Minimizing Transmission Cost for Third-Party Information Exchange with Network Coding

Xiumin Wang, Chau Yuen, Tiffany Jing Li, Wentu Song, and Yinlong Xu

Abstract—In wireless networks, getting the global knowledge of channel state information (CSI, e.g., channel gain or link loss probability) is always beneficial for the nodes to optimize the network design. However, the node usually only has the local CSI between itself and other nodes, and lacks the CSI between any pair of other nodes. To enable all the nodes to get the global CSI, in this paper, we propose a network-coded third-party information exchange scheme, with an emphasis on minimizing the total transmission cost for exchanging the CSI among the nodes. We show that for a network of N nodes, if and only if any k nodes ($1 \leq k < N$) send at least $\binom{k}{2}$ packets, a feasible solution exists for third-party information exchange. Formulating the problem of feasible and optimal solutions as an integer linear programming (ILP) problem, we compute the optimal number of packets that must be transmitted by every node. Guided by the necessary and sufficient condition, we construct two practical transmission schemes: fair load (FL) scheme and proportional load (PL) scheme. A deterministic encoding strategy based on XORs coding over GF(2) is further designed to guarantee that with FL or PL scheme, each node finally can decode the complete packets. It is shown that in two specific networks, these two schemes are optimal, achieving the minimum transmission cost. In more general networks, simulation results show that PL is still close to optimal with a high probability. Finally, a distributed transmission protocol is developed, which allows FL and PL schemes to be operated in a distributed and hence scalable manner.

Index Terms—Network coding, cooperative data exchange, channel state information

1 Introduction

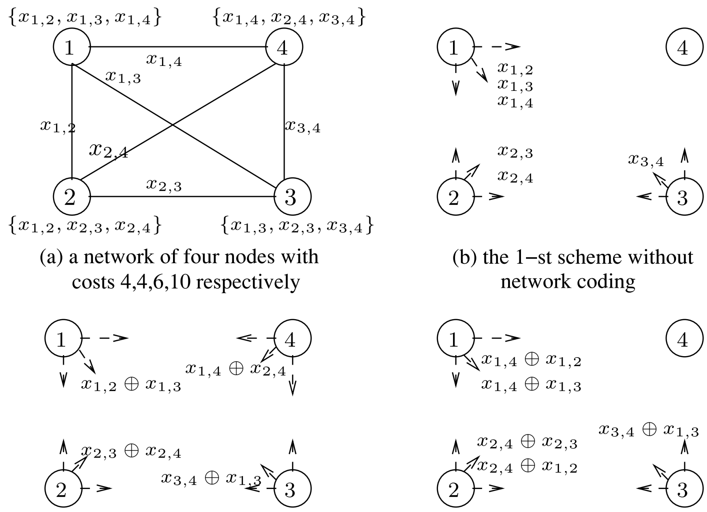
In wireless networks, it is always beneficial for the nodes to know the global knowledge of channel state information (CSI) such as the channel gain or the link loss probability. This is because global knowledge of CSI not only constitutes the necessary condition for the network design [1], [2], [3], [4], but also greatly eases the network optimization and improves the system performance in general. Consider the CSI of a connected link between node i and node j. This information is considered local to i and j, and can be obtained fairly easily [5]. However, for a *third-party* node k (where $k \neq i, j$), the channel information of link (i, j) becomes some sort of global knowledge and needs to take special arrangement to obtain it. The problem, referred to as "third-party information exchange", was first proposed in [1], [2]. The objective of third-party information exchange is to develop deterministic and efficient algorithms to enable all the nodes to obtain the complete global information by exchanging packets among themselves.

- X. Wang is with the School of Computer and Information, Hefei University of Technology, Hefei, China, and the State Key Laboratory of Integrated Services Networks, Xidian University, China. E-mail: wxiumin@hfut.edu.cn.
- C. Yuen and W. Song are with Singapore University of Technology and Design, Singapore. E-mail: {yuenchau, wentu_song}@sutd.edu.sg.
- T.J. Li is with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18105. E-mail: jingli@ece.lehigh.edu.
- Y. Xu is with Anhui Province Key Laboratory of High Performance Computing, School of Computer Science and Technology, University of Science and Technology of China, Hefei, China. E-mail: ylxu@ustc.edu.cn.

Manuscript received 25 Dec. 2013; revised 18 May 2014; accepted 14 Aug. 2014. Date of publication 21 Aug. 2014; date of current version 1 May 2015. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TMC.2014.2350989

Closely related to the problem of third-party information exchange is the problem of cooperative data exchange [7], [8]. In cooperative data exchange, each node is assumed to initially hold a subset of packets, and the objective is to ensure all the nodes will eventually obtain the whole set of packets by cooperatively exchanging data between themselves, over a lossless common broadcast channel [7]. The only difference between cooperative data exchange and third-party information exchange is the initial information held by the nodes. In the former, the initial subsets can be arbitrary, with or without much overlapping between nodes. In the latter, the information initially possessed by each node must and will only include the channel information from all of its neighbors to itself, and, with the assumption of channel reciprocity, each pair of neighboring nodes share only one piece of common information, and the shared information is different for every pair. Hence, the problem of third-party information exchange presents a special case of the general problem of cooperative data exchange.

Network coding, a cross-layer technology that was initially developed for static (wireline) networks [20], [21], has received extensive research attention in wireless community, due to its significant benefits in improving wireless performance [22], [23], [24], [25], [26], [27], [28] including throughput, reliability and etc. Recent studies show that network coding can also help reduce the number of transmissions or the transmission delay/cost for general cooperative data exchange [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. However, finding the deterministic code design to achieve these limits for cooperative data exchange can be non-trivial as it needs very high field size, and the complication comes, in part, from



(a) a network of four nodes with costs 4,4,6,10 respectively

(b) the 1–st scheme without network coding

(c) the 2–nd scheme with network coding

(d) the 3–rd scheme with network coding

	packets transmitted			
	cost per transmission	scheme in (b)	scheme in (c)	scheme in (d)
node 1	4	$x_{1,2}$ $x_{1,3}$ $x_{1,4}$	$x_{1,2} \oplus x_{1,3}$	$x_{1,4} \oplus x_{1,2}$ $x_{1,4} \oplus x_{1,3}$
node 2	4	$x_{2,3}$ $x_{2,4}$	$x_{2,3} \oplus x_{2,4}$	$x_{2,4} \oplus x_{2,3}$ $x_{2,4} \oplus x_{1,2}$
node 3	6	$x_{3,4}$	$x_{3,4} \oplus x_{1,3}$	$x_{3,4} \oplus x_{1,3}$
node 4	10		$x_{1,4} \oplus x_{2,4}$	
total cost		26	24	22

(e) total transmission cost comparison

Fig. 1. Third-party information exchange among four nodes.

the very general setup of cooperative data exchange. In comparison, instances of third-party information exchange present better opportunities for analytical and algorithmic tractability. An earlier study of third-party information exchange [1] demonstrated the existence of optimal solutions, where the optimality is measured in terms of minimizing the total number of transmissions. Since each node/device may be associated with a different transmission cost, e.g., different battery cost at the nodes, our previous work [6] extended the study by adding "weights" to transmissions, and tried to minimize the total transmission costs consumed for third-party information exchange.

In this paper, we further study network-coded third-party information exchange with the objective of minimizing the total transmission cost. Different from previous works, we aim to propose an efficient and scalable transmission scheme which can tell the exact number of packets that should be sent by each node. Besides that, we also aim to design an efficient deterministic encoding strategy, which not only can achieve good performance, but also has a very low encoding/decoding complexity (e.g., with a very low coding field size). We call a deterministic coding and transmission scheme feasible if it allows all the nodes to eventually deduce all the global CSI, and we call a feasible solution optimal if it also minimizes a certain cost metric. The goal of this paper is to develop constructive, feasible solutions (including how many packets for each node to transmit and how they are encoded and decoded) that are optimal with respect to the total transmission cost. Similar to the previous studies [1], [6], [7], [8], [9], [10], [11], [12], we assume that a common control channel is available which allows reliable broadcast by any node to all the other nodes. The main contributions of this paper can be summarized as follows:

- We formulate the problem of minimizing the total transmission cost as an integer linear programming (ILP) problem. We show that the optimal solution is one of "water-filling" nature, and a node with lower transmission cost should therefore send more packets than the node with higher transmission cost.
- We construct two efficient and feasible transmission schemes, and prove that the proposed two schemes will achieve optimality (with respect to the total transmission cost) under two specific scenarios.
- A deterministic encoding strategy is designed and can be used in combination with the proposed schemes to ensure the successful deduction of all desired packets at each node. Different from previous works on minimizing transmission cost, the proposed encoding strategy is based on XORs coding over only GF(2), which has very low complexity.
- We also develop a practical distributed transmission protocol that enables the proposed two transmission schemes to be operated in a distributed and hence scalable manner.

The rest of the paper is organized as follows. Section 2 presents the preliminary and problem description. Section 3 formulates the third-party information exchange problem as an integer linear programming problem. Section 4 presents two feasible transmission schemes and proves their optimality in two specific scenarios. Section 5 proposes a distributed transmission protocol and evaluates its performance using the two proposed transmission schemes. Finally, we conclude the paper in Section 6.

2 Preliminary and Problem Description

In this section, we first use an example to motivate the problem. Then, we describe the problem formulation, and the encoding and decoding processes. Finally, we introduce the related works.

