Linear Network Coding Strategies for the Multiple Access Relay Channel with Packet Erasures

Mingjun Dai, Ho Yuet Kwan, and Chi Wan Sung, *Member, IEEE*

***Abstract*—The multiple access relay channel (MARC) where multiple users send independent information to a single destination aided by a single relay under large-scale path loss and slow fading is investigated. At the beginning, the users take turns to transmit their packets. The relay is not aware of the erasure status of each packet at the destination but has the knowledge of the average signal-to-noise-ratio (SNR) of every communication link. With this knowledge, the relay applies network coded retransmission on the overheard packets so as to maximize the expected total number of recovered packets or minimize the average packet loss rate at the destination. Several network coding (NC) strategies at the relay are designed. In particular, for the case where the relay is given only one time slot for retransmission, an optimal NC construction is derived. For the multiple-slot case, three sub-optimal schemes are investigated, namely network coding with maximum distance separable (MDS) code (NC-MDS), the worst-user-first (WUF) scheme and a hybrid of NC-MDS and WUF. We prove that NC-MDS and WUF are asymptotically optimal in the high and low SNR regimes, respectively. A lower bound on the average packet loss rate has been derived. Numerical studies show that, in a cellular system, the hybrid scheme offers significant performance gain over a number of existing schemes in a wide range of SNR. We also observe that performance curves of both WUF and the hybrid scheme touch the derived lower bound in the low SNR regime.**

***Index Terms*—Multiple access relay channel, network coding, wireless relay network.**

I. Introduction

THE uplink of a wireless system, in which multiple users want to transmit their messages to a base station, is usually modeled by a multiple access channel (MAC). To improve system performance, a relay node may be added to assist the transmissions of the users. This is modeled by a multiple access relay channel (MARC) [1]. One particular case is that the each individual link is modeled by a packet erasure channel. For this type of channels, network coding (NC) [2] is a promising method to be used by the relay.

NC strategies at relays can be classified into two types, depending on whether there is feedback from the destination. It is not surprising that the use of feedback can lead to better performance, as demonstrated in [3]-[5]. On the other hand, the use of feedback increases control overhead and packet delay [6]. For this reason, some NC strategies which do not require any feedback have been designed [7]-[10]. For example, a simple binary NC method has been considered in [7]. For the two-user single-relay case, it has been shown to have lower spectrum cost than the traditional packet forwarding method [7]. This simple scheme, however, achieves only diversity order of two, even though there are more than two relays [8]. This poor diversity performance is due to its small field size. To achieve maximum diversity order, NC schemes with larger field sizes are investigated in [9], [10]. One common feature of those works is that the NC operations are all performed across all user packets, which is shown to be effective only in the high signal-to-noise ratio (SNR) regime. For the low or intermediate SNR regimes, whether NC should be applied or if applied, which user packets received at the relay should be network-coded remains unclear.

In this paper, we consider a MARC system without feedback. We assume that there are J users and one relay, and their transmissions suffer from independent, slow fading. All transmitters, including all J users and the relay, have no knowledge of the instantaneous channel state information (CSI) of its transmission link. The relay is assumed to know the statistics of all the communication links in the system. The transmissions of the users and the relay are performed in a time division multiple access manner. Each transmission frame consists of $J + S$ time slots, where $S \geq 1$ is a design parameter. In the first J time slots, each user takes turn to transmit his/her packet. The relay overhears K of them, where $K \leq J$. It should be noted that, owing to the random nature of the independent wireless links, the packets received successfully at the destination after the first J time slots are generally different from the packets overheard by the relay. That is how NC comes into play. After the first J time slots, the relay performs NC on the K overheard packets to generate S packets, and then transmit them in the next S time slots. These S network-coded packets can be regarded as *parity* packets. If these parity packets are properly designed, the expected number of recovered packets at the destination can be increased. Our objective is to determine an adaptive network code for the relay so as to maximize the expected number of recovered packets in a frame at the destination. The code is adaptive in the sense that the parity packets are chosen in a way which depends on the number of overheard packets, K.

For practical simplicity, we consider only *linear* network code. But even within this class, an optimal code for our

Manuscript received December 22, 2011; revised August 6, 2012; accepted September 5, 2012. The associate editor coordinating the review of this paper and approving it for publication was J. Luo.

This work was partially supported by a grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08).

M. Dai is with the College of Information Engineering, Shenzhen University, China (e-mail: mingjundai@hotmail.com). This work was partly done when he was at City University of Hong Kong.

H. Y. Kwan and C. W. Sung are with the Dept. of Electronic Engineering, City University of Hong Kong (e-mail: hykwan@gmail.com, albert.sung@cityu.edu.hk).

Digital Object Identifier 10.1109/TWC.2012.120412.112266

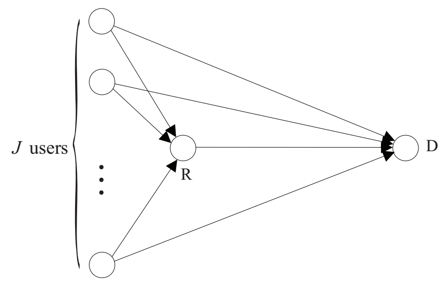


Fig. 1. The multiple access relay channel model.

problem is hard to find. Therefore, we first consider the special case where $S = 1$. It turns out that this case is tractable, and we derive an optimal NC scheme with polynomial-time complexity. For the general case where $S > 1$, we investigate three efficient sub-optimal schemes. The first one is a recently proposed scheme called NC with maximum distance separable (MDS) code(NC-MDS) [9], [10]. While the existing works only show that NC-MDS achieves full diversity, we provide a stronger result by proving its asymptotically optimality in terms of the expected number of recovered packets, or equivalently, the average packet loss rate, in the high SNR regime. For the low SNR regime, we propose the worst-user-first (WUF) scheme and prove that it is asymptotically optimal in the low SNR regime. The WUF scheme can be considered a degenerate case of NC, as the relay only retransmits some or all of its overheard packets without any network coding. The optimality of WUF in the low SNR regime reveals that NC may not be a good choice if the SNR is not high enough. By exploiting the advantages of WUF and NC-MDS, we develop a new scheme called Hybrid, which is a mixture of WUF and NC-MDS, for improving the performance in the intermediate SNR regime. A lower bound on the average packet loss rate has been derived. Numerical results show that Hybrid offers significant performance gain over a number of existing schemes in a wide range of SNR. We also find that the average packet loss rate curves of WUF and Hybrid touch the lower bound on the average packet loss rate in a wide low-SNR regime.

II. System Model and Problem Formulation

A. System Model

We consider a time-slotted MARC system as shown in Fig. 1, where there are J users, a relay R, and a destination D. The set of the users is denoted by $\mathcal{J}$. We group $J+S$ time slots together and call it a frame, where S is a design parameter. In each frame, each user in $\mathcal{J}$ wants to deliver a packet to D, possibly with the assistance of R. In other words, each user transmits packets at a rate of $1/(J+S)$ packets per slot duration. Transmissions within a frame are divided into two phases. The first phase consists of J time slots, in which each user in $\mathcal{J}$ takes turn to transmit his/her packet. The second phase consists of S time slots, which are used by R for packet retransmissions.

We model each wireless link from transmitter t to receiver r by a path loss model with independent Rayleigh fading, where $t \in \mathcal{J} \cup \{R\}$ and $r \in \{R, D\}$. Let L_{tr} be the path loss between t and r, which depends on the distance between them. Let $h_{tr} \sim \mathcal{CN}(0, \sigma_{tr}^2)$ be the Rayleigh-fading coefficient for the link between t and r. They are random but remain constant in each time slot. We assume that the instantaneous CSI of each link is available at its corresponding receiver, while the statistical CSI of all links is available at R.

