Efficient Network Coding-Based Loss Recovery for Reliable Multicast in Wireless Networks

Kaikai CHI[†,††a)], *Nonmember*, Xiaohong JIANG[††b)], *Member*, Baoliu YE[†††c)], *Nonmember*, *and* Susumu HORIGUCHI[††d)], *Member*

SUMMARY Recently, network coding has been applied to the loss recovery of reliable multicast in wireless networks [19], where multiple lost packets are XOR-ed together as one packet and forwarded via single retransmission, resulting in a significant reduction of bandwidth consumption. In this paper, we first prove that maximizing the number of lost packets for XOR-ing, which is the key part of the available network coding-based reliable multicast schemes, is actually a complex NP-complete problem. To address this limitation, we then propose an efficient heuristic algorithm for finding an approximately optimal solution of this optimization problem. Furthermore, we show that the packet coding principle of maximizing the number of lost packets for XOR-ing sometimes cannot fully exploit the potential coding opportunities, and we then further propose new heuristic-based schemes with a new coding principle. Simulation results demonstrate that the heuristic-based schemes have very low computational complexity and can achieve almost the same transmission efficiency as the current coding-based high-complexity schemes. Furthermore, the heuristic-based schemes with the new coding principle not only have very low complexity, but also slightly outperform the current high-complexity ones.
key words: wireless networks, network coding, reliable multicast, physical-layer broadcast, bandwidth saving

1. Introduction

Bandwidth continues to be one of the most precious resources in wireless networks. The *network coding* technique [1], which allows network nodes to perform coding operation in addition to the traditional routing function, has been proved promising for significantly reducing the bandwidth and energy consumptions in wireless networks. Formally, the network coding operation can be defined as follows: the network node encodes (mixes) several incoming packets of the same or different flows together and forwards the resulting encoded packet, rather than just relaying these incoming packets to output links one by one.

In wireless networks, the feasibility of using network coding to improve node transmission efficiency is basi-
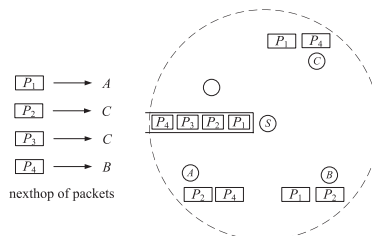

Fig. 1 A simple scenario of wirelesses network coding.

cally due to the existence of following two factors: 1) the physical-layer broadcast; 2) some nexthops of a node may possess some packets via overhearing, routing or other ways. Let us take the example in Fig. 1 to illustrate this basic idea. In this example, node S has packets P_1, P_2, P_3 and P_4 to forward. The packet P_1 has been forwarded from node C to node S and node B successfully overheard it. Thus nodes C and B already possess P_1. Similarly, via routing or overhearing, node A, B and C already possess some other packets, as shown in the figure. Due to the possession of some packets at nexthops A, B and C, if node S encodes P_1, P_2 and P_4 together to $P_1 \oplus P_2 \oplus P_4$ and transmits (locally broadcasts) this coding packet, nodes A, B and C can get the coding packet and retrieve P_1, P_2 and P_4, respectively. Therefore, by taking use of both physical-layer broadcast and packet possession at nexthops, network nodes can encodes multiple packets together and delivery them via a single transmission.

By now, considerable efforts have been devoted to demonstrate the benefits of using network coding in different specific communication paradigms, such as the unicast [2]–[11], multicast [12]–[14] and broadcast [15], [16]. For the unicast scenario, Wu et al. [2] showed that the exchange of independent information between two nodes in a wireless network can be efficiently performed by exploiting both the network coding and physical-layer broadcast. Recently, Katti et al. [4] proposed a practical network coding-based packet forwarding architecture (called COPE) to essentially improve the network throughput of multihop wireless networks. Based on the COPE-type coding scheme, the coding-aware routing was proposed in [6], [10]. Some efforts (e.g. [6]–[9]) have also been made to theoretically evaluate the throughput of COPE-type wireless networks. More recently, the physical-layer network coding was proposed to utilize wireless interference for network coding [5], [11]. As

for multicast case, Wu et al. [12] showed that in a mobile ad hoc network, adopting network coding for minimum-cost multicast can be formulated as a linear optimization problem and solved in polynomial time. The corresponding decentralized algorithms were further proposed in [13] to establish the minimum-cost multicast tree. Concerning the application of network coding for broadcast in wireless ad hoc networks, distributed probabilistic broadcast algorithms and deterministic broadcast algorithms have been proposed by Fragouli et al. [15] and Li et al. [16], respectively.

Reliable multicast [17], the lossless delivery of bulk data from one sender to a group of receivers, is widely used in many important applications such as the file distribution to a number of receivers and the dissemination of market data from a financial institution to its subscribers. Multicasting protocols can be classified as tree-based multicast which has only a single path between any source and receiver pair, or mesh-based multicast which may has multiple paths between any source and receiver pair. In this work, we focus on the tree-based multicasting.

For reliable multicast, packet error recovery can be divided into two major classes: (a) source-based recovery (i.e., only the source node of one multicast tree is responsible for the retransmission of lost packets) and (b) link-by-link recovery (i.e., each intermediate node of one multicast tree retransmits the packets which is not successfully received by its child nodes). In this paper, we focus on the latter class, i.e., aim to achieve the reliable one-hop multicast (or the reliable link-layer multicast). The source node of our multicast model is an arbitrary intermediate node of one multicast tree, which reliably multicasts the packets it received from its parent node to its child nodes.

Traditionally, the source adopts the Auto Repeat reQuest (ARQ) scheme to achieve the reliable link-layer multicast, i.e., simply retransmits one by one the *lost packets* (i.e. the packets that are not received yet by one or more receivers). Later, hybrid ARQ combining the FEC and ARQ has been proposed to achieve better performance [18]. The ARQ-based scheme and hybrid-ARQ-based scheme only transmit/retransmit the native packets (rather than the encoded packets including multiple native packets). During the last several years, erasure codes like Raptor codes, which combine a set of randomly selected packets with randomly generated coefficients for each transmission, have received significant attention and have been applied to reliable multicast [21], [22]. Recently, Nguyen et al. [19] applied the novel network coding technique to the reliable link-layer multicast in wireless networks and proposed two network coding-based schemes (a static one and a dynamic one) for it. Different from the erasure-code-based scheme with linear computational complexity, the network-coding-based schemes intelligently select a set of lost packets and determine the appropriate combination coefficients to achieve better performance, with a slightly higher computational complexity. The main idea of these coding-based reliable multicast schemes is to first buffer the lost packets for some time, then, instead of transmit these lost packets one by one,

the source XORs an optimal set of lost packets with distinct intended receivers together into one packet and transmits this XOR-ed packet in one retransmission[†]. The main difference between the static and dynamic schemes in [19] is that the static one will repeatedly retransmit the same XOR-ed packet until all its intended receivers receive it, while the dynamic one can dynamically update the XOR-ed packet in each retransmission for a further improvement in transmission efficiency at the cost of increasing the running time due to the dynamic update of XOR-ed packet.