2.1 Motivation Example

Consider a four-node network as shown in Fig. 1a, where the transmission costs of nodes $1, 2, 3, 4$ are $4, 4, 6, 10$ respectively. Here, the transmission cost can be the battery cost, e.g., different devices may consume different battery costs per single transmission, or it can also be set based on the status of the battery. In the latter example, if the device is currently with low battery, we can set it to have a higher transmission cost, and those devices with more battery, which have lower transmission cost, will in turn perform more transmissions, as such, this can prolong the network lifetime. Suppose that $x_{i,j}$ denotes the CSI between node i and node j, and the communication links are reciprocity, i.e., $x_{i,j} = x_{j,i}$ for $\forall i, j$. Initially, node i knows only the local information $x_{i,j}$, $\forall j \neq i$. We assume the availability of a lossless common broadcast channel among all the nodes, i.e., each node can receive the packets sent from any other node. Third-party information exchange let nodes exchange the packets among themselves to make sure that each node finally gets the complete packets $\{x_{i,j} \,|\, 1 \leq i \neq j \leq 4\}$.

Without network coding, each packet should be transmitted at least once, as each packet is required by at least one node. To minimize the total transmission cost, we can let node 1 send three packets, node 2 send two packets and node 3 send one packet, as shown in Fig. 1b. The total transmission cost with this approach is $3*4+2*4+1*6+0*10=26$. However, if we use network coding, the transmission cost can be reduced significantly. For example, the schemes in Figs. 1c and 1d reduce the transmission cost to 24 and 22 respectively, while all the nodes can still decode all their "required" packets, e.g., in Fig. 1c, node 1 can decode packet $x_{2,4}$ by $x_{1,4} \oplus (x_{1,4} \oplus x_{2,4}) = x_{2,4}$. The detail comparison can be seen in Fig. 1e.

From the above example, we can get that network coding can indeed reduce the total transmission cost during third-party information exchange. In addition, different transmission schemes and encoding strategies, e.g., Figs. 1c and 1d, also affect the total transmission cost, which motivate us to design the "best" third-party information exchange scheme and encoding strategy.

2.2 Problem Description

Consider a network with N nodes in $C = \{c_1, c_2, \ldots, c_N\}$, where each node $c_i \in C$ is associated with a transmission cost δ_i for sending a single packet over a lossless common broadcast channel [1], [7]. Let $x_{i,j}$ be the CSI (e.g., channel gain, link loss probability or delay) of the link between node c_i and node c_j. Initially, each node c_i only knows the local CSI, i.e., node c_i is only in possession of the set of packets $X_i = \{x_{i,j} \,|\, \forall j \in \{1, 2, \ldots, N\} \setminus \{i\}\}$. We assume that the communication links have reciprocity, i.e., $x_{i,j} = x_{j,i}$, $\forall i, j$, which implies that every pair of nodes c_i and c_j share one common packet $x_{i,j}$.

The purpose of third-party information exchange is to enable all the nodes to eventually retrieve the complete information $X = \{x_{i,j} \,|\, 1 \leq i, j \leq N, i \neq j, x_{i,j} = x_{j,i}\}$, where $|X| = \frac{N(N-1)}{2}$. Each packet transmitted by node c_i can only be a function of the packets X_i initially available to c_i, and in the context of linear network coding, this function is limited to a linear function, or, superposition of the elements in X_i. Let δ_i be the per-packet transmission cost for c_i and let y_i be the number of packets that c_i sends in a particular transmission scheme, where the transmission cost can be transmission power, remaining battery life and etc. Our goal is to develop a linear-network-coding-based feasible solution that enables all the nodes to extract the packets in X while at the same time minimizing the total transmission cost, i.e.,

$$\min \sum_{i=1}^{N} \delta_i y_i.$$

Without loss of generality, we assume that the nodes in C are ordered ascendingly by their transmission costs: $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$.

2.3 Encoding and Decoding Processes

The network-coding-based information exchange scheme involves distributed encoding and distributed decoding at all the nodes. To facilitate packet exchange, each node c_i generates a set of y_i encoded packets over a finite Galois field (GF) based on its initial information possession X_i. The kth encoded packet it sends, denoted by f_i^k, represents a linear combination of the packets in X_i:

$$f_i^k = \sum_{j=1, j\neq i}^{N} \zeta_{i,j}^k x_{i,j}, \tag{1}$$

where $\zeta_{i,j}^k \in \mathrm{GF}(q)$ denotes the encoding coefficient of $x_{i,j}$ in the encoded packet f_i^k. Without loss of generality, all the packets are regarded as equal-length vectors in GF(q), and the multiplications/additions in (1) conform to the GF(q) arithmetic. Then, we have

$$f_i^k = (x_{1,2}, x_{1,3}, \ldots, x_{1,N}, x_{2,3}, \ldots, x_{N-1,N}) \begin{pmatrix} \zeta_{1,2}^k \\ \zeta_{1,3}^k \\ \vdots \\ \zeta_{1,N}^k \\ \vdots \\ \zeta_{N-1,N}^k \end{pmatrix}.$$

Let v_i^k be the encoding vector of the encoded packet f_i^k, i.e., $v_i^k = (\zeta_{1,2}^k, \zeta_{1,3}^k, \ldots, \zeta_{1,N}^k, \ldots, \zeta_{N-1,N}^k)$. For example, in Fig. 1c, the encoding vector of the packet $x_{1,2} \oplus x_{1,3}$ sent by node 1 is $(1, 1, 0, 0, 0, 0)$, based on GF(2) (which is XOR operation and hence very simple). Note that, each original packet $x_{i,j}$ can also be represented as an encoding vector, where all the encoding coefficients are 0 except that the encoding coefficient of $x_{i,j}$ is 1. During data exchange, the encoding vector is sent along with the encoded packet.

After receiving the packets from other nodes, each node will get a coding matrix which includes the encoding vectors of the received packets and the original packets that are initially available at itself. Then, the node will check whether the rank of the coding matrix has reached $|X| = \frac{N(N-1)}{2}$ or not. If the rank is $\frac{N(N-1)}{2}$, the node can decode all the original packets in X by Gaussian Elimination method [29], otherwise, the node needs to receive more encoded packets before getting all the required packets.

2.4 Related Work

We now describe the related work on network-coded cooperative data exchange and third-party information exchange.

The existing researches on cooperative data exchange mainly focus on minimizing the total number of packets required to exchange [7], [8], [9], [10], [11], [12], the total transmission cost/delay [13], [14], [15], [16], or fixing the security issues [18], [19]. Specifically, the works in [7], [8], [9] provide some lower and upper bounds on the minimum number of transmissions and design some heuristic algorithms to get the minimum number of transmissions. The work in [12] further provides a divide and conquer algorithm to minimize the total number of transmissions. The cooperative data exchange with general or multi-hop topology was studied in [10], [11].

Close work related to our problem is cooperative data exchange to minimize the transmission cost. Specifically, [13] formulates the problem into a linear programming problem by assuming that the packet is allowed to be split into multiple sub-packets. Although packet splitting can achieve the optimal solution in polynomial time, it also incurs extra overheads, e.g., splitting packets and reassembling subpackets before/after transmission. The work in [14] presents a randomized algorithm, and prove that the optimal solution can be achieved with a very high probability. The work in [15] further derives an ILP to formulate the general problem of minimizing transmission cost and reduces the ILP to a polynomial-time algorithm. However, designing deterministic encoding strategy is non-trivial as it needs very high field size, which not only increases the decoding complexity but also incurs extra overhead for sending large encoding vectors. The work in [16] studies cooperative data exchange when the transmission cost is the delay. They consider two cases: with and without packet splitting. If packet splitting is allowed, the problem is formulated as a linear programming problem based on a graph model, and if splitting is not allowed, an efficient heuristic algorithm is designed to minimize the transmission delay. However, all the previous works lack the deterministic code design and need to know the global knowledge of the transmission costs of all the nodes.

The third-party information exchange was studied in [1], [6]. Specifically, the work in [1] presents an optimal solution with minimum number of transmissions, and designs a deterministic coding based on GF(2). Our previous work [6] extends the study by adding “weights” to transmissions of each node, and tries to minimize the total transmission costs consumed for third-party information exchange. An optimal solution based on ILP is designed and the performance of random linear network coding is analyzed. However, the solution based on ILP needs to know the global information of transmission costs of the nodes. Based on the previous work, in this paper, we further propose two deterministic transmission schemes, which can be implemented in a distributed manner and are optimal under two specific settings. In addition, a deterministic code is designed to support the proposed schemes.

3 A Centralized Optimal Solution for Third-Party Information Exchange

The case of identical δ_i ($\forall i$), as studied in [1], presents a special case of third-party information exchange. This section will first formulate the problem of third-party information exchange with minimum transmission cost as an integer linear programming problem. Then, we derive some theoretical results which will be used in the following sections.

3.1 Mathematical Formulation of the Proposed Problem

Based on our previous work [6] and the work in [15], we can get the necessary and sufficient condition of third-party information exchange problem as follows.

Theorem 1. *There exists a code design that enables any node to decode all of its desired packets, if and only if for* $\forall k \in \{1, 2, \ldots, N-1\}$ *and any cardinality* k *subset* $\widetilde{C}$ *of* C, *there exists*

$$\sum_{c_i \in \widetilde{C}} y_i \geq \binom{k}{2}, \tag{2}$$

namely, any k *(distinct) nodes will collectively send at least* $\binom{k}{2}$ *packets.*

Proof. For a detailed proof, readers are referred to [6]. □

According to the necessary and sufficient condition, we formulate the problem of finding an optimal solution for third-party information exchange to minimize the total transmission cost, as an integer linear programming problem as follows:

$$\min \sum_{i=1}^{N} \delta_i y_i, \tag{3}$$

subject to

$$\sum_{t=1}^{k} y_{i_t} \geq \binom{k}{2}, \forall i_t \in \{1, \ldots, N\}, 1 \leq k < N. \tag{4}$$

Still consider Fig. 1 as an example. As shown in Fig. 1d, a feasible transmission scheme derived from our ILP has a total transmission cost of 22. In comparison, as shown in Fig. 1c, the transmission scheme proposed by [1], which aims to minimize the total number of transmissions, has a total transmission cost of 24.