We assume that the transmit power of all transmitters are the same and is denoted by P. The variance of the additive white Gaussian noise at each receiver is denoted by δ^2. Let Γ be the transmit SNR, defined as P/δ^2. The received SNR for the link from t to r is denoted by Γ_{tr}, which is equal to $L_{tr}|h_{tr}|^2\Gamma$. A received packet at r is erased if Γ_{tr} falls below a certain threshold, Γ_{thd}. Otherwise, it is said to be intact. The erasure probability of a packet from t to r is given by $\Pr\{\Gamma_{tr} < \Gamma_{thd}\} = 1 - e^{-\lambda_{tr}/\Gamma}$, where $\lambda_{tr} \triangleq \Gamma_{thd}/(\sigma_{tr}^2 L_{tr})$. Correspondingly, we denote the probability that an intact packet from t is received at r as $o_{tr} \triangleq \Pr\{\Gamma_{tr} \geq \Gamma_{thd}\}$.

Packet Retransmissions by Relay: Suppose after the first phase of a frame, there are K packets received intact at R. For notational convenience, label the corresponding users as users 1, 2, ..., K such that $o_{1D} \leq o_{2D} \leq \cdots \leq o_{KD}$. This set of K users is denoted by $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$. The other $J-K$ users can be ignored, since the relay cannot decode their packets successfully and will not retransmit the information from them. From the point of view of R in each frame, it is just like there are K effective users in the system. Therefore, from now on, we consider only the K users in $\mathcal{K}$.

For this particular frame, we index the S time slots in phase II by a set $\mathcal{S} \triangleq \{K+1, \ldots, K+S\}$. Note that this slot indexing is for simplifying notations only and should not be confused with the global time slot index in a transmission frame.

The relay is allowed to apply linear NC to perform retransmissions. Every packet is thus regarded as an N-dimensional vector over GF(q), where q is a power of prime to be specified later. The packet from user k is represented by a column vector $\boldsymbol{m}_k \in \text{GF}(q)^N$. Let $\boldsymbol{M} \triangleq [\boldsymbol{m}_1 \cdots \boldsymbol{m}_K]$. At slot $s \in \mathcal{S}$, the relay transmits $\boldsymbol{m}_s = \boldsymbol{M}\boldsymbol{c}_s \in \text{GF}(q)^N$, where column vector $\boldsymbol{c}_s \in \text{GF}(q)^K$ is called the encoding vector. It should be noted that although retransmissions from the relay is based on the received packets from the K users in a frame, K itself is random across the frames. The relay indeed provides retransmission services for all J users in the system.

Packet Reception at Destination: As mentioned, we only consider the reception of the packets from users in $\mathcal{K}$. The reception of the packets from the other $J - K$ users is governed only by their links to D, and is independent of what the relay does. Among the K packets transmitted by users in $\mathcal{K}$ in phase I, let L_{I} of them be received intact at D. Among the S packets transmitted by R in phase II, let L_{II} of them be received intact at D. Let $L \triangleq L_{\text{I}} + L_{\text{II}}$ and $P_{L \geq K} \triangleq \Pr\{L \geq K\}$.

Let $\boldsymbol{T} \triangleq [\boldsymbol{I}_K \; \boldsymbol{C}]$, where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix and $\boldsymbol{C} \triangleq [\boldsymbol{c}_{K+1} \cdots \boldsymbol{c}_{K+S}]$. The $K+S$ packets transmitted by users in $\mathcal{K}$ and by R can be represented by the columns of $N \times (K+S)$ matrix $\boldsymbol{Y} \triangleq \boldsymbol{M}\boldsymbol{T}$. We call $\boldsymbol{T}$ the encoding matrix and its columns the encoding vectors. Note that an uncoded packet has an encoding vector with Hamming weight one.

Let the slot indices of the packets received intact at D be the set $\mathcal{S}_L \triangleq \{s_1, \dots, s_L\} \subseteq \mathcal{K} \cup \mathcal{S}$. Let $\boldsymbol{Y}(\mathcal{S}_L)$ and $\boldsymbol{T}(\mathcal{S}_L)$ be the $N \times |\mathcal{S}_L|$ and $K \times |\mathcal{S}_L|$ submatrices of $\boldsymbol{Y}$ and $\boldsymbol{T}$ whose columns are chosen according to the indices in $\mathcal{S}_L$, respectively. It is easy to see that the columns of $\boldsymbol{Y}(\mathcal{S}_L)$ represent the intact packets received at D and $\boldsymbol{Y}(\mathcal{S}_L) = \boldsymbol{MT}(\mathcal{S}_L)$. Given $\boldsymbol{T}(\mathcal{S}_L)$ and $\boldsymbol{Y}(\mathcal{S}_L)$, packet $\boldsymbol{m}_1 \dots \boldsymbol{m}_K$ can all be recovered by D iff $\boldsymbol{Y}(\mathcal{S}_L) = \boldsymbol{MT}(\mathcal{S}_L)$ can be solved, which is equivalent to the case where $\boldsymbol{T}(\mathcal{S}_L)$ consists of at least K linearly independent columns.

B. Problem Formulation

Consider one transmission frame. In phase I, K packets are received intact at R. For $k \in \mathcal{K}$, let I_k be the indicator function such that $I_k = 1$ if $\boldsymbol{m}_k$ can be recovered by D and $I_k = 0$ otherwise. Among these K packets, let G be the expected total number of recovered packets at D, which is equal to $\sum_{k=1}^{K} \mathrm{E}[I_k]$. Our objective is to maximize G with an appropriate choice of $\boldsymbol{T}$, or equivalently $\boldsymbol{C}$. The problem can be formulated as

$$\max_{\boldsymbol{C} \in \mathcal{C}} G \triangleq \sum_{k=1}^{K} \mathrm{E}[I_k], \tag{1}$$

where $\mathcal{C}$ is defined as the set of all $K \times S$ matrices with each entry drawn from GF(q). Clearly, $|\mathcal{C}| = q^{KS}$. Hence solving (1) by simply searching among all choices in $\mathcal{C}$ involves high computational complexity.

Observe that among the K packets under consideration, $\sum_{k=1}^{K} o_{kD}$ represents the expected total number of recovered packets at D after phase I. Since it does not depend on the choice of $\boldsymbol{C}$, maximizing G is equivalent to maximizing $F \triangleq G - \sum_{k=1}^{K} o_{kD}$, where F represents the gain in the expected total number of recovered packets due to retransmissions in phase II. The above two objectives are also equivalent to minimizing $P_{\text{loss}} \triangleq \left(1 - \frac{G}{K}\right)$, where P_{loss} represents the average packet loss rate for those K users. In the following, we will use the above three equivalent optimization targets interchangeably.

