents the Euclidian distance of  $a$  and  $b$ , and  $\arcsin x$  denotes the inverse function of the sine function.  $C_n^m$  is the operation of choosing  $m$  nodes from  $n$  nodes.

## 2. Background and System Model

### 2.1 Compressed Sensing

Compressed Sensing indicates that only a small fraction of data projections is needed to reconstruct all the original data which contain many zero elements [2]. Assume a data vector  $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ , has a  $K$ -sparse representation under a transform basis  $\Psi$

$$\mathbf{x} = \Psi \boldsymbol{\theta}, \quad (1)$$

where  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_N]^T$  with  $K$  non-zero entries,  $K \ll N$ . Note that  $\Psi \in \mathbf{C}^{N \times N}$  can be the Discrete Cosine Transform (DCT) basis, or the Discrete Wavelet Transform (DWT) basis, etc. Under certain conditions, only  $M$  samplings,  $M = O(K \log N)$ , are enough to be used to

recover the original  $N$ -dimension signal  $\mathbf{x}$ , i.e.,

$$\begin{aligned}\mathbf{y} &= \mathbf{\Phi}\mathbf{x} \\ &= \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta},\end{aligned}\quad (2)$$

where  $\mathbf{y}=[y_1, y_2, \dots, y_M]^T$  is the measurement vector, and  $\mathbf{\Phi} \in \mathbf{R}^{M \times N}$  is the measurement matrix whose row vector should be strongly incoherent with the sparse basis  $\mathbf{\Psi}$ , or the sensing matrix  $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$  has to satisfy the restricted isometry property (RIP) [19], that is

$$(1-\delta_k)\|\boldsymbol{\theta}\|_2^2 \leq \|\mathbf{A}\boldsymbol{\theta}\|_2^2 \leq (1+\delta_k)\|\boldsymbol{\theta}\|_2^2, \quad (3)$$

where  $\delta_k \in (0,1)$ . The reconstruction process will be a solving of  $l_1$ -norm convex optimization problem

$$\min_{\boldsymbol{\theta} \in \mathbf{R}^N} \|\boldsymbol{\theta}\|_{l_1} \quad \text{s.t. } \mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta}. \quad (4)$$

Using the estimated  $\hat{\boldsymbol{\theta}}$ , one can obtain the reconstructed data  $\hat{\mathbf{x}}$ , i.e.,

$$\hat{\mathbf{x}} = \mathbf{\Psi}\hat{\boldsymbol{\theta}}. \quad (5)$$

## 2.2 System Model

Consider a WSN with  $N$  nodes deployed uniformly and randomly in an area  $S = 1 \times 1$ , where two of them can communicate if their Euclidian distance is less than the identical transmission range,  $r_t$ , and the communication is synchronized and slotted. The density of the network can be adjusted by varying the transmission range,  $r_t$ , with  $r_t^2 > (S \ln N) / \pi N$  for connectivity. Each sensor in the network can function as a source node, an intermediate node or a destination node. Distributed data storage is considered and a mobile sink is employed to collect the compressed data from any arbitrary  $M$  sensors,  $M \ll N$ , and then reconstruct the source data finally.

The summary of notations can be seen from the [Table 1](#). The compressed coefficient matrix  $\mathbf{\Phi}$  which meets the RIP condition is generated during the data delivering and it is sparse, with entries

$$\varphi_{i,j} = \begin{cases} +1 & \text{with probability } \frac{1}{2s} \\ 0 & \text{with probability } 1 - \frac{1}{s} \\ -1 & \text{with probability } \frac{1}{2s} \end{cases}$$

where  $s$  is a parameter for sparseness. The reasons for employing such a measurement matrix are that: 1) it is simple to implement by just randomly adding or subtracting packets together; and 2) the sparse measurement matrix reduces the size of packet header while it still yields good reconstruction. The mobile sink queries any arbitrary  $M$  nodes in the network and then obtains the compressed data vector  $\mathbf{y}$ , which coincides with the CS.

**Table 1.** Summary of Notations

Notation	Description	Field
$\mathbf{x}$	Source data vector	$\mathbf{R}^N$
$x_i$	Source data at node $i$	$\mathbf{R}$
$\Psi$	Orthonormal transform matrix	$\mathbf{C}^{N \times N}$
$\theta$	Sparse transformed coefficient vector	$\mathbf{R}^N$
$\Phi$	Compression coefficient /Measurement matrix	$\mathbf{R}^{M \times N}$
$\varphi_{i,i}$	Compression coefficient of $x_i$	$\mathbf{R}$
$\mathbf{y}$	Compressed data vector	$\mathbf{R}^M$

### 3. Proposed Method

In this section, the proposed method is firstly described in subsection 3.1, and the next hop selection principle is discussed in subsection 3.2. In subsection 3.3, the corresponding packet format is given. After that, the detailed procedures of the proposed method are presented in subsection 3.4, and finally, the expressions for the broadcasting probability of the selected nodes and the numbers of transmissions and receptions are derived in subsection 3.5 and 3.6.

#### 3.1 Overall Description of the Proposed Method

In the proposed compressed network coding method, CS and RLNC are jointed to reduce the numbers of the transmissions and receptions based on the compressibility of the sensed data. Moreover, the next hop selection principle and a new packet format are designed to control the condition of the transmission and further reduce the number of the transmissions.

Specifically, each node  $i$  firstly multiplies its reading  $x_i$  by a random coefficient  $\varphi_{i,i}$ , which is chosen from  $\{+1, -1\}$  with equal probability, and then each node stores the packet. The packet is composed of the coefficient  $\varphi_{i,i}$ , the member node ID  $i$ , the next hop node ID  $f$ , and the data  $\varphi_{i,i}x_i$ . Let a structure variable  $\mathbf{P}(i)$  denotes the packet, where  $\mathbf{P}(i).\mathbf{nex}=[f]$ ,  $\mathbf{P}(i).\mathbf{coe}=[\varphi_{i,i}]$ ,  $\mathbf{P}(i).\mathbf{mem}=[i]$ , and  $\mathbf{P}(i).\mathbf{dat}=\varphi_{i,i}x_i$ .