After knowing the number of packets that each node should send, random linear network coding can be implemented at each node to generate the encoded packets. For the performance of random linear network coding, the readers are referred to our previous work [6].

3.2 Some Analytical Results

Based on Theorem 1, we can obtain the following three corollaries, which will be used in the following proofs.

Corollary 1. *To minimize the total transmission cost, the number of packets sent by the node with a lower transmission cost should be no less than the node with a higher transmission cost.*

Proof. The proof follows directly from the symmetry of the network. Suppose in a feasible coding scheme, node c_i has a higher transmission cost than node c_j and transmits more packets than c_j. Then we can transform this coding scheme by interchanging the index i with the index j. The results is a new feasible coding scheme with a reduced total transmission cost. □

Since we have assumed an ascending order for the transmission cost $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$, to minimize the total transmission cost, we will get an descending order for the number of transmitted packets $y_1 \geq y_2 \geq \cdots \geq y_N$. Fig. 1d also confirms Corollary 1, i.e., a node with a lower transmission cost must transmit no fewer packets than the node with a higher transmission cost.

The constraint in (4) can thus be simplified by the following corollary.

Corollary 2. *Given $\delta_1 \le \delta_2 \le \cdots \delta_N$, to minimize the total transmission cost, the constraint in (4) of the ILP can be reduced to*

$$\sum_{i=1}^{k} y_{N-i+1} \ge \binom{k}{2}, \forall k \in \{1, 2, \ldots, N-1\}, \tag{5}$$

$$y_{i-1} \ge y_i, \forall i \in \{2, 3, \ldots, N\}. \tag{6}$$

Proof. We prove that the constraint in (4) is equivalent to the constraints in (5) and (6).

First, we will prove that from constraint in (4), we can get constraints in (5) and (6).

- With (4), for $\forall k$ nodes in $\{c_{N-k+1}, c_{N-k+2}, \ldots, c_N\}$, where $k \in \{1, 2, \ldots, N-1\}$, the total number of packets sent by them should be at least $\binom{k}{2}$. Thus, we get constraint in (5).
- According to Corollary 1, given $\delta_1 \le \delta_2 \le \cdots \delta_N$, we can obtain $y_{i-1} \ge y_i$ for $\forall i$, which thus derives constraint in (6).

Second, we prove that with constraints in (5) and (6), we can get (4). For any $k < N$ nodes, we get from (5) that $y_{N-k+1} + y_{N-k+2} + \cdots + y_N \ge \binom{k}{2}$. At the same time, it is easy to see from (6) that any k nodes must have collectively sent no fewer than $y_{N-k+1} + y_{N-k+2} + \cdots + y_N$ packets. That is, for any k-subset $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, N\}$, we have

$$\begin{aligned} y_{i_1} + y_{i_2} + \cdots + y_{i_k} &\ge y_{N-k+1} + y_{N-k+2} + \cdots + y_N \\ &\ge \binom{k}{2}, \end{aligned} \tag{7}$$

which thus leads to the constraint in (4). □

Based on the above analysis, we can further derive the following Corollary.

Corollary 3. *To minimize the total transmission cost of third-party information exchange, the optimal solution $\{y_i^*\}$ always holds that $\sum_{i=2}^{N} y_i^* = \binom{N-1}{2}$.*

Proof. See Appendix A. □

4 Deterministic Algorithms with Minimum Transmission Cost for Two Special Cases

The ILP formulated in the previous section allows us to determine the optimal number of packets to be exchanged between nodes for any solution to be feasible, but the formulation is a centralized optimization problem that requires complexity to solve it. In addition, the centralized algorithm needs the knowledge of each node's transmission cost. In what follows, we present two deterministic transmission schemes that can be performed in a distributed manner (shown in Section 5) without the knowledge of the exact transmission cost of each node. We will show that these two schemes are optimal, i.e., capable of achieving the minimum transmission cost, in two specific cases: 1) $\sum_{i=1}^{k} \delta_i \ge (k-1)\delta_N$, for all $k \in \{2, 3, \ldots, N-1\}$, and 2) $\delta_1 + \delta_2 < \delta_N$, respectively.

4.1 The Fair Load (FL) and Proportional Load (PL) Transmission Schemes

The first transmission scheme developed here is termed the *FL* transmission scheme, where all the nodes will send the same number of packets, with the possible exception for the last node if there are odd number of nodes.

Definition 1 (The FL Scheme). *Suppose the cost of transmission satisfies $\delta_1 \le \delta_2 \le \cdots \le \delta_N$.*

- *When N is odd, any of the first $(N-1)$ nodes, e.g., c_i $(1 \le i \le N-1)$, sends $\frac{N-1}{2}$ packets, and the last node c_N sends nothing:*

$$y_i = \begin{cases} \frac{N-1}{2}, & if\ 1 \le i \le N-1, \\ 0, & otherwise. \end{cases} \tag{8}$$

- *When N is even, every node sends $\frac{N}{2} - 1$ packets:*

$$y_i = \frac{N}{2} - 1, \forall i \in \{1, 2, \ldots, N\}. \tag{9}$$

Lemma 1. *The FL scheme in Definition 1 provides a feasible solution to the third-party information exchange problem.*

Proof. See Appendix B. □

The second transmission scheme we propose is termed the *PL* transmission scheme.

Definition 2 (The PL Scheme). *Consider N nodes with non-decreasing per-node transmission cost $\delta_1 \le \delta_2 \le \cdots \le \delta_N$. In the PL scheme, the node c_i will transmit y_i packets, where*

$$y_i = \begin{cases} N - i, & if\ 2 \le i \le N, \\ N - 2, & if\ i = 1. \end{cases} \tag{10}$$

Lemma 2. *The PL scheme in Definition 2 is a feasible solution to our third-party information exchange problem.*

Proof. See Appendix C. □

4.2 Deterministic Code Design Based on GF(2)

In this section, we design a deterministic encoding strategy (based on XORs coding over GF(2)), which can be performed locally at each node and can be used in combination with the proposed FL and PL schemes to ensure the successful deduction of all the desired packets by all the nodes.

Definition 3 (Encoding Strategy). *Each node $c_i \in C$ sends y_i encoded packets. The nodes generates the encoded packets according to the following rules:*

- *For the FL scheme with odd N and the PL scheme, the jth packet sent by c_i is generated as follows:*

$$x_{i,N} \oplus x_{i,g(i,j)}, \tag{11}$$

where

$$g(i,j) = \begin{cases} i\%(N-1) + 1, & if\ j = 1, \\ g(i, j-1)\%(N-1) + 1, & otherwise. \end{cases} \tag{12}$$

TABLE 1
Encoding Strategy and Transmission Scheme When $N = 7$.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
The FL scheme	$x_{1,7} \oplus x_{1,2}$ $x_{1,7} \oplus x_{1,3}$ $x_{1,7} \oplus x_{1,4}$	$x_{2,7} \oplus x_{2,3}$ $x_{2,7} \oplus x_{2,4}$ $x_{2,7} \oplus x_{2,5}$	$x_{3,7} \oplus x_{3,4}$ $x_{3,7} \oplus x_{3,5}$ $x_{3,7} \oplus x_{3,6}$	$x_{4,7} \oplus x_{4,5}$ $x_{4,7} \oplus x_{4,6}$ $x_{4,7} \oplus x_{4,1}$	$x_{5,7} \oplus x_{5,6}$ $x_{5,7} \oplus x_{5,1}$ $x_{5,7} \oplus x_{5,2}$	$x_{6,7} \oplus x_{6,1}$ $x_{6,7} \oplus x_{6,2}$ $x_{6,7} \oplus x_{6,3}$	
The PL scheme	$x_{1,7} \oplus x_{1,2}$ $x_{1,7} \oplus x_{1,3}$ $x_{1,7} \oplus x_{1,4}$ $x_{1,7} \oplus x_{1,5}$ $x_{1,7} \oplus x_{1,6}$	$x_{2,7} \oplus x_{2,3}$ $x_{2,7} \oplus x_{2,4}$ $x_{2,7} \oplus x_{2,5}$ $x_{2,7} \oplus x_{2,6}$ $x_{2,7} \oplus x_{2,1}$	$x_{3,7} \oplus x_{3,4}$ $x_{3,7} \oplus x_{3,5}$ $x_{3,7} \oplus x_{3,6}$ $x_{3,7} \oplus x_{3,1}$	$x_{4,7} \oplus x_{4,5}$ $x_{4,7} \oplus x_{4,6}$ $x_{4,7} \oplus x_{4,1}$	$x_{5,7} \oplus x_{5,6}$ $x_{5,7} \oplus x_{5,1}$	$x_{6,7} \oplus x_{6,1}$	

- *For the FL scheme with even N, the jth packet sent by c_i is*

$$x_{i,i\%N+1} \oplus x_{i,(i+j)\%N+1}. \tag{13}$$

Note that the above encoding operation is based only on XORs coding over GF(2). Take a network with $N = 7$ nodes as an example. The encoded packets are shown in Table 1, where y_i is determined by either the FL or the PL scheme.

Lemma 3. *The encoding strategy described in Definition 3 makes sure that with the FL or the PL transmission scheme, each node can finally get/decode the complete packets in X.*

Proof. See Appendix D. □

4.3 Optimality Analysis of FL Transmission Scheme

We now show that the proposed FL scheme is actually an optimal solution to a specific network scenario.

Theorem 2. *Consider N nodes with per-node transmission cost $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$. The FL scheme in Definition 1 presents an optimal solution that minimizes the total transmission cost when $\sum_{i=1}^{k} \delta_i \geq (k-1)\delta_N$, for all $k \in \{2, 3, \ldots, N-1\}$.*

Proof. See Appendix E. □

Based on the above analysis, we can derive the following Theorem. (The proof is fairly straight-forward and is therefore omitted for simplicity.)

