Computer Networks 108 (2016) 345–356

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Compressive network coding for wireless sensor networks: Spatio-temporal coding and optimization design

Siguang Chen[a,c,*], Chuanxin Zhao[b,d], Meng Wu[a], Zhixin Sun[a], Haijun Zhang[c,1], Victor C.M. Leung[c,2]

[a] Key Lab of Broadband Wireless Communication and Sensor Network Technology of Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing, China
[b] School of Electrical Engineering and Computing, Curtin University, Perth, Australia
[c] Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, Canada
[d] Department of Network Engineering, Anhui Normal University, Wuhu, China

ARTICLE INFO

Article history:
Received 14 February 2016
Revised 18 July 2016
Accepted 13 September 2016
Available online 13 September 2016

Keywords:
Network coding
Compressed sensing
Spatio-temporal compression
Optimization
Distributed algorithm
Wireless sensor networks

ABSTRACT

Considering the temporal and spatial correlations of sensor readings in wireless sensor networks (WSNs), this paper develops a clustered spatio-temporal compression scheme by integrating network coding (NC), compressed sensing (CS) and spatio-temporal compression for correlated data. The proper selection of NC coefficients and measurement matrix is investigated for this scheme. This design ensures successful reconstruction of original data with a considerably high probability and enables successful deployment of NC and CS in a real field. Moreover, in contrast to other spatio-temporal schemes with the same computational complexity, the proposed scheme possesses lower reconstruction error by employing independent encoding in each sensor node (including the cluster head nodes) and joint decoding in the sink node. In order to further reduce the reconstruction error, we construct a new optimization model of reconstruction error for the clustered spatio-temporal compression scheme. A distributed algorithm is developed to iteratively determine the optimal solution. Finally, simulation results verify that the clustered spatio-temporal compression scheme outperforms other two categories of compression schemes significantly in terms of recovery error and compression gain and the distributed algorithm converges to the optimal solution with a fast and stable speed.

1. Introduction

Wireless sensor networks (WSNs) consisting of one (or few) sink node(s) and a large number of sensor nodes are usually deployed in a large region to monitor physical or environmental conditions, such as temperature, light, humidity, etc. Since the time-series data of a sensor node usually have temporal dependency, and the observed data of nearby nodes monitoring the same region at the same time slot are highly correlated, the sensor readings usually have both temporal and spatial correlations. Exploiting these correlations can reduce the number of transmissions, and therefore, decrease the energy consumption in WSNs. However, there exist challenges as sensor nodes have limited energy and low computational capability. Fortunately, compressed sensing (CS) [1–3] transfers most of the computational complexity into the sink node (i.e., reduces the computational burden of sensor nodes), and is considered an effective tool to explore the mutual correlation of sensor readings. Using CS, information can be reconstructed with a high probability of success from a small collection of measurements, which means it can prolong the lifetime of WSNs effectively.

On the other hand, network coding (NC) [4] allows the intermediate nodes to encode the incoming packets rather than simply forwarding them. This powerful theory can improve the network load and enhance network robustness by employing path diversity. So, in addition to prolonging the lifetime of WSNs, NC improves data security. As a result, combining NC and CS for exploiting the correlations of sensor readings in WSNs has become an attractive topic.

The existent research regarding the temporal and spatial correlations in WSNs can be classified into the following four categories.

The first category consists of schemes which exploit either temporal or spatial correlation but not both, such as [5–9]. Xie et al.

* Corresponding author.
E-mail addresses: sgchen@njupt.edu.cn (S. Chen), chuanxin.zhao@curtin.edu.au (C. Zhao), wum@njupt.edu.cn (M. Wu), sunzx@njupt.edu.cn (Z. Sun), haijunzhang@ece.ubc.ca (H. Zhang), vleung@ece.ubc.ca (V.C.M. Leung).
[1] Member, IEEE.
[2] Fellow, IEEE.

http://dx.doi.org/10.1016/j.comnet.2016.09.007
1389-1286/

[5] used hybrid CS method to propose an analytical model and centralized clustering algorithm for obtaining the minimum number of transmissions in sensor networks. However, the sensing data have considerable redundancy in temporal dimension, and they do not exploit it. The works [6–9] investigated the correlation of sensing data by combining the NC and CS. Luo et al. [6] proposed a compressive NC for approximate sensor data gathering via exploring the temporal correlation of sensing data. This paper overcomes the all-or-nothing property of NC and achieves graceful degradation in data precisions. Yang et al. [7] designed a compressed NC-based distributed data storage scheme by utilizing the spatial correlation of sensor readings. This scheme possesses an energy-efficient property by reducing the total number of transmissions and receptions. Nabaee et al. [8,9] constructed a data gathering technique by mining the spatial correlation of sensor data. This technique can achieve a good approximation of the original data with small amount of data received. Similarly, these four works do not consider the temporal and spatial correlations simultaneously which has a significant impact on network efficiency.

The second category includes schemes which study the joint sparsity model-based (JSM-based) spatio-temporal correlations [10–12], where the temporal and spatial correlations are integrated. In [10], the authors presented a balanced spatio-temporal compression scheme for WSNs. This scheme can reduce energy consumption and prevent overloading of nodes. Chen et al. [11] developed a compressive NC for error control in WSNs. This encoding mechanism can achieve considerable compression ratio and tolerate finite erasures and errors at the same time. In [12], Kong et al. proposed a novel CS-based approach to reconstruct the massive missing data and develop an environmental space time improved CS algorithm to enhance the reconstruction accuracy. The feature of the second category is that the spatial and temporal signals are transformed into a long vector. Although the spatial and temporal correlations are exploited fully, the computational complexity of reconstruction process is high.

The works [13,14] belong to the third category where the spatial and temporal correlations are both considered and investigated separately. Feizi et al. [13] conceived a power efficient sensing scheme by combining source channel NC and CS. The main merits of this scheme are the low decoding complexity, independent structure and the continuous rate distortion performance. Nevertheless, it assumes that the sampling data of original time-series data still have spatial dependency which is not enough to be convincing. This assumption was eliminated in [14], in which Lee et al. constructed a low complexity sensing for spatio-temporal data. The principle of this scheme is that it samples time-series data in the temporal dimension randomly, and then measures the data in the spatial dimension. It is simple and easy to implement. However, the reconstruction error of this scheme will not be low if the sensor readings fluctuate remarkably among faraway nodes in the same time slot.

Gong et al. [15] formulated the fourth category scheme where the spatial and temporal correlations are both considered and investigated as a unity. The NC scheme in [15] is a spatiotemporal compressive scheme for distributed data storage in WSNs. This scheme can reduce the number of transmissions and receptions, but involves a high computational complexity in the reconstruction process.

Despite the fact that the schemes mentioned above did a lot of meaningful research work in exploration of correlations of sensor readings, all of them (except [6]) only focus on the design of encoding/decoding methods and neglect the optimization of network resource allocation which can improve the network performance significantly.

The network optimization scenario considered in this paper is similar with the resource optimization schemes in NC-based wireless network. Currently, the existing works on optimization of NC-based network resource mainly address the problems of achieving the maximum throughput [16,17], the maximum lifetime [18], the minimum energy consumption [19,20], the minimum packet delay [21] and the tradeoff between two randomly former metrics [22–24].

In [16,17], the maximum throughput of networks was studied by developing joint congestion control and scheduling with NC. Tan et al. in [18] maximized the network lifetime by optimizing the network flow control and video encoding bit rate jointly. The main goal of works [19,20] is to minimize the network energy consumption. The work [21] minimized the packet delay in a TDMA-based wireless networks by utilizing NC and successive interference cancelation techniques. The authors in [22] attempted to make a tradeoff between network throughput and energy consumption. In [23], the issue of throughput-delay tradeoff in NC was studied. Also, the tradeoff between network throughput and lifetime was investigated in [24]. These schemes mainly optimize the conventional network performance metrics, however, the more benefits will be obtained when the networks formulate new optimization objective by combining the conventional performance metrics with CS theory. Based on the compressive NC scheme, the work [6] constructed a new optimization objective to improve the performance of compressive NC flows. Although it achieves the optimal network utility, wireless interference will be a big challenge for this scheme.

Motivated by the shortcomings of prior literatures on mining the spatial and temporal correlations and optimizing the network resources, we propose a clustered spatio-temporal compression scheme by combining the NC and CS in WSNs and formulate a new optimization model to make link capacity assignment. The main contributions can be summarized as follows.

One main contribution of our work is that we integrate the CS, NC and spatio-temporal compression into an unified and new system, the NC coefficients and measurement matrix are designed properly for this new system. This design ensures successful reconstruction of original data with a considerably high probability and enables successful deployment of NC and CS in a real field.

The second main contribution is that in contrast to other spatio-temporal schemes with the same computational complexity, the proposed scheme demonstrates lower reconstruction error by developing a new spatio-temporal coding method which employing independent encoding in each sensor node (including the cluster head nodes) and joint decoding in the sink node. At the same time, it has lower computational complexity as compared with JSM-based spatio-temporal scheme and the fourth category scheme by exploiting the temporal and spatial correlations of original sensing data step by step.

Our third main contribution is that we construct a new optimization model for minimizing reconstruction error of the proposed clustered spatio-temporal compression scheme, in which the unreliability of wireless links and the effect of wireless interference are taken into account. The minimization of reconstruction error can be achieved in a distributed manner by utilizing dual decomposition, subgradient algorithm and low-pass filtering method.

Finally, the proposed compression scheme has been verified to have considerable compression gain and lower reconstruction error, and the optimization problem has been validated to converge to the optimal solution with a fast and stable speed.

