

Research Article

Network Coded Wireless Cooperative Multicast with Minimum Transmission Cost

Xiumin Wang,¹ Jin Wang,² and Shukai Zhang^{2,3}

¹ School of Computer and Information, Hefei University of Technology, Hefei 230009, China

² School of Computer Science and Technology, Soochow University, Suzhou 215006, China

³ State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China

Correspondence should be addressed to Jin Wang, wjin1985@suda.edu.cn

Received 27 July 2012; Accepted 6 September 2012

Academic Editor: Yong Sun

Copyright © 2012 Xiumin Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study multicasting over wireless lossy links. Instead of downloading all the data from the source node, we allow the destination nodes themselves to locally exchange the packets, as local communication within a cluster achieves higher packet reception probability with less transmission cost. However, when shall we stop the transmission from the source node? If the source stops too early, the destination nodes locally cannot reconstruct all the original packets, while if the source stops too late, the benefit of cooperative data exchange cannot be fully exploited. In this paper, we propose a network coded hybrid source and cooperative exchange scheme to determine when to stop the source sending and start the exchange process, so as to minimize the total transmission cost. For the case when the clusters are predefined, we derive the expected total transmission cost with our hybrid scheme. Our theoretical results show that under a special condition, the source node should keep sending the packets until all the destinations get the complete information. For the case when the clusters are not predefined, we propose a cluster division algorithm such that the destination nodes within each cluster can conduct data exchange locally with energy efficiency. Finally, simulation results demonstrate the effectiveness of the proposed scheme.

1. Introduction

Over the past decades, wireless sensor networks have attracted a great deal of research attentions [1–3]. Once deployed, sensors are expected to operate for a long period of time, and it is impractical to reach these sensors physically. However, it is quite often necessary to update the software running on those sensors or add new functionalities to the sensors [4, 5], which needs to reliably multicast large data objects with energy efficiency [5, 6]. Particularly, in wireless lossy sensor networks, multicasting packets from a single source node is still a challenge problem due to the heterogeneous lossy links to the destination nodes. To satisfy all the destination nodes, the source node needs to keep sending until the destination node with the worst packet reception link successfully receives all the packets, which is inefficient for the source node.

Recently, cooperative data exchange [7–11] has become a promising approach to achieve the efficient data communications. Instead of downloading all the packets from

the source node (e.g., the server) [12, 13], cooperative data exchange allows the destination nodes to cooperatively exchange their received packets among themselves, once the destination nodes collectively hold all the packets. Compared with pure source-dominated multicast, cooperative data exchange has two main benefits: (a) short-range communication among destination nodes is often more reliable and consumes less transmission cost, (b) the bandwidth saved at the source node can serve more other nodes in the system.

However, most of the existing works on cooperative data exchange problem never consider when to stop downloading the data from the source node. On the one hand, if the source node stops too early, the destination nodes themselves may be unable to collectively reconstruct the complete packets. On the other hand, if the source node stops too late, the benefit of cooperative data exchange cannot be fully utilized. In addition, although the average transmission cost with cooperative data exchange is lower than with source transmission, the sum of the transmission costs within multiple clusters may exceed the cost with pure source

sending. To sum up, for wireless multicasting, it is important for us to consider when to stop the source sending and start the cooperative data exchange, so as to reduce the total transmission cost or energy consumption [1–3, 14, 15].

So far, only the work in [16] studies the hybrid source transmission and cooperative data exchange for wireless multicasting, and it stops the source sending exactly once the destination nodes within each cluster collectively hold the complete information. Although the scheme in [16] performs well in reducing the transmission delay, it is still possible that the cooperative data exchange by multiple clusters may consume more transmission cost/energy than source-dominated scheme. As discussed above, the cooperative data exchange needs to be conducted separately in each cluster, and thus the sum of their transmission costs may exceed the transmission cost with pure source transmission. In addition, the predefined clusters in [16] might not perform well, as the destination nodes in some clusters may collect the complete information more quickly than the destinations in other clusters, which may stop the source transmission too late for the other clusters.

Recent work also shows that network coding [17–19] can improve the network throughput and reliability, especially for wireless lossy networks. Instead of sending/forwarding the original packets directly, network coding allows the source/transmitting node to linearly combine multiple packets together. With this approach, each transmitted packet has almost the same contribution in reconstructing the original packets, and hence improves wireless reliability [20]. Considering the benefits of network coding, we assume that the packets sent from the source node or the packets exchanged among the destination nodes are all linear encoded at the source/transmitting nodes before sending.

In this paper, given the heterogeneous link loss probability and the transmission cost of the source transmission and data exchange among the destinations, we aim to design a hybrid source and cooperative data exchange scheme to minimize the total transmission cost consumed during multicasting. For the case when the clusters are predefined, we determine when the source should stop sending the packets to the destination nodes and the cooperative data exchange should start. For the case when the clusters are not predefined, we consider how to divide the destinations into the clusters. The main contribution of the paper can be concluded as follows.

- (i) We theoretically derive the expected total transmission cost required with traditional source-dominated scheme and our hybrid transmission scheme.
- (ii) Our analysis shows that under a special condition, the source node should keep sending the packets until all the destinations get the complete information, so as to reduce the total transmission cost.
- (iii) We also propose an efficient algorithm to determine how to group the destination nodes into the clusters, so as to make sure that the destination nodes within each cluster can collectively recover all the original packets with energy efficiency.
- (iv) We compare the performance of the proposed scheme with some existing schemes. Simulation results show that the proposed scheme can significantly reduce the total transmission cost.

The rest of the papers are organized as follows. In Section 2, we introduce the background and some related works. The system model and the problem description are presented in Section 3. The expected total transmission costs with source-dominated and our scheme are discussed in Section 4. In Section 5, we consider how to group the destinations into clusters after receiving the sufficient number of packets from the source. The simulation results are presented in Section 6. We conclude the paper in Section 7.

2. Background and Related Work

In this section, we provide a brief introduction to the existing wireless network coded multicast scheme and summarize some related works.

2.1. Wireless Network Coding. Network coding was originally proposed in information theory [17] and recently has become a promising approach to improve the network performance in throughput [21, 22], reliability [19, 23], security [18, 24], and so forth. Instead of forwarding the original packets directly, network coding allows the source node/intermediate node to combine multiple packets together before sending it out. The work in [25] ensures that all the encoded packets generated by the same peer are linearly independent with a high probability, if we use linear network coding based on a sufficient large field size.