III. Optimal Algorithm for $S = 1$

For the special case where R is given one slot for retransmission, we need to find a column vector $\boldsymbol{C} = \boldsymbol{c}_{K+1} \triangleq [c_{(K+1)1} \cdots c_{(K+1)K}]^T \in \text{GF}(q)^K$ so as to maximize G. We claim that there is no performance loss even if we restrict the field size, q, to be two. To see this, consider an encoding vector $\boldsymbol{c}_{K+1} \in \text{GF}(q)^K$, where $q > 2$. If $\boldsymbol{c}_{K+1}$ is replaced by the binary vector $\tilde{\boldsymbol{c}}_{K+1}$ whose component is equal to one whenever the corresponding component in $\boldsymbol{c}_{K+1}$ is non-zero, then the number of decodable packets remains the same no matter what $\mathcal{S}_L$ is, with reasons described below. Let $\mathcal{U}$ be the set of indices of the non-zero elements in $\boldsymbol{c}_{K+1}$. Since packets $\boldsymbol{m}_k$'s for $k \in \mathcal{K} \setminus \mathcal{U}$ are not involved in the NC process, $\sum_{k \in \mathcal{K} \setminus \mathcal{U}} I_k$ does not depend on the choice of the encoding vector. Now consider the sum for the users who involve in the NC process, $\sum_{k \in \mathcal{U}} I_k$. If the packet in slot $K+1$ is erased, then it does not matter which encoding vector is chosen. If not, then one more packet can be recovered iff $|\mathcal{S}_L \cap \mathcal{U}| = |\mathcal{U}| - 1$. Using $\boldsymbol{c}_{K+1}$ is the same as using $\tilde{\boldsymbol{c}}_{K+1}$, since the cardinalities of the two sets in the condition remain unchanged when $\boldsymbol{c}_{K+1}$ is replaced by $\tilde{\boldsymbol{c}}_{K+1}$. As a result, there is no loss of optimality in restricting the field to be binary.

For $q = 2$, there are 2^K choices for $\boldsymbol{c}_{K+1}$ totally. To find the optimal binary $\boldsymbol{c}_{K+1}$ for maximizing G, a simple brute-force exhaustive search may involve $O(2^K)$ computational complexity. In the following, we propose an algorithm that can find the optimal $\boldsymbol{c}_{K+1}$ in polynomial time.

Let $n \in \mathcal{K}$ be the Hamming weight of $\boldsymbol{c}_{K+1}$. Given the value of n, our objective becomes

$$\max_{\boldsymbol{c}_{K+1} \in \mathcal{C}} G \text{ s.t. the Hamming weight of } \boldsymbol{c}_{K+1} \text{ is } n, \tag{2}$$

where $\mathcal{C} = \text{GF}(2)^K$. The above problem can be considered a sub-problem of the original optimization problem in (1) for a given n. Note that there are K choices for n, namely $1, 2, \dots, K$, so the original optimization problem can be divided into K sub-problems. If the computational complexity of each sub-problem can be reduced, the computational complexity for the original optimization problem can be reduced as well. We investigate the sub-problems for $n = 1$, $1 < n < K$, and $n = K$ in the following.

For $n = 1$: Suppose all entries in $\boldsymbol{c}_{K+1}$ are zero except $c_{(K+1)u} = 1$ for an element $u \in \mathcal{K}$. Problem in (2) for $n = 1$ can be interpreted as retransmitting one packet without NC that maximizes G or F.

We first consider the recovery of $\boldsymbol{m}_u$ at D. If there is no packet erasure at slot u or slot $K+1$, $\boldsymbol{m}_u$ can be recovered. Hence, I_u equals 1 if $\{\max(\Gamma_{uD}, \Gamma_{RD}) \geq \Gamma_{thd}\}$ and 0 otherwise. Averaging over all channel realizations, we then have

$$\begin{aligned}\mathrm{E}[I_u] &= 1 - \Pr\{\max(\Gamma_{uD}, \Gamma_{RD}) < \Gamma_{thd}\} \\ &= 1 - \Pr\{\Gamma_{uD} < \Gamma_{thd}\}\Pr\{\Gamma_{RD} < \Gamma_{thd}\} \\ &= o_{uD} + o_{RD}(1 - o_{uD}).\end{aligned} \tag{3}$$

Clearly, $\mathrm{E}[I_k] = o_{kD}$ for $k \in \mathcal{K} \backslash \{u\}$. Then, summing up $\mathrm{E}[I_k]$ for all $k \in \mathcal{K}$ and subtracting phase I's contribution, we obtain $F = o_{RD}(1 - o_{uD})$. In this case, maximizing F is equivalent to selecting $u \in \mathcal{K}$ such that o_{uD} is the minimum. In other words, the optimal $u^* = \arg\min_{k \in \mathcal{K}} o_{kD}$. As a result, for the case $n = 1$, we have to choose the user who has the worst user-to-destination channel.

For $1 < n < K$: Recall that we use $\mathcal{U}$ to represent the users whose packets are chosen for linear combinations in the NC process, so $|\mathcal{U}| = n$. Given the value of n, there are $\binom{K}{n}$ possible choices for $\mathcal{U}$. Problem (2) is equivalent to finding the optimal $\mathcal{U}$ among these $\binom{K}{n}$ choices for $\boldsymbol{c}_{K+1}$.

We first consider the recovery of $\boldsymbol{m}_u$ for $u \in \mathcal{U}$. For each channel realization, there are two possibilities that $\boldsymbol{m}_u$ can be recovered at D: either there is no packet erasure at slot u or there is no packet erasure at the slots with indices in the set $\mathcal{U} \cup \{K+1\} \backslash \{u\}$ but packet erasure occurs in slot u. Hence, I_u equals 1 if $\{\Gamma_{uD} \geq \Gamma_{thd}\}$ or $\{\Gamma_{uD} < \Gamma_{thd}, \Gamma_{RD} \geq \Gamma_{thd}, \min_{v \in \mathcal{U}_{-u}} (\Gamma_{vD}) \geq \Gamma_{thd}\}$ and 0 otherwise, $u \in \mathcal{U}$, where $\mathcal{U}_{-u} \triangleq \mathcal{U} \backslash \{u\}$. Averaging over all channel realizations, we have

$$\mathrm{E}[I_u] = o_{uD} + o_{RD}(1 - o_{uD}) \prod_{v \in \mathcal{U}_{-u}} o_{vD}, \ u \in \mathcal{U}. \quad (4)$$

Next, for other users, we have $\mathrm{E}[I_k] = o_{kD}$ for $k \in \mathcal{K} \backslash \mathcal{U}$. Then, summing up $\mathrm{E}[I_k]$ for all $k \in \mathcal{K}$, we obtain

$$F = \sum_{u \in \mathcal{U}} \Big(o_{RD}(1 - o_{uD}) \prod_{v \in \mathcal{U}_{-u}} o_{vD} \Big). \quad (5)$$

To find the optimal $\mathcal{U}$ that maximizes the above F, we assume that a single element, denoted as u, is to be determined while the other $n-1$ elements have already been chosen from $\mathcal{K}$. Denote the set of those $n-1$ elements by $\mathcal{V}$. We want to find $u \in \mathcal{K} \setminus \mathcal{V}$ that maximizes F, which can now be rewritten as

$$F = \sum_{v \in \mathcal{V}} \Big(o_{RD}(1 - o_{vD}) o_{uD} \prod_{w \in \mathcal{V} \backslash \{v\}} o_{wD} \Big) + o_{RD}(1 - o_{uD}) \prod_{v \in \mathcal{V}} o_{vD}. \quad (6)$$

Regarding the above F as a function of o_{uD} and differentiating it with respect to o_{uD}, we obtain

$$\frac{\partial F}{\partial o_{uD}} = \sum_{v \in \mathcal{V}} \Big(o_{RD}(1 - o_{vD}) \prod_{w \in \mathcal{V} \backslash \{v\}} o_{wD} \Big) - o_{RD} \prod_{v \in \mathcal{V}} o_{vD} = o_{RD} \Big(\prod_{v \in \mathcal{V}} o_{vD} \Big) \Big[\sum_{v \in \mathcal{V}} \frac{1}{o_{vD}} - n \Big]. \quad (7)$$

Observe that $\frac{\partial F}{\partial o_{uD}}$ does not depend on o_{uD}. Given $\mathcal{V}$, F can either be a monotone function of o_{uD} or a constant. If F is a constant, then any choice of $u \in \mathcal{K} \setminus \mathcal{V}$ is optimal. Otherwise, the optimal u must be $u^* = \arg\min_{u \in \mathcal{K} \setminus \mathcal{V}} \{o_{uD}\}$ or $\arg\max_{u \in \mathcal{K} \setminus \mathcal{V}} \{o_{uD}\}$.