Theorem 3. *When $\sum_{i=1}^{k} \delta_i \geq (k-1)\delta_N$ for all $k \in \{2, 3, \ldots, N-1\}$, the minimum transmission cost for third-party information exchange, denoted as δ, can be formulated as follows:*

$$\delta = \begin{cases} \frac{N-2}{2} \sum_{i=1}^{N} \delta_i, \; if \; N \; is \; even \\ \frac{N-1}{2} \sum_{i=1}^{N-1} \delta_i, otherwise, \end{cases} \tag{14}$$

where $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$.

Our analysis and results generalize the results given in [1]. When $\delta_1 = \delta_2 = \cdots = \delta_N$, the problem of minimizing the total transmission cost reduces to the problem of minimizing the total number of transmissions, as studied in [1], and our results in Theorems 2 and 3 confirm the result in [1].

4.4 Optimality Analysis of PL Transmission Scheme

We now show that the proposed PL scheme is optimal for a specific network scenario.

Theorem 4. *When $\delta_1 + \delta_2 < \delta_N$, the PL transmission scheme described in Definition 2 is an optimal solution with minimum transmission cost, where $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$.*

Proof. See Appendix F. □

From the above theorem, we can see that irrespective of the number of nodes, the condition depends only on the first two nodes c_1, c_2, and the last node c_N (assuming non-decreasing order of nodes' transmission costs). This condition has advantage of the scalability of the network in terms of its size, e.g., if a new node joins the network, and its transmission cost is more than the previous second node, then the transmission scheme need not change. It also applies if some nodes walk away from the network.

The above analysis also leads to the following theorem. (The detailed proof is omitted for simplicity.)

Theorem 5. *When $\delta_1 + \delta_2 < \delta_N$, the minimum transmission cost for our third-party information exchange is*

$$\delta = \delta_1(N-2) + \sum_{i=2}^{N} \delta_i(N-i), \tag{15}$$

where $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$.

5 Distributed Transmission Protocol

In a distributed system, each node may be unaware of the transmission costs of the other nodes. Under such circumstances, it is impractical to run the ILP proposed in Section 3.1 (especially in a distributed manner) to determine the number of packets that every node should send. In this section, we propose a distributed transmission protocol such that the proposed FL and PL schemes can both be conducted in a distributed manner.

5.1 Distributed Transmission

The proposed distributed transmission protocol does not require the knowledge of the exact transmission costs of the nodes. Only the knowledge of the order of the transmission costs suffices. The distributed transmission protocol can be operated with FL or PL scheme as follows:

- A synchronization timer is first generated independently at every node c_i according to a predefined function $f(\delta_i)$, where δ_i is the transmission cost, and f is a strictly decreasing function that ensures that a higher transmission cost will lead to a smaller synchronization timer. The exact

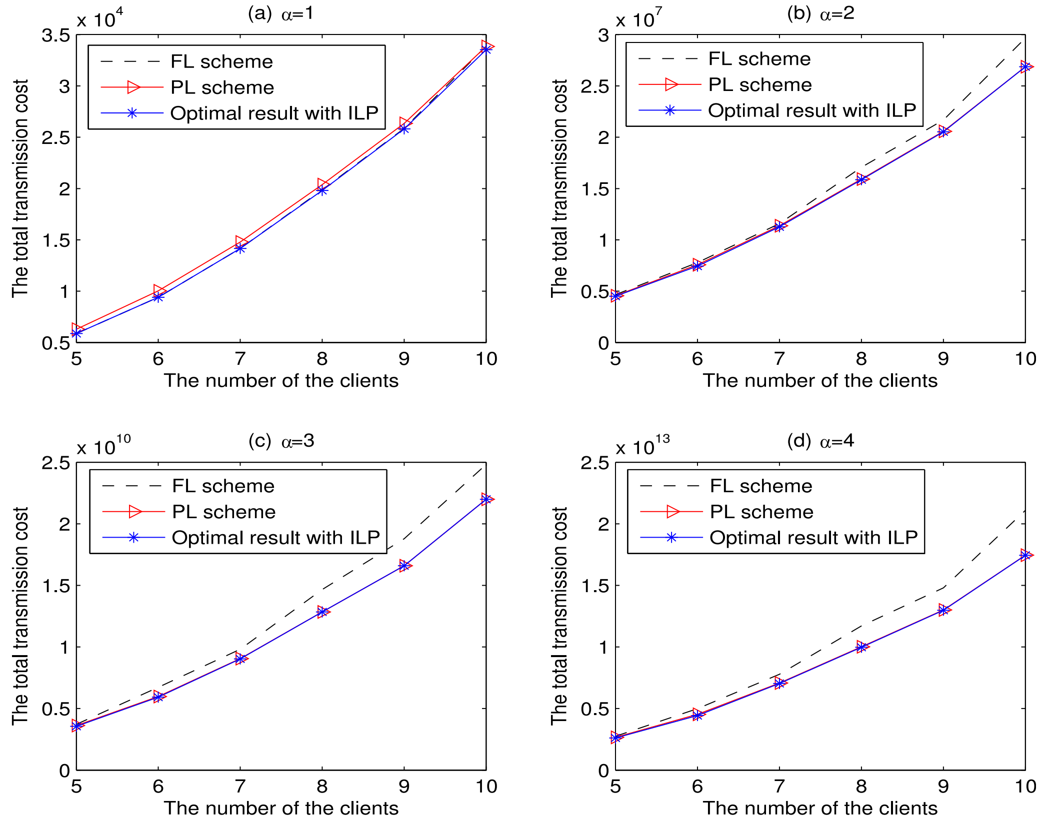


Fig. 2. Comparison of transmission cost for FL and PL.

form of f may be flexible, and may be determined according to the specific system settings. Additionally, to avoid packet collision due to a tie in the transmission cost, a random dithering value may be added to the synchronization timer. For example, a possible form of the synchronization timer may look like:

$$t = C' - a_1 * \delta_i + a_2 * rand(), \tag{16}$$

where $rand()$ is a random generator that randomly produces a value between -1 and 1, and C', a_1 and a_2 are properly-chosen scalers.

- Initially, each node sends its first packet as soon as its synchronization timer times out. In some cases (such as (8) or (10)), the node with the smallest synchronization timer (i.e., c_N) sends nothing except a small message to denote its index.
- After the above step, every node will know the order of their transmission costs. In the following transmission, they will send the remaining packets according to the scheme described in Definitions 1 or 2, e.g., node c_i will send $y_i - 1$ packets (as the first packet is sent in the previous step).

Hence, by assuming that every node only knows the total number of nodes N in the network, the distributed transmission protocol described above will enable every node to know the order of their transmission cost in a distributed way. Once the node c_i knows the order of its transmission cost, it can then send y_i packets, where y_i is determined by the designed transmission scheme, e.g. FL scheme in Definition 1 or PL scheme in Definition 2.

5.2 Performance Evaluation

We now evaluate the performance of the proposed transmission scheme used in combination with FL (in Definition 1) and PL (in Definition 2) in a realistic network scenario.

Consider randomly casting N nodes in a $1{,}000\ \mathrm{m} \times 1{,}000\ \mathrm{m}$ square. Suppose that the transmission cost of node c_i is determined by $\delta_i = d_i^{\alpha} + c$ [31], where d_i is the largest distance from c_i to any other node, α is the positively-valued path loss exponent, and c is the positive number indicating the fixed-base signal processing cost.

We first consider $c = 0$, which describes the case when the signal processing cost is negligibly small compared to the cost for amplification (to combat propagation attenuation). As shown in Fig. 2, when $\alpha = 1$, i.e., the transmission cost δ_i is linearly proportional to the distance d_i (which does not happen in a wireless context), the total transmission cost with the FL scheme is better than that with PL scheme, and close to the ILP result. This is because, when $\delta_i = d$, the probability that $\sum_{i=1}^{k} \delta_i \geq (k-1)\delta_N$ for $\forall k \leq N-1$ is higher than $\delta_1 + \delta_2 < \delta_N$. However, as soon as $\alpha \geq 2$, we observe that $\delta_1 + \delta_2 < \delta_N$ holds with a high probability, and hence the PL transmission scheme becomes noticeably advantageous over the FL scheme, exhibiting the total cost very close to the optimal result obtained by the ILP.

When the signal processing cost c is non-negligible, the performances of FL and PL schemes depend on the actual value of c. We investigate the performance by adopting a setting used in [32] where we vary c in $[1{,}000, 10{,}000]$ and $[10^8, 10^9]$ for $\alpha = 2$ and $\alpha = 4$ respectively. As shown in Fig. 3, for most values of c, we observe that $\delta_1 + \delta_2 < \delta_N$ for $\alpha \geq 2$ still holds with a high probability, and hence the total cost of PL scheme is still rather close to the optimal cost

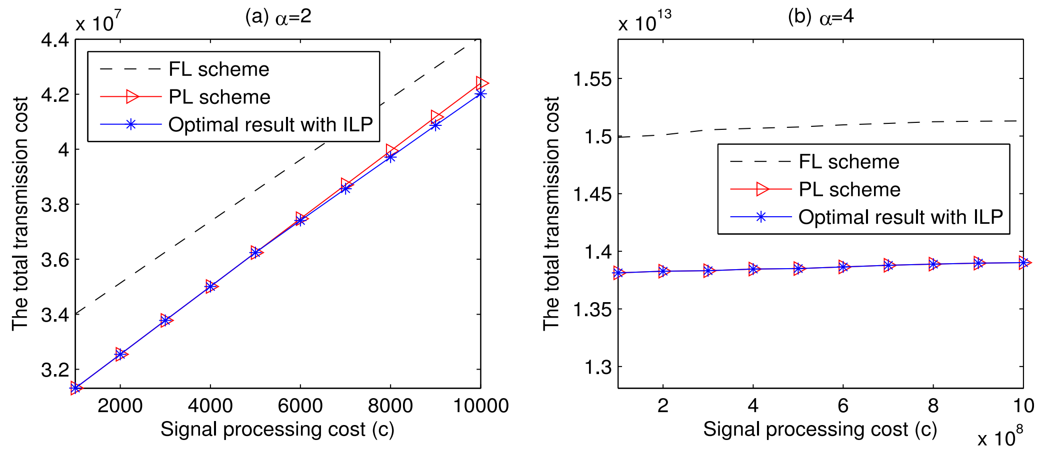


Fig. 3. Comparison of transmission cost for FL and PL with $N = 10$.

calculated from ILP. When $\alpha = 4$, we also can see that the signal processing cost is almost negligible, as the signal processing cost is negligibly small compared to d_i^4.