Network coding in wireless lossy networks was also considered in the literature. The work in [21] proposed a first wireless network coding architecture, named COPE. By exploiting the broadcast nature of wireless medium, each node stores the overheard packets for a while. When a node transmits a packet, it uses its knowledge of what its neighbors have overheard to perform opportunistic coding [21, 22]. After receiving an encoded packet, multiple neighbors can decode their wanted packets with their overheard packets. In other words, the sender or transmitting node can deliver “multiple packets” to different neighbors in a single transmission, which thus improves the throughput. The work in [19, 23] theoretically shows that network coding significantly reduces the expected number of retransmissions in lossy networks compared to traditional ARQ scheme. The work in [5] considers the impact of both wireless unreliable communication and sleep scheduling of sensor nodes and proposes a deterministic code design at the source node so as to accomplish the data dissemination process at the earliest time.

2.2. Reliable Multicast in Lossy Wireless Networks. Traditionally, wireless single-hop multicast mainly focuses on source-dominated transmission, where the source node keeps sending the packets until all the destination nodes in the system obtain the complete information. However, the performance of such a source-dominated transmission

degrades significantly when the packet reception probabilities of the destination nodes are heterogeneous. In this case, the source node cannot stop sending the packets until the destination node with the worst link state successfully gets the packets, even if all the other destination nodes received the packets in a much earlier time.

Recently, cooperative data exchange [7] among the users (e.g., mobile users) becomes one of the most promising approaches in designing efficient data transmissions. In cooperative data exchange, each client initially holds a subset of the packets and wants all the packets that others have. In the literature, it is assumed that there is a common communication channel among the users to receive/send the packets from/to all other users. Cooperative data exchange is to make sure that each client finally can get the complete packets by exchanging packets among themselves through the common channel. Compared with source-dominated transmissions, cooperative data exchange appears to have two benefits: (1) short-range communications among the users are often more reliable and faster, (2) the bandwidth saved at the server can serve more other clients in the system.

Most recent works on cooperative data exchange mainly focus on how to minimize the total number of packets to be exchanged/transmitted among the users [7–9], or the total transmission cost consumed at the users [10, 11], such that all the users in the system can finally recover/receive the complete packets. Current works also show that network coding can reduce the number of transmissions or the total transmission cost required for cooperative data exchange process [7–11].

However, in the literature, most of the works either focus on pure source-dominated multicast or focus on cooperative data exchange by assuming that each user initially holds a subset of packets. The only work that considers the hybrid architecture of source transmission and data exchange transmission is in [16]. Specifically, the source node is set to stop sending the packets once the destinations in each cluster can collectively reconstruct the original packets, followed by the cooperative data exchange within each cluster. The numerical results show that with this hybrid transmission, the transmission delay of the multicast can be reduced compared with pure source-dominated multicast. However, the source node may stop too early, as the sum of the energy consumed by the cooperative data exchange within multiple clusters may consume more energy, which is inefficient for wireless sensor networks.

3. System Model and Problem Description

In this section, we first introduce the system model of our problem. Then, we discuss the problem to be solved in two different cases: with predefined clusters and without predefined clusters.

3.1. System Model. In this paper, we consider a multicast application, where a source node s needs to send n packets to m destination nodes in $D = \{d_1, d_2, \dots, d_m\}$. Before transmitting the packets, the source node may generate more

than n encoded packets over the original n packets. It is typical to assume that with network coding, after receiving any n encoded packets, the destination node can recover the original n packets.

Suppose that the destination nodes are formed into multiple clusters in $\mathcal{C} = \{C_1, C_2, \dots\}$, where $C_k \in \mathcal{C}$ denotes the set of destination nodes in cluster k . Without loss of generality, we assume that each pair of clusters maintain disjoint destination nodes, that is, $C_{k_1} \cap C_{k_2} = \emptyset$, for all $k_1 \neq k_2$ and $\bigcup_{C_k \in \mathcal{C}} C_k = D$. As in cooperative data exchange, we assume that local data exchange within each cluster is conducted through a common channel.

Our multicasting process consists of two transmission stages. In the first stage, the source node sends the packets to all the destination nodes in D . In the second stage, the destination nodes within the same cluster perform the cooperative data exchange among themselves. These two stages need to make sure that each destination finally can reconstruct the original n packets. Due to unreliable wireless communications, suppose that the packet loss probability from the source node to every destination node is l_s , and the packet loss probability between the destination nodes in cluster C_k is l_k . We also assume that the transmission cost of the source node is t_s , and the transmission cost for the data exchange within cluster C_k is $t(C_k)$. Without loss of generality, we assume that the transmission cost of the source node is higher than the transmission cost of the destination node, that is, $t_s \geq t_{C_k}$ for $\forall k$, and the reception probability of the packet sent from s is lower than the reception probability of the packet exchanged within the same cluster, that is, $l_s \geq l_k$.

In this paper, we aim to minimize the total transmission cost consumed during transmissions, while we ensure that all the destination nodes in D can successfully reconstruct the original n packets. Let x_{hs} be the number of packets that the source node sends, and x_k be the number of packets sent by the destination nodes within the cluster C_k . Thus, the total transmission cost can be written as

$$x_{hs}t_s + \sum_{C_k \in \mathcal{C}} x_k t(C_k) \quad (1)$$

which needs to be minimized.

An example of the system model is shown in Figure 1. The source node s needs to send three packets to the destinations in $D = \{d_1, d_2, \dots, d_7\}$, and the original packets are generated into three encoded packets p_1, p_2 , and p_3 . The set of the packets received at each destination is given near the node after the source node sends three packets p_1, p_2, p_3 , for example, the set of packets received by node d_4 is $\{p_3\}$. In the next subsections, we will discuss the problems to be considered in this example.

3.2. Problem Description with Predefined Clusters. In this section, we consider the case when the destination nodes in each cluster are predetermined.

Due to unreliable wireless communications, some packets may be lost at some destination nodes. So, one challenge problem is to determine when the source node s stops

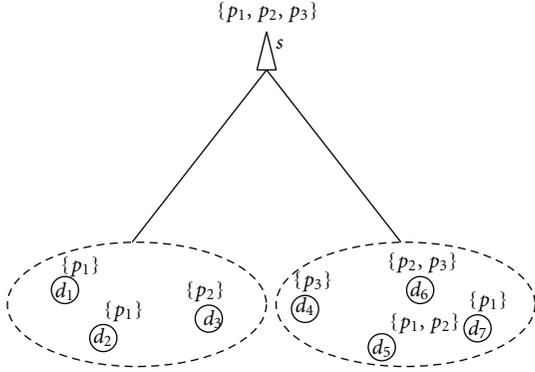


FIGURE 1: An example of system model after sending three packets in $\{p_1, p_2, p_3\}$ from source node s to receiver nodes in $D = \{d_1, d_2, \dots, d_7\}$.

sending the packets such that the destination nodes within each cluster can exchange the packets among themselves, and the total transmission cost defined in (1) is minimum.