By intelligently XOR-ing multiple lost packets together, the current coding-based multicast schemes can result in a significant improvement on the transmission efficiency of reliable link-layer multicast. We prove that, however, finding the optimal set of lost packets for XOR-ing, which is the key part of the current two coding-based schemes, is a complex NP-complete problem. Therefore, these two schemes are not scalable to large number of multicast receivers. Then we propose an efficient heuristic algorithm for this optimization problem, obtaining heuristic-based multicast schemes. Up to this step, the work is relatively complete. However, we find that the packet coding principle of maximizing the number of lost packets for XOR-ing sometimes cannot fully exploit the potential coding opportunities, and we then further propose new heuristic-based schemes (also a static one and a dynamic one) by applying a new coding objective, maximizing the total number of intended receivers of the coding packet.

In summary, the main contributions of this work are as follows:

1. We first prove that in the current two coding-based reliable multicast schemes, the problem of finding the maximum set of lost packets for XOR-ing is NP-complete. We then present an efficient heuristic algorithm for finding an approximately optimal solution of this optimization problem.
2. We further propose two heuristic-based schemes by applying a new coding objective which maximizes the total number of intended receivers of the coding packet.
3. We demonstrate through extensive simulation that the new heuristic-based schemes not only have very low computational complexity, but also can provide slightly higher transmission efficiency than the current complex coding-based schemes.

The rest of this paper is organized as follows. Section 2 briefly reviews two available coding-based multicast schemes. Section 3 presents two new heuristic-based multicast schemes. Section 4 further presents two heuristic-based multicast schemes with the new objective. Simulation results are presented in Sect. 5. Section 6 presents some issues that should be addressed in the future. Finally, Sect. 7 concludes this paper.

[†]The intended receivers of a packet are the receivers which have not received this packet.

Fig. 2   Packet-loss table inside the source node.

Fig. 3   An example XOR-ing lost packets together for retransmission.

2.   Available Network Coding-Based Multicast Schemes

In this section, we briefly review two available coding-based schemes proposed in [19] for the reliable link-layer multicast and also their main limitation.

2.1   Available Static Scheme and Dynamic Scheme

To achieve the reliable link-layer multicast, traditionally the source simply retransmits the lost packets one by one. Rather than one by one retransmission of lost packets, the basic idea of the coding-based schemes is to first buffer the lost packets for some time and then encode multiple lost packets together into one new packet for retransmission, such that multiple lost packets can be delivered via one retransmission. In detail, two available network coding-based schemes are as follows.

Static scheme: This coding-based scheme consists of a *transmission phase* and a *retransmission phase*. In the transmission phase, the source R_0 transmits (physical-layer broadcasts) a fixed number of N packets one by one to M receivers, and stores the lost packets to a buffer of size N (called *lost-packet buffer* in this paper). The R_0 also maintains a table whose entry $e_{i,j}$ is used to indicate whether the receiver R_i has correctly received P_j or not, as shown in Fig. 2. Here, $e_{i,j} = 0$ if R_i correctly received P_j and $e_{i,j} = 1$ otherwise. In the retransmission phase, the R_0 first finds the optimal set of lost packets (in terms of the number of lost packets) without common intended receivers to XOR and then repeatedly transmits this XOR-ed packet until all its intended receivers receive it. After finishing the transmission of the current set of lost packets, the R_0 continues to find a new optimal set of lost packets and repeats the above process. In this way, the source keeps sending out the encoded packets until no lost packet is left in the list, and then starts the transmission of next N packets.

Dynamic scheme: Different from the static scheme, in this scheme the source R_0 will update (i.e. to find) the optimal set of lost packets for XOR-ing once the last XOR-ed packet is received by one or more intended receivers (i.e. once the packet-loss table is changed), such that lost packets may be delivered to their intended receivers in a more efficient way.

Let us take the example in Fig. 3 to illustrate how these two schemes work. In this example, both lost packets P_1 and P_4 have one intended receiver R_1 and both lost packets P_3 and P_5 have one intended receiver R_2. Traditionally,
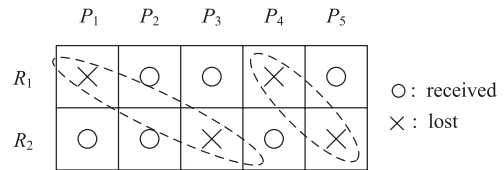
each one of P_1, P_3, P_4 and P_5 is retransmitted alone and each one of them has only one intended receiver. When using the above static scheme, the source can XOR P_1 and P_3 together to $P_C = P_1 \oplus P_3$, which totally has two intended receivers R_1 and R_2 (i.e. $P_1 \oplus P_3$ is useful for both R_1 and R_2). Once R_1 receives P_C, it can recover P_1 by $P_1 = P_C \oplus P_3$. Similarly, R_2 can recover P_3 by $P_3 = P_C \oplus P_1$. The source repeatedly transmits P_C until both R_1 and R_2 receive it, and then starts the transmission of next group of lost packets $\{P_4, P_5\}$. When using the dynamic scheme, however, the source dynamically changes the XOR-ed packet. Suppose $P_1 \oplus P_3$ is transmitted and only received by R_1, then the source will XOR P_3 and P_4 together for next transmission. From this example we can know that, by XOR-ing lost packets together to increase the average number of intended receivers per packet, the number of retransmissions can be effectively reduced.

2.2   NP-Completeness of the Optimization Problem

Despite the lower bandwidth requirements than the traditional non-coding scheme, both two current coding-based schemes actually suffer from the high-complexity problem as shown below.

In both of the current schemes, there is an optimization problem of maximizing the number of lost packets with distinct intended receivers for XOR-ing. Let L be the number of lost packets and assume without loss of generality that $P_1, P_2, \ldots, P_L$ are lost packets. Then, this optimization problem can be mathematically formulated as follows.

Given: values of $e_{i,j}$'s: $i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, L\}$.
Encoded packet: $P = a_1 P_1 \oplus \cdots \oplus a_L P_L$
Maximize: $\sum_{i=1}^{L} a_i$
Over variables: $a_i \in \{0, 1\} : 1 \leq i \leq L$
Subject to: $\sum_{i=1}^{L} a_i e_{1,i} \leq 1,$
$\qquad\qquad \sum_{i=1}^{L} a_i e_{2,i} \leq 1,$
$\qquad\qquad \cdots$
$\qquad\qquad \sum_{i=1}^{L} a_i e_{M,i} \leq 1.$

Now we show that this maximum-lost-packet coding (MLPC) problem is NP-complete based on the reduction from the NP-complete maximum independent set (MIS) problem [20].