The remainder of this paper is organized as follows. In Section 2, we introduce basic theory of CS. The network model is defined in Section 3. In Section 4, the proposed compression scheme is given in detail. Section 5 formulates the optimization problem with the goal of minimizing reconstruction error. In Section 6, we analyze the performance of the proposed scheme. Finally, conclusions are drawn in Section 7.

2. Compressed sensing background

The basic fundamental of CS is that the information can be compressed into a small amount of equivalent information, and then reconstructed successfully with a high probability. For example, consider a signal x of length N that can be represented as $x = \Psi\theta$ for a given matrix $\Psi \in R^{N \times N}$ and column vector $\theta \in R^N$. The vector θ is called the coefficient vector. To measure the signal x, we obtain a sampling vector $y \in R^n$ by means of a $n \times N$ measurement (projection) matrix Φ

$$y = \Phi x = \Phi\Psi\theta = \Theta\theta \tag{1}$$

where $n \ll N$.

Now the objective is to reconstruct the original signal x accurately or approximately given y, Φ and Ψ. The reconstruction is performed by finding the solution of the following l_0 minimization problem

$$\tilde{\theta} = \arg\min_{\theta} \|\theta\|_0 \quad s.t.\ \Theta\theta = y. \tag{2}$$

After obtaining $\tilde{\theta}$, the reconstructed signal can be calculated as

$$\tilde{x} = \Psi\tilde{\theta}. \tag{3}$$

Many algorithms have been proposed recently to solve this minimization problem, Some examples are Orthogonal Matching Pursuit (OMP) [25] and Stage-wise Orthogonal Matching Pursuit (StOMP) [26].

Next, three definitions are given below.

Definition 1. Sparse signal [27]: we define that signal x is s-sparse over dictionary Ψ when coefficient vector θ has at most s non-zero values, i.e., $\|\theta\|_0 \le s$, where $\|\|_0$ counts the number of non-zero elements.

Definition 2. Compressible signal [27]: we define that signal x is s-compressible if its sorted coefficient magnitudes in dictionary Ψ decay rapidly, i.e., signal x can be well-approximated by a s-sparse signal which can be shown to satisfy

$$\|x - x_s\|_2 \le \kappa \tag{4}$$

where κ denotes a small number and x_s is a s-sparse signal.

Definition 3. Let Φ be a $n \times N$ matrix and let $s < N$ be an integer. There exists a constant δ_s that holds the following condition for all s-sparse vectors $x \in R^N$

$$(1 - \delta_s)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1 + \delta_s)\|x\|_2^2 \tag{5}$$

then the matrix Φ satisfies the restricted isometry property (RIP) with restricted isometry constants (RIC) δ_s [28].

3. Network model

In this section, we define the network structure and describe the function of different type nodes.

In our network scenario, we construct a clustered wireless sensor network where a sink node is used to collect the sensing data observed by sensor nodes. The whole sensor network has multiple clusters, and the nearby sensor nodes are allocated into the same cluster where each cluster has a cluster head node. A sensor node in one cluster which has the highest residual energy will be selected as cluster head node. Each sensor node has the power to sample the data and obtain the sampling data by exploiting the temporal correlation of the original data.

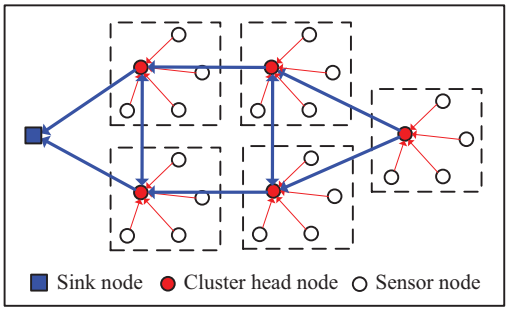

Fig. 1. Data collection in a clustered WSN with spatio-temporal compression.

The cluster head nodes are responsible for collecting sampled data from their inner sensor nodes. At the same time, they also act as relay nodes for forwarding data from other cluster head nodes. Because the observed data of nearby sensors often exhibit spatial correlation, the cluster head nodes generate random projections of the sampled data of sensor nodes for transmitting instead of transmitting the original sampled data. We exploit the temporal and spatial correlations of original data simultaneously in our network model, but the temporal and spatial correlations are considered step by step.

The linear NC is performed before the cluster head nodes forward their own and the incoming packets. The NC operation is performed over a real field. Finally, the cluster head nodes transmit the encoded data to the sink node via a one hop or multihop manner. The whole transmission process of sensing data are showed in Fig. 1.

In this network scenario, we only consider communication with one sink node for simplicity, but the proposed scheme can be generalized to multiple sink nodes. The topology of the cluster head nodes and sink node is represented by the directed graph $G(V, E)$, where V is the set of cluster head nodes and the sink node, and E is the set of directed links. Let $\Gamma^+(k)$ be the links emanated from a node k, and $\Gamma^-(k)$ be the links entering into a node k. Meanwhile, we define that there are L clusters in the network, and each cluster has M sensor nodes where the original data length (i.e., temporal dimension) of each sensor node is N. The data transmitted over link e at time t can be denoted by $h_t(e)$.

4. Clustered spatio-temporal compression scheme

A clustered spatio-temporal compression scheme is developed by extending the work in [8] to a more realistic scenario where sensor readings exhibit both spatial and temporal correlations. The CS, NC and spatio-temporal compression are integrated into an unified and new system. The proper design of NC encoding coefficients and measurement matrix ensures successful reconstruction of original data with a considerably high probability. Meanwhile, the reconstruction error and computational complexity of this scheme is lower due to employing independent encoding in each sensor node and joint decoding in sink node. The processing and transmitting of original data in this compression scheme are illustrated in Fig. 2.

In the proposed framework, each sensor projects its original signal to a lower dimensional space by exploiting the temporal correlation of the original signal. Also, the cluster head node randomly projects the sampled signals of its inner sensors to a lower dimensional space by exploiting the spatial correlation. The linear NC is performed in the communication between different cluster head nodes. The sink node uses the CS-based spatial and temporal decoding algorithms to reconstruct the original signals.

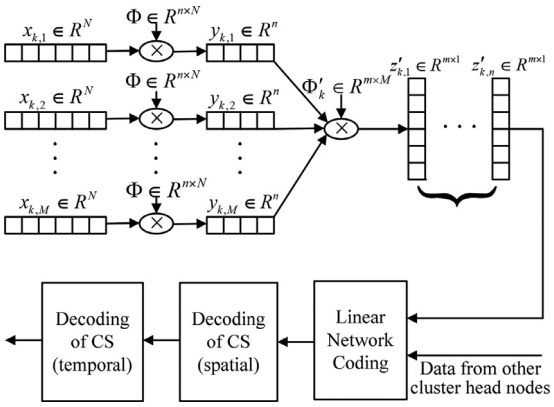

Fig. 2. Clustered spatio-temporal compression framework.

4.1. Independent encoding

In this subsection, the encoding methods for sensor nodes and cluster head nodes are presented, and the preconditions for successful decoding in the sink node are also described.

Sensor node: the original signal of sensor node i in cluster $k \in \{1, 2, \ldots, L\}$ can be denoted as

$$x_{k,i} = [x_{k,i1} \quad x_{k,i2} \quad \ldots \quad x_{k,iN}]^T \tag{6}$$

where $x_{k,\,ij}$ represents original data (temporal dimension) of sensor node i at time j and the sensor node i belongs to cluster k, and $x_{k,\,i} \in R^{N\times 1}$.

Then sensor node $i$samples its original signal by measurement matrix $\Phi \in R^{n\times N}$ $(n \ll N)$ to obtain

$$y_{k,i} = \Phi x_{k,i} \tag{7}$$

where matrix Φ consists of n rows selected randomly from an $N \times N$ identity matrix, and $y_{k,\,i} \in R^{n\times 1}$.

Finally, sensor node i sends its sampling signal $y_{k,\,i}$ to the cluster head node k.

Cluster head node: we assume that cluster head node k receives M sampled signals from its M inner cluster sensor nodes, and these received signals are organized into a matrix. The cluster head node uses a measurement matrix Φ'_k to measure these received signals, the measurement result of jth ($j \in \{1, 2, \ldots, n\}$) column of that matrix can be shown as

$$\begin{aligned} z'_{k,j} &= [z_{k,1j} \quad z_{k,2j} \quad \cdots \quad z_{k,mj}]^T = \Phi'_k y'_{k,j} \\ &= \Phi'_k [y_{k,1j} \quad y_{k,2j} \quad \cdots \quad y_{k,Mj}]^T, \Phi'_k \in R^{m\times M} \end{aligned} \tag{8}$$

where $y_{k,\,ij}$ represents the jth sampled data of sensor node i and the sensor node i belongs to cluster k, $y'_{k,j} \in R^{M\times 1}$ and $z'_{k,j} \in R^{m\times 1}$.

As the relay node, the output data of cluster head node k over link e during time slot t can be represented by

$$h_t(e) = \sum_{e' \in \Gamma^-(k)} \beta_t(e, e').h_{t-1}(e') + \sum_{i \in \{1,2,\ldots,m\}} \alpha_t(e, i).z_{k,i} \tag{9}$$

where $z_{k,i} = [z_{k,i1} \quad z_{k,i2} \quad \cdots \quad z_{k,in}]$, the links e and e' denote the outgoing and incoming links of cluster head node k respectively, and $\beta_t(e, e')$ and $\alpha_t(e, i)$ are the linear combination coefficients of NC at time slot t which are chosen from real numbers. The cluster head node k's signals are included in the encoding process. In order to track the transformation process of NC encoding coefficients, we append the $[(k-1)m+i]$th row of a $L.m$-dimensional identity matrix to the transmitted packet of $z_{k,\,i}$.