Subsequently, each node selects itself as a source node with probability  $p_0$  [8]. Each selected source  $i$  chooses a potential farthest node  $j$  as a transmitter for the next time interval by the next hop selection principle which is discussed in subsection 3.2, and the node  $i$  changes the value of  $\mathbf{P}(i).\mathbf{nex}$  with  $j$ , i.e.,  $\mathbf{P}(i).\mathbf{nex}=[j]$ . Afterwards, all source nodes broadcast their initial packets. There are two cases about the reception nodes of the transmitters should be considered.

**Case 1:** If the reception node  $j$  at the time  $q$  does not share any member node ID with the transmission node  $i$ , i.e.,

$$[\mathbf{P}(i).\mathbf{mem}] \cap [\mathbf{P}(j).\mathbf{mem}] = \emptyset, \quad (6)$$

then  $\mathbf{P}(i)$  is merged into  $\mathbf{P}(j)$ , with

$$\begin{aligned}\mathbf{P}(j).\text{mem} &= [\mathbf{P}(j).\text{mem}, \mathbf{P}(i).\text{mem}] \\ \mathbf{P}(j).\text{coe} &= [\mathbf{P}(j).\text{coe}, \mathbf{P}(i).\text{coe}] \\ \mathbf{P}(j).\text{dat} &= \mathbf{P}(j).\text{dat} + \mathbf{P}(i).\text{dat} .\end{aligned}\quad (7)$$

Next, the reception node  $j$  needs to verify that if it has been chosen as a next forwarder by the node  $i$ . if

$$\mathbf{P}(i).\text{nex} = j , \quad (8)$$

the reception node  $j$  will broadcast its packet at the time  $(q+1)$ . Meanwhile, the next node  $j$  needs to determine its next hop by the next hop selection principle in subsection 3.2 before broadcasting its packet.

**Case 2:** If the reception node  $j$  of the node  $i$  doesn't meet the condition in (6), it will not broadcast its packet at the time  $(q+1)$  even it has been chosen as the next hop by the node  $i$ , which means that the transmission of this path is end with the node  $i$ .

The above two cases indicate that only the reception nodes which meet the conditions in (6) and (8) simultaneously will broadcast the packets at the next time, in other words, the selected node at the time  $q$  broadcasts its packet with a probability  $p_{q+1}$  at the time  $(q+1)$ , where the value of  $p_{q+1}$  is discussed in subsection 3.5. Therefore, the transmission will end after the broadcast processing repeats a few times.

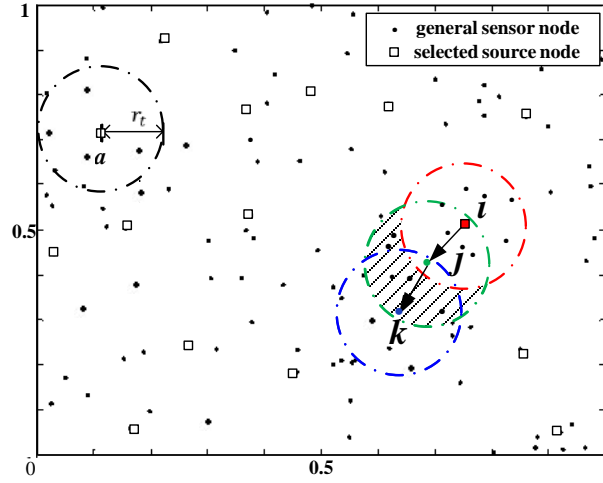
After broadcasting, the mobile sink queries any arbitrary  $M$  nodes to obtain  $M \times N$  measurement matrix  $\Phi$  and the  $M \times 1$  measurement vector  $\mathbf{y} = \Phi \mathbf{x}$ .

Finally, the sink or the data processing center reconstructs the sensor readings  $\mathbf{x}$  from  $\mathbf{y}$  with the CS recovery algorithm, such as basis pursuit (BP), and orthogonal matching pursuit (OMP).

### 3.2 Next Hop Selection Principle

In the available probability broadcasting based transmission scheme [18], each reception node  $j$  selects itself as a transmitter with a probability  $p_a$ , which means that the forwarding decision of the node  $j$  is blind without considering the distance to the transmission node  $i$ . If the node  $j$  is very close to  $i$ , most neighbors of the node  $j$  are the neighbors of the node  $i$ , in the case, the forwarding of the node  $j$  will be redundant. To solve this problem, the next hop selection principle is designed in the proposed method.

Consider the nodes in Fig. 1, where the node  $j$  will broadcast its packet at the time  $q$ , and the node  $i$  is the parent of the node  $j$ . Obviously, all nodes including node  $j$  in  $\Omega(i)$  have received a packet from the node  $i$  at the time  $(q-1)$ . Moreover, the packet  $\mathbf{P}(i)$  has been merged into  $\mathbf{P}(j)$  according to the condition in (6).



**Fig. 1.** Illustration of the next hop selection  
(the red square is the node  $i$ , the green dot is the node  $j$ , and the blue dot is the node  $k$ )

When the node  $j$  performs the broadcasting, all nodes in  $\Omega(j)$  will receive  $\mathbf{P}(j)$  which contains the information from  $\mathbf{P}(i)$ . However, according to the merging condition in (6), only the nodes in the shaded area in **Fig. 1** will merge the received packet  $\mathbf{P}(j)$  into their original packets potentially, and the others in  $\Omega(j)$  will drop the received packet. Therefore, the next hop  $k$  of the node  $j$  should be selected from the shaded area.