Theorem 1: *The MLPC problem is NP-complete.*

Proof: It is easy to know that the MLPC problem belongs to NP. Therefore, it is enough to show a polynomial-time reduction from the MIS problem described below to the MLPC problem.
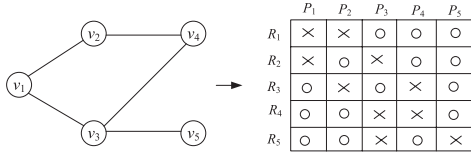
	P_1	P_2	P_3	P_4	P_5
R_1	×	×	○	○	○
R_2	×	○	×	○	○
R_3	○	×	○	×	○
R_4	○	○	×	×	○
R_5	○	○	×	○	×

Fig. 4 An example used to illustrate the reduction from a graph to a packet-loss table.

Maximum Independent Set Problem:
Instance: A graph $G(V, E)$ and a positive integer $K \leq |V|$.
Question: Does G contain a subset of vertices with cardinality K such that no two vertices in this subset are adjacent in G?

Here is the reduction. Given an instance $G = (V, E)$ of the MIS problem, construct an instance of the MLPC problem as follows. Label the nodes in G by $v_1, v_2, \cdots, v_{|V|}$. Then the lost packet set is defined as $\{P_1, P_2, \cdots, P_{|V|}\}$, where P_i corresponds to the vertex v_i in the MIS problem. At the beginning, set each $e_{i,j}$ to zero and set parameter k to zero. Now, in the order from $i = 1$ to $i = |V|$, we define the packet-loss table of Fig. 2 in the following way: corresponding to each v_i's neighbor v_j with $j > i$, let $k = k + 1$, $e_{k,i} = 1$ and $e_{k,j} = 1$. Figure 4 shows an example of the reduction from a graph to a packet-loss table. From this figure we can see that, for each neighbor pair (v_i, v_j) in G, there is a corresponding packet pair (P_i, P_j) which have exactly one common intended receiver.

The above transformation is clearly polynomial. Based on this transformation, we can know that the answer to the instance of the MIS problem is "YES" if and only if there is a set of K lost packets with distinct intended receivers in the MLPC problem. ∎

No polynomial-time algorithms are available to obtain the optimal solution of an NP-hard optimization problem. Thus, for the available coding-based schemes, their computational complexity is a serious problem in multicast with a number of receivers. In many real wireless networks, the number of multicast receivers might be quite large, such as the cellular networks and high-density sensor networks. In these networks, the application of the available coding-based reliable multicast schemes is impractical. In addition, the computational complexity of available coding-based schemes might become unacceptable in some networks whose node computation capacity is quite limited, such as sensor networks, even though the number of multicast receivers is not large. Therefore, it is necessary to develop an efficient heuristic algorithm to deal with it, such that the current two coding-based reliable multicast schemes can be scalable to large number of receivers.

3. Heuristic-Based Static Scheme and Dynamic Scheme

In this section, we propose an efficient heuristic algorithm for the MLPC problem and then get the corresponding heuristic-based static scheme and dynamic scheme.

3.1 Heuristic Algorithm for the MPLC Problem

For a lost packet P_i, call the lost packets having one or more common intended receivers with it as its *neighbor packets*. Let d_i be the number of P_i's neighbor packets.

Our heuristic algorithm is to find lost packets with distinct intended receivers as many as possible. Its main idea is to select coding packets from the set of candidate packets one by one so as to achieve the low complexity. Notice that once a packet P_i is selected for coding (i.e. XOR-ing), any one of its d_i neighbor packets cannot be selected for coding. Thus, at each selection step, our algorithm selects out the lost packet P_i which has the smallest value of d_i and removes P_i's d_i neighbor packets from the set of candidate packets. Such a selection rule can guarantee that after each packet selection, maximal candidate packets are left for further selection.

Specifically, the heuristic algorithm is as follows. Let $S_t = \{P_1, \ldots, P_L\}$ be the set of lost packets in the MLPC problem. The heuristic algorithm selects a packet P_i with the smallest value of d_i, then removes from S_t both the P_i and its d_i neighbor packets, and iterates this process on the remaining S_t until S_t is empty. The set of selected packets is the output of this algorithm. Formally, it is illustrated as follows.

Heuristic algorithm for the MLPC problem
Output: set C of lost packets.
Steps:
1     $S_t := S$, $C := \phi$ and $c_{i,j} = 0$ for $1 \leq i, j \leq N$.
2   For each $P_i \in S_t$, obtain d_i, and let $c_{i,j} = c_{j,i} = 1$ if P_i and P_j have common intended receivers.
3   **while** $S_t \neq \phi$ **do**
4        $A := \{P_i\} \bigcup \{P_j : c_{i,j} = 1\}$, where P_i is the packet in S_t having the minimin value of d_i.
5        $S_t := S_t \backslash A$
6        $C := C \bigcup \{P_i\}$.
7   Update d_i according to A and $c_{i,j}$.
8   **end while**

Now, we analyze the computational complexity of the above heuristic algorithm. To obtain the set C of lost packets for XOR-ing, the source first takes time $O(N^2)$ to initialize S_t, C and $c_{i,j}$. Step 2 takes time $O(MN^2)$ to calculate d_i and $c_{i,j}$. Steps 4, 5 and 6 take time $O(N)$, and Step 7 takes time $O(MN)$. The iteration of Steps 4, 5, 6 and 7 will be conducted $O(N)$ times. Thus, the overall computational complexity is $O(MN^2)$.

3.2 Heuristic-Based Static Scheme and Dynamic Scheme

By adopting the above heuristic algorithm for finding an approximately optimal solution (i.e. a set of lost packets) of the MLPC problem, we can get the corresponding heuristic-based static scheme and dynamic scheme, which are illustrated in Figs. 5 and 6, respectively. The heuristic-based dynamic scheme has higher transmission efficiency than the

Heuristic-based static scheme
Steps:
1   Transmit N native packets one by one and build the packet-loss table.
2   Let S be the set of lost packets.
3   For each $P_i \in S$, obtain its weight $w_i = \sum_{j=1}^{M} e_{j,i}$.
4   **while** $S \neq \phi$ **do**
5   Obtain a set C of lost packets by using the proposed heuristic algorithm.
6   XOR the packets in C together and get the resulting encoded packet P_C.
7   Repeatedly transmit packet P_C until all its intended receivers receive it.
8     $S := S \backslash C$.
9   **end while**

Fig. 5   Heuristic-based static scheme.