According to Nabaee and Labeau [8], the transmitting data of all links at time slot t can be written as

$$h_t = B_t.h_{t-1} + A_t.Z \tag{10}$$

where the matrices $B_t \in R^{|E|\times|E|}$, $A_t \in R^{|E|\times Lm}$ and $Z \in R^{Lm\times n}$ can be respectively represented as

$$B_t : \{B_t(e, e')\} = \begin{cases} \beta_t(e, e'), e' \in \Gamma^-(k) \& e \in \Gamma^+(k), k \in V \\ 0, \text{otherwise} \end{cases} \tag{11}$$

$$A_t : \{A_t(e, i)\} = \begin{cases} \alpha_t(e, i), e \in \Gamma^+(k), k \in V, i \in \{1, 2, \ldots, m\} \\ 0, \text{otherwise} \end{cases} \tag{12}$$

$$Z = \begin{bmatrix} z'_{1,1} & z'_{1,2} & \cdots & z'_{1,n} \\ z'_{2,1} & z'_{2,2} & \cdots & z'_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ z'_{L,1} & z'_{L,2} & \cdots & z'_{L,n} \end{bmatrix}. \tag{13}$$

Sink node: from (10), the received data of the sink node at time slot t can be organized as follows:

$$\begin{aligned} u_t &= \left[h_t(e) : e \in \Gamma^-(D)\right] = F.h_t \\ &= F.B_t.h_{t-1} + F.A_t.Z \\ &= \Omega_t.Z \end{aligned} \tag{14}$$

where u_t denotes the received data, and the matrices F and Ω_t are respectively defined as

$$F : \{F(e, i)\} = \begin{cases} 1, i \text{ denotes data index of } e, e \in \Gamma^-(D) \\ 0, \text{otherwise} \end{cases} \tag{15}$$

$$\Omega_t = F.\sum_{t'=1}^{t} \prod_{t''=t'+1}^{t} B_{t''}.A_{t'}. \tag{16}$$

Meanwhile, we can find out that the value of Ω_t can be obtained by retrieving the appended $L.m$-dimensional matrix from the received packets.

Thus the whole real data which the sink node receives can be given by

$$\begin{aligned} U &= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \end{bmatrix} = \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \vdots \\ \Omega_t \end{bmatrix}.Z \\ &= \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \vdots \\ \Omega_t \end{bmatrix}.\begin{bmatrix} \Phi'_1 & 0 & \cdots & 0 \\ 0 & \Phi'_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \Phi'_L \end{bmatrix} \\ &\quad .\begin{bmatrix} y'_{1,1} & y'_{1,2} & \cdots & y'_{1,n} \\ y'_{2,1} & y'_{2,2} & \cdots & y'_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{L,1} & y'_{L,2} & \cdots & y'_{L,n} \end{bmatrix} \\ &= \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \vdots \\ \Omega_t \end{bmatrix}.\begin{bmatrix} \Phi'_1 & 0 & \cdots & 0 \\ 0 & \Phi'_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \Phi'_L \end{bmatrix} \\ &\quad .\begin{bmatrix} \Psi'_1 & 0 & \cdots & 0 \\ 0 & \Psi'_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \Psi'_L \end{bmatrix}.\begin{bmatrix} \theta'_1 & \theta'_2 & \cdots & \theta'_n \end{bmatrix} \\ &= \Omega.\Phi'.\Psi'.\begin{bmatrix} \theta'_1 & \theta'_2 & \cdots & \theta'_n \end{bmatrix} \\ &= \Theta.\Psi'.\begin{bmatrix} \theta'_1 & \theta'_2 & \cdots & \theta'_n \end{bmatrix} \end{aligned} \tag{17}$$

where the matrix Φ' is a block-diagonal matrix which is composed of L measurement matrices, Ψ' is also a block-diagonal matrix which is composed of L orthonormal bases, the signal $[\ y'_{1,i} \quad y'_{2,i} \quad \cdots \quad y'_{L,i}\ ]^T, i \in \{1, 2, \ldots, n\}$ is a batch of compressible signals over Ψ', and θ'_i is the coefficient vector.

Theorem 1. *Suppose that the entries of measurement matrix $\Phi'_k, k \in \{1, 2, \ldots, L\}$ and coding coefficients $\beta_t(e, e')$ are selected independently from a zero-mean Gaussian distribution, and $\alpha_t(e, i) = 0, t > 1$, then the entries of matrix Θ are independent zero-mean Gaussian variables.*

Proof. If the coding coefficient $\beta_t(e, e')$ are selected independently from a zero-mean Gaussian distribution, and $\alpha_t(e, i) = 0, t > 1$, similar to the proof of Theorem 3.1 in [8], we can show that the entries of Ω are zero-mean Gaussian random variables, and the entries of different rows are independent.

Because the entries of matrix Θ are linear combinations of the entries of Φ', and a linear combination of independent Gaussian random variables is still a Gaussian variable; the entries of matrix Θ are zero-mean Gaussian random variables as long as the entries of matrix $\Phi'_k, k \in \{1, 2, \ldots, L\}$ are selected independently from zero-mean Gaussian distribution.

Subsequently, as the entries of different columns in matrix Θ are obtained by linear combinations of variables from two independent columns of Φ', the entries of different columns in Θ are independent. Furthermore, since the entries of same columns in matrix Θ can be considered as the linear combinations of variables from two independent rows of Ω, the entries of the same columns in Θ are independent.

According to the above results, we can conclude that the entries of matrix Θ are the independent zero-mean Gaussian variables. □

4.2. Joint decoding

The received data of sink node can be organized as in Eq. (17) and then decoded jointly.

Roughly speaking, a Gaussian matrix with independent zero-mean entries is incoherent with regard to any orthogonal basis. Moreover, the zero-mean Gaussian matrix with independently and identically distributed entries satisfies RIP with high probability [29]. So if matrix Θ satisfies the conditions described in Theorem 1, then resulting $\Theta.\Psi'$ will have the RIP with high probability.

When the matrices Θ, Ψ' and U are provided in the sink node, and the matrix Θ holds the conditions in Theorem 1, an approximate value $\tilde{\theta}'_i$ of $\theta'_i, i \in \{1, 2, \ldots, n\}$ can be calculated by the reconstruction algorithm of CS.

After the approximate value $\tilde{\theta}'_i$ is found, the approximate matrix $\tilde{Y}'$ of Y' can be easily recovered by computing

$$\tilde{Y}' = \Psi'.\begin{bmatrix}\tilde{\theta}'_1 & \tilde{\theta}'_2 & \cdots & \tilde{\theta}'_n\end{bmatrix} \tag{18}$$

where $Y' = \begin{bmatrix} y'_{1,1} & y'_{1,2} & \cdots & y'_{1,n} \\ y'_{2,1} & y'_{2,2} & \cdots & y'_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{L,1} & y'_{L,2} & \cdots & y'_{L,n} \end{bmatrix}$.

Then we have the following equation

$$Y' = \tilde{Y}' + E' \tag{19}$$

where E' denotes the error occurred in the reconstruction process of Y'.

Subsequently, we can obtain an equation as follows:

$$\begin{aligned} Y'_{(k-1)M+i} &= \tilde{Y}'_{(k-1)M+i} + E'_{(k-1)M+i} \\ &= \left[\Phi x_{k,i}\right]^T, k \in \{1, 2, \ldots, L\}, i \in \{1, 2, \ldots, M\} \end{aligned} \tag{20}$$

Table 1
Notations.

E	The set of network links
E_C	The set of network cliques
V	The set of cluster head nodes and sink node
g^l	Total flow rate of flow l
H^l	A routing matrix of data flow l
p_{ij}	The probability of link being reliable (i, j)
f^l_{ij}	The transmission rate of flow l over link (i, j)
C_{ij}	The associated capacity of link (i, j)

where the subscript $(k\text{-}1)M+i$ denotes the $[(k\text{-}1)M+i]$-*th* row of a matrix.

Eq. (20) can also be written as

$$y_{k,i} = \left[\tilde{Y}'_{(k-1)M+i} + E'_{(k-1)M+i}\right]^T = \Phi x_{k,i} = \Phi\Psi\theta_{k,i} \tag{21}$$

where $y_{k,i}$ is a batch of sampling data, which is compressible over orthonormal basis Ψ, and $\theta_{k,i}$ is the coefficient vector.

Afterward, an approximate value $\tilde{\theta}_{k,i}$ of $\theta_{k,i}$ can be recovered by the reconstruction algorithm of CS. Finally, the original signal of sensor node i in cluster k can be obtained by calculating the following approximate value

$$\tilde{x}_{k,i} = \Psi\tilde{\theta}_{k,i}, k \in \{1, 2, \ldots, L\}, i \in \{1, 2, \ldots, M\}. \tag{22}$$

5. Optimization design

Reconstruction error is an important performance metric for above compression scheme. Currently, the existent publications improve the reconstruction error mainly by training the sparse dictionary or optimizing the measurement matrix. These papers focus on the design of sparsifying transform and deterministic measurement matrix, which do not consider the network resources and scenario that can impact the reconstruction error significantly. In this section, we formulate a new model to optimize the reconstruction error of the proposed compression scheme from the perspective of network resources. The focus of this optimization is on the communication between cluster head nodes and one sink node, among which the cluster head nodes transmit data to the sink node, in the meantime, they also take as intermediate forwarders. In this network model, we only consider one sink node for collecting sensor readings, but our design can be generalized to multiple sink nodes. The observed data of different sensor nodes at a specific time slot is referred to as a data flow. The random NC operation implemented only among packets of the same flow (intra-flow NC).

The notations used in our optimization problem are explained in Table 1.