- (i) If the source node s stops too early, the destination nodes in some clusters may be unable to collectively recover all the original packets.
- (ii) If the source node s stops too late, the low-cost transmission among the destination nodes themselves cannot be fully utilized, which may incur high transmission cost because of high packet loss probability and high transmission cost of the source transmission.
- (iii) The sum of the transmission costs by the data exchange by all the clusters may exceed the transmission cost by the pure source transmission. In this case, the source node should keep sending until all the destination nodes obtain the complete information.

As shown in Figure 1, if the source node s stops after sending p_1, p_2, p_3 , the destination nodes in the left cluster (within a circle) cannot collectively reconstruct the original three packets. On contrary, if the source node s sends too many packets, all the destination nodes within each cluster may obtain the complete packets, which does not utilize the low-cost cooperative data exchange. However, if the sum of the transmission costs in these two clusters are higher than the transmission cost from the source node, it is better to let the source node keep sending until all the destination nodes get the complete information.

We refer to the above problem of determining when to stop the source transmission, as minimizing the multicast transmission cost problem with predefined clusters.

3.3. Problem Description with Nonpredefined Clusters. In the above section, we assume that the destination nodes in each cluster are given. However, the predefined clusters may not perform well, as the destination nodes in some clusters may collectively hold the complete information much later than the destination nodes in other clusters, which thus delays

TABLE 1: Main notations and their descriptions.

\mathcal{C}	The set of the clusters
C_k	The k th cluster in \mathcal{C}
D	The set of the destination nodes in the system
d_i	The i th destination node in D
l_s	Packet loss probability on the link from s to each destination
l_k	Packet loss probability of the data exchange within cluster C_k
m	The total number of destination nodes in D
n	The total number of original packets
s	The source node
t_s	The transmission cost of the source node s
$t(C_k)$	The transmission cost of the data exchange within cluster C_k
x_{hs}	Number of packets sent by s in the first stage of hybrid scheme
x_k	The number of packets exchanged within cluster C_k

the starting time of the cooperative data exchange in other clusters.

Still take Figure 1 as an example, if we group the destinations in $\{d_1, d_2, \dots, d_7\}$ into two clusters, that is, the destinations in the same circle belong to the same cluster. After the source node s sends three packets p_1, p_2, p_3 , the destination nodes in the left cluster cannot recover the original packets, as they collectively have received only two packets so far. However, if we move destination node d_4 into the left cluster, the destination nodes in both clusters can reconstruct all the three original packets, and then s can stop sending the packets.

Thus, after the source node sends sufficient number of packets, we need to consider how to group the destination nodes into multiple clusters such that the destination nodes within each cluster collectively can reconstruct the original packets with energy efficiency. In this case, the number of clusters in \mathcal{C} and the set of the destinations in each cluster C_k need to be determined by our algorithm.

4. Minimum Multicast Transmission Cost with Predefined Clusters

In this section, we consider the case when the clusters are predefined. We first analyze the expected total transmission costs with source-dominated scheme and our hybrid transmission scheme. We then formulate the problem of minimizing the total transmission cost with our hybrid transmission scheme as an integer programming.

To ease the understanding, the main notations used in the paper are given in Table 1.

4.1. Transmission Cost with Source-Dominated Transmission. With source-dominated scheme, the source node s keeps sending the encoded packets that are generated based on n original packets, until all the destination nodes in D

```

begin
  //after sending  $x_{hs}$  packets from the source node  $s$ 
  if the destinations in  $D$  cannot recover the original packets
  then
    The source  $s$  should send more packets before the
    second stage;
  end
  else
     $U = D$ ;
     $C_1 = \emptyset$ ;
    not_finish=1;
     $k = 1$ ;
    while not_finish do
      find a destination  $d_i \in U$  which satisfies the rules
      (1), (2), and (3) and incurs the least cost to  $C_k$ ;
      if There does not exist a feasible  $d_i \in U$  then
         $C_k = C_k \cup U$ ;
        not_finish=0;
      end
      else
        add  $d_i$  into  $C_k$ ;
        delete  $d_i$  from  $U$ ;
        if If the destinations in  $C_k$  can collectively
        recover the original packets then
           $k = k + 1$ ;
           $C_k = \emptyset$ ;
        end
      end
    end
  end
end

```

ALGORITHM 1: Algorithm design for cluster determination.

successfully decode the original n packets. With linear network coding, after receiving any n encoded packets from s , the destination node can decode the original n packets.

Suppose that \bar{x}_s is the expected transmission cost required with the source-dominated scheme. To make sure all the destinations can recover the complete packets, the number of packets that s sends should be at least the maximum number of packets to be sent to satisfy the decoding requirement of each destination. That is

$$\bar{x}_s = \max\{\bar{x}'_1, \bar{x}'_2, \dots, \bar{x}'_m\}, \quad (2)$$

where \bar{x}'_i means the expected number of packets required to be sent by s , so as to guarantee that destination d_i can decode all the n original packets. Since the packet loss probability from s to each destination node is l_s , we can derive \bar{x}'_i as follows:

$$\bar{x}'_i = \frac{n}{1 - l_s}. \quad (3)$$

In other words, we have

$$\bar{x}_s = \max\{\bar{x}'_i \mid \forall i\} = \frac{n}{1 - l_s}. \quad (4)$$

With the above equations, we can obtain the expected transmission cost as follows:

$$\bar{x}_s t_s = \frac{nt_s}{1 - l_s}. \quad (5)$$

4.2. Transmission Cost with Hybrid Transmission Scheme.

With hybrid transmission scheme, there are two stages. In the first stage, the source node s sends x_{hs} encoded packets. In the second stage, the destination nodes in the same cluster cooperatively exchange the packets that they received before, until each of them can decode all the packets.

We now consider the first stage. For the success of the second stage, the number of packets sent by the source s should make sure that the destination nodes within each cluster can collectively reconstruct all the original packets. The probability that at least one of the destinations in cluster C_k can receive the current transmission is $1 - l_k^{|C_k|}$. Thus, to guarantee at least n transmissions of the packets are successfully received by the collection of the destination nodes in C_k , the expected number of packets sent by the source node s , denoted as \bar{x}_{hs} , should satisfy the following condition:

$$\bar{x}_{hs} \geq \frac{n}{1 - l_k^{|C_k|}}. \quad (6)$$

When considering all the clusters, (6) can be written as

$$\bar{x}_{hs} \geq \max_{C_k \in \mathcal{C}} \left\{ \frac{n}{1 - l_k^{|C_k|}} \right\}. \quad (7)$$

Note that if we set $\bar{x}_{hs} = \min_{C_k \in \mathcal{C}} \{n/(1 - l_k^{|C_k|})\}$, it means the source node s stops sending the packets exactly once the destination nodes in each cluster collectively can recover the complete packets.