Heuristic-based dynamic scheme
Steps:
1   Transmit N native packets one by one and build the packet-loss table.
2   Let S be the set of lost packets.
3   **while** $S \neq \phi$ **do**
4   For each $P_i \in S$, obtain its weight $w_i = \sum_{j=1}^{M} e_{j,i}$.
5   Obtain a set C of lost packets by using the proposed heuristic algorithm.
6   XOR the packets in C together and get the resulting encoded packet P_C.
7   Repeatedly transmit packet P_C until one or more intended receivers receive it.
8   For each packet $P_i \in C$ which has been recoved by all its intended receivers, $S := S \backslash \{P_i\}$.
9   **end while**

Fig. 6   Heuristic-based dynamic scheme.

heuristic-based static scheme due to its dynamic update of the set of lost packets for XOR-ing.

It is easy to know that the heuristic-based dynamic scheme and the heuristic-based static scheme have the same order of computational complexity $O(MN^2)$. The difference between them is that in the static scheme, the same encoded packet will be repeatedly transmitted until all intended receivers obtain it, whereas in the dynamic scheme, the encoded packet needs to be updated once it is received by one or more intended receivers. Thus, the dynamic scheme takes more time than the static one, although they have the same order of complexity.

4. Heuristic-Based Static and Dynamic Schemes with a New Coding Objective

The current coding-based schemes aim at maximizing the number of lost packets with distinct intended receivers for XOR-ing. However, this optimization objective sometimes cannot fully exploit the potential coding opportunities, since

	P_1	P_2	P_3	P_4	P_5	P_6
R_1	×	×	×	○	○	○
R_2	○	○	×	×	×	○
R_3	×	○	○	○	×	×
R_4	×	○	○	○	○	○

Fig. 7   An example used to illustrate the limitation of current coding-based schemes.

it is the average number of intended receivers per XOR-ed packet that directly determines how much the transmission efficiency can be improved.

Consider the example in Fig. 7. In this example, if we aim at maximizing the number of lost packets with distinct intended receivers for XOR-ing, then P_2, P_4 and P_6 (three lost packets) will be XOR-ed together. The resulting $P_2 \oplus P_4 \oplus P_6$ has only three intended receivers: R_1, R_2 and R_3. However, if we XOR P_1 and P_4 (only two lost packets) together, the resulting XOR-ed packet is useful for all four receivers.

4.1   Maximizing the Total Number of Intended Receivers

In the following we study the hardness of maximizing the total number of intended receivers. Formally, this optimization problem can be formulated as follows.

Given: values of $e_{i,j}$'s: $i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, L\}$.
Encoded packet: $P = a_1 P_1 \oplus \cdots \oplus a_L P_L$
Maximize: $\sum_{j=1}^{M} \sum_{i=1}^{L} a_i e_{i,j}$
Over variables: $a_i \in \{0, 1\} : 1 \leq i \leq L$
Subject to: $\sum_{i=1}^{N} a_i e_{1,i} \leq 1$,
              $\sum_{i=1}^{L} a_i e_{2,i} \leq 1$,
              $\cdots$
              $\sum_{i=1}^{L} a_i e_{M,i} \leq 1$.

The following theorem shows that this maximum-intended-receiver coding (MIRC) problem is also NP-complete based on the reduction from the NP-complete MIS problem in 3-regular graphs (where each node has a degree of three) [20].

Theorem 2: *The MIRC problem is NP-complete.*
Proof: It is easy to know that the MIRC problem belongs to NP. Therefore, it is enough to show a polynomial-time reduction from the MIS problem in 3-regular graphs described below to the MIRC problem.
MIS Problem in 3-Regular Graphs:
Instance: A 3-regular graph $G(V, E)$ and a positive integer $K \leq |V|$.
Question: Does G contain a subset of vertices with cardinality K such that no two vertices in this subset are adjacent in G?

Here is the reduction. Given an instance $G = (V, E)$ (a 3-regular graph) of the MIS problem, construct an instance of the MIRC problem as follows. Label the nodes
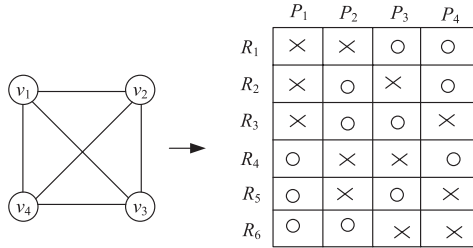
Fig. 8 An example used to illustrate the reduction from a graph to a packet-loss table.

in G by $v_1, v_2, \cdots, v_{|V|}$. Then the lost packet set is defined as $\{P_1, P_2, \cdots, P_{|V|}\}$, where P_i corresponds to the vertex v_i in the MIS problem. At the beginning, set each $e_{i,j}$ to zero and set parameter k to zero. Now, in the order from $i = 1$ to $i = |V|$, we define the packet-loss table of Fig. 2 in the following way: corresponding to each v_i's neighbor v_j with $j > i$, let $k = k + 1$, $e_{k,i} = 1$ and $e_{k,j} = 1$. Figure 8 shows an example of the reduction from a 3-regular graph to a packet-loss table. From this figure we can see that, for each neighbor pair (v_i, v_j) in G, there is a corresponding packet pair (P_i, P_j) which have exactly one common intended receiver.

The above transformation is clearly polynomial. Based on this transformation, we can know that the answer to the instance of the MIS problem is "YES" if and only if there is a set of lost packets totally with $3K$ intended receivers in the MLPC problem. ∎

4.2 Heuristic-Based Schemes with the New Coding Objective

Since the MIRC problem is NP-complete, there is no polynomial-time algorithm for finding the optimal solution. However, by slightly revising the heuristic algorithm proposed for the MLPC problem, we can get an efficient heuristic algorithm for finding an approximately optimal solution of this problem.

The difference between the MLPC problem and the MIRC problem is the optimization objective function. Thus, we can get the heuristic for the MIRC problem through the following slight change of the current heuristic: in Step 4, we select the packet P_i with the maximum value of w_i/d_i in the current S for XOR-ing. The reason for this is that selecting the packet with the maximum value of w_i/d_i can effectively increase the total number of intended receivers (since w_i is the increment of the number of intend receivers when including P_i for XOR-ing) and also leave as many lost packets as possible in S_t for further selection.

Formally, this heuristic algorithm is illustrated as follows.