We assume that the data from cluster head nodes are transmitted to the neighboring cluster head nodes through multiple links and transmitted to the sink node along a multipath network predefined by a multipath routing protocol. The pair (i, j) denote the link from cluster head node i to its neighboring node j. Also $H^l_{ij} = 1$ indicates that the flow l traverses link (i, j), on the contrary, $H^l_{ij} = 0$ indicates that the flow l doesn't traverse link (i, j). The different links may interfere with each other due to overlapping in the communication range, so we define a clique as a basic conflict set that consists of several links which interfere with each other and at most one of them can carry data at the same time and avoid the interference. The set E_C consists of all the cliques in the network, i.e., a random clique E^S_C is an element of E_C.

Due to the unreliability of wireless links, we assume that the cluster head nodes forward their encoded data over links successfully with a certain probability, i.e., each link (i, j), $i, j \in V$ has a reliability probability p_{ij}, e.g., p_{ij}=0.9 means that only 90% of the

transmitted data over link (i, j) can be received successfully by the next hop node. The transmission rate f_{ij}^l varies from link to link.

5.1. Optimization problem formulation

In this optimization model, we will minimize the reconstruction error for a group flow G from above proposed compression scheme. As we known, the reconstruction error of a CS algorithm is determined by the amount of final received measurements and compressible (or sparse) degree of original data over orthonormal basis when the measurement matrix satisfies the RIP. According to [30], the average squared reconstruction error of spatial compression in data flow $l \in G$ is upper bounded by a constant times $(Lm/\log LM)^{-2\mu_l}$ where μ_l governs the compression degree of spatial dimension of data flow l. In addition, we assume that different data flows from G are all compressible signals in spatial dimension, i.e., different data flows from G have the same compression degree. Therefore, we define the reconstruction error of spatial compression in data flow l as a function which can be written as

$$R(g^l) = C_0 (g^l)^{-2\mu} \tag{23}$$

where the constant $C_0 > 0$, $\mu = \mu_l \geq 0$ is a constant related to the compression degree of data, and g^l is the total flow rate of flow l which is measured as the amount of received measurements per unit of time in the sink node.

Based on the network scenario defined above and the objective of our optimization problem, we can formulate the reconstruction error minimization problem of compression flows as follows:

$$\begin{aligned}
\min_{g,f} \quad & \sum_{l\in G} R(g^l) \\
\text{Subject to} \quad & g^l = \sum_{i\in V} g_i^l, \forall l \in G \\
& \sum_{j:H_{ij}^l=1} p_{ij} f_{ij}^l - \sum_{j:H_{ji}^l=1} p_{ji} f_{ji}^l = \sigma_i^l, \forall i \in V, \forall l \in G \\
& \frac{\sum_{l\in G}\sum_{(i,j)\in E_C^S} f_{ij}^l}{C_{ij}} \leq 1, \forall E_C^S \in E_C \\
& g^l, g_i^l, f_{ij}^l \geq 0, \forall i, j \in V, \forall l \in G.
\end{aligned} \tag{24}$$

The first constraint demonstrates that the total flow rate of received flow l at the sink node equals to the sum of the flow rates of all cluster head nodes, where the flow rate $g_D^l = 0$, and D denotes the sink node.

The second constraint is a flow conservation law that needs to be satisfied for transmitting data. Roughly speaking, the sum of incoming flow rates should be equal to the total outgoing flow rates, but the source and sink nodes are two exceptions. Because all the cluster head nodes are source nodes, the value of σ_i^l is given by

$$\sigma_i^l = \begin{cases} g_i^l, & if \quad i \in \{V - D\} \\ -g^l, & if \quad i = D \end{cases} \tag{25}$$

where $f_{Dj}^l = 0, j \in V$, and D denotes the sink node.

Intra-flow coding allows the same flow from different cluster head nodes within a link to share capacity by coding together. Meanwhile, the sum of the transmission rate of all flows $l \in G$ over link (i, j) should not exceed the overall capacity of link (i, j). It means that this optimization model should satisfy the following constraint

$$\sum_{l\in G} f_{ij}^l \leq C_{ij}, \forall i \in V. \tag{26}$$

However, the third constraint applies a more strict restriction on variable f_{ij}^l to take into account the effect of wireless interference. This constraint indicates that the mutual interfering links cannot transmit the information at the same time, i.e., the sum of occupancy rate of all links belonging to the same clique E_C^S must not exceed unity. This constraint also expresses the requirement that each cluster head node cannot transmit and receive data simultaneously, but it can either transmit or receive data at the same time. The symbol C_{ij} in the third constraint denotes the capacity of random link $(i, j) \in E_C^S$, and we assume that all links of network have equal capacity in this network scenario.

The other constraints are simple, g^l, g_i^l and f_{ij}^l are defined as non-negative.

5.2. Distributed algorithm

Since the objective function of the above optimization problem is a continuous and strictly convex function, and the constraints are linear, the optimization problem (24) is a convex optimization problem. Therefore, there is no duality gap, and we apply a dual decomposition method to solve this optimization problem. Firstly, we relax the second and third constraints in (24) to construct a Lagrangian function with multipliers λ_i^l and $\varepsilon_{E_C^S}$ as follows:

$$\begin{aligned}
& L(g, f, \lambda, \varepsilon) \\
& = \sum_{l\in G} R(g^l) - \sum_{i\in V, l\in G} \lambda_i^l \left(\sum_{j:H_{ij}^l=1} p_{ij} f_{ij}^l - \sum_{j:H_{ji}^l=1} p_{ji} f_{ji}^l - \sigma_i^l \right) \\
& \quad - \sum_{E_C^S \in E_C} \varepsilon_{E_C^S} \left[\left(\sum_{l\in G} \sum_{(i,j)\in E_C^S} f_{ij}^l \right) \Big/ C_{ij} - 1 \right] \\
& = \sum_{l\in G} \left[R\left(\sum_{i\in V} g_i^l \right) + \sum_{i\in\{V-D\}} \lambda_i^l g_i^l - \lambda_D^l \sum_{i\in V} g_i^l \right] \\
& \quad + \sum_{l\in G} \left\{ \sum_{i\in V} \sum_{j:H_{ij}^l=1} p_{ij} f_{ij}^l (\lambda_j^l - \lambda_i^l) \right. \\
& \quad \left. - \sum_{E_C^S \in E_C} \left[\left(\sum_{(i,j)\in E_C^S} \varepsilon_{E_C^S} f_{ij}^l \right) \Big/ C_{ij} \right] \right\} \\
& \quad + \sum_{E_C^S \in E_C} \varepsilon_{E_C^S}
\end{aligned} \tag{27}$$

where the Lagrange multipliers λ_i^l and $\varepsilon_{E_C^S}$ can be interpreted as the congestion price and interference price, respectively.

Next, the primal optimization problem can be formulated as follows:

$$D(\lambda, \varepsilon) = \min_{g,f\geq 0} L(g, f, \lambda, \varepsilon). \tag{28}$$

In order to solve the above primal problem we define the following dual problem.

$$\begin{aligned}
\max_{\lambda,\varepsilon} \quad & D(\lambda, \varepsilon) \\
\text{Subject to } & \lambda \geq 0; \varepsilon \geq 0.
\end{aligned} \tag{29}$$

Based on the separable property of Lagrange function and the definition of $D(\lambda, \varepsilon)$, $D(\lambda, \varepsilon)$ can be decomposed into two subproblems. The first subproblem can be considered as a congestion control problem which is written as

$$\begin{aligned}
& \min_{g} \quad L_1(g, \lambda) \\
& = \min_{g} \quad R\left(\sum_{i\in V} g_i^l \right) + \sum_{i\in\{V-D\}} \lambda_i^l g_i^l - \lambda_D^l \sum_{i\in V} g_i^l \\
& \text{Subject to } g_i^l \geq 0.
\end{aligned} \tag{30}$$

The second subproblem represents the combined constraints flow conservation and wireless interference, and it can be written as

$$\min_{f} \quad L_2(f,\lambda,\varepsilon)$$

$$= \min_{f} \quad \sum_{i\in V}\sum_{j:H_{ij}^l=1} p_{ij}f_{ij}^l(\lambda_j^l-\lambda_i^l) - \sum_{E_C^S\in E_C}\left[\left(\sum_{(i,j)\in E_C^S}\varepsilon_{E_C^S}f_{ij}^l\right)\Big/C_{ij}\right]$$

$$\text{Subject to } f_{ij}^l \geq 0. \tag{31}$$

The dual problem (29) can be solved by utilizing a subgradient algorithm. The dual variables can be updated as follows:

$$\lambda_i^l(r+1) = \left[\lambda_i^l(r)+\rho_\lambda\left(\sum_{j:H_{ij}^l=1} p_{ij}f_{ij}^l(r) - \sum_{j:H_{ji}^l=1} p_{ji}f_{ji}^l(r) - \sigma_i^l(r)\right)\right]^+ \tag{32}$$

$$\varepsilon_{E_C^S}(r+1) = \left[\varepsilon_{E_C^S}(r)+\rho_\varepsilon\left(\frac{\sum_{l\in G}\sum_{(i,j)\in E_C^S} f_{ij}^l(r)}{C_{ij}} - 1\right)\right]^+ \tag{33}$$

where the values of $g_i^l(r)$ and $f_{ij}^l(r)$ are selected from the optimal solution of subproblems (30) and (31) at the rth iteration, ρ_λ and ρ_ε are small step sizes corresponding to the dual variables, and $[g]^+ = g$ when g is a non-negative value, otherwise 0.

The solution of the above subproblems are described in the following parts.