6 Conclusion

We have studied network-coded cooperative information exchange scheme to minimize the total transmission cost for exchanging third-party information. A necessary and sufficient condition is first developed for the existence of the feasible solution. Formulating the problem of minimizing the total transmission cost as an integer linear programming problem, we develop useful results for feasible and optimal solutions. We have further designed two efficient transmission schemes, FL and PL schemes, and evaluated their performances for specific and randomized networks. We also construct a deterministic encoding strategy, which is based only on XORs coding over GF(2), to support FL and PL transmission schemes. A distributed transmission protocol is proposed that enables FL and PL schemes to be operated in a decentralized manner. Simulations show that the proposed PL scheme can in general perform very close to the optimal solution in practical wireless contexts.

Appendix A
Proof of Corollary 3

(Proof by contradiction.) According to Theorem 1, we have

$$\sum_{i=2}^{N} y_i^* \geq \binom{N-1}{2}. \tag{17}$$

Thus, we only need to prove that if the number of packets sent by the nodes in $\{c_2, c_3, \ldots, c_N\}$ is more than $\binom{N-1}{2}$, i.e., $\sum_{i=2}^{N} y_i^* > \binom{N-1}{2}$, it must contradict the assumption that the solution $\{y_i^*\}$ has the minimum total transmission cost.

Suppose that

$$\sum_{i=2}^{N} y_i^* > \binom{N-1}{2}. \tag{18}$$

According to Theorem 1, we can get $y_i^* + y_{i+1}^* + \cdots + y_N^* \geq \binom{N-i+1}{2}$, where $i \in \{3, 4, \ldots, N\}$. Suppose that k is the smallest index of $\{y_i^*\}$ such that $y_i^* + y_{i+1}^* + \cdots + y_N^* = \binom{N-i+1}{2}$, i.e.,

$$k = \arg\min_{3 \leq i \leq N} \left\{ \sum_{i'=i}^{N} y_{i'}^* = \binom{N-i+1}{2} \right\}. \tag{19}$$

In other words, for $\forall i < k$, we have

$$y_i^* + y_{i+1}^* + \cdots + y_N^* > \binom{N-i+1}{2}. \tag{20}$$

It then follows that

$$\begin{aligned} y_{k-1}^* + y_k^* + y_{k+1}^* + \cdots + y_N^* &> \binom{N-k+2}{2} \\ y_k^* + y_{k+1}^* + \cdots + y_N^* &= \binom{N-k+1}{2}. \end{aligned} \tag{21}$$

Thus, we have

$$y_{k-1}^* > \binom{N-k+2}{2} - \binom{N-k+1}{2} = N-k+1. \tag{22}$$

Since $y_{k+1}^* + y_{k+2}^* + \cdots + y_N^* \geq \binom{N-k}{2}$, according to (21), we also have

$$y_k^* \leq \binom{N-k+1}{2} - \binom{N-k}{2} = N-k. \tag{23}$$

Comparing y_{k-1}^* and y_k^*, we get $y_{k-1}^* > y_k^*$.

Now consider another solution $\{y_i'\}$, where $y_i' = y_i^*$ for all i's except that $y_{k-1}' = y_{k-1}^* - 1$. Since $y_{k-1}^* > y_k^*$, we know $y_{k-1}' \geq y_k'$. Following (19) and considering $\sum_{i'=i}^{N} y_{i'}^* > \binom{N-i+1}{2}$, we have

$$\sum_{i'=i}^{N} y_{i'}' = \sum_{i'=i}^{N} y_{i'}^* - 1 \geq \binom{N-i+1}{2}, \text{ when } i < k, \tag{24}$$

and

$$\sum_{i'=i}^{N} y_{i'}' = \sum_{i'=i}^{N} y_{i'}^* \geq \binom{N-i+1}{2}, \text{ when } i \geq k. \tag{25}$$

It is also easy to verify that $y_i' \geq y_{i'}'$ when $i < i'$.

Hence, according to Corollary 2, $\{y_i'\}$ must be a feasible solution to our third-party information exchange problem. However, the total transmission cost with solution $\{y_i'\}$ is smaller than that of solution $\{y_i^*\}$, which contradicts the assumption that $\{y_i^*\}$ is the optimal solution.
This contradiction suggests the invalidity of the assumption $\sum_{i=2}^{N} y_i^* > \binom{N-1}{2}$. Since Corollary 2 states $\sum_{i=2}^{N} y_i^* \geq \binom{N-1}{2}$, it then follows that $\sum_{i=2}^{N} y_i^* = \binom{N-1}{2}$.

Appendix B
Proof of Lemma 1

Here we prove for the case of even N. The case of odd N follows from a very similar manner, and is therefore omitted for simplicity. When N is even, the number of packets each node sent is $\frac{N}{2}-1$, and we wish to show that such an arrangement can lead to a feasible solution. For any $k \in \{1, 2, \ldots, N-1\}$, we have

$$\begin{aligned}\sum_{i=1}^{k} y_{N-i+1} - \binom{k}{2} &= \frac{(N-2)k}{2} - \frac{k(k-1)}{2} \\ &= \frac{k(N-k-1)}{2} \\ &\geq 0,\end{aligned} \tag{26}$$

which satisfies the first constraint in (5). In addition, since $y_1 = y_2 = \cdots = y_N$, the second constraint in (6) can also be satisfied. It then follows from Corollary 2 and Theorem 1 that this FL scheme guarantees the existence of a feasible code design for every node to successfully deduce its desired packets.

Appendix C
Proof of Lemma 2

To show the feasibility of PL transmission scheme, it is sufficient to check the conditions in (5) and (6). For any $k \in \{1, 2, \ldots, N-1\}$ nodes, the total number of packets they send sums up to

$$\begin{aligned}\sum_{i=1}^{k} y_{N-i+1} &= \sum_{i=1}^{k}(N-(N-i+1)) \\ &= \sum_{i=1}^{k}(i-1) = \frac{k(k-1)}{2} \\ &= \binom{k}{2},\end{aligned} \tag{27}$$

which satisfies the first condition for a feasible solution. Additionally, for each node $c_i \in C$, we have $y_i = N-i > y_{i+1} = N-i-1$ for $i \geq 2$, and $y_1 = y_2$, which also satisfy the second condition.

Appendix D
Proof of Lemma 3

We consider the FL and PL transmission schemes separately.

Case I. FL. We separate the cases of odd N and even N. In what follows, we will detail the case of odd N. The case of even N follows in a similar way.

Suppose N is odd. Following the FL transmission scheme and the proposed encoding strategy, node c_i will send the following set of encoded packets:

- when $1 \leq i \leq \frac{N-1}{2}$, the set of $\frac{N-1}{2}$ packets sent by c_i are $\{x_{i,N} \oplus x_{i,i+1}, x_{i,N} \oplus x_{i,i+2}, \ldots, x_{i,N} \oplus x_{i,i+\frac{N-1}{2}}\}$;
- when $\frac{N-1}{2} < i \leq N-2$, the set of $\frac{N-1}{2}$ packets sent by c_i are $\{x_{i,N} \oplus x_{i,i+1}, x_{i,N} \oplus x_{i,i+2}, \ldots, x_{i,N} \oplus x_{i,N-1}, x_{i,N} \oplus x_{i,1}, x_{i,N} \oplus x_{i,2}, \ldots, x_{i,N} \oplus x_{i,i-\frac{N-1}{2}}\}$;
- when $i = N-1$, the set of $\frac{N-1}{2}$ packets sent by c_{N-1} are $\{x_{N-1,N} \oplus x_{N-1,1}, \ldots, x_{N-1,N} \oplus x_{N-1,\frac{N-1}{2}}\}$;
- when $i = N$, node c_i sends nothing, since $y_N = 0$.

Based on the above encoded packets, we first prove that each packet $x_{i,j} \in X$ must be encoded in at least one packet sent by the nodes. We now consider $i \leq \frac{N-1}{2}$:

- for packet $x_{i,j}$ (or $x_{j,i}$), where $i < j \leq i + \frac{N-1}{2}$, it must be encoded in one packet sent by node c_i;
- for packet $x_{i,j}$ (or $x_{j,i}$), where $j < i$, it must be encoded in one packet sent by node c_j;
- for packet $x_{i,j}$ (or $x_{j,i}$), where $i + \frac{N-1}{2} < j \leq N-1$, it must be encoded in one packet sent by node c_j.

In a similar way, we can prove that when $i > \frac{N-1}{2}$, packet $x_{i,j}$ is also encoded in at least one packet sent by c_i or c_j. That is for $\forall x_{i,j} \in X$, it must be encoded in at least one packet sent by the nodes.

To show the feasibility of the solution, we verify that every node can successfully deduce its desired packets. First, consider node c_N. We can easily obtain that c_N must be able to decode all the original packets that are involved in the packets sent by node c_i for $\forall i \neq N$, as $x_{i,N}$ is encoded in all the packets sent by c_i. Since all the original packets are involved in some packets sent by some nodes, node c_N can thus decode all of its desired packets.