Heuristic algorithm for the MIRC problem
Output: set C of lost packets.
Steps:
1     $S_t := S$, $C := \phi$ and $c_{i,j} = 0$ for $1 \le i, j \le N$.
2   For each $P_i \in S_t$, obtain d_i, and let $c_{i,j} = c_{j,i} = 1$ if P_i

and P_j have common intended receivers.
3   **while** $S_t \ne \phi$ **do**
4        $A := \{P_i\} \bigcup \{P_j : c_{i,j} = 1\}$, where P_i is the packet in S_t having the maximum value of w_i/d_i.
5        $S_t := S_t \backslash A$
6        $C := C \bigcup \{P_i\}$.
7   Update d_i according to A and $c_{i,j}$.
8   **end while**

With these heuristic-based schemes, we can not only achieve the low computational complexity, but also take full advantage of network coding to improve the transmission efficiency, as shown in the next section.

5. Simulation Results

In this section, through extensive simulation we investigate the average search time for finding a set of lost packets for XOR-ing and also the average number of transmissions per packet (henceforth called *transmission bandwidth*) for different reliable multicast schemes. For each scenario of parameter setting (number M of receivers, size N of the lost-packet buffer and link packet loss probabilities), our simulation conducts the multicast transmission of $N \times 10^5$ packets.

In our simulation, the packets loss at each receiver follows the Bernoulli distribution. The packet losses at different receivers are independent. At the MAC layer, all nodes have a fair access to the shared wireless medium, i.e., after the source node transmits one data packet, each receiver has the chance to return ACK/NACK to the source node when necessary. It is assumed that all ACK/NACKs are instantaneous and lossless, which is reasonable because ACK/NACKs have very small size (say 32 bytes) and can be transmitted in short time with a very small loss probability.

5.1 Coding Solution Search Time

Here, the average search time of a scheme is the average elapsed CPU time for obtaining the coding solution when this scheme is adopted. The simulations are carried out on a computer running Windows XP with Intel 3.0 GHz CPU and 2 GB memory. The average elapsed CPU time is obtained by averaging all coding solution searches during the multicast transmission of $N \times 10^5$ packets.

In [19], for the optimization problem in the available two coding-based schemes (a static one and a dynamic one), no explicitly algorithm is given for finding the optimal solution. In our simulation, the following two algorithms are adopted for finding the optimal solution such that we can evaluate how much the optimal-solution search time is: (1) the exhaustive search method which is simple to implement but time consuming and (2) a well-design algorithm which first finds all pairs of lost packets which definitely cannot be encoded together and then skips checking all those coding solutions which involves such a pair of lost packets.
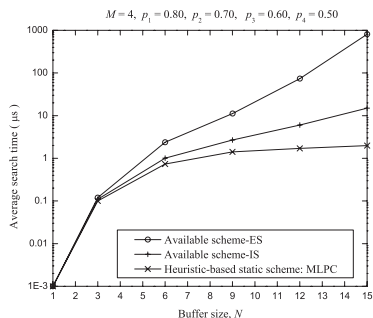
Fig. 9 Average coding solution search time versus the size of lost-packet buffer, in the cases of four receivers. p_i is the packet deliver ratio of the link from the source to the i'th receiver.
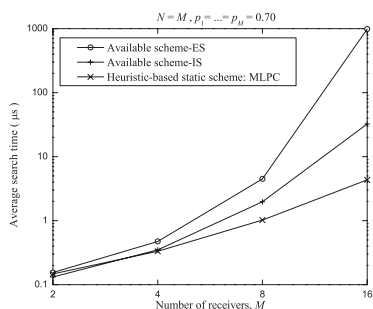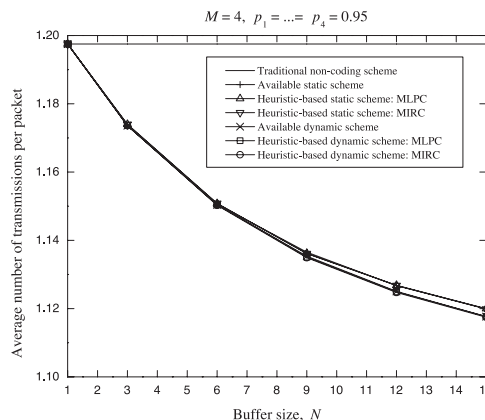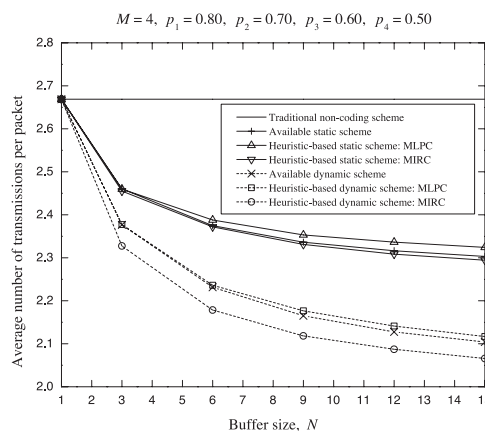

Fig. 10 Average coding solution search time versus the number of receivers.

(a) The low-packet-loss environment.

(b) The high-packet-loss environment.

Fig. 11 Transmission bandwidth versus lost-packet buffer size.

The proposed heuristic algorithms for the MLPC problem and MIRC problem take almost the same time to find a coding solution. Thus, here we only show the average solution search time of the available coding-based scheme and the heuristic algorithm for the MLPC problem, as summarized in Fig. 9. The curve labeled "Available scheme-ES" corresponds to the available scheme adopting the above algorithm 1 and the curve labeled "Available scheme-IS" corresponds to the available scheme adopting the above algorithm 2. From this figure, it can be seen that as the buffer size increases, the search time of the available coding-based schemes exponentially increases, whereas the search time of the proposed heuristic-based schemes approximately linearly increases. The reason for this behavior is explicit. As discussed previously, in the available two coding-based schemes, the search for the exactly maximum set of lost packets with distinct intended receivers is an NP-complete problem, which does not have polynomial-time algorithms. In two improved schemes, however, the low-complexity heuristic algorithm is adopted to find an approximately optimal set of lost packets. Thus they take far less time to obtain the coding solution than the available ones.

Figure 10 shows the average search time under different numbers of receivers. From this figure, we can arrive at the same conclusion that the heuristic-based schemes are much fastest than the available schemes to find a coding solution. It can be predicted that, when the number of receivers is large (say over 32), the previous coding-based schemes have extremely high computational complexity and thus im-

practical for real applications. However, the computational complexity of our heuristic-based schemes increases slowly (actually polynomially) as the number of receivers increases and thus are scalable to large number of multicast receivers.

5.2 Transmission Bandwidth

For all network coding-based schemes, their transmission bandwidths greatly depend on the lost-packet buffer size, so we first investigate the transmission bandwidth of different schemes under different sizes of the lost-packet buffer. Figure 11 shows the transmission bandwidths of different schemes.