Subproblem (30): According to the objective function of subproblem $L_1(g, \lambda)$, the optimal flow rate l of cluster head node $i \in \{V-D\}$ can be denoted as

$$g_i^{l*} = \arg\min_{g} L_1(g,\lambda) = \arg\min_{g}\left[R\left(\sum_{i\in\{V-D\}} g_i^l\right) + \sum_{i\in\{V-D\}}\lambda_i^l g_i^l - \lambda_D^l\sum_{i\in\{V-D\}} g_i^l\right]. \tag{34}$$

At the rth iteration, the cluster head node $i \in \{V-D\}$ adjusts its flow rate l according to the congestion prices $\lambda_i^l(r)$, $\lambda_D^l(r)$ and the flow rate $g^l(r)$, and the optimization problem (34) can be solved by the following method.

As explained in [31], the low-pass filtering method can remove oscillation and accelerate the convergence speed. So, we use the same concept to solve the optimization problem (34). In this way, the cluster head node $i \in \{V-D\}$ applies the following low-pass filtering method to update the flow rate g_i^l.

$$g_i^l(r+1) = \left\{(1-\gamma_1)g_i^l(r) + \gamma_1\bar{g}_i^l(r) + \gamma_1\left[C_0\left(\sum_{i\in\{V-D\}} g_i^l(r)\right)^{-2\mu-1} + \lambda_i^l(r) - \lambda_D^l(r)\right]\right\}^+ \tag{35}$$

$$\bar{g}_i^l(r+1) = (1-\gamma_1)\bar{g}_i^l(r) + \gamma_1 g_i^l(r) \tag{36}$$

where the augmented variable $\bar{g}_i^l$ is defined as the optimal estimation of g_i^l, and γ_1 is a small step size. Finally, to calculate the g^l:

$$g^l(r+1) = \sum_{i\in\{V-D\}} g_i^l(r+1). \tag{37}$$

Subproblem (31): According to the objective function of subproblem $L_2(f, \lambda, \varepsilon)$, the optimal transmission rate of flow l over link (i, j) can be denoted as

$$f_{ij}^{l*} = \arg\min_{f} L_2(f,\lambda,\varepsilon) = \arg\min_{f}\left\{\sum_{i\in V}\sum_{j:H_{ij}^l=1} p_{ij}f_{ij}^l(\lambda_j^l-\lambda_i^l) - \sum_{E_C^S\in E_C}\left[\left(\sum_{(i,j)\in E_C^S}\varepsilon_{E_C^S}f_{ij}^l\right)\Big/C_{ij}\right]\right\}. \tag{38}$$

At the rth iteration, each link (i, j) adjusts its transmission rate of flow l according to the congestion prices $\lambda_i^l(r)$, $\lambda_j^l(r)$ and interference price $\varepsilon_{E_C^S}(r)$, and the optimization problem (38) can be solved by the following joint method.

First, we obtain the partial derivative of function $L_2(f, \lambda, \varepsilon)$ with respect to the variable f_{ij}^l:

$$\frac{\partial L_2(f,\lambda,\varepsilon)}{\partial f_{ij}^l} = p_{ij}(\lambda_j^l-\lambda_i^l) - \varepsilon_{E_C^S}/C_{ij} \tag{39}$$

and then we construct a joint method by employing the first-order Lagrangian algorithm and low-pass filtering method. Finally, the link (i, j) updates the transmission rate of flow l in the following format

$$f_{ij}^l(r+1) = \{(1-\gamma_2)f_{ij}^l(r) + \gamma_2\bar{f}_{ij}^l(r) + \gamma_2[p_{ij}(\lambda_j^l(r)-\lambda_i^l(r)) - \varepsilon_{E_C^S}(r)/C_{ij}]\}^+_{H_{ij}^l=1,(i,j)\in E_C^S} \tag{40}$$

$$\bar{f}_{ij}^l(r+1) = (1-\gamma_2)\bar{f}_{ij}^l(r) + \gamma_2 f_{ij}^l(r) \tag{41}$$

where the augmented variable $\bar{f}_{ij}^l$ is defined as the optimal estimation of f_{ij}^l, and γ_2 is a small step size.

Integrating the content described above, we can obtain the following distributed optimization algorithm.

Algorithm 1. $r = 1, 2, 3, \ldots$, when the rth iteration is executing, this distributed optimization algorithm can be shown as

(1) Cluster head node $i \in \{V-D\}$ solves the optimization problem (34) according to the congestion prices $\lambda_i^l(r)$, $\lambda_D^l(r)$ and the flow rate $g^l(r)$;
(2) Each link (i, j), $i \in \{V-D\}$ solves the optimization problem (38) according to the congestion prices $\lambda_i^l(r)$, $\lambda_j^l(r)$ and interference price $\varepsilon_{E_C^S}(r)$;
(3) Each link (i, j), $i \in V$ updates the congestion price $\lambda_i^l(r+1)$ according to Eq. (32) and the solutions of step (1) and (2);
(4) Cluster head node i updates the interference price $\varepsilon_{E_C^S}(r+1)$ according to Eq. (33) and the solution of step (2);
(5) Setting r=r+1, then the procedure goes back to step (1) and executes repeatedly until iterations end.

Although Algorithm 1 is a distributed algorithm, it still needs some feedback messages. For example, in order to solve the optimization problem (38), the transmission rates of i's neighbor nodes should be sent to cluster head node i, and the congestion price $\lambda_j^l(r)$ of next hop node j should be fed back to node i. Fortunately, the communication cost of feedback messages is low for only transmitting a small amount of digits and most of messages can be included in ACK packet.

5.3. Convergence analysis

In this subsection, we analyze the convergence of Algorithm 1. Assume that $(\lambda^*, \varepsilon^*)$ denotes the optimal solution of dual problem, and (g^*, f^*) denotes the optimal solution of primal problem. We have the following convergence result by employing the convergence analysis method in [32].

Theorem 2. *If step sizes ρ_λ and ρ_ε are sufficiently small, and the initial values of g, f, λ and ε are all not less than 0, then the dual variables of Algorithm 1 converge statistically to the optimal solution $(\lambda^*, \varepsilon^*)$ of dual problem (29).*

Proof. We define the $\nabla(r)$ as the subgradient vector of dual problem (29), then $\nabla_i^l(r)$ and $\nabla_{E_C^S}(r)$ can be shown as

$$\nabla_i^l(r) = \sum_{j:H_{ij}^l=1} p_{ij} f_{ij}^l(r) - \sum_{j:H_{ji}^l=1} p_{ji} f_{ji}^l(r) - \sigma_i^l(r) \tag{42}$$

$$\nabla_{E_C^S}(r) = \frac{\sum_{l\in G}\sum_{(i,j)\in E_C^S} f_{ij}^l(r)}{C_{ij}} - 1. \tag{43}$$

Meanwhile, we define the Lyapunov function as follows:

$$\dot{V}(\lambda(r), \varepsilon(r)) = \frac{1}{2\rho_\lambda}\sum_{i,l}(\lambda_i^{l*} - \lambda_i^l(r))^2 + \frac{1}{2\rho_\varepsilon}\sum_{E_C^S}(\varepsilon_{E_C^S}^* - \varepsilon_{E_C^S}(r))^2. \tag{44}$$

According to Eqs. (32) and (33), we have

$$\begin{aligned}
&\dot{V}(\lambda(r+1), \varepsilon(r+1)) \\
&\le \frac{1}{2\rho_\lambda}\sum_{i,l}[\lambda_i^{l*} - \lambda_i^l(r) - \rho_\lambda\nabla_i^l(r)]^2 \\
&\quad + \frac{1}{2\rho_\varepsilon}\sum_{E_C^S}[\varepsilon_{E_C^S}^* - \varepsilon_{E_C^S}(r) - \rho_\varepsilon\nabla_{E_C^S}(r)]^2 \\
&= \frac{1}{2\rho_\lambda}\sum_{i,l}\{[\lambda_i^{l*} - \lambda_i^l(r)]^2 - 2\rho_\lambda\nabla_i^l(r)[\lambda_i^{l*} - \lambda_i^l(r)] + \rho_\lambda^2\nabla_i^{l2}(r)\} \\
&\quad + \frac{1}{2\rho_\varepsilon}\sum_{E_C^S}\{[\varepsilon_{E_C^S}^* - \varepsilon_{E_C^S}(r)]^2 - 2\rho_\varepsilon\nabla_{E_C^S}(r)[\varepsilon_{E_C^S}^* - \varepsilon_{E_C^S}(r)] \\
&\quad + \rho_\varepsilon^2\nabla_{E_C^S}^2(r)\} \\
&= \dot{V}(\lambda(r), \varepsilon(r)) + \sum_{i,l}\nabla_i^l(r)[\lambda_i^l(r) - \lambda_i^{l*}] \\
&\quad + \sum_{E_C^S}\nabla_{E_C^S}(r)[\varepsilon_{E_C^S}(r) - \varepsilon_{E_C^S}^*] \\
&\quad + \sum_{i,l}\frac{\rho_\lambda}{2}\nabla_i^{l2}(r) + \sum_{E_C^S}\frac{\rho_\varepsilon}{2}\nabla_{E_C^S}^2(r).
\end{aligned} \tag{45}$$

Based on the subgradient property, we can obtain

$$\begin{aligned}
&\sum_{i,l}\nabla_i^l(r)[\lambda_i^l(r) - \lambda_i^{l*}] + \sum_{E_C^S}\nabla_{E_C^S}(r)[\varepsilon_{E_C^S}(r) - \varepsilon_{E_C^S}^*] \\
&\le D(\lambda(r), \varepsilon(r)) - D(\lambda^*, \varepsilon^*).
\end{aligned} \tag{46}$$

Substituting the corresponding part with the formula (46), then the inequality (45) can be reformulated as

$$\begin{aligned}
&\dot{V}(\lambda(r+1), \varepsilon(r+1)) \\
&\le \dot{V}(\lambda(r), \varepsilon(r)) - [D(\lambda^*, \varepsilon^*) - D(\lambda(r), \varepsilon(r))] + \frac{\rho_\lambda}{2}Q_1 + \frac{\rho_\varepsilon}{2}Q_2
\end{aligned} \tag{47}$$

where $\sum_{i,l}\nabla_i^{l2}(r) \le Q_1$ and $\sum_{E_C^S}\nabla_{E_C^S}^2(r) \le Q_2$.