Next, consider node c_1. When $i > \frac{N-1}{2}$, packet $x_{i,N} \oplus x_{i,1}$ must be sent by c_i. As c_1 possesses packet $x_{i,1}$, it can decode $x_{i,N}$. Correspondingly, c_1 can decode all the original packets involved in the packets sent by c_i with $x_{i,N}$. Particularly, from the packets sent by node $c_{j+\frac{N-1}{2}}$, where $j \leq \frac{N-1}{2}$, c_N can decode packet $x_{j+\frac{N-1}{2},j}$, as $x_{j+\frac{N-1}{2},j}$ is encoded in one packet sent by node $c_{j+\frac{N-1}{2}}$. In addition, since $x_{j,N} \oplus x_{j,j+\frac{N-1}{2}}$ must be sent by node c_j, c_1 can thus decode $x_{j,N}$ with $x_{j+\frac{N-1}{2},j}$. Accordingly, c_1 can decode all the original packets that are involved in the packets sent by c_j, where $j \leq \frac{N-1}{2}$. That is, c_1 can decode all the original packets involved in the packets sent by other nodes. Thus, node c_1 can deduce all of its desired packets. In the same way, we can show that any node c_i, where $i \neq 1, N$, can deduce all the original packets.

Case II. PL. Following the PL transmission scheme described in Definition 2 and the proposed encoding strategy, node c_i will send the following set of encoded packets:

- when $i = 1$, the set of $(N-2)$ packets sent by c_1 are $\{x_{1,N} \oplus x_{1,2}, x_{1,N} \oplus x_{1,3}, \ldots, x_{1,N} \oplus x_{1,N-1}\}$;
- when $2 \leq i \leq N-1$, the set of $(N-i)$ packets sent by c_i are $\{x_{i,N} \oplus x_{i,i+1}, x_{i,N} \oplus x_{i,i+2}, \ldots, x_{i,N} \oplus x_{i,N-1}, x_{i,N} \oplus x_{i,1}\}$;
- when $i = N$, node c_i sends nothing, since $y_N = 0$.

From the above encoding procedure, we get the following observations: 1) each packet $x_{i,j}$ must be encoded into at least one of the packets sent by node c_i or node c_j, 2) $\forall j$, packet $x_{1,j}$ must be encoded into one of the packets sent by node c_1, and 3) for any node $c_i \in C$, where $i \neq 1, N$, packet $x_{i,1}$ must be encoded into one of the packets sent by node c_i.

We now show that every node can deduce its desired packets. First consider node c_1. From Observation 3, we know that node c_1 can decode all the other packets that are encoded into the packets sent by any other node, and will thus deduce the complete information. For any other node c_i, where $i > 1$, since $x_{1,i}$ is encoded into some packet(s) sent by c_1 (Observation 2), it can decode packet $x_{1,N}$. Subsequently, with the availability of $x_{1,N}$, node c_i can decode all the other packets in $\{x_{1,j} \,|\, \forall j \neq 1\}$. In addition, according to Observation 3, from the packets sent by node c_j, where $j \neq i, 1$, node c_i can decode packet $x_{j,N}$ using packet $x_{1,j}$, and thus retrieve all the packets that c_j has. As such, node c_i has deduced the complete information.

Appendix E
Proof of Theorem 2

We separate the case of even N and odd N. We use $\{y_i\}$ to denote the solution corresponding to FL scheme in Definition 1.

Case I. Even N. (Proof by contradiction.)

Suppose there exists an optimal transmission scheme, $\{y_i^*\}$, with a smaller total transmission cost than $\{y_i\}$. Now consider the three possibilities for y_N^*: i) $y_N^* > y_N$, ii) $y_N^* < y_N$, and iii) $y_N^* = y_N$.

Case I. i) According to Corollary 1, when $\delta_1 \leq \delta_2 \leq \cdots \leq \delta_N$, we have $y_1^* \geq y_2^* \geq \cdots \geq y_N^*$. Thus, in Case I.i) when $y_N^* > \ y_N = \frac{N}{2} - 1$, we have $y_i^* \geq y_N^* > y_N = \frac{N}{2} - 1$, $\forall i \in \{1, 2, \ldots, N\}$. This leads to $\sum_{i=1}^{N} \delta_i y_i^* > \sum_{i=1}^{N} \delta_i y_i$, which contradicts with the the assumption that $\{y_i^*\}$ has a smaller total transmission cost.

Case I. ii) According to Corollary 3, we have

$$\sum_{i=2}^{N} y_i^* = \binom{N-1}{2} = \sum_{i=2}^{N} y_i. \tag{28}$$

Since in ii) we have $y_N^* < y_N = \frac{N}{2} - 1$, there must exist an i where $2 \leq i \leq N-1$, $y_i^* > \frac{N}{2} - 1 = y_i$. Let k be the largest index for which $\{y_i^*\} > y_i$, i.e., $k = \arg\max_i \{y_i^* > \ y_i, \forall i \in \{2, 3, \ldots, N-1\}\}$. When $k+1 \leq i \leq N-1, y_i^* \leq y_i = \frac{N}{2} - 1$. Since $y_{i'}^* \geq y_k^*$ and $y_k^* > y_k = \frac{N-1}{2} = y_{i'}$, $\forall i' \leq k$, we have $y_{i'}^* > y_{i'} = \frac{N}{2} - 1$. We show that it is not possible to have $k = N-1$. Suppose $k = N-1$; we will then have $y_i^* > \frac{N}{2} - 1$, $\forall i \in \{2, 3, \ldots, N-1\}$. Since y_i^* is an integer, we have $y_i^* \geq \frac{N}{2}$. From the above analysis, we get

$$\begin{aligned} \sum_{i=2}^{N} y_i^* &= \sum_{i=2}^{N-1} y_i^* + y_N^* \\ &\geq (N-2)\frac{N}{2} + y_N^* \\ &\geq \binom{N-1}{2} + \frac{N-2}{2} + y_N^* \\ &> \binom{N-1}{2}, \end{aligned} \tag{29}$$

which contradicts with the result in (28). Hence, $k \neq N-1$, which implies $k < N-1$.

Let $z_i^* = |y_i^* - y_i|, \forall i = 1, 2, \ldots, N$. We have

$$z_i^* = \begin{cases} y_i^* - y_i, & \text{if } 1 \leq i \leq k, \\ y_i - y_i^*, & \text{if } k+1 \leq i \leq N. \end{cases} \tag{30}$$

Since $y_i^* \geq y_{i+1}^*$ and $y_i = y_{i+1}$, we have $z_i^* \geq z_{i+1}^*$, when $1 \leq i \leq k$.

From (28) we obtain

$$\sum_{i=2}^{k} z_i^* - \sum_{i=k+1}^{N} z_i^* = 0. \tag{31}$$

Since $y_N^* < y_N$ and $y_i^* \leq y_i$ when $k+1 \leq i \leq N-1$, we have

$$\sum_{i=k+1}^{N} z_i^* \geq 1. \tag{32}$$

In addition, since $y_k^* > y_k$, we get $z_k^* \geq 1$. This gives rise to

$$z_1^* \geq z_2^* \geq \cdots \geq z_k^* \geq 1. \tag{33}$$

We now compare the total transmission cost between the proposed solution $\{y_i\}$ and the optimal solution $\{y_i^*\}$.

$$\begin{aligned} &\sum_{i=1}^{N} \delta_i y_i^* - \sum_{i=1}^{N} \delta_i y_i \\ &= \delta_1 z_1^* + \sum_{i=2}^{k} \delta_i z_i^* - \sum_{i=k+1}^{N} \delta_i z_i^* \\ &\geq \delta_1 z_2^* + \sum_{i=2}^{k} \delta_i z_i^* - \delta_N \sum_{i=k+1}^{N} z_i^* \\ &= z_k^* \sum_{i=1}^{k} \delta_i + \left(z_{k-1}^* - z_k^*\right) \sum_{i=1}^{k-1} \delta_i + \cdots + \left(z_2^* - z_3^*\right) \sum_{i=1}^{2} \delta_i \\ &\quad - \delta_N \sum_{i=k+1}^{N} z_i^* \\ &\geq z_k^*(k-1)\delta_N + \left(z_{k-1}^* - z_k^*\right)(k-2)\delta_N + \cdots + \left(z_2^* - z_3^*\right)\delta_N \\ &\quad - \delta_N \sum_{i=k+1}^{N} z_i^* \\ &= \delta_N \sum_{i=2}^{k} z_k^* - \delta_N \sum_{i=k+1}^{N} z_i^* = 0. \end{aligned} \tag{34}$$

We can observe that the transmission cost of $\{y_i\}$ is not larger than that of the optimal solution $\{y_i^*\}$. Hence, in Case I.ii when $y_N^* < y_N$, the proposed solution $\{y_i\}$ must also be optimal with minimum transmission cost.

Case I. iii). In this case, we assume $y_N^* = y_N = \frac{N}{2} - 1$. Since $y_i^* \geq y_N^* = \frac{N}{2} - 1$, y_i^* must be equal to $y_i = \frac{N}{2} - 1$, as in the proposed solution. (Otherwise, the transmission cost with the optimal solution would exceed that of the propose solution.) Hence, in this case I.iii) when $y_N^* = y_N$, the proposed solution is also optimal.

Case II. Odd N. (Proof by contradiction.)

Again suppose $\{y_i^*\}$ is an optimal solution with minimum transmission cost. Since $y_N = 0$, there are only two possibilities for y_N^*: i) $y_N^* > y_N = 0$, or ii) $y_N^* = y_N = 0$.