From this figure, first we can observe that in general the transmission bandwidth of each network coding-based scheme decreases as the lost-packet buffer size increases, and when the lost-packet buffer size is not very small, the coding-based multicast schemes can substantially outperform the non-coding multicast scheme. For example, in the environment of Fig. 11(b), using buffer size $N = 9$, compared to the traditional reliable multicast scheme the average number of transmissions per packet can be reduced by over 20.6% when using the improved heuristic-based dy-
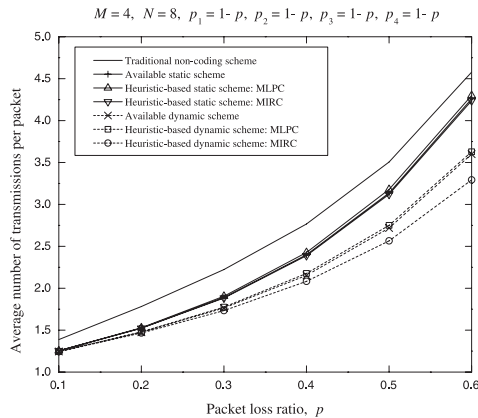
$M = 4, \ N = 8, \ p_1 = 1\text{-} p, \ p_2 = 1\text{-} p, \ p_3 = 1\text{-} p, \ p_4 = 1\text{-} p$

Fig. 12 Transmission bandwidth versus packet loss probability.

namic scheme. Additionally, results in Fig. 11 show that the dynamic schemes greatly outperform the static schemes, at the cost of increasing the running time due to the dynamic update of XOR-ed packet.

In Fig. 11(a) it can be observed that, when the packet loss rates are very small, the heuristic-based schemes (labeled with MLPC) achieve the same performance as the previous coding-based schemes. That is to say, in the low-packet-loss scenarios, our proposed heuristic almost always finds the coding solution with the maximum number of lost packets. Also, we can see that in the low-packet-loss scenarios the heuristic-based schemes with the new coding objective (labeled with MIRC) do not provide further improvement, as compared to the heuristic-based schemes with the old coding objective. This is because when the packet loss rates are very small (say 0.05), for each packet the probability that two or more receivers did not receive this packet is very small. We can know that, for such a case, maximizing the number of intended receivers is usually same as maximizing the number of lost packets with distinct intended receivers. Thus, for the low-packet-loss link-level multicast, our proposed heuristic-based schemes only exhibit the advantage of low computational complexity. When the packet loss rates are not small (refer to Fig. 11(b)), the proposed heuristic-based static scheme and dynamic scheme almost have the same performance as the available static scheme and dynamic scheme, respectively. However, when using the heuristic-based schemes with the new coding objective, the transmission bandwidth can be further reduced and even smaller than that of the available corresponding coding-based schemes. For example, in the environment of Fig. 11(b), compared to the available dynamic scheme the heuristic-based dynamic scheme with the new coding objective can further reduce the transmission bandwidth by 2.4% when $N = 6$.

We further investigate the transmission bandwidth under different link-packet-loss probabilities and different numbers of receivers, as summarized in Fig. 12 and Fig. 13, respectively. From these two figures, we can draw the same conclusion that the heuristic-based static scheme per-

forms almost as well as the available static scheme, and the heuristic-based dynamic scheme performs almost as well as the available dynamic scheme. When adopting the heuristic algorithm with the new coding objective, we can even achieve lower transmission bandwidth than the available coding-based complex schemes.

From Fig. 13(a), the same conclusions as that from Fig. 11(a) can be obtained. That is to say, for the low-packet-loss link-level multicast, our proposed heuristic-based schemes exhibit the advantage of low computational complexity and achieve the same transmission efficiency as the previous coding-based schemes. However, when the packet loss rates are medium or large (refer to Fig. 13(b)), the transmission efficiency improvement achieved by our schemes with the new coding objective is clear. We can see that this improvement increases as the number of receivers increases. Thus, it can be expected that when the number of receivers is large the improvement will be particularly remarkable.

5.3 Comparison between Network Coding and Erasure Code

In some literature, erasure codes have been applied to wireless multicast to achieve high transmission efficiency. It is well known that the network coding-based scheme can also be considered to be an erasure code. Here we investigate the *bandwidth efficiency* (the ratio of the number of successfully transmitted data bits to that of actually transmitted bits) of conventional erasure codes and network coding in wireless reliable multicast. Among erasure codes, Raptor code having good performance is selected to compare with network coding [21], [22]. For Raptor code, we simulate its performance under different settings of k, which is the number of input symbols of Raptor code. The pre-code of the Raptor code is with rate 0.95. The inner LT code uses the degree distribution $\Omega(x) = 0.007969x + 0.493570x^2 + 0.166622x^3 + 0.072646x^4 + 0.082558x^5 + 0.056058x^8 + 0.037229x^9 + 0.055590x^{19} + 0.025023x^{65} + 0.003135x^{66}$. As k increases, the performance of Raptor code increases, but the required buffer size also increases. Among network coding-based schemes, our MIRC-based dynamic scheme with the NACK-based feedback mechanism in [23] is selected and the buffer size N is set to 200 such that this scheme can achieve good performance. According to the slightly increasing trend of computation complexity shown in Fig. 9 (theoretically, following $O(MN^2)$), we can know that this value setting of N is reasonable (i.e., not too large). Additionally, data packets have the size of 1000 bytes and ACK (NACK) packets have the size of 32 bytes.

The simulation results are summarized in Fig. 14. First, it can be observed from both two subfigures that for Raptor code-based multicast, the performance increases as k increases. Second, from Fig. 14(a), we can clearly see that in the low-packet-loss environment, network coding-based multicast outperforms Raptor code-based multicast even when $k = 1000$. Third, in the high-packet-loss environ-
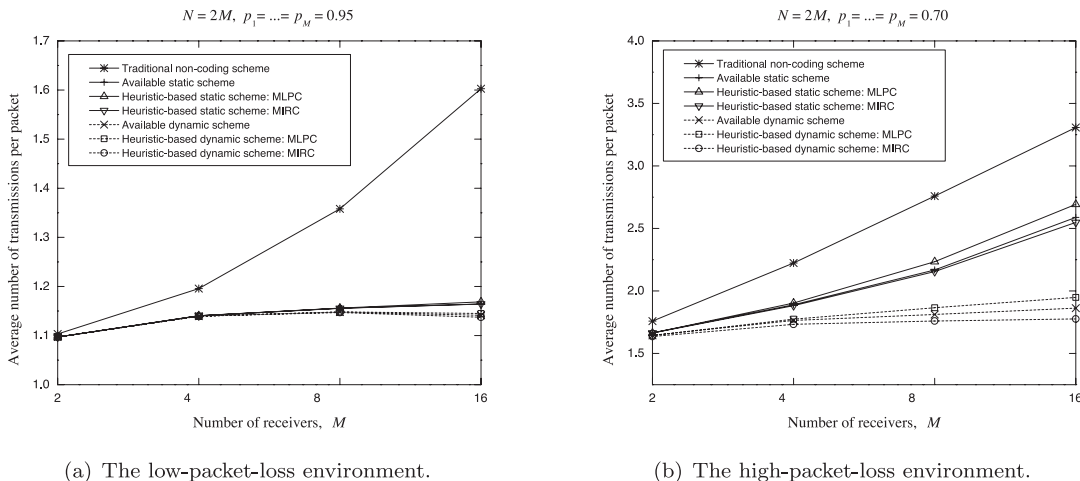
(a) The low-packet-loss environment.