As the above argument applies to arbitrary u, there are only $n+1$ possible candidates for the optimal $\mathcal{U}$, namely, $\{1, 2, \ldots, n\}, \{1, 2, \ldots, n-1, K\}, \ldots, \{K-n+1, \ldots, K\}$. To find the optimal solution, relay R computes the corresponding F for each of the above $n+1$ choices and chooses the one with the largest F. As a result, the number of choices can be reduced from $\binom{K}{n}$ to $n+1$. An example for $n = 3$ is: There are 4 choices for $\mathcal{U}$, including $\{1, 2, 3\}$, $\{1, 2, K\}$, $\{1, K-1, K\}$, and $\{K-2, K-1, K\}$, where $K \geq 4$.

For $n = K$: there is only one choice that is $\boldsymbol{c}_{K+1} = (1, 1, \ldots, 1)$.

With the results of the above three types of sub-problems, we only need to generate the $1 + (3 + 4 + \ldots + K) + 1 = \frac{K(K+1)}{2} - 1$ choices of $\mathcal{U}$ instead of considering 2^K choices in an exhaustive search and compute their corresponding values of F. Given $\mathcal{U}$, F can also be written as

$$F = o_{RD} \Big(\prod_{u \in \mathcal{U}} o_{uD} \Big) \Big[\sum_{u \in \mathcal{U}} \frac{1}{o_{uD}} - |\mathcal{U}| \Big], \quad (8)$$

which involves $O(K)$ multiplications and $O(K)$ additions. The optimal $\mathcal{U}$ can then be obtained by choosing the one among the $\frac{K(K+1)}{2} - 1$ choices that maximizes F. The overall complexity of the algorithm is $O(K^3)$.

IV. Algorithms for $S > 1$

Unlike the case of $S = 1$ where the optimal solution of $\boldsymbol{C}$ can be found in polynomial time, things become more involved when $S > 1$. As explained before, in general, such problems may involve searching among all q^{KS} possible choices, and it may not be easy to derive an efficient algorithm for the optimal choices of $\boldsymbol{C}$. As a result, schemes with good performance and low computational complexity are usually sensible choices in practice. We first perform a complete analysis on an existing scheme, namely, NC with maximum distance separable (MDS) code scheme (NC-MDS), which is proposed in [9], [10]. In particular, we prove that NC-MDS is asymptotically optimal in the high SNR regime. We then propose the worst user first (WUF) scheme for the low SNR environment. Then, by hybridizing these two schemes, we construct a new scheme that promises good performance in the intermediate SNR regime.

A. NC-MDS

At the high SNR environment, the probability that at least K intact packets are received at D (i.e. $P_{L \geq K}$) is high. In this case, the MDS code could be exploited to generate network coded packets. The MDS code is a class of erasure codes which has the property that, given a $(K+S, K)$ MDS code, the K original packets can be recovered from any K arbitrary intact encoded packets. More specifically, the $K \times (K+S)$ encoding matrix $\boldsymbol{T}$ of a $(K+S, K)$ MDS code has the following mathematical property:

Property 1: Any K columns of the $K \times (K+S)$ encoding matrix $\boldsymbol{T}$ are linearly independent.

MDS codes can be employed at R with S retransmission slots by applying MDS encoding on the intact packets received at R in phase I. We call this retransmission scheme NC with MDS code (NC-MDS). With the NC-MDS scheme at R and a high SNR environment for a high $P_{L \geq K}$, D is able to recover all packets of the users in $\mathcal{K}$ with high probability. In order to apply NC-MDS at R, an encoding matrix $\boldsymbol{T}$ with Property 1 should be provided. An example is the well-known Reed Solomon (RS) code [11]. In fact, for $q > K+S$, MDS codes of any rate exists. A given encoding matrix $\boldsymbol{T}$ can always be transformed into a systematic form $\boldsymbol{T}'$ by elementary row operations and column permutations. The encoding vectors for generating the network coded packets can then be obtained by selecting the last S columns of $\boldsymbol{T}'$.

Existing works on NC-MDS only show its optimality in terms of diversity order. In the following, we provide a stronger result on NC-MDS, namely, its optimality in terms of average packet loss rate.

Theorem 1: The achieved G by an NC-MDS is given by:

$$G_{\text{NC-MDS}} = K P_{L \geq K} + \mathrm{E}[L_{\mathrm{I}} | L < K](1 - P_{L \geq K}). \quad (9)$$

Proof: If $L \geq K$ at D, according to Property 1, all packets can be recovered at D. Therefore, we only need to consider the case where $L < K$. For that case, it is clear that the L_{I} packets received correctly by D in phase I can be recovered, since those packets are uncoded. We are going to prove that the L_{II} packets received by D in phase II are useless for recovering the missing packets in phase I. Note that Property 1 implies the L columns of the encoding matrix corresponding to the L intact packets are linearly independent. Let the packet of user $k \in \mathcal{K}$ be an erased packet in phase I. According to Property 1, the L columns of the encoding matrix corresponding to the L intact packets and the k-th column of the matrix are all linearly independent, since the total number of columns under concern is $L+1 \leq K$. Hence, the packet of user $k \in \mathcal{K}$ cannot be recovered. This argument applies to any arbitrary erased packet and therefore (9) follows. ■

Intuitively, it is likely that $L \geq K$ in the high SNR environment but not so in the low SNR environment. Indeed, NC-MDS is optimal when the SNR is high enough, as shown in the following theorem.

Theorem 2: There exists an SNR threshold, above which, an NC-MDS scheme outperforms any linear NC scheme that does not belong to the class of NC-MDS.

The proof is relegated to the appendix.

B. WUF

When SNR is low, the probability that at least K intact packets are successfully received at D (i.e. $P_{L \geq K}$) is low, hinting that the NC-MDS scheme is a poor choice for the low SNR regime. For such a poor channel environment, we propose another method which utilizes each slot in phase II to retransmit a packet received by R in phase I. In other words, the packets transmitted by the relay is *uncoded*, and the number of non-zero entries in vector $\boldsymbol{c}_s$ is one for all $s \in \mathcal{S}$. It can be shown that uncoded retransmission would be a good option, especially when SNR is very low.

For a particular channel realization, define $\mathcal{H}_k \subseteq \mathcal{S}$ as the set of slot indices within phase II that are allocated to user k with cardinality $n_k \in \{0, 1, 2, \ldots, S\}$, for $k \in \mathcal{K}$.