We then consider the second stage. After the source node sends \bar{x}_{hs} packets, the expected number of packets required by each destination in cluster C_k is

$$\bar{L} = n - \bar{x}_{hs}(1 - l_s). \quad (8)$$

Then, the expected number of packets to be sent by other destinations through cooperative data exchange, denoted as \bar{M}_k , such that one specific destination in cluster C_k can decode, is

$$\bar{M}_k = \frac{\bar{L}}{1 - l_k}. \quad (9)$$

According to the work in [16], the expected number of packets required to be exchanged among cluster C_k such that all the destination nodes in C_k can obtain its required \bar{M}_k packets is

$$\bar{x}_k = \frac{|C_k|}{|C_k| - 1} \left(\bar{M}_k + \bar{W}_{|C_k|} \sqrt{\sigma_{\bar{M}_k}^2} \right), \quad (10)$$

where $\bar{W}_{|C_k|}$ is the expectation of the $|C_k|$ -th order statistic of a sequence of $|C_k|$ normal random variables, and $\sigma_{\bar{M}_k}^2$ (given in [16]) is the variance of \bar{M}_k .

Thus, the total transmission cost with hybrid transmission scheme can be formulated as follows:

$$\bar{x}_{hs} t_s + \sum_{C_k \in \mathcal{C}} t(C_k) \bar{x}_k. \quad (11)$$

If we assume that the packet loss probability within each cluster is the same, that is, $l_k = l_r$ for for all $C_k \in \mathcal{C}$, and the transmission cost within each cluster is also the same, that is, $t(C_k) = t_r$ for $\forall C_k \in \mathcal{C}$, we can obtain the following theorem.

Theorem 1. *If $|\mathcal{C}| t_r (1 - l_s) \geq t_s (1 - l_r)$, to minimize the expected total transmission cost, it is better to let the source node keep sending the packets until all the destination nodes in D successfully obtain the original n packets, that is, $x_k = 0$.*

Proof. Let \bar{x}_{hs} be the number of packets sent by the source node s in a hybrid transmission scheme. Since we assume that $|\mathcal{C}| t_r (1 - l_s) \geq t_s (1 - l_r)$, we have

$$t_r \geq \frac{t_s (1 - l_r)}{|\mathcal{C}| (1 - l_s)}. \quad (12)$$

Based on the above equation (12), we then compare the expected total transmission cost required by source-dominated scheme with that by the hybrid transmission scheme for any $x_k > 0$ as follows:

$$\begin{aligned} & \bar{x}_{hs} t_s + \sum_{C_k \in \mathcal{C}} t(C_k) \bar{x}_k - \bar{x}_s t_s \\ &= \bar{x}_{hs} t_s + \sum_{C_k \in \mathcal{C}} t(C_k) \bar{x}_k - \frac{nt_s}{1 - l_s} \\ &\geq \bar{x}_{hs} t_s + |\mathcal{C}| t_r \bar{M}_k - \frac{nt_s}{1 - l_s} \\ &= \bar{x}_{hs} t_s + |\mathcal{C}| t_r \frac{n - \bar{x}_{hs} (1 - l_s)}{1 - l_r} - \frac{nt_s}{1 - l_s} \\ &\geq \bar{x}_{hs} t_s + |\mathcal{C}| \frac{n - \bar{x}_{hs} (1 - l_s)}{1 - l_r} \frac{t_s (1 - l_r)}{|\mathcal{C}| (1 - l_s)} - \frac{nt_s}{1 - l_s} \\ &= \bar{x}_{hs} t_s + \frac{(n - \bar{x}_{hs} (1 - l_s)) t_s}{1 - l_s} - \frac{nt_s}{1 - l_s} \\ &= \frac{\bar{x}_{hs} t_s (1 - l_s) + (n - \bar{x}_{hs} (1 - l_s)) t_s - nt_s}{1 - l_s} \\ &= 0. \end{aligned} \quad (13)$$

From the above equation, we can see that the expected total transmission cost with hybrid scheme is not lower than source-dominated transmission scheme when $|\mathcal{C}| t_r (1 - l_s) \geq t_s (1 - l_r)$. In other words, to reduce the total transmission cost, the source node should keep sending until all the destination nodes can successfully decode the complete information, that is, $x_k = 0$, which thus proves the theorem. \square

4.3. Hybrid Transmission Scheme with Minimum Total Transmission Cost. As discussed above, in some cases, the sum of the transmission costs within all the clusters may be more than transmission cost of the source transmission. Under such circumstances, the source node should keep sending the packets until all the destination nodes get the complete information, which is a special case of our hybrid scheme, that is, $x_{hs} = x_s$.

To determine x_{hs} , the problem of minimizing the total transmission cost with hybrid transmission scheme can be formulated as follows:

$$\min \bar{x}_s t_s + \sum_{C_k \in \mathcal{C}} t(C_k) \bar{x}_k \quad (14)$$

subject to (7).

With the above integer formulation, we can obtain the best time to stop the source transmission and start the cooperative data exchange process.

5. Cluster Determination in Cooperative Data Exchange Stage

As discussed in Section 3.3, after the source transmissions in the first stage, the destination nodes in some clusters may

be unable to reconstruct the original packets if the clusters are predefined. In this section, we consider after the source node sends x_{hs} packets in the first stage, where $x_{hs} \geq n$, how to group the destination nodes into multiple clusters so as to make sure that the destination nodes in each cluster collectively can recover all the original packets.

Assume that after the source node sends x_{hs} packets, the set of the packets received by destination node d_i is denoted as H_i , for example, $H_6 = \{p_2, p_3\}$ in Figure 1. Since the total number of packets sent by the source node s in the first stage is x_{hs} , we have $H_i \subseteq \{p_1, p_2, \dots, p_{x_{hs}}\}$. Before describing how to group the clusters, we first discuss given the exact packet reception states of the destination nodes, how many transmissions are required by the data exchange process within a local cluster.

5.1. Number of Transmissions Required for Data Exchange Process within a Local Cluster. We now discuss the number of transmissions required for data exchange within a specific cluster C_k , given the exact packet reception sets of the destination nodes in C_k .

Without loss of generality, we assume that after the source s sends x_{hs} packets, all the destination nodes in cluster C_k collectively can recover the original n packets. As in the above section, we assume that the packets sent by the source node s are linear independent with each other, that is, after receiving any n packets from s , the node can decode the original n packets. In other words, after the first stage, the total number of packets collectively received by the destination nodes within each cluster C_k should satisfy

$$\left| \bigcup_{d_i \in C_k} H_i \right| \geq n. \quad (15)$$

Before describing the number of transmissions for data exchange process within cluster C_k , we first introduce a useful existing result in [26].