(b) The high-packet-loss environment.

Fig. 13   Transmission bandwidth versus number of receivers.

(a) The low-packet-loss environment.
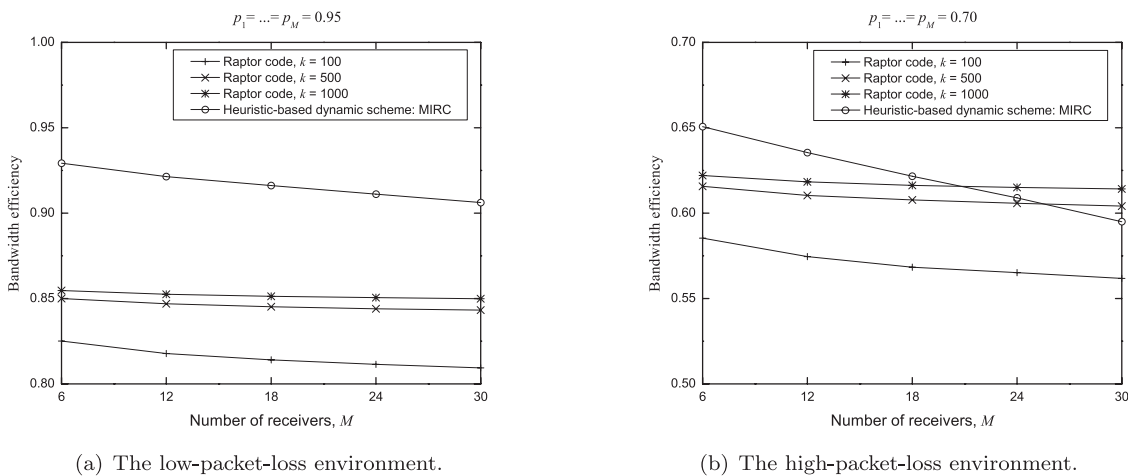
(b) The high-packet-loss environment.

Fig. 14   Comparison between network coding and erasure code in wireless reliable multicast.

ment (refer to Fig. 14(b)), when the number of receivers is not large, network coding-based multicast has higher transmission efficiency than Raptor code-based multicast. When the number of receivers is large, say 32, the network coding-based scheme has approximately the same performance as Raptor code-based scheme with a large k. This is because, for network coding, when the packet loss rates at receivers are high, the probability that any two lost packets have common intended receivers is large and thus the coding opportunities reduce as compared to the low-packet-loss case. However, the coding opportunities can be increased by adopting a larger buffer for lost packets.

On the whole, network coding-based multicast scheme achieves higher transmission efficiency than Raptor code-based multicast scheme. It should be noticed that Raptor code-based scheme has linear time complexity, lower than that of network coding-based one. Thus, both two kinds of schemes have their own advantage and potential applications.

6.   Discussion

In this preliminary paper, we present our efficient network coding-based multicast schemes. However, there are still some important issues to be further considered.

It is necessary to design efficient algorithms to select intermediate nodes responsible for coding. From the network coding point of view, there are two types of networks: 1) all nodes are implemented with functions for buffing and coding packets; 2) only a few nodes are implemented with functions for buffing and coding packets. In the first type of networks, for multicast communication, the selection of intermediate coding nodes is relatively simple. The available algorithms (like Steiner-tree algorithms) can be used to build a multicast tree and the intermediate nodes having two or more child node on the tree will conduct the network coding-based link-level multicast. In the second type of networks, due to the fact that not each node has the coding function, the selection of intermediate coding nodes should be careful. The overall goal is to select as many nodes with

coding function as possible but still guarantee that the overall tree cost is low, which might be an optimization problem. Thus, it deserves our further effort to design the algorithms for selecting the intermediate coding nodes.

In our multicast schemes, a receiver will send an ACK (or NACK) to the source once it correctly receives one packet. It is well-known that there is an ACK explosion problem in the conventional multicast schemes. Clearly, this problem still exists in the network coding-based multicast schemes. Thus, how to avoid the feedback implosion is also one important issue for the network coding-based schemes, which we will further study in the future.

Another issue is about file transfer time. Although the file transfer time issue is not well considered in the MPLC and MIRC problems, we briefly discuss this issue here. As the MPLC problem aims at maximizing the number of lost packets in the encoded packets, the MPLC-based scheme may achieve a smaller average packet delay (i.e., achieve a smaller average file transfer time) than the MIRC-based scheme. On the contrary, the MIRC aims at maximizing the number of intended receivers in the encoded packets. Thus, it is expected that when using the MIRC-based scheme, the difference of file transfer time among receivers is smaller than that of the MLPC-based scheme.

7. Conclusion

In this paper, we have proved the NP-completeness of the optimization problem in the available network coding-based reliable multicast schemes, and proposed an efficient heuristic algorithm for finding an approximately optimal solution of this problem. By adopting a new coding principle to fully take use of potential coding opportunities, we further enhanced the heuristic-based schemes. Compared with the available coding-based schemes, the enhanced heuristic-based low-complexity schemes can not only be scalable to larger number of multicast receivers (which is particularly significant since many multicast applications involve a great number of receivers), but also achieve slightly higher transmission efficiency.

It has also been shown that the transmission efficiency improvement from using network coding increases with both the size of lost-packet buffer and the number of multicast receivers. This improvement can be very significant when the lost-packet buffer size and number of receivers are large enough. E.g., for the case that the number of receivers is six and the buffer size is twelve packets, the transmission efficiency can be improved by as far as 29.0% when the enhanced heuristic-based dynamic scheme is adopted. Thus, the network coding provides us a new dimension for a more efficient transmission of reliable link-layer multicast in wireless networks.

Notice that all the previous and our coding-based reliable multicast schemes do not take the packet delay issue into account. Therefore, one interesting and important future work is to further extend the current work and design the delay-guaranteed reliable multicast scheme, although the applications of reliable multicast generally can tolerate delay.