Case II. i): When $y_N^* > 0$, according to Corollary 3, there must exist an y_i^* such that $y_i^* < y_i = \frac{N-1}{2}$. Suppose that k is the smallest index for which $\{y_i^*\} < y_i$, i.e., $k = \arg\min_i \{y_i^* < y_i\}$. In other words, $y_i^* \geq y_i$, $\forall i \in \{1, 2, \ldots, k-1\}$. Now if $k = N-1$, then there exists a solution $\{y_i'\}$ where all $y_i' = y_i^*$, except that $y_{N-1}' = y_{N-1}^* + 1, y_N' = y_N^* - 1$. It is easy to verify that the solution $\{y_i'\}$ satisfies all the constraints of the feasible solution, but its transmission cost falls below that of the optimal solution, which contradicts with the assumption. Hence, we have $k < N-1$.

In addition, when $i > k$, since $y_k^* < y_k = \frac{N-1}{2}$ and $y_k^* \geq y_i^*$, we have $y_i^* < \frac{N-1}{2} = y_i$. That is,

$$y_i^* < y_i, \text{ if } k \leq i \leq N-1, \tag{35}$$

and

$$y_i^* \geq y_i, \text{ if } 1 \leq i \leq k-1. \tag{36}$$

Since $y_k^* < y_k = \frac{N-1}{2}$ and $y_{k-1}^* \geq y_{k-1} = \frac{N-1}{2}$, we have

$$y_{k-1}^* > y_k^*. \tag{37}$$

According to Corollary 2, $y_i^* + y_{i+1}^* + \ldots + y_N^* > \binom{N-i+1}{2}$ or $y_i^* + y_{i+1}^* + \cdots + y_N^* = \binom{N-i+1}{2}$, $\forall i \in \{k, k+1, \ldots, N\}$. If we always have $y_i^* + y_{i+1}^* + \cdots + y_N^* > \binom{N-i+1}{2}$, the transmission cost with optimal solution can be reduced by decreasing y_N^* by 1 but increasing y_k^* by 1, which contradicts with the optimal solution. Thus, there must exist i, where $k \leq i \leq N$, such that $y_i^* + y_{i+1}^* + \cdots + y_N^* = \binom{N-i+1}{2}$. Suppose that t is the largest index of $\{y_t^*\}$ where $y_t^* + y_{t+1}^* + \cdots + y_N^* = \binom{N-t+1}{2}$, $k \leq t \leq N$.

From Corollary 2, we can easily obtain

$$\begin{aligned} y_t^* + y_{t+1}^* + \cdots + y_N^* &= \binom{N-t+1}{2} \\ &> y_t^* + \binom{N-t}{2}. \end{aligned} \tag{38}$$

It then follows that

$$y_t^* < N - t. \tag{39}$$

Since

$$\begin{aligned} y_{t-1}^* + y_t^* + \cdots + y_N^* &= y_{t-1}^* + \binom{N-t+1}{2} \\ &\geq \binom{N-t+2}{2}, \end{aligned} \tag{40}$$

we get

$$y_{t-1}^* \geq N - t + 1 > y_t^* + 1. \tag{41}$$

This implies that we have a better solution $\{y_i'\}$ with less transmission cost, where $y_i' = y_i^*$ for all other i's except $y_N^* = y_N^* - 1$, and $y_t' = y_t^* + 1$. This contradicts with the assumption that $\{y_i^*\}$ is optimal, and hence it is impossible to have $y_N^* > y_N$.

Case II. ii). As discussed in the above, $y_N^* > y_N$ does not hold, which leaves only one possibility $y_N^* = y_N = 0$.

When $i < N$, we can follow the same train of thoughts as the case of even N, to prove that the solution $\{y_i\}$ can achieve the minimum transmission cost as the solution $\{y_i^*\}$.

Appendix F
Proof of Theorem 4

Let $\{y_i^*\}$ be an arbitrary optimal solution with minimum transmission cost. We prove the case of $i = N$ by contradiction. We start by assuming that $y_N^* = y_N = 0$ does not hold, that is, we assume that for any optimal solution $\{y_i^*\}$, y_N^* must be larger than zero.

Let k be the largest index of $\{y_i^*\}$ such that $y_i^* + y_{i+1}^* + \cdots + y_N^* = \binom{N-i+1}{2}$, i.e.,

$$k = \arg \max_{2 \leq i \leq N} \left\{ \sum_{i'=i}^{N} y_{i'}^* = \binom{N-i+1}{2} \right\}. \tag{42}$$

In other words, for $\forall i \geq k+1$, we can obtain that

$$y_i^* + y_{i+1}^* + \cdots + y_N^* > \binom{N-i+1}{2}. \tag{43}$$

Suppose $k = 2$. Since $\{y_i^*\}$ is the optimal solution with minimum transmission cost, we know that $y_1^* = y_2^*$. We can find another transmission scheme $\{y_i'\}$, where $y_i' = y_i^*$ for all other i's, except that $y_1' = y_2' = y_2^* + 1$ and $y_N' = y_N^* - 1$. According to (43), $\forall i \in \{2, 3, \ldots, N\}$, we have

$$y_i' + y_{i+1}' + \cdots + y_N \geq \binom{N-i+1}{2}. \tag{44}$$

Since $y_i' \geq y_{i+1}'$ for $\forall i \in \{1, 2, \ldots, N-1\}$, the second constraint in (6) can also be satisfied. Thus, the solution $\{y_i'\}$ is one of the feasible solutions of our third-party information exchange problem. We now compare its transmission cost with the optimal solution $\{y_i^*\}$:

$$\begin{aligned} &\sum_{i=1}^{N} \delta_i y_i' - \sum_{i=1}^{N} \delta_i y_i^* \\ &= \delta_1 y_1' + \delta_2 y_2' + \delta_N y_N' - (\delta_1 y_1^* + \delta_2 y_2^* + \delta_N y_N^*) \\ &= \delta_1 + \delta_2 - \delta_N < 0. \end{aligned} \tag{45}$$

This contradicts with the assumption that $\{y_i^*\}$ achieves the minimum transmission cost. Hence $k = 2$ does not hold.

When $k > 2$, from (42) and (43), we get

$$\begin{aligned} y_k^* &= \binom{N-k+1}{2} - (y_{k+1}^* + y_{k+2}^* + \cdots + y_N^*) \\ &< \binom{N-k+1}{2} - \binom{N-k}{2} = N - k. \end{aligned} \tag{46}$$

In addition, according to Corollary 2, we have $y_{k-1}^* + y_k^* + \cdots + y_N^* \geq \binom{N-k+2}{2}$. Thus,

$$\begin{aligned} y^*_{k-1} &\geq \binom{N-k+2}{2} - (y^*_k + \cdots + y^*_N) \\ &= \binom{N-k+2}{2} - \binom{N-k+1}{2} \\ &= N-k+1. \end{aligned} \tag{47}$$

It then follows from (46) and (47) that

$$y^*_{k-1} > y^*_k. \tag{48}$$

We consider two possibilities: $\delta_k = \delta_N$ and $\delta_k > \delta_N$. If $\delta_k = \delta_N$, since $\delta_k \leq \delta_{k+1} \leq \cdots \leq \delta_N$, we can obtain that $\delta_k = \delta_{k+1} = \cdots = \delta_N$. Thus, $\sum_{i=k}^N \delta_i y^*_i = \delta_N \binom{N-k+1}{2}$. We then construct another solution $\{y'_i\}$ such that

$$y'_i = \begin{cases} y^*_i, & \text{if } 1 \leq i < k, \\ y_i = N-i, & \text{if } k \leq i \leq N. \end{cases} \tag{49}$$

It is easy to verify that $\{y'_i\}$ is also an optimal solution, since for $\forall i \geq k$, $y'_i + y'_{i+1} + \cdots + y'_N = \binom{N-i+1}{2}$, and for $\forall i$, $y'_i \geq y'_{i+1}$. In addition, the transmission cost with solution $\{y'_i\}$ is the same as $\{y^*_i\}$:

$$\begin{aligned} &\sum_{i=1}^N \delta_i y'_i - \sum_{i=1}^N \delta_i y^*_N \\ &\quad = \sum_{i=k}^N \delta_i y'_i - \sum_{i=k}^N \delta_i y^*_i \\ &\quad = \delta_N \sum_{i=k} y'_i - \delta_N \sum_{i=k} y^*_i = 0. \end{aligned} \tag{50}$$

That is, if $\delta_k = \delta_N$, then the solution $\{y'_i\}$, where $y'_N = 0$, is an optimal solution. This contradicts with the assumption that an optimal solution cannot have the Nth transmission number being 0.

If $\delta_k > \delta_N$, we can find another solution $\{y''_i\}$, where all $y''_i = y^*_i$ except that $y''_k = y^*_k + 1$ and $y''_N = y^*_N - 1$. From (42) and (43), we have

$$\begin{aligned} y''_i + y''_{i+1} + \cdots + y''_N &\geq \binom{N-i+1}{2}, i \neq k, \\ y''_k + y''_{k+1} + \cdots + y''_N &= \binom{N-k+1}{2}, i = k. \end{aligned} \tag{51}$$

In addition, we can easily verify that $y''_i \geq y''_{i+1}$, $\forall i \in \{1, 2, \ldots, N-1\}$. Thus, $\{y''_i\}$ is a feasible solution. However, since $\delta_k > \delta_N$, the transmission cost of solution $\{y''_i\}$ is less than that of solution $\{y^*_i\}$, which contradicts with the optimality of $\{y^*_i\}$.

Hence, the assumption y^*_N must be greater than 0 does not hold, which means that there exists an optimal solution with $y^*_N = 0$.

We then consider such an optimal solution where $y^*_N = 0$. To satisfy $y^*_{N-1} + y^*_N \geq \binom{2}{2} = 1$, we have $y^*_{N-1} \geq 1$. Following a similar line of thoughts, we can prove that there exists an optimal solution with $y^*_{N-1} = 1$, based on $y^*_N = 0$. Correspondingly, we have proven that the solution $y^*_i = y_i, \forall i$ remains an optimal solution.