Define $\Gamma_k \triangleq \max_{s \in \mathcal{H}_k} L_{RD}(|h_{RD}(s)|^2)\Gamma$ when $\mathcal{H}_k \neq \emptyset$ and $\Gamma_k \triangleq 0$ when $\mathcal{H}_k = \emptyset$, where $h_{RD}(s)$ represents the fading coefficient from R to D at slot $s \in \mathcal{S}$. Since packet k is received correctly iff the $n_k + 1$ slots assigned to user k (in phases I and II) has no erasure, we have I_k equals 1 if $\{\max(\Gamma_{kD}, \Gamma_k) \geq \Gamma_{thd}\}$ and 0 otherwise. Averaging over all channel realizations, we have

$$\begin{aligned} \mathrm{E}[I_k] &= 1 - \Pr\{\max(\Gamma_{kD}, \Gamma_k) < \Gamma_{thd}\} \\ &= 1 - (1 - o_{kD})(1 - o_{RD})^{n_k}. \end{aligned} \tag{10}$$

Summing up $\mathrm{E}[I_k]$ for all $k \in \mathcal{K}$, F achieved by uncoded retransmission can be expressed as:

$$F_{\text{uncoded}} = \sum_{k=1}^{K} (1 - o_{kD})\big(1 - (1 - o_{RD})^{n_k}\big). \tag{11}$$

Since F_{uncoded} depends on $\mathcal{H}_k$'s only through their cardinalities n_k's, the uncoded retransmission problem can be formulated as

$$\max_{n_1, n_2, \ldots, n_K} F_{\text{uncoded}} \text{ s.t. } \sum_{k=1}^{K} n_k = S. \tag{12}$$

This problem can be solved by the greedy approach. The S time slots are assigned one by one in an iterative manner. Denote $\boldsymbol{n}^{(t)} \triangleq (n_1^{(t)}, \ldots, n_K^{(t)})$ as a vector whose k-th element $n_k^{(t)}$ represents the number of retransmission slots allocated for retransmitting the packet of user $k \in \mathcal{K}$ at the t-th iteration step, where $t \geq 0$. Furthermore, we define $F_{\text{uncoded}}^{(t)}$ as the achieved F_{uncoded} at the t-th iteration:

$$F_{\text{uncoded}}^{(t)} \triangleq \sum_{k=1}^{K} (1 - o_{kD})\Big(1 - (1 - o_{RD})^{n_k^{(t)}}\Big). \tag{13}$$

Suppose at the $(t+1)$-th iteration, the packet of user u is selected for retransmission. Then we have $n_k^{(t+1)}$ equals $n_k^{(t)} + 1$ for $k = u$ and $n_k^{(t)}$ otherwise. Define the increment of F_{uncoded} at the t-th iteration step as $\Delta F_{\text{uncoded}}^{(t)} \triangleq F_{\text{uncoded}}^{(t+1)} - F_{\text{uncoded}}^{(t)}$. Then we have

$$\Delta F_{\text{uncoded}}^{(t)} = o_{RD}(1 - o_{uD})(1 - o_{RD})^{n_u^{(t)}}, \tag{14}$$

which can be maximized by choosing u in the following way: $u^* = \arg\max_{u \in \mathcal{K}} \bar{o}_{uD}^{(t)}$, where $\bar{o}_{uD}^{(z)} \triangleq (1 - o_{uD})(1 - o_{RD})^{n_u^{(z)}}$. That means, R transmits the packet of user u^* at slot $(t+1) \in \mathcal{S}$ as an uncoded retransmission packet. Having defined the iterative relation, we start the algorithm by setting $\boldsymbol{n}^{(0)}$ as the zero vector, and the algorithm terminates after S iterations.

Recall that the term $1 - o_{uD}$ is the erasure probability of a packet of user u at D in phase I. After the t-th iteration, if R has sent that packet by n_u times, the erasure probability of that packet will be discounted by $(1 - o_{RD})^{n_u}$. Therefore, $\bar{o}_{uD}^{(t)}$ can be interpreted as the updated erasure probability of that packet after iteration t, and $u^* = \arg\max_{u \in \mathcal{K}} \bar{o}_{uD}^{(t)}$ as helping the user with the largest updated erasure probability. That is why we call this scheme *Worst User First* (WUF).

Theorem 3: WUF is optimal for the uncoded retransmission problem.

Proof: Suppose we have already assigned the first $S-1$ time slots and there is only one time slot left. The erasure probabilities have been updated as mentioned above. The scenario after assigning the first $S-1$ time slots is equivalent to the case where $S = 1$. According to our analysis in Subsec. III-A, it is optimal is to retransmit the packet with the largest erasure probability, which is the same as WUF.

Consider the scenario that $S - \tau$ time slots have been assigned and there are τ time slots left, where $1 \leq \tau \leq S-1$. Suppose that WUF is optimal. We are going to show that WUF is optimal when there are $\tau + 1$ time slots left. Let $\boldsymbol{a}^* = (a_1^*, a_2^*, \ldots, a_\tau^*)$ be an optimal assignment vector, where $a_j^* \in \mathcal{K}$ represents that the j-th time slot in the remaining τ slots is used to retransmit the packet of user a_j^*. Similarly, let $\boldsymbol{a}'$ be the assignment vector obtained by using WUF. Note that a permutation of a given assignment vector does not change its performance, since the fading coefficients from R to D in different time slots are i.i.d. First consider the case where $a_i^* = a_1'$ for some i. We permute $\boldsymbol{a}^*$ to obtain $\tilde{\boldsymbol{a}} \triangleq (a_i^*, a_1^*, a_2^*, \ldots, a_{i-1}^*, a_{i+1}^*, \ldots, a_\tau^*)$. Note that $\boldsymbol{a}'$ performs at least as well as $\tilde{\boldsymbol{a}}$, since their first components are the same and WUF is optimal when applying to the remaining $\tau - 1$ slots according to the hypothesis. Hence, $\boldsymbol{a}'$ has the same performance as $\boldsymbol{a}^*$ and is thus optimal. Next consider the case where $a_i^* \neq a_1'$ for all i's. $\boldsymbol{a}^*$ cannot be optimal, since replacing a_τ^* by a_1' yields a better performance according to our analysis in subsec. III.

The statement follows by combining the above two cases and using mathematical induction. ■

While MC-NDS is optimal at the high SNR regime, WUF is optimal at the low SNR regime.

Theorem 4: There exists an SNR threshold, below which, the WUF scheme outperforms any other linear NC schemes.

The proof is relegated to the appendix.

C. Hybrid

As NC-MDS and WUF are designed for the high and low SNR regimes, respectively, they may not be a good choice for the intermediate SNR regime. For this reason, we propose combining these two schemes for exploiting the advantages from both. This strategy is a hybrid of NC-MDS and WUF, and we call it Hybrid for short.

We divide packets received at R into two groups. Group I contains the n packets that have the smallest erasure probabilities, where $0 \leq n \leq K$. Group II contains the remaining $K - n$ packets. By dividing packets in this way, there are $K + 1$ possible group patterns. For a given group pattern, R applies NC-MDS to packets in group I and WUF to group II.

Recall that R has S slots for transmissions. Among these S slots, S_1 slots are allocated to group I and the remaining $S - S_1$ slots to group II, where $0 \leq S_1 \leq S$. Totally there are $S + 1$ slot-allocation patterns.

Considering the group patterns and slot-allocation patterns together, there should be $(K + 1)(S + 1)$ possible choices to consider. However, for the two degenerate cases where one of the groups has no users, the allocation of slots becomes trivial and has only one choice. Hence, the number of choices we need to consider reduces to $(K - 1)(S + 1) + 2$. For each choice, we compute the corresponding F. Finally, we select the group pattern and slot-allocation pattern that attains the maximum of F. Clearly, the Hybrid scheme subsumes both NC-MDS and WUF as special cases.