To avoid transmission redundancy, the size of the shaded area in **Fig. 1** should be the largest, in other words, the distance between  $j$  and  $i$  should be the farthest. Since to find the farthest neighbor  $j$  of the node  $i$  by GPS [20] is not practical, the two-hop neighbor information is employed in the proposed method. Here, the two-hop neighbor information means the neighbor information of the neighbor, and the information is easy to be obtained. If each node broadcasts its one-hop neighbor list, every node obtains the two-hop neighbor information.

Assume the node  $j$  needs to choose a next transmitter  $k$ , there are two cases needed to be considered:

**Case 1:** The node  $j$  is an intermediate node, which means it has a parent node. Assume that the node  $i$  is the parent of the node  $j$ . Since all nodes in  $\Omega(i)$  have received a packet from the node  $i$ , to ensure the largest new covered communication region, the node  $k$  should be selected from  $\Omega(j) \setminus \Omega(i)$  (the shaded area in **Fig. 1**). If the node  $k$  in  $\Omega(j) \setminus \Omega(i)$  has the maximum number of different neighbors with node  $i$ , that is,

$$\begin{aligned} k &= \arg \max_k |\Omega(k) \setminus \Omega(i)| \\ \text{s.t. } k &\in \Omega(j) \setminus \Omega(i), \end{aligned} \quad (9)$$

the node  $k$  is probably the farthest neighbor of the node  $j$  and it is the next hop of the node  $j$ .

**Case 2:** The node  $j$  is a source node, which has no parent node  $i$ . The next hop  $k$  of the node  $j$  will be selected from  $\Omega(j)$ , and if a neighbor  $k$  of the node  $j$  has the maximum number of different neighbors with the node  $j$ , that is,

$$\begin{aligned} k &= \arg \max_k |\Omega(k) \setminus \Omega(j)| \\ \text{s.t. } k &\in \Omega(j), \end{aligned} \quad (10)$$

the node  $k$  is probably the farthest neighbor and it is the next hop of the node  $j$ .

### 3.3 Designing of Packet Format

To conveniently broadcast the next hop node ID selected by the current node, a new packet format is designed in the proposed method. The new designed packet contains the packet header and the data, which can be seen in Fig. 2. The packet header consists of three parts. The first part contains a  $\lceil \log_2^N \rceil$ -bit slot for the next hop node ID, and the second and third ones store the sets of 2-bit coefficients  $\varphi_{i,j}$  (+1 or -1) and  $\lceil \log_2^N \rceil$ -bit node IDs respectively.

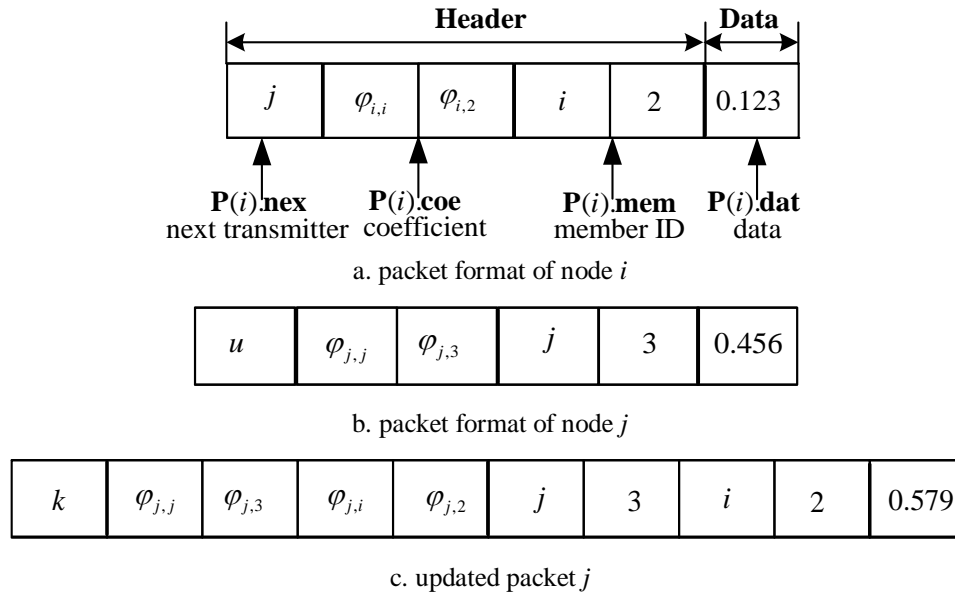


Fig. 2. Examples of packet format and pattern to update packet

An example of the structure of the updating packet is given in Fig. 2. Assume that the node  $j$  with original packet shown in Fig. 2(b) hears a packet  $\mathbf{P}(i)$ , and  $\mathbf{P}(i)$  shares no same member ID with  $\mathbf{P}(j)$ . Moreover, the node  $j$  verifies that it has been selected by the node  $i$ . By the merging condition in (6), the updated packet  $\mathbf{P}(j)$  is illustrated in Fig. 2(c).

### 3.4 Procedures of the proposed method

Table 2 shows the procedures of the proposed method.

Table 2.. Procedures of the Proposed Method

**Inputs:**  $\mathbf{G}(\mathbf{V}, \mathbf{E})$ ,  $\mathbf{x}=[x_1, x_2, \dots, x_N]^T$

Each node initials its packet and selects itself as a source node with probability  $p_0$ .