Utilizing the inequality recursively, the inequality (47) can be recast in the following form:

$$\begin{aligned}
&\dot{V}(\lambda(r+1), \varepsilon(r+1)) \\
&\le \dot{V}(\lambda(1), \varepsilon(1)) - \sum_{\tau=1}^{r}[D(\lambda^*, \varepsilon^*) - D(\lambda(\tau), \varepsilon(\tau))] + \frac{\rho_\lambda}{2}rQ_1 \\
&\quad + \frac{\rho_\varepsilon}{2}rQ_2.
\end{aligned} \tag{48}$$

By reason of $\dot{V}(\lambda(r+1), \varepsilon(r+1)) \ge 0$, we have

$$\begin{aligned}
&\sum_{\tau=1}^{r}[D(\lambda^*, \varepsilon^*) - D(\lambda(\tau), \varepsilon(\tau))] \\
&\le \dot{V}(\lambda(1), \varepsilon(1)) + \frac{\rho_\lambda}{2}rQ_1 + \frac{\rho_\varepsilon}{2}rQ_2.
\end{aligned} \tag{49}$$

Furthermore, since $D(\lambda(\tau), \varepsilon(\tau))$ is a concave function with respect to variables λ and ε, we can get the following inequality via Jensen's inequality.

$$D(\lambda^*, \varepsilon^*) - D(\bar{\lambda}(r), \bar{\varepsilon}(r)) \le \frac{\dot{V}(\lambda(1), \varepsilon(1))}{r} + \frac{\rho_\lambda}{2}Q_1 + \frac{\rho_\varepsilon}{2}Q_2 \tag{50}$$

where $\bar{\lambda}(r) = \frac{1}{r}\sum_{\tau=1}^{r}\lambda(r)$ and $\bar{\varepsilon}(r) = \frac{1}{r}\sum_{\tau=1}^{r}\varepsilon(r)$.

Finally, we can conclude that

$$\limsup_{t\to\infty}[D(\lambda^*, \varepsilon^*) - D(\bar{\lambda}(r), \bar{\varepsilon}(r))] \le \frac{\rho_\lambda}{2}Q_1 + \frac{\rho_\varepsilon}{2}Q_2. \tag{51}$$

According to the definition of statistical convergence in [32], given the small enough step sizes ρ_λ and ρ_ε, the dual variables of Algorithm 1 converge statistically to the optimal solution $(\lambda^*, \varepsilon^*)$ of dual problem (29). □

Since the primal optimization problem (24) is a convex programming problem, there is no duality gap. In other words, the optimization problems (34) and (38) converge statistically to the optimal solution (g^*, f^*) of primal problem (24) when dual variables converge to the optimal solution $(\lambda^*, \varepsilon^*)$.

6. Performance evaluation

This section provides the performance analysis to show the robustness and efficiency of the proposed clustered spatio-temporal compression scheme and the distributed optimization model in the previous sections.

6.1. Performance analysis: compression schemes

In this subsection, we compare the simulation results of relative recovery error and compression gain of our method with those of two other categories of compression schemes. The algorithm CoSaMP [33] is used as the reconstruction algorithm of CS for spatial and temporal compression data in this simulation. The simulation uses the real data collected by Intel Berkeley research lab [34]. 40 × 500 (i.e., $L.M\times N$) temperature readings are selected from this WSN, where we divide these sensors into 4 clusters and set the spatial and temporal readings as s and w-compressible signals, respectively. We also assume that the value of s and w equals to 5 and 60, respectively.

In the simulation figures, 'Spatial' refers to the category of compression schemes which only considers the spatial compression, such as [7–9]; 'Spatio-temporal' denotes the category of compression schemes which considers the spatial and temporal compression separately, such as [13,14]; 'Clustered spatio-temporal' denotes the proposed scheme in this paper.

The relative recovery error and compression gain are defined to be $\|x - \tilde{x}\|_2/\|x\|_2$ and $(L.M.N)/[L.m(n+L.m)]$, respectively, where the original 40 × 500 readings are reshaped into a 20,000 × 1

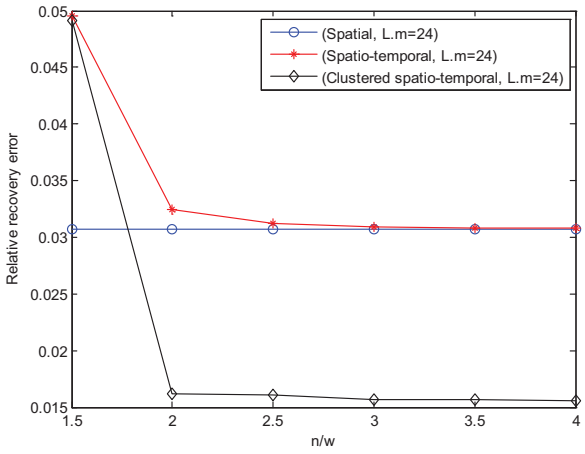

Fig. 3. Relative recovery error with different overmeasuring factor.

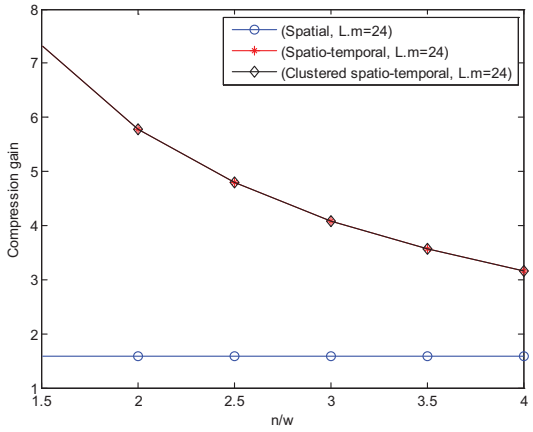

Fig. 4. Compression gain with different overmeasuring factor.

vector x, and $\tilde{x}$ is an approximate reconstruction vector of x. The overmeasuring factor n/w of temporal dimension varies from 1.5 to 4.

In this simulation scenario, the orthonormal bases Ψ and $\Psi_i', i \in \{1, 2, \ldots, L\}$ are defined as discrete cosine transform. The measurement matrix Φ consists of n rows selected randomly from an $N \times N$ identity matrix. The elements of measurement matrix $\Phi_i', i \in \{1, 2, \ldots, L\}$ are selected from Gaussian distribution $N(0, 1)$. The entries of Ω are drawn randomly from Gaussian distribution $N(0, 1/(L.m))$.

Fig. 3 displays the relationship between relative recovery error and overmeasuring factor in the temporal dimension when the row size of matrix Φ' is fixed to 24. The curves show that the relative recovery error decreases gradually with the increase of value n/w, and the reconstruction error will stay stable around a fixed value when the value of overmeasuring factor n/w is more than 3. Since the selected sensor readings are compressible signals rather than sparse signals, when the recovery error approaches a minimum value, further increase of the value of n/w will has a negligible effect on the reconstruction error.

Meanwhile, observe from this figure that the recovery error of our proposed scheme is lower than other two schemes except for n/w=1.5. Since the compression ratio of temporal dimension is too large when the overmeasuring factor n/w=1.5, the recovery error of spatial compression scheme is lower than our proposed scheme. However, with the increase of overmeasuring factor, the proposed scheme shows lower recovery error. For example, while the overmeasuring factor $n/w \geq 3$, the relative recovery error of our scheme is around 0.0156 compared to 0.0308 of other schemes. As the spatial correlation can be explored more completely and deeply in clustered spatio-temporal compression scheme, the recovery error is much lower than conventional spatio-temporal schemes [13,14] especially when the temperature readings fluctuate remarkably among faraway nodes in the same time slot.

In addition, the proposed scheme holds another advantage in that it has lower computational complexity as compared with JSM-based spatio-temporal schemes [10–12] and the work [15] in the reconstruction process of CS, because the temporal and spatial correlations of original sensing data are explored step by step. Thereinto, the computational complexities of the proposed scheme and the works [10–12,15] in the reconstruction process are $\max\{O(L^2mM), O(nN)\}$ and $O(L^2mMnN)$, respectively, and the proposed scheme has the same computational complexity as the conventional spatio-temporal schemes [13,14].

With the growth of overmeasuring factor, there is diminishing return in the compression gain as the number of measurements increases. Fig. 4 demonstrates that the compression gain of spatio-temporal compression scheme is higher than the spatial compression scheme as the spatial scheme only investigates the spatial correlation. In this simulation scenario, the clustered spatio-temporal scheme has the same compression gain as the conventional spatio-temporal schemes [13,14].

Combining Figs. 3 and 4 we can derive that the relative recovery error of our scheme is much lower than the conventional spatio-temporal schemes [13,14] while they have the same compression gain (also have the same computational complexity). Besides, the performance of our scheme is better than spatial compression schemes [7–9] in terms of both recovery error and compression gain. Moreover, the decrease in the amount of redundant information is considerable by utilizing the spatial and temporal correlations step by step. It also reveals that the clustered spatio-temporal scheme can reduce the energy consumption of communications significantly. Although the transmission of NC encoding coefficients will decrease the compression gain, this effect will be weakened by extending the packet length.

Overall, the clustered spatio-temporal compression scheme outperforms other two categories of compression schemes significantly in terms of recovery error and compression gain.