Acknowledgment

References

[1] R. Ahlswede, N. Cai, S.-Y.R. Li, and R.W. Yeung, "Network information flow," IEEE Trans. Inf. Theory, vol.46, no.2, pp.388–404, July 2000.

[2] Y. Wu, P.A. Chou, and S.-Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," Proc. 39th Annual Conf. Infom. Sci. and Systems (CISS), 2005.

[3] Z. Li and B. Li, "On increasing end-to-end throughput in wireless ad hoc networks," Proc. Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, 2005.

[4] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XOR in the air: Practical wireless network coding," ACM SIGCOMM, Pisa, 2006.

[5] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," ACM SIGCOMM, 2007.

[6] S. Sengupta, S. Rayanchu, and S. Banerjee, "An analysis of wireless network coding for unicast sessions: The case for coding-aware routing," IEEE INFOCOM, 2007.

[7] J. Liu, D. Goeckel, and D. Towsley, "Bounds on the gain of network coding and broadcasting in wireless networks," IEEE INFOCOM, 2007.

[8] J. Le, J.C.S. Lui, and D.M. Chiu, "How many packets can we encode?—An analysis of practical wireless network coding," IEEE INFOCOM, 2008.

[9] K. Chi, X. Jiang, and S. Horiguchi, "Network coding opportunity analysis of COPE in multihop wireless networks," IEEE WCNC'08, Las Vegas, 2008.

[10] B. Ni, N. Santhapuri, Z. Zhong, and S. Nelakuditi, "Routing with opportunistically coded exchanges in wireless mesh networks," WiMesh, 2006.

[11] S. Zhang, S.C. Liew, and P.P. Lam, "Hot topic: Physical-layer network coding," MOBICOM, 2006.

[12] Y. Wu, P.A. Chou, and S.-Y. Kung, "Minimum-energy multicast in mobile ad hoc networks using network coding," IEEE Trans. Commun., vol.53, no.11, pp.1906–1918, Nov. 2005.

[13] D.S. Lun, N. Ratnakar, M. Médard, R. Koetter, D.R. Karger, T. Ho, E. Ahmed, and F. Zhao, "Minimum-cost multicast over coded packet networks," IEEE Trans. Inf. Theory, vol.52, no.6, pp.2608–2623, June 2006.

[14] J.S. Park, D. Lun, F. Soldo, M. Gerla, and M. Medard, "Performance of network coding in ad hoc networks," IEEE MILCOM, 2006.

[15] C. Fragouli, J. Widmer, and J.-Y.L. Boudec, "Efficient broadcasting using network coding," IEEE/ACM Trans. Netw., vol.16, no.2, pp.450–463, April 2008.

[16] L. Li, R. Ramjee, M. Buddhikot, and S. Miller, "Network coding-based broadcast in mobile ad hoc networks," IEEE INFOCOM, 2007.

[17] J. Nonnenmacher, M. Lacher, M. Jung, G. Carl, and E.W. Biersack,

"How bad is reliable multicast without local recovery?" IEEE IN-FOCOM, 1998.

[18] A. Shiozaki, "Adaptive type-ii hybrid broadcast arq system," IEEE Trans. Commun., vol.44, no.4, pp.420–422, April 1996.

[19] D. Nguyen, T. Nguyen, and B. Bose, "Wireless broadcasting using network coding," Third Workshop on Network Coding, Theory, and Applications, Jan. 2007.

[20] M. Garey and D. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, 1979.

[21] M. Luby, M. Watson, T. Gasiba, T. Stockhammer, and W. Xu, "Raptor codes for reliable download delivery in wireless broadcast systems," Proc. IEEE CCNC'06, Jan. 2006.

[22] A. Shokrollahi, "Raptor Codes," IEEE Trans. Inf. Theory, vol.52, no.6, pp.2551–2567, June 2006.

[23] Z. Zhang, K. Chi, X. Jiang, and S. Horiguchi, "Network coding-based reliable multicast scheme in wireless networks," IEICE Technical Report, RCS2009-161, Nov. 2009.
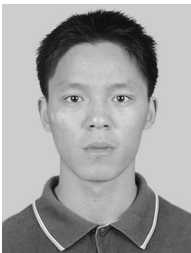
Kaikai Chi received the B.S. and M.S. degrees from Xidian University, China, in 2002 and 2005, respectively. He received his Ph.D. degree from Tohoku University, Japan, in 2009. He is now a postdoctoral fellow in Graduate School of Information Sciences, Tohoku University, Sendai, Japan. His current research focuses on the application of network coding in both wired and wireless networks. In 2008, he received Best Paper Award of IEEE WCNC. Now, he is a student member of the IEEE.

Xiaohong Jiang received his B.S., M.S. and Ph.D. degrees in 1989, 1992, and 1999 respectively, all from Xidian University, Xian, China. He is currently an Associate Professor in the Department of Computer Science, Graduate School of Information Science, TOHOKU University, Japan. Before joining TOHOKU University, Dr. Jiang was an assistant professor in the Graduate School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), from Oct. 2001 to Jan. 2005. Dr. Jiang was a JSPS (Japan Society for the Promotion of Science) postdoctoral research fellow at JAIST from Oct. 1999–Oct. 2001. He was a research associatein the Department of Electronics and Electrical Engineering, the University of Edinburgh from March 1999–Oct. 1999. Dr. Jiang's research interests include optical switching networks, routers, network coding, WDM networks, VoIP, interconnection networks, IC yield modeling, timing analysis of digital circuits, clock distribution and fault-tolerant technologies for VLSI/WSI. He has published over 130 referred technical papers in these areas. He is a senior member of IEEE.

Baoliu Ye received the Ph.D. degree in computer science from Nanjing University, China in 2004. He is now an associate professor at the department of Computer Science and Technology, Nanjing University, China. He served as a visiting researcher of the University of Aizu, Japan from March 2005 to July 2006. His current research interests include peer-to-peer computing, pervasive computing, wireless network and streaming services. He has published over 20 technical papers in these areas. He is a member of IEEE, ACM and CCF.

Susumu Horiguchi received the Beng., MEng., and Ph.D. degrees from Tohoku University in 1976, 1978, and 1981, respectively. He was a Full Professor in the Graduate School of Information Sciences, Tohoku University. He was a visiting scientist at the IBM T.J. Watson Research Center from 1986 to 1987. He was also a professor in the Graduate School of Information Science, JAIST (Japan Advanced Institute of Science and Technology). He has been involved in organizing international workshops, symposia, and conferences sponsored by the IEEE, IEICE, IASTED, and IPS. He has published more than 150 papers technical papers on optical networks, interconnection networks, parallel algorithms, high performance computer architectures, and VLSI/WSI architectures. He is a senior member of the IEEE and member of IPS and IASTED.