Add the source nodes to the set **Next**;

**while** the set **Next** is not empty

**Current**←**Next**;

**Next**←  $\emptyset$

**for** each node  $i$  in the set **Current**

        selects the next hop node  $k$  of the node  $i$  using (9) or (10);



---

```

P(i).nex=k;
broadcasts P(i);
for each node j which receives P(i)
  if [P(i).mem] ∩ [P(j).mem]= ∅
    merges P(i) into P(j) using (7);
    if (j=P(i).nex )
      adds node j to the set Next;
    end
  end
end
Current← ∅ ;
end
end

```

---

A mobile sink randomly queries  $M$  nodes in the network to obtain the  $M \times N$  measurement matrix  $\Phi$  and the  $M \times 1$  measurement vector  $\mathbf{y} = \Phi \mathbf{x}$ .

Reconstruct  $\mathbf{x}$  from  $\mathbf{y}$  using BP or OMP algorithm.

---

### 3.5 The expression for probability $p_{q+1}$

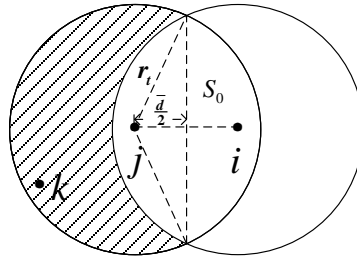
Consider a large WSN with  $N$  nodes described in section 2, the total area  $S=1 \times 1$ , there is a link between the node  $i$  and  $j$  if  $d(i, j) < r_i$ . Let  $k$  be the next hop selected by the node  $j$  at the time  $q$  ( $q > 1$ ). The node  $k$  will broadcast its packet at the time  $(q+1)$  with the probability  $p_{q+1}$ ,

$$\begin{aligned}
 p_{q+1} = 1 - & \frac{C_{N_t^q}^2 \frac{(14\pi - 15\sqrt{3})r_i^2}{36} \cdot \frac{\pi r_i^2}{6}}{N_t^q \left( \frac{\bar{d}}{2} \sqrt{4r_i^2 - \bar{d}^2} + 2r_i^2 \arcsin \frac{\bar{d}}{2r_i} \right)} \\
 & - \frac{\frac{(4\pi - 3\sqrt{3})}{4} \frac{r_i^2}{6} + \frac{3\sqrt{3}}{16\pi} \left( \pi - \frac{15\sqrt{3}}{16} \right) r_i^2}{\frac{\bar{d}}{2} \sqrt{4r_i^2 - \bar{d}^2} + 2r_i^2 \arcsin \frac{\bar{d}}{2r_i}}, \quad (11)
 \end{aligned}$$

where  $N_t^q$  is the number of transmissions at the time  $q$ .  $\bar{d}$  is the distance between the current node  $j$  and its farthest neighbor node  $k$ , and according to [8],  $\bar{d}$  can be obtained by Eq. (12)

$$N = \frac{\bar{d}}{(r_i - \bar{d}) \left( \frac{\bar{d}}{2} \sqrt{4r_i^2 - \bar{d}^2} + 4r_i^2 \arcsin \frac{\bar{d}}{2r_i} \right)}. \quad (12)$$

*Proof:* Let  $N_m^q$  be the total number of the nodes which update their packets at the time  $q$  ( $q > 1$ ), and  $N_t^q$  is the number of transmitters at the time  $q$  and  $S_1$  is the area of the shaded part in Fig. 3.



**Fig. 3.** Illustration of  $S_1$

Assume node  $j$  is a transmitter at the time  $q$ , the average number of neighbors which update their original packets with  $\mathbf{P}(j)$  at the time  $q$  is  $N_m^q / N_t^q$ . Obviously, the  $N_m^q / N_t^q$  neighbors are located in the shaded area in **Fig. 3**. The node  $j$  will select a next hop node  $k$  from the nodes in the shaded area in **Fig. 3** as discussed in subsection 2.

Based on the conditions in (6) and (8), the selected node  $k$  will broadcast its packet at the time  $(q+1)$  with the probability  $p_{q+1}$ , i.e.,

$$p_{q+1} = (N_m^q / N_t^q) / N_1, \tag{13}$$

where  $N_1$  is the expected number of nodes in the shaded area, which can be obtained by

$$\begin{aligned} N_1 &= NS_1 \\ &= N(\pi r_t^2 - S_0) \\ &= N \left\{ \pi r_t^2 - 2 \left[ r_t^2 \arccos \frac{\frac{\bar{d}}{2} - \frac{\bar{d}}{2}}{r_t} - \frac{\bar{d}}{2} \sqrt{r_t^2 - \left(\frac{\bar{d}}{2}\right)^2} \right] \right\} \\ &= N \left( 2r_t^2 \arcsin \frac{\bar{d}}{2r_t} + \frac{\bar{d}}{2} \sqrt{4r_t^2 - \bar{d}^2} \right). \end{aligned} \tag{14}$$

In (13),  $N_m^q$  can be approximately written as (whose derivation is given in Appendix)

$$\begin{aligned} N_m^q &= N_t^q N \left( \frac{\bar{d}}{2} \sqrt{4r_t^2 - \bar{d}^2} + 2r_t^2 \arcsin \frac{\bar{d}}{2r_t} \right) \\ &\quad - C_{N_t^q}^2 \frac{(14\pi - 15\sqrt{3})}{36} r_t^2 \times N \frac{\pi r_t^2}{6} \\ &\quad - N_t^q \frac{(4\pi - 3\sqrt{3})}{4\pi} N \frac{\pi r_t^2}{6} \\ &\quad - N_t^q \frac{3\sqrt{3}}{16\pi} \times N \left( \pi - \frac{15\sqrt{3}}{16} \right) r_t^2. \end{aligned} \tag{15}$$

Substituting (14) and (15) into (13), Eq. (11) can be obtained.