ACKNOWLEDGMENTS

This work was supported by the National Nature Science Foundation of China (under Grant No. 61300212, No. 61379038), and the Doctoral Fund of Ministry of Education of China (No. 20130111120010). Tiffany Jing Li's work was supported by the US National Science Foundation under Grants No. EMT-0829888, CMMI-0928092, OCI-1133027, and ECCS-1343372. It was also supported in part by Singapore University of Technology and Design, and the open research fund of the State Key Laboratory of Integrated Services Networks, Xidian University (No. ISN15-10).

REFERENCES

[1] D. J. Love, B. M. Hochwald, and K. Krishnamurthy, "Exchanging third-party information in a network," in *Proc. UCSD Inf. Theory Appl. Workshop*, Jan. 2007, pp. 1–6.
[2] O. Aluko, B. Clerckx, D. Love, and J. Krogmeier, "Enhanced limited-coordination strategies for multi-user mimo systems," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1181–1185.
[3] J. Wu, B. Cheng, C. Yuen, Y. Shang, and J.-L. Chen, "Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, in press, vol. 99, Jun. 2014.
[4] X. Chen, Z. Zhang, and H.-H. Chen, "On distributed antenna systems with limited feedback precoding: Opportunities and challenges," *IEEE Wireless Commun.*, vol. 17, no. 2, pp. 80–88, Feb. 2010.
[5] J. Wu, C. Yuen, B. Cheng, Y. Shang, and J. Chen, "Goodput-aware load distribution for real-time traffic over multipath networks," *IEEE Trans. Parallel Distrib. Syst.*, in press, vol. 99, Aug. 2014.
[6] X. Wang, W. Song, C. Yuen, and J. L. (Tiffany), "Exchanging third-party information with minimum transmission cost," in *Proc. IEEE Global Commun. Conf.*, Dec. 2012, pp. 2438–2443.
[7] S. El Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," in *Proc. IEEE Inf. Theory Workshop*, Jan. 2010, pp. 1–5.
[8] T. Courtade, B. Xie, and R. Wesel, "Optimal exchange of packets for universal recovery in broadcast networks," in *Proc. Military Commun. Conf.*, Nov. 2010, pp. 2250–2255.
[9] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb, "A randomized algorithm and performance bounds for coded cooperative data exchange," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 1888–1892.
[10] M. Gonen and M. Langberg, "Coded cooperative data exchange problem for general topologies," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2606–2610.
[11] T. A. Courtade and R. D. Wesel, "Coded cooperative data exchange in multihop networks," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1136–1158, Feb. 2014.
[12] N. Milosavljevic, S. Pawar, S. El Rouayheb, M. Gastpar, and K. Ramchandran, "Deterministic algorithm for the cooperative data exchange problem," in *Proc. IEEE Int. Symp. Inf. Theory*, Aug. 2011, pp. 410–414.
[13] S. Tajbakhsh, P. Sadeghi, and R. Shams, "A generalized model for cost and fairness analysis in coded cooperative data exchange," in *Proc. IEEE Int. Symp. Netw. Coding*, Jul. 2011, pp. 1–6.
[14] D. Ozgul and A. Sprintson, "An algorithm for cooperative data exchange with cost criterion," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2011, pp. 1–4.
[15] T. A. Courtade and R. D. Wesel, "Weighted universal recovery, practical secrecy, and an efficient algorithm for solving both," in *Proc. 49th Annu. Allerton Conf. Commun., Control Comput.*, Sep. 2011, pp. 1349–1357.
[16] X. Wang and C. Yuen, "Delay minimization for network coded cooperative data exchange with rate adaptation," in *Proc. IEEE VTC-fall*, Sep. 2013, pp. 1–5.
[17] Z. Dong, S. H. Dau, C. Yuen, J. Gu, and X. Wang, "Delay minimization for relay-based cooperative data exchange with network coding," *IEEE/ACM Trans. Netw.*, in press, vol. 99, Jul. 2014.
[18] M. Yan and A. Sprintson, "Weakly secure network coding for wireless cooperative data exchange," in *Proc. IEEE Global Commun. Conf.*, Dec. 2011, pp. 1–5.

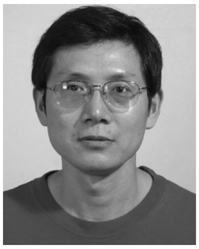
[19] W. Song, X. Wang, C. Yuen, J. Li, and R. Feng, "Error correction for cooperative data exchange," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1856–1859, Nov. 2012.
[20] R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
[21] S. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.
[22] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 497–510, Jun. 2006.
[23] S. El Rouayheb, M. Chaudhry, and A. Sprintson, "On the minimum number of transmissions in single-Hop wireless coding networks," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2007, pp. 120–125.
[24] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, "Wireless broadcast using network coding," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 914–925, Feb. 2009.
[25] C. Fragouli, D. Katabi, A. Markopoulou, M. Medard, and H. Rahul, "Wireless network coding: Opportunities & challenges," in *Proc. IEEE Military Commun. Conf.*, Oct. 2007, pp. 1–8.
[26] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability gain of network coding in lossy wireless networks," in *Proc. 27th IEEE Conf. Comput. Commun.*, Apr. 2008, pp. 2171–2179.
[27] A. Eryilmaz, A. Ozdaglar, and M. Medard, "On delay performance gains from network coding," in *Proc. Conf. Inf. Sci. Syst.*, Jan. 2006, pp. 1–5.
[28] X. Wang, C. Yuen, and Y. Xu, "Coding based data broadcasting for time critical applications with rate adaptation," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2429–2442, Jun.2014.
[29] M. Medard and A. Sprintson, Network coding: *Fundamentals and Applications*. Waltham, MA, USA: Academic, 2011.
[30] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
[31] J. Wieselthier, G. Nguyen, and A. Ephremides, "On the construction of energy-efficient broadcast and multicast trees in wireless networks," in *Proc. IEEE 19th Annu. Joint Conf. Comput. Commun. Soc.*, 2000, vol. 2, pp. 585–594.
[32] I. Stojmenovic and X. Lin, "Power-aware localized routing in wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 12, no. 11, pp. 1122–1133, Nov. 2001.

Xiumin Wang received the BS degree from the Department of Computer Science, Anhui Normal University, Wuhu, China, in 2006. She received the PhD degree from both the Department of Computer Science, University of Science and Technology of China, Hefei, China, and the Department of Computer Science, City University of Hong Kong, Hong Kong, under a joint PhD program. She is currently an associate professor in the School of Computer and Information, Hefei University of Technology, Hefei, China. Her current research interests include wireless networks, routing design, and network coding.

Chau Yuen received the BEng and PhD degrees from Nanyang Technological University, Singapore, in 2000 and 2004, respectively. He was a post doc fellow in Lucent Technologies Bell Labs, Murray Hill, NY, during 2005. He was a visiting assistant professor at the Hong Kong Polytechnic University, Hong Kong, in 2008. From 2006 to 2010, he was at the Institute for Infocomm Research (Singapore) as a senior research engineer. He joined Singapore University of Technology and Design as an assistant professor from June 2010. He serves as an associate editor for the *IEEE Transactions on Vehicular Technology*. In 2012, he received the IEEE Asia-Pacic Outstanding Young Researcher Award.

Tiffany Jing Li received the BS degree in computer science from Peking University, Beijing, China, and the master's and PhD degrees in electrical engineering from Texas A&M University, College Station, in 1997, 1999, and 2002, respectively. After obtaining the PhD degree, she joined the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, and was promoted to an associate professor in 2008. Her current research interests include coding and communication theory, wireless communications and networks, and digital data storage systems. He served as an editor for the *IEEE Transaction on Wireless Communications* from 2008 to 2011, an associated editor for the *IEEE Communications Letters* from 2004 to 2008, and as an editorial board member for the *IEEE Communications Surveys and Tutorials* journal from 2008 to 2010. She also served as a Symposium co-chair in ICC'2013, ICC'2008, GLOBECOM'2005, Chinacom'2006, Wirelesscom'2005. She was the IEEE COMSOC Data Storage Technical Committee chair 2011-2012, and an IEEE Distinguished lecturer 2012-2013. She received Lehigh University P.C. Rossin Assistant Professorship 2005, the TAMU Ethel Ashworth-Tsutsui Memorial Award for Research 2001, the Chinese Academy of Science Striving Scholarship 1994-1997, the candidacyship and third place prize in the '93 Chinese Mathematics Olympiad (CMO, the Winter Camp), the first place prize in '92 Chinese National Mathematics Competition for High School Students, and the first place prize in '92 Chinese National Physics Competition for High School Students.

Wentu Song received the BS and MS degrees in Mathematics from Jilin University, Changchun, China, in 1998 and 2006, respectively, and the PhD degree in mathematics from Peking University, Beijing, China, in 2012. He is currently a post doc research fellow in SUTD-MIT International Design Centre, Singapore University of Technology and Design, Singapore, where he is currently working on coding theory, network coding, and network distributed storage.



Yinlong Xu received the BS degree in mathematics from Peking University, Beijing, China, in 1983, and the MS and PhD degrees in computer science from University of Science and Technology of China (USTC), Hefei, China, in 1989 and 2004, respectively. He is currently a professor with the School of Computer Science and Technology at USTC. Prior to that, he served in the Department of Computer Science and Technology at USTC as an assistant professor, a lecturer, and an associate professor. Currently, he is leading a group of research students in doing some networking and high performance computing research. His current research interests include network coding, wireless network, combinatorial optimization, design and analysis of parallel algorithm, parallel programming tools, etc. He received the Excellent PhD Advisor Award of Chinese Academy of Sciences in 2006.