D. Complexity Analysis

In this subsection, we discuss the encoding and decoding complexity of our proposed schemes, and have a comparison with random linear network code (RLNC) [12]. In general, the number of time slots used in the second phase, S, can depend on K, the number of packets received successfully at R. We assume that S is bounded above by a constant, $S_{\max}$.

To implement NC-MDS, we may first determine J matrices, $\boldsymbol{T}_1, \boldsymbol{T}_2, \ldots, \boldsymbol{T}_J$. For $1 \leq j \leq J$, $\boldsymbol{T}_j$ is a $j \times (j + S_{\max})$ matrix, which represents a systematic encoding matrix for a $(j + S_{\max}, j)$ MDS code. Those matrices are pre-computed and stored. When K packets have been received successfully at R in phase I, the matrix $\boldsymbol{T}_K$ will be used for encoding. It is shortened to $K+S$ columns by removing the last $S_{\max} - S$ columns. The encoding can then be done by right-multiplying the $K \times (K + S)$ shortened matrix to the K received packet vector. As we only need to compute the parity packets, $O(KS)$ multiplications are involved. The decoding of an MDS code is $O(K^2)$.

As for WUF, it does not involve encoding complexity while involves finding the worst user in each of the S iterations. Decoding is not needed, since no linear combination of packets are involved.

For the Hybrid, we need to consider $O(KS)$ different choices of group patterns and slot-allocation patterns. We choose the group and slot-allocation pattern that maximizes F. For this particular pattern, both NC-MDS and WUF are performed. The encoding complexity is $O(K^2)$. The decoding complexity is the same as NC-MDS, i.e., $O(K^2)$. Different from other schemes, Hybrid requires the calculations of F for the $O(KS)$ different choices of group patterns and slot-allocation patterns. While exact calculations are very time consuming, one may use Monte-Carlo simulations to obtain approximate values of F.

RLNC has an encoding complexity of $O(KS)$, which is the same as NC-MDS. It differs from NC-MDS in that the coefficients of the encoding matrix are generated randomly. For decoding, Gaussian elimination is needed, which has a complexity of $O(K^3)$ [13], which is higher than our proposed methods.

V. Upper Bounds

We give the performance upper bound on G and on the total number of recovered packets for each channel realization, respectively. These two automatically translate into lower bounds on the average packet loss rate that will be used for evaluating the schemes in next section.

Analytical Upper Bound: Given the number of retransmission slots S, the expected total number of the recovered packets at D can be increased by an amount at most $o_{RD}S$. So we immediately have

Theorem 5: The expected total number of recovered packets at D, also known as G is upper bounded by $G' = o_{RD}S + \sum_{k=1}^{K} o_{kD}$.

Note that the above bound is very tight in the low SNR regime but very loose in the high SNR regime, as will be demonstrated in the next section.

Numerical Upper Bound: For each channel realization, we propose an upper bound on G_{CH}, which is defined as the total number of recovered packet at D. It can be obtained in the following manner: When $L \geq K$, K is the upper bound for G_{CH} in these cases. On the other hand, when $L < K$, at most L packets can be recovered at D. Formally, the upper bound on G_{CH} for each channel realization is given by $G_{\text{CH}} \leq \min(L, K)$. Note that this upper bound is rather optimistic especially in the high SNR regime.

VI. Numerical Study

We perform numerical study on the proposed schemes together with some existing schemes in a mobile system where the base station (BS) serves as destination. We adopt the COST-Hata-Model [14, Chap. 2] as the path loss model between a BS, denoted by b, and a user, denoted by u, with distance d_{ub} (km). The path loss factor in dB is given by $L_{ub} = 46.3 + 33.9 \log_{10} f - 13.82 \log_{10} h_b - a(h_u) + (44.9 - 6.55 \log_{10} h_b) \log_{10} d_{ub}$, where $a(h_u) \triangleq (1.1 \log_{10} f - 0.7)h_u - (1.56 \log_{10} f - 0.8)$, f is the carrier frequency in MHz, and h_b and h_u are the effective antenna heights of

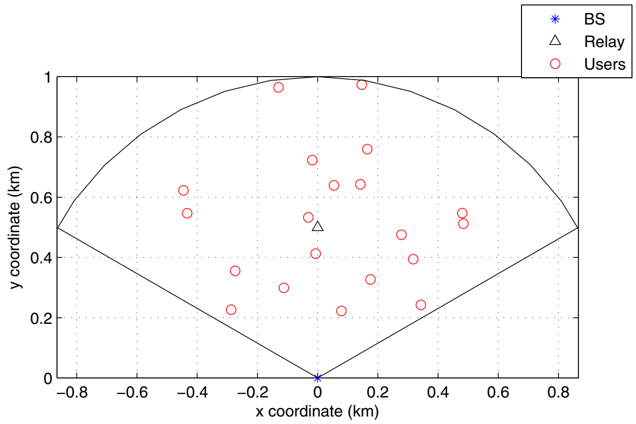


Fig. 2. Location of the nodes.

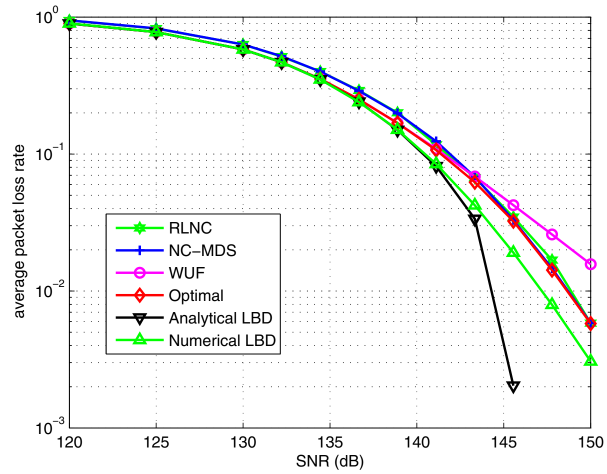


Fig. 3. Average packet loss rate comparison among WUF, NC-MDS, Optimal, two lower bounds, and RLNC, $J = 20$, $S = 1$.

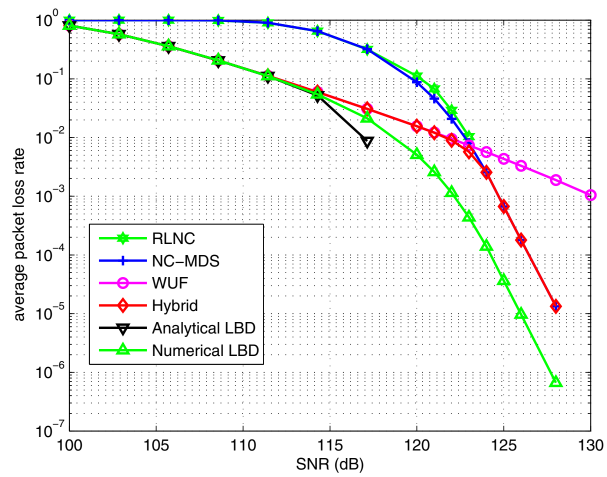


Fig. 4. Average packet loss rate comparison among WUF, NC-MDS, Hybrid, two lower bounds, and RLNC, $J = 20$, $S = 20$.

the BS and the user in meter, respectively. The relay plays two different roles during the two phases. During phase I, it plays the role of a "BS". During phase II, it plays the role of a "user". Correspondingly, the propagation loss in the computation of L_{ub} is calculated according to the role of the relay. We consider a 120° sector of a cell shown in Fig. 2. One unit length in the figure represents 1 km. The BS, which corresponds to D in the MARC model, locates at $(0, 0)$ and relay R locates at $(0, 0.5)$. In our simulations, we assume that users are uniformly distributed in the sector. We further assume that the carrier frequency is 2 GHz, the effective antenna heights of the users, relay R, and the BS are 1, 10, and 30 meters, respectively. Besides, we set $J = 20$ and $\sigma_{xR} = \sigma_{xD} = \sigma_{RD} = 1$ for all $x \in \mathcal{X}$. Fig. 2 shows one realization of the locations of the users in a sector.