### 3.6 The expressions for $N_t^{tot}$ and $N_r^{tot}$

By derivation, the total number of transmissions  $N_t^{tot}$  and the total number of receptions  $N_r^{tot}$  of the proposed method are respectively

$$N_t^{tot} = \sum_{q=1}^{N_f} N_t^q, \quad (16)$$

$$N_r^{tot} = \sum_{q=1}^{N_f} N_t^q \pi r_t^2 N, \quad (17)$$

where  $N_f$  is the number of repetitions of the broadcast processing, and the number of transmitters at the time  $q$  is

$$N_t^q = \begin{cases} p_0 N, & q = 1, 2 \\ p_q N_t^{q-1}, & q > 2. \end{cases} \quad (18)$$

*Proof:* At the time  $q=1$ , the source nodes broadcast their packets, i.e.,

$$N_t^1 = p_0 N. \quad (19)$$

Obviously, all reception nodes including the  $N_t^1$  selected nodes of the sources update their packets with the received ones which only have one node ID. Therefore, all of the nodes selected at the time  $q=1$  will broadcast their packets at the time  $q=2$ , i.e.,

$$N_t^2 = N_t^1 = p_0 N. \quad (20)$$

When  $q>2$ , at the time  $(q-1)$ , the  $N_t^{q-1}$  transmitters selected  $N_t^{q-1}$  next hop nodes for the next time  $q$ , the selected  $N_t^{q-1}$  nodes will broadcast their packet at the time  $q$  with the probability  $p_q$ . Therefore, the number of transmitters at the time  $q$  is

$$N_t^q = p_q N_t^{q-1}, \quad (q > 2). \quad (21)$$

Combining (19), (20) and (21), Eq. (18) can be obtained.

The total number of receptions is

$$N_r^{tot} = \sum_{q=1}^{N_f} N_t^q N_{nei}, \quad (22)$$

where  $N_{nei}$  is the average number of neighbors of one node, and

$$\begin{aligned} N_{nei} &= (\pi r_t^2 / S) N \\ &= \pi r_t^2 N. \end{aligned} \quad (23)$$

Substituting (23) into (22), Eq. (17) can be obtained.

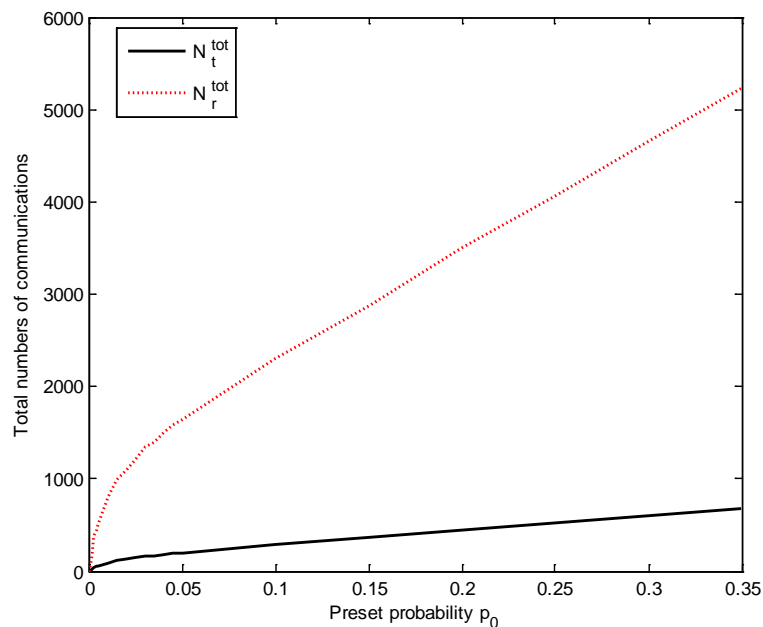
Particularly, the additional overhead of the proposed method focus on the calculation for each node to find an intermediate node  $k$  and the calculation complexity is about  $O(N_{nei}^3)$ , where  $N_{nei}$  can be obtained by equation (23).

## 4. Simulation Results

To evaluate the performance of the proposed method, a WSN with  $N=1000$  nodes uniformly and randomly deployed in a unit area is considered. In the network, the communication range  $r_t$  takes value from 0.05 to 0.085 according the connectivity condition. In the simulation, the DCT basis and the OMP algorithm are employed to reconstruct the sensor readings. In comparison with the proposed method, the available schemes in [8], [17] and [18] are also simulated.

**Fig. 4** shows the total number of communications curves versus  $p_0$ . In **Fig. 4**,  $N_t^{tot}$  and  $N_r^{tot}$  are the total numbers of transmissions and receptions respectively. It can be seen that both the total number of transmissions  $N_t^{tot}$  and the total number of receptions  $N_r^{tot}$  increase as the preset probability  $p_0$  increases. The higher  $p_0$  is, the more source nodes are selected, which will result in the increasing of the energy cost of communications. Therefore, the probability  $p_0$  should be as small as possible under the premise of the reasonable reconstruction.