6.2. Performance analysis: optimization design

In this subsection, we provide a certain network to test the convergence of Algorithm 1, and evaluate the effects of network and control parameters on performance. The considered network consists of five cluster head nodes and one sink node as shown in Fig. 1. The network has ten links and all links have equal capacity. There are five cliques in the network.

The parameters are initialized as follows. The step sizes of different optimization problem are unified. The reliable probabilities of different links are unified to be 0.9. The constant C_0 and μ are equal to 1 (or 0.01) and 0.9, respectively. The initial prices are set to be 1 for both congestion and interference prices. The initial value of transmission rates are set to be 0.2, 0.3, 0.3, 0.6, 0.6 for node 1 to 5, respectively.

Figs. 5 and 6 show the evolution of congestion and interference prices respectively with link capacity C_{ij}=1.5 and step size =0.15. It can be seen from both figures that the congestion and interference prices converge gradually along with the number of iterations, and both of them approach the optimal value within 200 iterations. Thus, the convergence speed of the dual problem is fast, which further confirms the validity of Theorem 2.

From Fig. 5, we can find out that the congestion price of node 6 (i.e., sink node) is bigger than other nodes. Since node 6 is the only

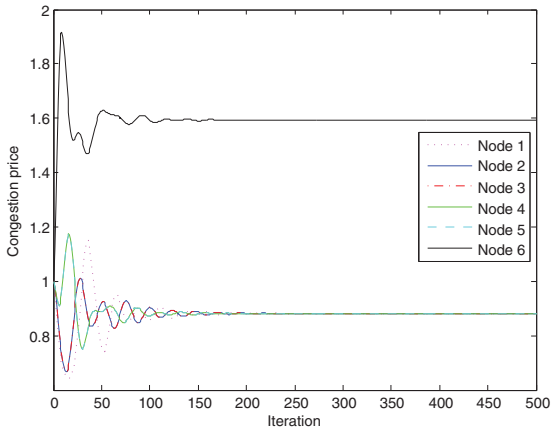

Fig. 5. Evolution of congestion prices with C_{ij}=1.5 and step size=0.15.

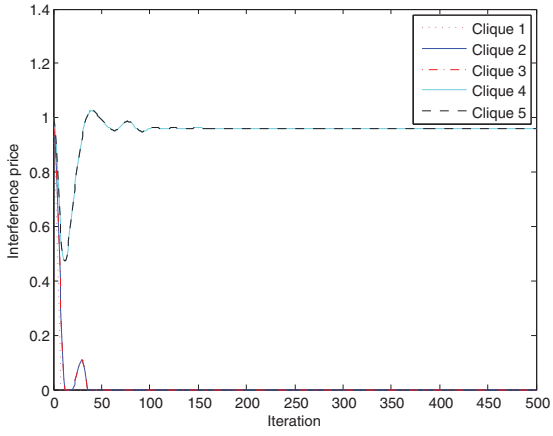

Fig. 6. Evolution of interference prices with C_{ij}=1.5 and step size=0.15.

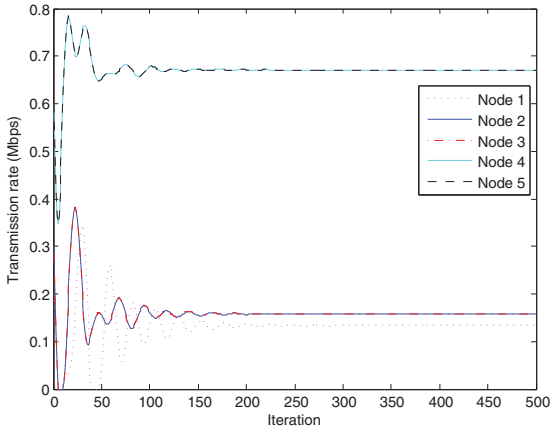

Fig. 7. Evolution of transmission rates with C_{ij}=1.5 and step size=0.15.

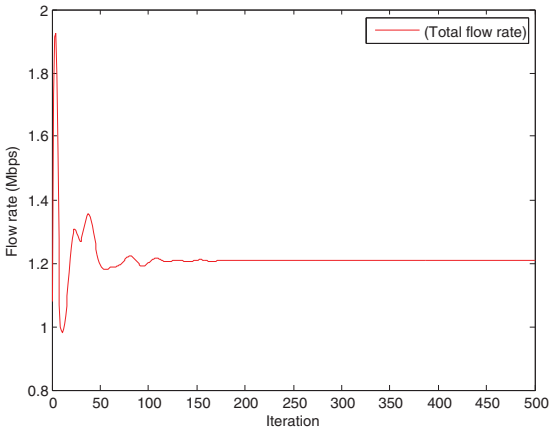

Fig. 8. Evolution of flow rate with C_{ij}=1.5 and step size=0.15.

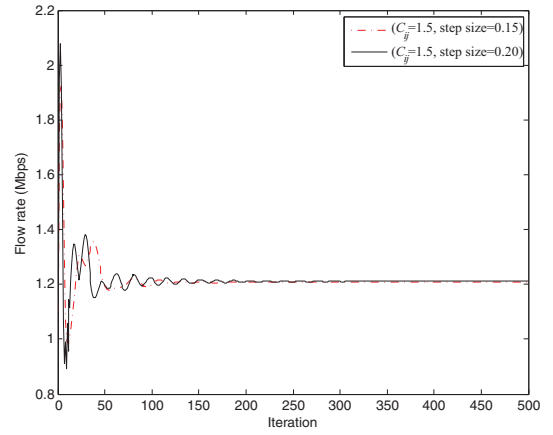

Fig. 9. Evolution of flow rates with different step sizes.

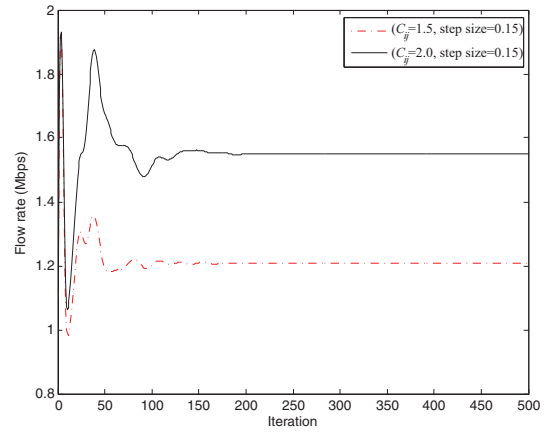

Fig. 10. Evolution of flow rates with different link capacities.

sink node for collecting the data from all the cluster head nodes, the links entering into node 6 are very busy which causes the high congestion price of node 6. Similarly, the interference prices of cliques 4 and 5 are high in Fig. 6 due to including node 6 in their cliques.

Figs. 7 and 8 plot the evolution of transmission rates and flow rate with link capacity C_{ij}=1.5 and step size =0.15. The convergence speed of transmission rates and flow rate is fast, which is similar to the evolution of congestion and interference prices in Figs. 5 and 6. Meanwhile, the transmission rates and flow rate converge to the corresponding optimal value while the congestion and interference prices approach the optimal value. This result also confirms the correctness of theory described in previous section that the primal variables (g^*, f^*) is the optimal solution of primal problem when dual variables converge to the optimal solution (λ^*, ε^*).

It can be also observed in Fig. 7 that the transmission rates of nodes 4 and 5 are higher than other nodes. This is due to the fact that nodes 4 and 5 are near the sink node and need to relay data from other nodes.

Fig. 9 presents the effect of different step sizes on flow rate. It shows that the larger the step size, the bigger the oscillation and the faster the convergence speed. Although a small step size has lower convergence speed, it will obtain a solution with higher ac-

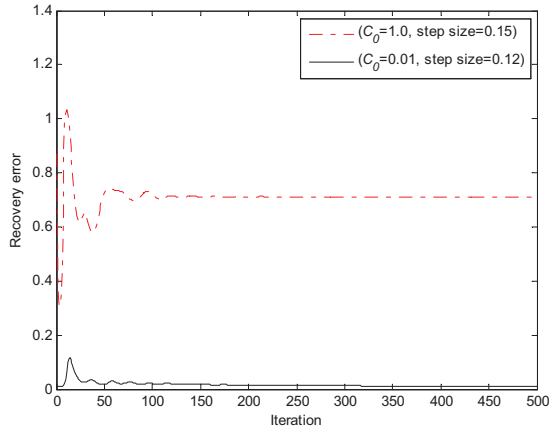

Fig. 11. Evolution of reconstruction errors with different C_0 and step sizes.

curacy. The users can dynamically set the step size, and select a proper size by making a tradeoff between convergence speed and precision of the optimal solution.

Fig. 10 illustrates the evolution of flow rates with different link capacities. Obviously, the flow rate will enhance with the increase of link capacity. Under different link capacities, both flow rates can converge to optimal value which further confirms the stability of Algorithm 1.

Fig. 11 shows the evolution of recovery errors with different C_0 and step sizes. Although the value of constant C_0 will affect the value of recovery error, the convergence property of recovery error is constant. The curves demonstrate that the recovery error converges to a stable value around 200 iterations, which also confirm that the objective function of primal problem (24) achieves the optimal value when the flow rate approaches the optimal value, i.e., the reconstruction error can be minimized by optimizing the network resource allocation.