In the following, we consider two settings. In the first setting, we assumes perfect user-relay links, i.e., relay R can recover all the J users' packets. The second setting is a more realistic one that the received packet at relay R could be erased due to fading. Let N_{sim} denote the number of times/channel realizations that we run the simulation. Each data point we present in our performance chart involves $N_{\text{sim}} = 3 \times 10^5$ random channel realizations.

A. Perfect User-relay Links

We set $J = K = 20$, $q = 2^{\lceil \log_2(K+S) \rceil}$, and the field size of RLNC to be q for a fair comparison with NC-MDS. The average packet loss rate P_{loss}, as defined in Subsec. II-B, of WUF, NC-MDS, the Optimal algorithm for $S = 1$, Hybrid, RLNC, and the two lower bounds derived from the two upper bounds in Sec. V are plotted in Fig. 3 for $S = 1$ and Fig. 4 for $S = 20$, respectively.

We first examine the case for $S = 1$. In Fig. 3, we observe that the performance of RLNC and NC-MDS are close to each other. In addition, the performance of WUF is almost the same as the optimal algorithm for $S = 1$ in low SNR region, while NC-MDS has almost the same performance as the optimal algorithm for $S = 1$ in high SNR region. In Fig. 3, we observe that there is a performance gap between the optimal algorithm for $S = 1$ and the numerical LBD in the intermediate and high SNR regime. It is mainly due to the fact that the numerical LBD is too optimistic as explained in Sec. V.

Next, we examine the case for $S > 1$. In Fig. 4, it can be seen that the performance of RLNC and NC-MDS are almost the same. Besides, Hybrid is able to take the advantages of both WUF and NC-MDS in the low and the high SNR regions, respectively, and even offers a marginal performance gain over WUF and NC-MDS in the intermediate SNR region. Furthermore, the curve of Hybrid overlaps with the analytical lower bound in the low SNR regime. At high SNR, there is a 2 dB performance gap between Hybrid and the numerical LBD. It can be attributed to the fact that the numerical LBD is too optimistic as explained before. Note that due to the high decoding complexity of RLNC, we do not plot the data points of RLNC for SNR larger than 125 dB.

B. Imperfect User-relay Links

In the imperfect user-relay link case, K is a random variable. The performance metric adopted in previous subsection

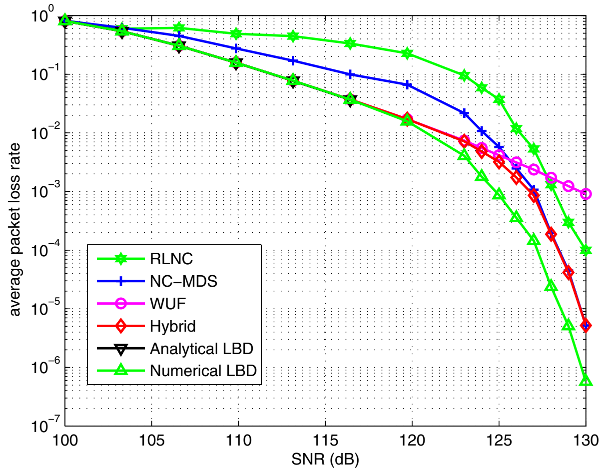


Fig. 5. Average packet loss rate comparison among WUF, NC-MDS, Hybrid, two lower bounds, and RLNC, $J = 20$.

cannot be directly applied due to the random nature of K. Therefore, we have to define another performance metric. For $i = 1, \ldots, N_{\text{sim}}$, let K_i be the number of packets correctly received by the relay at the i-th realization, and G_i be the number of recovered packets at the destination out of the K_i packets. Then we define $P_{\text{loss}} \triangleq \left(1 - \frac{\sum_{i=1}^{N_{\text{sim}}} G_i}{\sum_{i=1}^{N_{\text{sim}}} K_i}\right)$, which measures the proportion of packets that have been received by the relay but cannot be recovered by the destination over the N_{sim} simulation runs.

For the i-th realization, we set $S = K_i$, $q = 2^{\lceil \log_2(K_i+S) \rceil}$, and the field size of RLNC to be q. We examine WUF, NC-MDS, Hybrid, RLNC, and the two lower bounds by considering their performance of P_{loss} via Monte-Carlo simulations in Fig. 5. It is found that the results are similar to the results of all the concerned schemes in perfect user-relay link cases in Fig. 4 except that NC-MDS outperforms RLNC significantly in a wide range of SNR. The exception is due to the following reason. The field size adopted by NC-MDS and RLNC in the imperfect case is less than that of the perfect case. Note that when the field size of RLNC is small, there is a high chance that several received packets may be linearly dependent, which is spectral inefficient. This indicates poor performance of RLNC in case of obtained G, which leads to a gap in terms of P_{loss} between RLNC and NC-MDS in the imperfect user-relay setting.

VII. Conclusion

We investigate linear NC construction at the relay in the MARC system so as to minimize the average packet loss rate at the destination. For the case where the relay is given a single slot to transmit, we derive an efficient optimal linear NC construction. For the case where the relay is given multiple slots to transmit, we investigate three efficient sub-optimal algorithms, including NC-MDS, WUF, and Hybrid. We prove the asymptotic optimality of NC-MDS and WUF in the high and low SNR regimes, respectively. On the other hand, the Hybrid scheme, which takes the advantages of WUF and NC-MDS, promises performance gain over NC-MDS and WUF especially in the intermediate SNR regime. Compared to RLNC, all the three schemes enjoy more efficient decoding with WUF being the most efficient. With respect to encoding complexity, all the three schemes involve the same complexity as RLNC, while Hybrid requires additional calculations so as to choose the best group and slot-allocation pattern. We derive a lower bound on the average packet loss rate. Numerical results show that for $S > 1$ the average packet loss rate curves of WUF and Hybrid overlap with the lower bound on the average packet loss rate in a wide low-SNR region. That shows WUF and Hybrid are excellent choices for the low SNR region. In particular, Hybrid is shown to be an all-round scheme that offers an excellent performance in a wide range of SNR.

The development of retransmission schemes at the relay for the MARC is the main contribution of this paper. In this study, the physical layer issues are simplified by assuming that the transmission links are packet erasure channels. A more general assumption is to replace the erasure channel model by the fading channel model. In that case, physical-layer techniques can be adopted, which will further improve system performance. For example, if uncoded packet retransmission like WUF is used at the relay, the base station may employ maximal ratio combining to combine the packet from the user with the packet from the relay before decoding. When network coding is used, new cross-layer methods have to be designed, which is an interesting extension of the work in this paper.