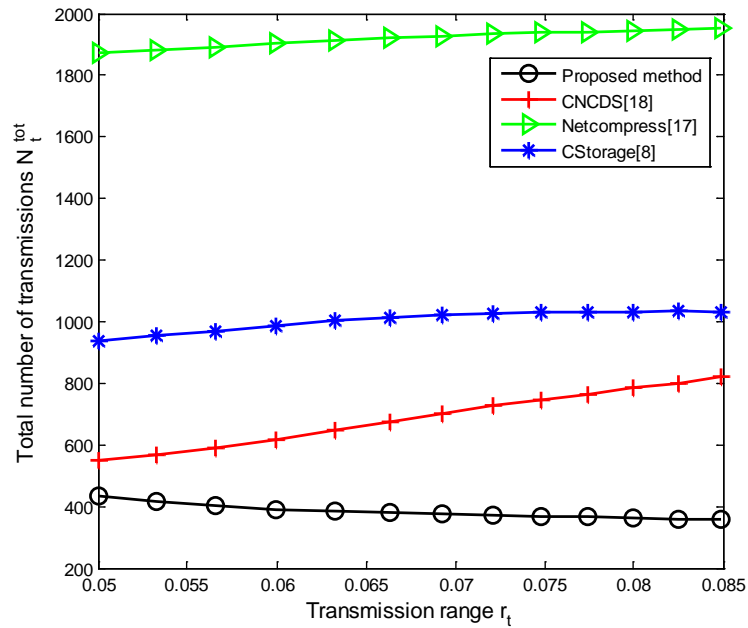
**Fig. 5** gives the total number of communications curves versus the transmission range  $r_t$  for the different methods. In **Fig. 5**, the ‘CStorage’, ‘Netcompress’ and ‘CNCDS’ are the methods given in [8], [17] and [18] respectively, and the forwarding probability  $p_a$  of the CNCDS in [17] is set to an optimum value 0.24. From **Fig. 5**, one can see that both the total number of transmissions and the total number of receptions of the proposed method are smaller than those of the available methods in [8], [17] and [18]. Compared with the CStorage in [8], the proposed method reduces the total number of communications by 60% at least, where the total number of communications is the one of the transmissions or the receptions. Moreover, the proposed method can reduce the total number of communications about by 80% and 30% when compared with the algorithms in [17] and [18] respectively. Therefore, the proposed method can greatly improve the energy efficiency. In addition, as the transmission range  $r_t$  increases, the total number of transmissions decreases slightly due to the decreasing of  $p_{q+1}$  and the total number of receptions of the proposed method increase slowly due to the increasing of the number of neighbors for each node, while the available schemes in [8], [17] and [18] increase rapidly. It demonstrates that the proposed method can perform well in the large-scale and dense networks.



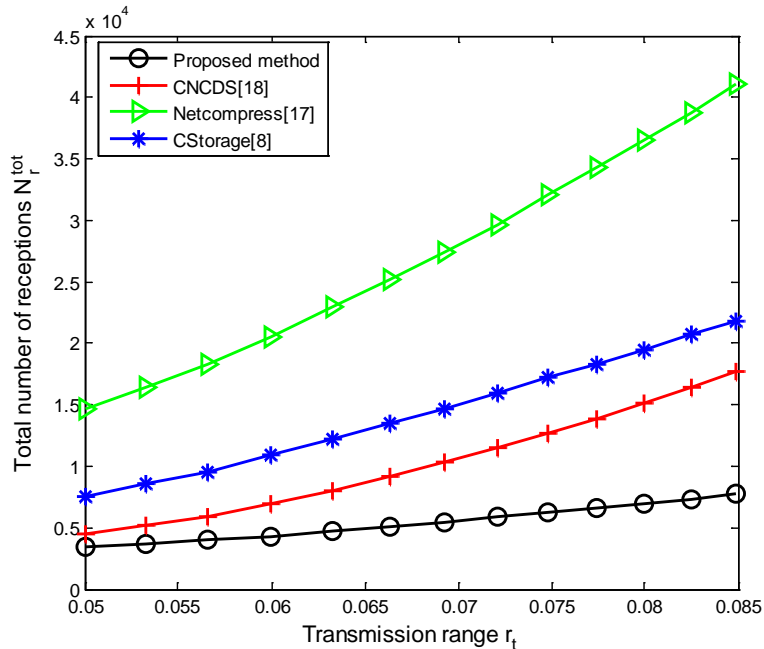
**Fig. 4.** Total numbers of communications curves versus  $p_0(r_t=0.05)$

**Fig. 6** shows the mean square error (MSE) performance of the reconstruction against  $M$  for the different methods. It can be seen from **Fig. 6** that the MSE performance of the

Netcompress in [17] is a little better, but its energy efficiency is the worst, which can be seen from Fig. 5. Compared with the CNCDS in [18] and the CStorage in [8], the proposed method has a better MSE performance when  $M$  is small. Moreover, the MSE performance of the whole schemes tends to be the same when  $M$  is greater than 130.



a. Total number of transmissions



b. Total number of receptions

Fig. 5. Total numbers of communications against transmission range  $r_t$  ( $p_0=0.18$ ).

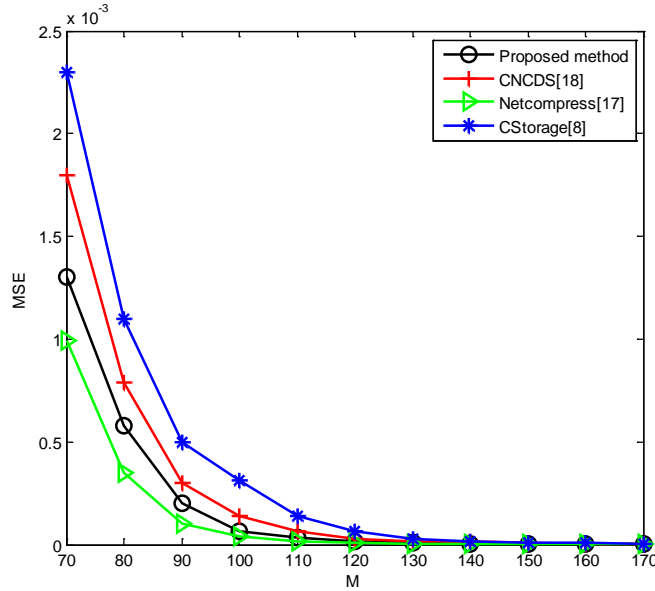


Fig. 6. MSE comparison against  $M$  for different methods ( $p_0=0.18, r_t=0.05$ )

To verify the correctness and the efficiency of the proposed method, the comparison between the theoretical value and the simulation result of the total numbers of communications for the proposed method is given in Fig. 7. From Fig. 7, one can see that the simulation result is almost coincident with the theoretical value for the total number of transmissions  $N_t^{tot}$  or the total number of receptions  $N_r^{tot}$ , where the small gap is caused by the approximate calculations of  $\bar{d}$  and  $N_m^q$  in (12).