Theorem 2 (see [26]). *Provided that the encoding field size is large enough, the minimum number of packets to be exchanged is given by*

$$x = \text{rank} \left(\bigcup_{d_i \in C_k} H_i \right) - \min_P \frac{S_R(P) - \text{rank}(\bigcup_{d_i \in C_k} H_i)}{|P| - 1}, \quad (16)$$

where $P = \{S_1, S_2, \dots, S_{|P|}\}$ denotes a disjoint partition of the node sets and $S_j \in P$ is the subset of the nodes in C_k , where $2 \leq P \leq |C_k|$, and $\text{rank}(\bigcup_{d_i \in C_k} H_i)$ means the rank of the matrix including the encoding vectors of the packets in $\bigcup_{d_i \in C_k} H_i$, $S_R(P) = \sum_{S_j \in P} \text{rank}(\bigcup_{d_i \in S_j} H_i)$.

Note that in our problem, the packets sent by the source node s are linear independent with each other. Thus, for a

given cluster C_k , we can obtain that

$$\begin{aligned} \text{rank} \left(\bigcup_{d_i \in C_k} H_i \right) &= n, \\ S_R(P) &= \sum_{S_j \in P} \text{rank} \left(\bigcup_{d_i \in S_j} H_i \right) \\ &= \sum_{S_j \in P} \min \left\{ \left| \bigcup_{d_i \in S_j} H_i \right|, n \right\}. \end{aligned} \quad (17)$$

Given the packets received by the destination nodes in C_k , according to Theorem 2, we can then derive the number of packets to be exchanged among the destinations in C_k as follows:

$$x_k = n - \min_P \frac{\sum_{S_j \in P} \min \left\{ \left| \bigcup_{d_i \in S_j} H_i \right|, n \right\} - n}{|P| - 1}, \quad (18)$$

where P is a disjoint partition of the nodes in C_k and $S_j \in P$ is the j th subset of the nodes by the partition P .

5.2. Optimal Cluster Division with Minimum Transmission Cost in the Data Exchange Process. We now consider how to divide all the destination nodes into multiple clusters, so as to minimize the total transmission cost in the second stage.

For m destination nodes in D , the maximum number of clusters that can be formed is m . The problem of minimizing the total transmission cost of the data exchange within the clusters by cluster division can be formulated as follows:

$$\min_{C_k} \sum_{k=1}^m t(C_k) x_k \quad (19)$$

subject to

$$C_k \cap C_{k'} = \emptyset, \quad (20)$$

$$\bigcup_{k=1}^m C_k = D, \quad (21)$$

$$\left| \bigcup_{d_i \in C_k} H_i \right| = n. \quad (22)$$

In (19), $t(C_k)$ means the transmission cost if the destination nodes in C_k form a cluster, and x_k is number of packets to be sent by the data exchange within cluster C_k , which can be calculated with (18). Thus, the objective is to minimize the total transmission cost of the data exchange within all the clusters, by determining C_k . The constraints in (20) and (21) denote that each destination node should be located in one and only one cluster. The constraint in (22) represents that the destination nodes can form a cluster if and only if these destination nodes collectively can recover all the n original packets.

Although the above formulation can derive the optimal cluster division to minimize the total transmission cost in

the data exchange process, the complexity of calculating x_k is too high, as we need to enumerate all the possible partitions of the nodes in C_k , $\{P\}$. It makes the formulation difficult to be solved. Thus, in the next subsection, we propose a suboptimal algorithm to divide the clusters.

5.3. Algorithm Design. In this section, given the packet reception state of each destination node, we consider to group all the destination nodes into multiple clusters.

Without loss of generality, we suppose that after x_{hs} transmissions from source node s , all the destinations collectively can recover the original n packets. Otherwise, the source node s should send more packets before starting the second stage (i.e., cooperative data exchange). Let U be the set of all the destinations left in the system, that is, $U = D$.

We then introduce how to select the destinations in U into the k th cluster C_k . When adding a new destination d_i from U into C_k , we make the following rules.

- (1) The destination d_i should be able to contribute at least one “innovative” packet to the destination nodes that have been added to C_k so far.
- (2) Deleting d_i from U will not sacrifice the decoding capability of the destinations left in U , that is, (22).
- (3) The sum of the transmission costs of the new cluster C_k (including d_i) and the cluster formed by the left destinations in U (excluding d_i) is smaller than the transmission cost of the cluster formed by $U \cup C_k$.

We then describe the process of adding new destinations to C_k as follows.

- (i) We start from the destination node $d_i \in U$ that satisfies the above three rules and incurs the least transmission cost if it is added into C_k . If there does not exist such a node in U , we can obtain that the destinations in U can only form one cluster, that is, $C_k = U$. In this case, the algorithm terminates.
- (ii) If we can find a feasible destination d_i , we add d_i into C_k and correspondingly delete d_i from U . We then check if the destinations in C_k can collectively recover all the original packets. If they can, we finish the cluster C_k and repeat the above process to find the $(k + 1)$ th cluster C_{k+1} from U .
- (iii) If they still cannot recover all the original packets, we continue the above two steps.
- (iv) The algorithm continues until we cannot find a feasible node $d_i \in U$ that satisfies the rules (1), (2), and (3). In this case, the destination nodes left in U should be all added to the current considered cluster C_k , that is, $C_k = C_k \cup U$.

The detailed process of the above algorithm is shown in Algorithm 1. With the above algorithm, we can make sure that the destination nodes within each cluster $C_k \in C$ collectively can reconstruct all the original packets.

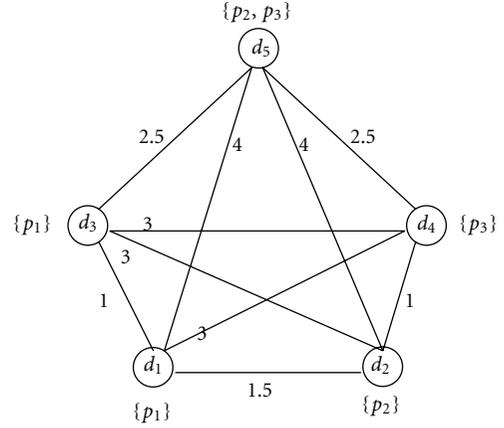


FIGURE 2: An example of five destination nodes.

5.4. Illustration Example. We take Figure 2 as an example to illustrate how Algorithm 1 works, where five destination nodes in $\{d_1, d_2, \dots, d_5\}$ need to get three original packets. Assume that the first stage stops after the source node s sends three packets $\{p_1, p_2, p_3\}$. The set of the packets received by each destination in the first stage is given near the node. To simplify the understanding, we define the transmission cost of the transmission within a cluster C_k as the square of the maximum distance between every two destinations within the same cluster [25, 27], that is,

$$t(C_k) = \max_{d_i, d_{i'} \in C_k} \{|d_i - d_{i'}|^2\}, \quad (23)$$

where $|d_i - d_{i'}|$ denotes the distance between two nodes d_i and $d_{i'}$ (given on the edge between two nodes). Note that the above model is only an example, and our algorithm does not restrict the model of the transmission cost.