7. Conclusion

Based on the temporal and spatial correlations of sensor readings, this paper proposed a clustered spatio-temporal compression scheme to reduce the number of transmissions and formulated a new optimization model to minimize the reconstruction error. The compression scheme could reduce the number of transmissions significantly. In the meantime, the design of NC encoding coefficients and measurement matrix was given for guaranteeing the reconstruction of clustered compression data successfully with an overwhelming probability. The proposed scheme also had lower reconstruction error and computational complexity by employing independent encoding in each sensor node and joint decoding in the sink node. In addition, in order to minimize the reconstruction error, a distributed algorithm was developed to achieve the optimal solution. Finally, the simulation results further confirmed the properties of the clustered spatio-temporal compression scheme and optimization model.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (Nos. 61201160, 61471025, 61373135, 61672299), the Natural Science Foundation of Jiangsu Province (No. BK20131377), the China Scholarship Council Project (No. 201408320091), Six Talented Eminence Foundation of Jiangsu Province (No. XYDXXJS-044) and the Scientific Research Foundation of Nanjing University of Posts and Telecommunications (No. NY213048).

References

[1] E.J. Candes, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theor. 52 (2) (2006) 489–509.
[2] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theor. 52 (4) (2006) 1289–1306.
[3] R.G. Baraniuk, Compressive sensing, IEEE Signal Process. Mag. 24 (4) (2007) 118–120+124.
[4] R. Ahlswede, N. Cai, S.R. Li, R.W. Yeung, Network information flow, IEEE Trans. Inf. Theor. 46 (5) (2000) 1204–1216.
[5] R. Xie, X. Jia, Transmission-efficient clustering method for wireless sensor networks using compressive sensing, IEEE Trans. Parallel Distrib. Syst. 25 (3) (2014) 806–815.
[6] C. Luo, J. Sun, F. Wu, Compressive network coding for approximate sensor data gathering, in: Proceedings of IEEE GLOBECOM, 2011, pp. 1–6.
[7] X. Yang, X. Tao, E. Dutkiewicz, et al., Energy-efficient distributed data storage for wireless sensor networks based on compressed sensing and network coding, IEEE Trans. Wireless Commun. 12 (10) (2013) 5087–5099.
[8] M. Nabaee, F. Labeau, Restricted isometry property in quantized network coding of sparse messages, in: Proceedings of IEEE GLOBECOM, 2012, pp. 112–117.
[9] M. Nabaee, F. Labeau, Quantized network coding for correlated sources, EURASIP J. Wireless Commun. Netw. 2014 (2014) 1–17.
[10] M. Mahmudimanesh, A. Khelil, N. Suri, Balanced spatio-temporal compressive sensing for multi-hop wireless sensor networks, in: Proceedings of IEEE International Conference on Mobile Ad-Hoc SenssorSystems, 2012, pp. 389–397.
[11] S. Chen, M. Wu, K. Wang, et al., Compressive network coding for error control in wireless sensor networks, Wireless Netw. 20 (8) (2014) 2605–2615.
[12] L. Kong, M. Xia, X. Liu, et al., Data loss and reconstruction in wireless sensor networks, IEEE Trans. Parallel Distrib. Syst. 25 (11) (2014) 2818–2828.
[13] S. Feizi, M. Medard, A power efficient sensing/communication scheme: joint source-channel-network coding by using compressive sensing, in: Proceedings of 49th Annual Allerton Conference on Communication, Control, and Computing, 2011, pp. 1048–1054.
[14] D. Lee, J. Choi, Low complexity sensing for big spatio-temporal data, in: Proceedings of IEEE International Conference on Big Data, 2014, pp. 323–328.
[15] B. Gong, P. Cheng, N. Liu, et al., Spatiotemporal compressive network coding for energy-efficient distributed data storage in wireless sensor networks, IEEE Commun. Lett. 19 (5) (2015) 803–806.
[16] L. Chen, T. Ho, M. Chiang, et al., Congestion control for multicast flows with network coding, IEEE Trans. Inf. Theor. 58 (9) (2012) 5908–5921.
[17] R. Hou, K. Lui, J. Li, Joint congestion control and scheduling in wireless networks with network coding, IEEE Trans. Veh. Technol. 63 (7) (2014) 3304–3317.
[18] C. Tan, J. Zou, M. Wang, et al., Network lifetime optimization for wireless video sensor networks with network coding/ARQ hybrid adaptive error-control scheme, Comput. Netw. 55 (9) (2011) 2126–2137.
[19] R. Niati, A.H. Banihashemi, T. Kunz, Throughput and energy optimization in wireless networks: joint MAC scheduling and network coding, IEEE Trans. Veh. Technol. 61 (3) (2012) 1372–1382.
[20] Z. Chen, T.J. Lim, M. Motani, Digital network coding aided two-way relaying: energy minimization and queue analysis, IEEE Trans. Wireless Commun. 12 (4) (2013) 1947–1957.
[21] M. Alvandi, M. Mehmet-Ali, J.F. Hayes, Delay optimization and cross-layer design in multihop wireless networks with network coding and successive interference cancelation, IEEE J. Sel. Areas Commun. 33 (2) (2015) 295–308.
[22] C. Zhao, Y. Luo, F. Chen, et al., Energy effective congestion control for multicast with network coding in wireless ad hoc network, Math. Probl. Eng. 2014 (2014) 1–12.
[23] P. Sadeghi, M. Yu, N. Aboutorab, On throughput-delay tradeoff of network coding for wireless communications, in: Proceedings of ISITA, 2014, pp. 689–693.
[24] J. Liu, C.R. Lin, Cross-layer optimization for performance trade-off in network code-based wireless multi-hop networks, Comput. Commun. 52 (2014) 89–101.
[25] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theor. 53 (12) (2007) 4655–4666.
[26] D.L. Donoho, Y. Tsaig, I. Drori, et al., Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit, IEEE Trans. Inf. Theor. 58 (2) (2012) 1094–1121.
[27] E.J. Candes, Compressive sampling, in: Proceedings of International Congrgrss of Mathematicians, 2006, pp. 1433–1452.
[28] E.J. Candes, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theor. 51 (12) (2005) 4203–4215.
[29] R. Baraniuk, M. Davenport, R. Devore, et al., A simple proof of the restricted isometry property for random matrices, Constructive Approx. 28 (3) (2008) 253–263.
[30] J. Haupt, R. Nowak, Signal reconstruction from noisy random projections, IEEE Trans. Inf. Theor. 52 (9) (2006) 4036–4048.
[31] J. Jin, W. Wang, M. Palaniswami, Utility max-min fair resource allocation for communication networks with multipath routing, Comput. Commun. 32 (17) (2009) 1802–1809.
[32] L. Chen, S.H. Low, M. Chiang, et al., Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks, in: Proceedings of INFOCOM, 2006, pp. 1–13.
[33] D. Needell, J.A. Tropp, CosaMP: iterative signal recovery from incomplete and inaccurate samples, Commun. ACM 53 (12) (2010) 93–100.
[34] Intel Lab Data, http://db.csail.mit.edu/labdata/labdata.html.

Siguang Chen is currently an Associate Professor at Nanjing University of Posts and Telecommunications. He received his Ph.D. in information security from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2011. He finished his Postdoctoral research work in City University of Hong Kong in 2012. From 2014 to 2015, he also was a Postdoctoral Fellow in the University of British Columbia. His research interests are in the area of dependable and secure network coding, wireless network resource optimization and the interplay between network coding and compressed sensing.

Chuanxin Zhao is currently a Research Associate at Curtin University and an Associate Professor at Anhui Normal University. He received his Ph.D. in computer science from Suzhou University, Suzhou, China, in 2012. His main research interests include network coding and wireless network resource management and Optimization.

Meng Wu received his B.S., M.S. and Ph.D. degrees in communication engineering and computer science from Zhenjiang University, Shanghai Jiaotong University, Southeast University, in 1985, 1990 and 1993, respectively. Currently, he is a Professor of Nanjing University of Posts and Telecommunications. His main research areas are wireless communications, secure network coding, sensor network and the related information security.

Zhixin Sun received his Ph.D. from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1998. He finished his postdoctoral research work in Seoul National University in 2002. Currently, he is a Professor of Nanjing University of Posts and Telecommunications. His current research interests include computer network and security, network multimedia communications and network management and protocol.

Haijun Zhang received his Ph.D. degree from Beijing University of Posts Telecommunications (BUPT). He hold a Postdoctoral Research Fellow position in Department of Electrical and Computer Engineering, the University of British Columbia (UBC). He was an Associate Professor in College of Information Science and Technology, Beijing University of Chemical Technology. From 2011 to 2012, he visited Centre for Telecommunications Research, King's College London, London, UK, as a joint PhD student and Research Associate. He has published more than 50 papers and has authored 2 books. He serves as the editors of Journal of Network and Computer Applications, Wireless Networks (Springer), and KSII Transactions on Internet and Information Systems. He served as Symposium Chair of GAMENETS'2014 and Track Chair of ScalCom'2015. He also serves or served as TPC members of many IEEE conferences, such as Globecom and ICC. His current research interests include 5G, Resource Allocation, Heterogeneous Small Cell Networks and Ultra-Dense Networks.

Victor C.M. Leung is a Professor of Electrical and Computer Engineering and holder of the TELUS Mobility Research Chair at the University of British Columbia (UBC). His research is in the areas of wireless networks and mobile systems. He has coauthored more than 800 technical papers in archival journals and refereed conference proceedings, several of which had won best paper awards. Dr. Leung is a Fellow of IEEE, a Fellow of the Royal Society of Canada, a Fellow of the Canadian Academy of Engineering and a Fellow of the Engineering Institute of Canada. He is serving or has served on the editorial boards of JCN, IEEE JSAC, Transactions on Computers, Wireless Communications, and Vehicular Technology, Wireless Communications Letters, and several other journals. He has provided leadership to the technical program committees and organizing committees of numerous international conferences. Dr. Leung was the recipient of the 1977 APEBC Gold Medal, NSERC Postgraduate Scholarships from 1977 to 1981, a 2012 UBC Killam Research Prize, and an IEEE Vancouver Section Centennial Award.