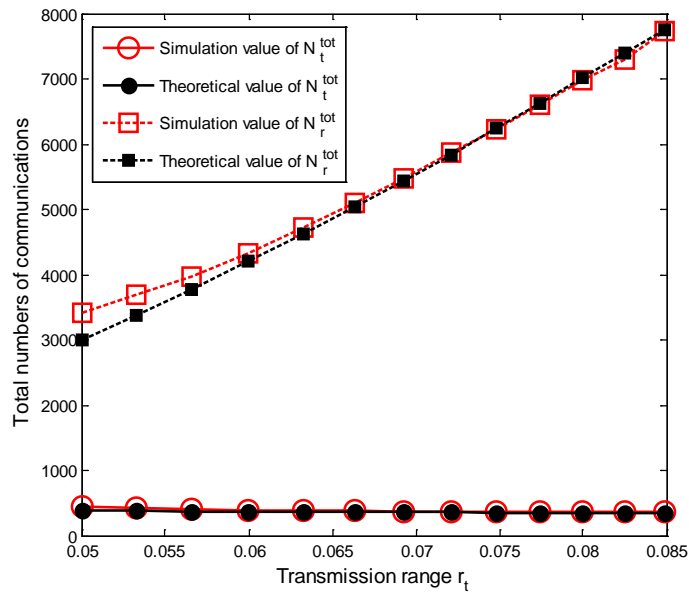


Fig. 7. Comparison between the theoretical value and simulation result of the total numbers of communications

## 5. Conclusion

In this paper, an improved energy-efficient data delivering method based on compressed sensing and network coding is proposed for the WSNs. In the proposed method, the next hop selection principle is employed to determine the next transmitter to further reduce the number of the transmissions. Based on the random graph theory, we derive the expression for  $N_i^{tot}$  and  $N_r^{tot}$  to verify the correctness and the efficiency of the proposed method. Compared with the available schemes, the proposed method can reduce the communication consumption greatly. Moreover, the proposed method can be used in the large-scale and dense networks to improve the energy efficiency of the networks.

## Appendix

### Calculation of $N_m^q$ in Section 3.5

Assume node  $j$  is a transmitter at the time  $q(q>1)$ , and the node  $i$  is its parent node. The node  $j$  will select a next hop node  $k$  from the nodes in the shaded area in Fig. 3 as discussed in section 3.2 and  $S_1$  is the area of the shaded part.

If the node  $j$  broadcasts its packet at the time  $q$ , only the nodes in the shaded area in Fig. 3 may update their packets with the received packet  $\mathbf{P}(j)$ . There are  $N_i^q$  nodes broadcasting their packets at the time  $q$ , the total number of nodes which may update their packets is

$$N_{m1}^q = N_i^q N_1 = N_i^q N \left( \frac{\bar{d}}{2} \sqrt{4r_i^2 - \bar{d}^2} + 2r_i^2 \arcsin \frac{\bar{d}}{2r_i} \right). \quad (24)$$

However, some of the  $N_{m1}^q$  nodes will not update their packets with the received packet when the following two cases happen.

**Case 1:** At the time  $(q-1)$ , another node  $j'$  and the node  $j$  update their packets with the same packet  $\mathbf{P}(i)$ , and the node  $j'$  is also a transmitter at the time  $q$ . The nodes in the shaded area ( $S_2$ ) in Fig. 8(a) will merge only one of the two packets. However, the number of nodes in the shaded area  $N_{m2}^q$  is calculated twice in  $N_{m1}^q$ . Therefore, it should be subtracted from  $N_{m1}^q$ .

**Case 2:** At the time  $(q-1)$ , the node  $j$  updates their packet with not only the  $\mathbf{P}(i)$  but also a packet  $\mathbf{P}(i')$  from the node  $i'$ . According to the condition in (6), the nodes in the shaded area ( $S_3$  and  $S_4$ ) in Fig. 8(b) and Fig. 8(c) which are the neighbors of the node  $i'$  will not merge the received packet  $\mathbf{P}(j)$  into their original packets. The number of these nodes  $N_{m3}^q$  should be subtracted from  $N_{m1}^q$  too.

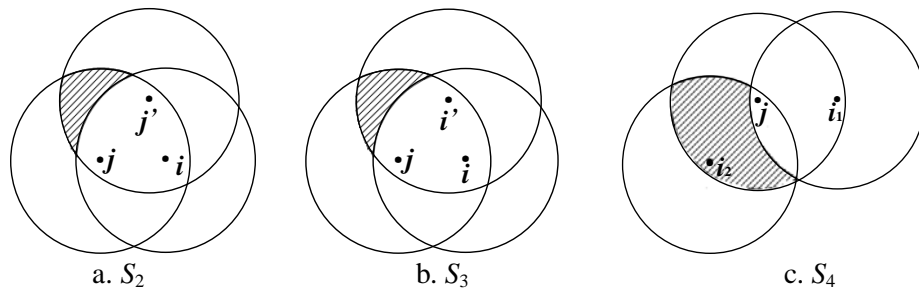


Fig. 8. Illustrations of  $S_2$ ,  $S_3$  and  $S_4$

Therefore,

$$N_m^q = N_{m1}^q - N_{m2}^q - N_{m3}^q, \quad (25)$$

According to [18],

$$N_{m2}^q = C_{N_t^q}^2 \frac{(14\pi - 15\sqrt{3})}{36} r_t^2 \times N \frac{\pi r_t^2}{6}, \quad (26)$$

$$N_{m3}^q = N_t^q \frac{(4\pi - 3\sqrt{3})}{4\pi} N \frac{\pi r_t^2}{6} + N_t^q \frac{3\sqrt{3}}{16\pi} \times N \left(\pi - \frac{15\sqrt{3}}{16}\right) r_t^2. \quad (27)$$

Substituting (24), (26) and (27) into (25), Eq.(15) can be obtained.