According to the Algorithm 1, initially we have $U = \{d_1, d_2, \dots, d_5\}$. We can easily check that if all the nodes in U form only one cluster, its transmission cost is $4^2 = 16$.

We first construct the cluster C_1 , by starting from any node d_1 . Since d_1 satisfies all the three rules defined in the above section, we add d_1 into C_1 and delete it from U , that is, $C_1 = \{d_1\}$, $U = \{d_2, d_3, \dots, d_5\}$. Since the nodes in C_1 cannot reconstruct the original three packets, we need to add more nodes to C_1 . Note that although adding d_3 incurs the least transmission cost of C_1 , d_3 violates the rule (1), as the packet it has is not innovative to the nodes in C_1 . We then find that d_2 can be added to C_1 and it incurs the least transmission cost among the nodes left in $U \setminus \{d_3\}$, that is, $C_1 = \{d_1, d_2\}$, $U = \{d_3, d_4, d_5\}$. With a similar approach, d_4 will be added to C_1 , that is, $C_1 = \{d_1, d_2, d_4\}$, $U = \{d_3, d_5\}$. Since the destinations in C_1 now can collectively reconstruct the original three packets, the construction of cluster C_1 terminates, and the transmission cost of C_1 is $t(C_1) = 3^2 = 9$.

We then construct the cluster C_2 . We can easily check that any node in $U = \{d_3, d_5\}$ cannot satisfy the rule (2). Thus, the destination nodes in U cannot be divided into multiple clusters. In other words, all the destinations in U are added into C_2 , that is, $C_2 = \{d_3, d_5\}$, and the

transmission cost of C_2 is $t(C_2) = 2.5^2 = 6.25$. With the above operation, the destination nodes in D form two clusters, $C_1 = \{d_1, d_2, d_4\}$, $C_2 = \{d_3, d_5\}$. Note that, the algorithm can start from any node, for example, if starting from node d_2 , we can also get two clusters $\{d_2, d_4, d_1\}$, $\{d_3, d_5\}$.

6. Simulation Results

In the simulation, we study a connected network graph, where nodes are randomly deployed in a two-dimensional (2D) space. We use l_s to simulate the packet loss probability on the link from the source node s to the destination nodes, and l_k to be the packet loss probability by the data exchange within cluster C_k . For the transmission cost, we use t_s and $t(C_k)$ to denote the transmission cost of the packet sent from source node and the packet exchanged within cluster C_k , respectively. Generally, we set $l_s \geq l_k$ and $t_s \geq t(C_k)$.

To demonstrate the performance of our proposed scheme, we also conduct two baseline algorithms: source-dominated scheme, and the hybrid scheme by [16]. As introduced before, in source-dominated scheme, the source node keeps sending until all the destination nodes get the complete packets. The difference between the hybrid scheme proposed by [16] and ours is that the scheme in [16] stops the source sending exactly once the destination nodes in each cluster can collectively reconstruct the full packets.

We define cost gain, denoted as δ , as the ratio of the transmission cost difference with source-dominated scheme and the hybrid scheme (our hybrid scheme or the scheme in [16]) to the total transmission cost with source-dominated scheme, for example, the cost gain by our hybrid scheme can be written as

$$\delta = \frac{x_s t_s - (x_{hs} t_s + \sum_{C_k \in \mathcal{C}} x(C_k) t_k)}{x_s t_s}. \quad (24)$$

6.1. The Impact of the Transmission Cost. In this section, we conduct the simulation to investigate the impact of the transmission cost on the cost gain. In this setting, we set $|\mathcal{C}| = 2$, $l_s = 0.5$, $l_k = 0.2$, $t_s = 6$, $|C_k| = 5$ and vary the transmission cost of the data exchange within each cluster in [1, 6].

As shown in Figure 3, the cost gain with our hybrid scheme is much better than the scheme proposed by [16]. We can observe that in some cases, for example, $t(C_k) \geq 3$, the total transmission cost with the scheme [16] is even more than source-dominated scheme. This is because, when the transmission cost of the packet exchange within a cluster is high, the sum of the transmission costs in multiple clusters exceeds the cost consumed by pure source sending. In other words, the hybrid scheme in [16] stops the source transmission too early. From the figure, we can also find that when $t(C_k)$ is no more than 3, the transmission cost with our hybrid scheme is better than the source-dominated scheme. However, when $t(C_k) \geq 4$, it is the same as the source-dominated scheme, which verifies the results of Theorem 1. In this case, the programming used in our hybrid scheme can

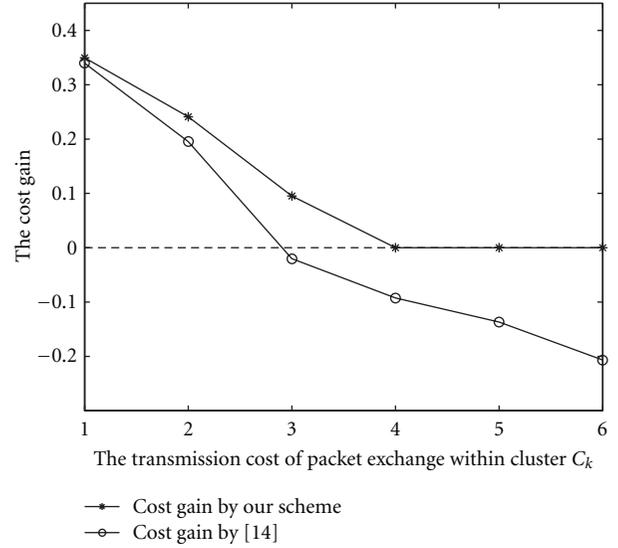


FIGURE 3: The impact of transmission cost of each cluster on the transmission cost gain ratio.

detect and keep the source sending until all the nodes receive the complete packets.

From Figure 3, we can also observe that with the increase of the transmission cost within a cluster, the cost gain decreases. This is reasonable, as when the transmission cost within a cluster increases, the sum of the total transmission costs of multiple clusters increases quickly and correspondingly reduces the gain.

6.2. The Impact of Packet Loss Probability. We now investigate the impact of packet loss probability on the performance of the total transmission cost. We fix $n = 100$, $l_k = 0.2$, $|\mathcal{C}| = 2$, $t_s = 6$, $t(C_k) = 3$ and vary the packet loss probability on the link from the source to the destination in [0.4, 0.9].

As shown in Figure 4, the cost gain with our hybrid scheme is better than with the scheme in [16], and our hybrid scheme also always consumes less transmission cost compared with source-dominated transmission scheme. From the figure, we observe that with the increase of the packet loss probability from the source to the destination nodes, the transmission cost gains with both hybrid schemes increase. This is because, when the links from the source to the destinations nodes are too bad, the source node should stop sending the packet as early as possible.