## References

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no.4, pp. 393-422, March 2002. [Article \(CrossRef Link\)](#)
- [2] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. of Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, October 30-November 2 2005. [Article \(CrossRef Link\)](#)
- [3] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *Proc. of INFOCOM , Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 2571 – 2582, March 7-11 2004. [Article \(CrossRef Link\)](#)
- [4] S. He, J. Chen, D. K. Y. Yau, Y. Sun, "Cross-layer optimization of correlated data gathering in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 11, pp.1678-1691, July 2012. [Article \(CrossRef Link\)](#)
- [5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006. [Article \(CrossRef Link\)](#)
- [6] M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Universal distributed sensing via random projections," in *Proc. of 5th International Conference on Information Processing in Sensor Networks*, pp. 177-185, April 19-21 2006. [Article \(CrossRef Link\)](#)
- [7] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: sparse signal modeling and monitoring framework," *IEEE Transactions on Wireless Communications*, vol. 11, no.10, pp. 3447 – 3461, October 2012. [Article \(CrossRef Link\)](#)
- [8] A. Talari, N. Rahnavard, "CStorage: decentralized compressive data storage in wireless sensor networks," *Ad Hoc Networks*, vol. 37, no. 2, pp. 475–485, February 2016. [Article \(CrossRef Link\)](#)
- [9] R. Ahlswede, N. Cai, S.Y.R. Li, and R.W. Yeung, "Network Information Flow," *IEEE Transactions on Information Theory*, vol.46, no. 4, pp. 1204-1216, July 2000. [Article \(CrossRef Link\)](#)
- [10] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking* , vol. 16, no. 3, pp. 497 - 510, June 2008. [Article \(CrossRef Link\)](#)
- [11] S. Sengupta, S. Rayanchu, and S. Banerjee, "Network coding-aware routing in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1158 – 1170, August 2010. [Article \(CrossRef Link\)](#)
- [12] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413 - 4430, October 2006. [Article \(CrossRef Link\)](#)
- [13] X. Liu, X. Gong, and Y. Zheng, "Reliable cooperative communications based on random network coding in multi-hop relay WSNs," *IEEE sensors journal*, vol. 14, no. 8, pp. 2514 – 2523, August 2014. [Article \(CrossRef Link\)](#)



- [14] S. Katti, S. Shintre, S. Jaggi, D. Katabi, and M. Medard, "Real network codes," in *Proc. of Allerton Conference on Communication, Control, and Computing*, pp. 389-395, September 26-28 2007.
- [15] C. Luo, J. Sun, and F. Wu, "Compressive network coding for approximate sensor data gathering," in *Proc. of GLOBECOM*, pp. 1-6, December 5-9 2011. [Article \(CrossRef Link\)](#)
- [16] M. Kwon, H. Park and P. Frossard, "Compressed network coding: overcome all-or nothing problem in finite fields," in *Proc. of Wireless Communications and Networking Conference (WCNC), IEEE*, pp. 2851-2856, April 6-9 2014. [Article \(CrossRef Link\)](#)
- [17] N. Nguyen, D. L. Jone, and S. Krishnamurthy, "Netcompress: coupling network coding and compressed sensing for efficient data communication in wireless sensor networks," in *Proc. of IEEE Workshop on Signal Processing Systems*, pp. 356 – 361, October 6-8 2010. [Article \(CrossRef Link\)](#)
- [18] X. Yang, X. Tao, E. Dutkiewicz, X. Huang, Y. J. Guo, and Q. Cui, "Energy-Efficient Distributed Data Storage for Wireless Sensor Networks Based on Compressed Sensing and Network Coding," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5087 - 5099, October 2013. [Article \(CrossRef Link\)](#)
- [19] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589-592, May 2008. [Article \(CrossRef Link\)](#)
- [20] M. Heissenbuttel, T. Braun, M. Walchli, and T. Bernoulli, "Optimized stateless broadcasting in wireless multi-hop networks," in *Proc. of INFOCOM, 25th IEEE International Conference on Computer Communications*, pp. 1-12, April 23-29 2006. [Article \(CrossRef Link\)](#)



**Ying Zhou** was born in Jiangsu, China in 1990. She received her B.S. degree in Optical Information Science and Technology at Nanjing University of Posts and Telecommunications (NUPT), 2012 and her M.S. degree in Communication and Information Systems at NUPT, 2013. She is currently studying towards her Ph.D. degree in the College of Telecommunications and Information Engineering, NUPT. Her research interests focus on the compressed sensing and network coding, wireless communication theories, key technologies of mobile communication systems, and internet of things.



**Lihua Yang** was born in Jiangsu, China in 1984. She received her B.S. degree in Electronics and communications engineering at Xi'an University of Science and Technology, 2007 and her M.S. degree in signal processing at Xidian University 2008. She received the Ph.D. degree in the communication and information systems at Xidian University on March 2013. She is currently a Lecturer of the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. Her research interests focus on the key technologies of the physical layer in the mobile wireless communications, and the network coding.



**Longxiang Yang** is with the college of Communications and Information Engineering, Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China. He is a full professor and a doctoral supervisor with NUPT. He has fulfilled multiple National Natural Science Foundation projects of China. He has authored and coauthored more than 100 technical papers published in various journals and conferences. His research interests include cooperative communication and network coding, wireless communication theories, key technologies of mobile communication systems, ubiquitous networks, and internet of things.