In addition, when the packet loss probability $l_s \geq 0.8$, the cost gains with both hybrid schemes are almost the same, as our hybrid scheme may stop the source transmission at almost the same time as in [16], that is, after the destination nodes in each cluster can reconstruct the complete packets.

6.3. The Impact of the Number of Clusters. We now conduct the simulation to investigate the impact of the number of the clusters on the total transmission cost of the proposed schemes. In this setting, we fix $n = 100$, $l_s = 0.6$, $l_k = 0.2$, $t_s = 6$, $t(C_k) = 2$ and the number of clusters are varied from 1 to 7.

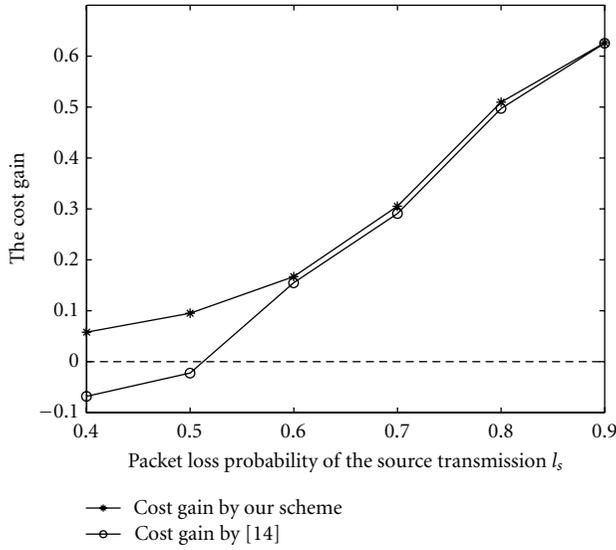


FIGURE 4: The impact of packet loss probability on the transmission cost gain ratio.

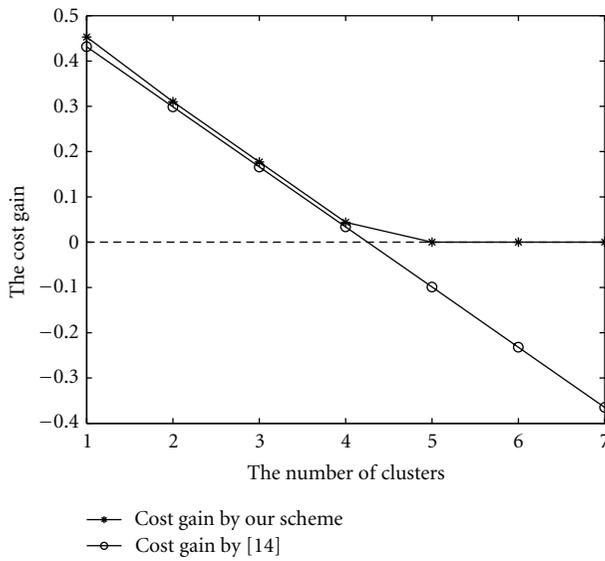
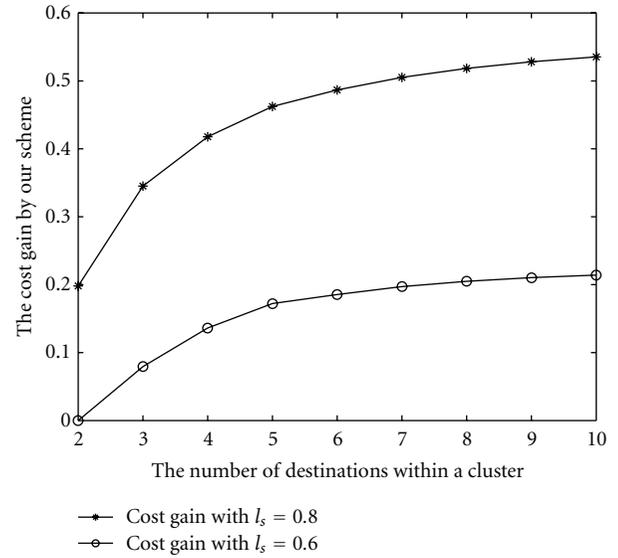
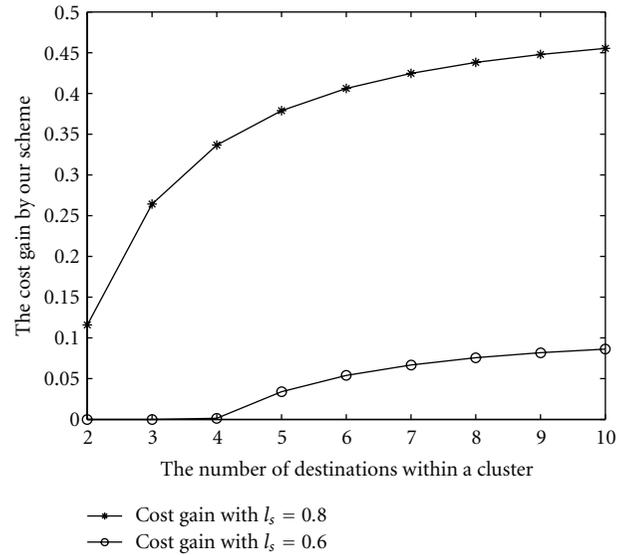


FIGURE 5: The impact of the number of clusters on the transmission cost gain ratio.

As shown in Figure 5, when the number of clusters is few, the total transmission costs with both hybrid schemes are less than with the source-dominated scheme. This is because, when the number of clusters is few, the sum of their transmission costs does not exceed the source transmission cost. However, when the number of clusters increases, for example, more than 5, pure source transmission is better, since the sum of the transmission costs in the clusters is more than the transmission cost from the source. However, we can observe that, the transmission cost with our scheme is not higher than with source-dominated scheme in all the cases, as our hybrid scheme always chooses the best time to stop the source transmission.



(a) $|\mathcal{C}| = 3$



(b) $|\mathcal{C}| = 4$

FIGURE 6: The impact of the number of destinations within the cluster on cost gain by our scheme.

6.4. *The Impact of the Number of Receiver Nodes within Clusters.* Finally, we investigate the impact of the number of destination nodes within each cluster on the performance of our transmission scheme.

As shown in Figure 6, we fix $n = 100$, $t_s = 6$, $t(C_k) = 2$, $l_k = 0.2$ and vary the number of destination nodes within each cluster in $[2, 10]$. From the figure, we can see that with the increase of the number of the destination nodes within each cluster, the cost gain with our hybrid scheme increases. This is because, with more destination nodes in a cluster, it is much more earlier that the destinations in the cluster collectively can reconstruct the original n packets. Thus, the second stage of our hybrid scheme, data exchange process,

can start earlier, which decreases the number of high-cost transmissions from the source node s .

From Figure 6, we also observe that the cost gain with the setting $l_s = 0.8$ is much higher than with the setting $l_s = 0.6$. The reason is that with source-dominated scheme, a large number of the transmissions will be wasted when the packet reception state on the links from s to the destination nodes is bad. In this case, the cooperative data exchange process in our hybrid scheme should start as early as possible, as it consumes less transmission cost. In other words, our hybrid scheme performs well especially when the packet reception from the source node is bad.

By comparing Figures 6(a) and 6(b), we can also obtain that when the number of clusters in the network increases, the cost gain of our hybrid scheme decreases. As each cluster consumes its independent transmission cost in the cooperative data exchange process, the sum of the transmission costs of the data exchange by all the clusters increases, which thus decreases the cost gain.

7. Conclusion

In this paper, we proposed a hybrid source and cooperative data exchange transmission scheme for reliable multicasting over wireless lossy links. Our hybrid scheme determines when to stop the source sending and start the cooperative data exchange, so as to minimize the total transmission cost. We theoretically derive the total transmission cost required with traditional source-dominated scheme and our hybrid scheme. We give a condition under which the source node should not stop sending the packets until all the destination nodes successfully get the complete information, so as to reduce the total transmission cost. If the clusters are predefined, we propose an efficient algorithm to divide the destination nodes into multiple clusters, such that the destination nodes within each cluster can conduct the cooperative data exchange separately with energy efficiency. Finally, simulation results demonstrate the effectiveness of the proposed scheme in reducing the total transmission cost.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (no. 2012HGBZ0640). It was also supported in part by National Natural Science Foundation of China under Grants no. 61202378, 61070169, Natural Science Foundation of Jiangsu Province under Grant no. BK2011376, Specialized Research Foundation for the Doctoral Program of Higher Education of China no. 20103201110018, and Application Foundation Research of Suzhou of China no. SYG201118.

References

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [2] H. M. N. D. Bandara, A. P. Jayasumana, and T. H. Illangasekare, "A top-down clustering and cluster-tree-based routing scheme for wireless sensor networks," *International Journal of Distributed Sensor Networks*, Article ID 940751, 17 pages, 2011.
- [3] H. Liu, X. Chu, Y. Leung, X. Jia, and P. Wan, "General maximal lifetime Sensor-Target surveillance problem and its solution," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 10, pp. 1757–1765, 2011.
- [4] S. S. Kulkarni and L. Wang, "MNP: multihop network reprogramming service for sensor networks," in *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems*, pp. 7–16, USA, June 2005.
- [5] X. Wang, J. Wang, and Y. Xu, "Data dissemination in wireless sensor networks with network coding," *Eurasip Journal on Wireless Communications and Networking*, vol. 2010, Article ID 465915, 2010.
- [6] Y. F. Wen and W. Liao, "Minimum power multicast algorithms for wireless networks with a Lagrangian relaxation approach," *Wireless Networks*, vol. 17, no. 6, pp. 1401–1421, 2011.
- [7] S. El Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," in *Proceedings of the IEEE Information Theory Workshop 2010 (ITW '10)*, pp. 1–5, January 2010.
- [8] T. A. Courtade, B. Xie, and R. D. Wesel, "Optimal exchange of packets for universal recovery in broadcast networks," in *Proceedings of the IEEE Military Communications Conference (MILCOM '10)*, pp. 2250–2255, November 2010.
- [9] N. Milosavljevic, S. Pawar, S. El-Rouayheb, M. Gastpar, and K. Ramchandran, "Deterministic algorithm for the cooperative data exchange problem," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '11)*, pp. 410–414, 2011.
- [10] S. Tajbakhsh, P. Sadeghi, and R. Shams, "A generalized model for cost and fairness analysis in coded cooperative data exchange," in *Proceedings of the IEEE International Symposium on Network Coding (NetCod '11)*, pp. 1–6, 2011.
- [11] X. Wang, W. Song, C. Yuen, and J. T. Li, "Exchanging third-party information with minimum transmission cost," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, 2012.
- [12] K. Liu and V. C. S. Lee, "RSU-based real-time data access in dynamic vehicular networks," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC '10)*, pp. 1051–1056, September 2010.
- [13] K. Liu and V. C. S. Lee, "Adaptive data dissemination for timeconstrained messages in dynamic vehicular networks," *Transportation Research C*, vol. 21, pp. 214–229, 2012.
- [14] M. Al-Ameen, S. M. R. Islam, and K. Kwak, "Energy saving mechanisms for MAC protocols in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2010, Article ID 163413, 16 pages, 2010.
- [15] M. Elkin, Y. Lando, Z. Nutov, M. Segal, and H. Shpungin, "Novel algorithms for the network lifetime problem in wireless settings," *Wireless Networks*, vol. 17, no. 2, pp. 397–410, 2011.
- [16] B. Shrader and T. Royster, "Cooperative multicast strategies under heterogeneous link loss rates," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '11)*, pp. 1–5, December 2011.
- [17] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.
- [18] J. Wang, J. Wang, K. Lu, B. Xiao, and N. Gu, "Optimal linear network coding design for secure unicast with multiple

- streams,” in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '10)*, pp. 1–8, March 2010.
- [19] M. Ghaderi, D. Towsley, and J. Kurose, “Network coding performance for reliable multicast,” in *Proceedings of the Military Communications Conference (MILCOM '07)*, October 2007.
- [20] X. Wang, K. Wu, J. Wang, and Y. Xu, “CAPF: coded anycast packet forwarding for wireless mesh networks,” *Wireless Networks*, vol. 17, no. 5, pp. 1273–1285, 2011.
- [21] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, “XORs in the air: practical wireless network coding,” *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 497–510, 2008.
- [22] C. Fragouli, J. Y. Le Boudec, and J. Widmer, “Network coding: an instant primer,” *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 1, pp. 63–68, 2006.
- [23] M. Ghaderi, D. Towsley, and J. Kurose, “Reliability gain of network coding in lossy wireless networks,” in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 2171–2179, April 2008.
- [24] N. Cai and R. W. Yeung, “Secure network coding on a wiretap network,” *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 424–435, 2011.
- [25] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, “On the construction of energy-efficient broadcast and multicast trees in wireless networks,” in *Proceedings of the IEEE 19th Annual Joint Conference of Computer and Communications Societies (INFOCOM '00)*, pp. 585–594, March 2000.
- [26] C. Chan, *Generating secret in a network [Ph.D. dissertation]*, Massachusetts Institute of Technology, 2010.
- [27] Y. Kim and G. De Veciana, “Is rate adaptation beneficial for inter-session network coding?” in *Proceedings of the IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 635–646, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