Appendix A

A. Proof of Theorem 2

Proof: Let $\boldsymbol{D} \triangleq (D_1, D_2, \ldots, D_{K+S})$, where D_i equals 0 if slot i experiences erasure and 1 otherwise, $i \in \mathcal{K} \cup \mathcal{S}$. There are $N \triangleq \sum_{L=0}^{K+S} \binom{K+S}{L} = 2^{K+S}$ realizations of $\boldsymbol{D}$ totally. For ease of description, we index these N realizations by $\{1, 2, \ldots, N\} \triangleq \mathcal{N}$. The indices are ordered such that a realization of $\boldsymbol{D}$ having more number of ones is assigned a smaller index; ties are broken arbitrarily. Denote the n-th realization by $\boldsymbol{d}^{(n)} = \{d_1^{(n)}, d_2^{(n)}, \ldots, d_{K+S}^{(n)}\}$, where $d_i^{(n)}$ indicates the erasure status of slot $i \in \mathcal{K} \cup \mathcal{S}$, $n \in \mathcal{N}$. Define $p^{(n)} \triangleq \Pr\{\boldsymbol{D} = \boldsymbol{d}^{(n)}\} = \prod_{i=1}^{K+S} \Pr\{D_i = d_i^{(n)}\}$.

We first consider the case where $L \geq K$. There are $N_1 \triangleq \sum_{L=K}^{K+S} \binom{K+S}{L}$ realizations of $\boldsymbol{D}$ totally, which are indexed by $\{1, 2, \ldots, N_1\} \subset \mathcal{N}$. Define $r^{(n)} \triangleq \Pr\{\boldsymbol{D} = \boldsymbol{d}^{(n)} | L \geq K\} = \frac{p^{(n)}}{P_{L \geq K}}, n \in \{1, 2, \ldots, N_1\}$. Consider a scheme which does not belong to the class of NC-MDS. There must exist $M > 0$ among these N_1 realizations, under which destination D can recover at most $K-1$ users' packets since the scheme does not satisfy Property 1. Let the set of the indices of these M realizations be $\mathcal{M}$. Define $r \triangleq \sum_{n \in \mathcal{M}} r^{(n)}$ and $n_{\min} \triangleq \arg\min_{n \in \mathcal{M}} r^{(n)}$. The expected total number of recovered packets by destination D when not all K users' packets can be recovered is bounded above by $r(K-1) + (1-r)K = K - r \leq K - Mr^{(n_{\min})}$.

Next consider the case where $L < K$. Totally, there are

$N_2 \triangleq \sum_{L=0}^{K-1} \binom{K+S}{L}$ realizations, which are indexed as the last N_2 indices in $\mathcal{N}$. Define $n_{\max} \triangleq \arg \max_{n=N_1+1}^{N} p^{(n)}$. Clearly, $1 - P_{L\geq K} \leq N_2 p^{(n_{\max})}$. Besides, the number of packets that can be recovered for each channel realization is at most L. Combining with the result for the case $L \geq K$, we have the following upper bound on the achievable G by any non-NC-MDS scheme:

$$G_{\text{non-NC-MDS}} \leq (K - Mr^{(n_{\min})})P_{L\geq K} + \text{E}[L|L < K](1 - P_{L\geq K}) \quad (15)$$
$$\leq (K - Mr^{(n_{\min})})P_{L\geq K} + \text{E}[L|L < K]N_2 p^{(n_{\max})}. \quad (16)$$

We want to prove that there exists a certain Γ_0, such that for $\Gamma \geq \Gamma_0$, $G_{\text{NC-MDS}}$, which can be expressed in (9), is greater than the above upper bound, that is equivalent to,

$$\frac{p^{(n_{\max})}}{p^{(n_{\min})}} \leq \frac{M}{N_2 \text{E}[L_{\text{II}}|L < K]}. \quad (17)$$

A sufficient condition for the above inequality to hold is that

$$\frac{p^{(n_{\max})}}{p^{(n_{\min})}} \leq \frac{M}{N_2 K}. \quad (18)$$

Note that the numerator and the denominator in the LHS of (18) are the occurrence probabilities of two different events; the first event has more erasures than the second. Under Rayleigh fading, we can write down the probability of erasure and non-erasure for slot $i \in \mathcal{K} \cup \mathcal{S}$ as

$$\Pr\{D_i = d_i\} = \begin{cases} e^{-\lambda(i)/\Gamma}, & \text{if slot } i \text{ experiences non-erasure,} \\ 1 - e^{-\lambda(i)/\Gamma}, & \text{if slot } i \text{ experiences erasure,} \end{cases} \quad (19)$$

where $\lambda(i) \triangleq \lambda_{iD}$ if $i \in \mathcal{K}$ and $\lambda(i) \triangleq \lambda_{RD}$ if $i \in \mathcal{S}$. Substituting (19) into (18) for all $i \in \mathcal{K}\cup\mathcal{S}$, we obtain that the numerator of LHS of (18) has more $1 - e^{-\lambda/\Gamma} \approx 0$ terms and fewer $e^{-\lambda/\Gamma} \approx 1$ terms than the denominator. This indicates that $\frac{p^{(n_{\max})}}{p^{(n_{\min})}} \to 0$ as $\Gamma \to 0$. Since the RHS of (18) does not depend on the SNR and is strictly greater than zero, we are able to find a threshold Γ_0 on the SNR, above which (18) is true. ■

B. Proof of Theorem 4

Proof: We say that a linear network code is non-trivial if not all packets are uncoded. We first prove that uncoded retransmission schemes perform better than non-trivial linear NC schemes in the low SNR regime. Consider an arbitrary uncoded scheme and an arbitrary non-trivial coding scheme. Denote their achievable G by G_{uncoded} and G_{coded}, respectively. We have

$$G_{\text{uncoded}} \geq \Pr[L = 1] + \sum_{x=2}^{K+S} \Big(\text{E}[L_{\text{I}}|L_{\text{II}} = 0, L = x]\Pr[L_{\text{II}} = 0, L = x] + \text{E}[L_{\text{I}}|L_{\text{II}} > 0, L = x]\Pr[L_{\text{II}} > 0, L = x]\Big), \quad (20)$$

$$G_{\text{coded}} \leq \big(1 \times p_{\text{uncoded}} + 0 \times p_{\text{coded}}\big)\Pr[L = 1] + \sum_{x=2}^{K+S} \Big(\text{E}[L_{\text{I}}|L_{\text{II}} = 0, L = x]\Pr[L_{\text{II}} = 0, L = x] + \text{E}[L|L_{\text{II}} > 0, L = x]\Pr[L_{\text{II}} > 0, L = x]\Big), \quad (21)$$

where p_{coded} is defined as the probability that non-erased packet is coded under the non-trivial code under condition $L = 1$, and $p_{\text{uncoded}} \triangleq 1 - p_{\text{coded}}$.

Therefore, $G_{\text{coded}} \leq G_{\text{uncoded}}$ if

$$p_{\text{coded}}\Pr[L = 1] \geq \sum_{x=2}^{K+S} \text{E}[L_{\text{II}}|L_{\text{II}} > 0, L = x]\Pr[L_{\text{II}} > 0, L = x] \quad (22)$$
$$= \sum_{x=2}^{K+S} \Bigg(\sum_{y=1}^{\min(x,S)} y\Pr[L_{\text{II}} = y]\Pr[L_{\text{I}} = x - y]\Bigg). \quad (23)$$

Since the non-trivial network code has at least one coded packet in phase II, we have

$$p_{\text{coded}}\Pr[L = 1] \geq \frac{o_{RD}}{1 - o_{RD}} \prod_{j=1}^{K+S} (1 - o_j), \quad (24)$$

where $o_j \triangleq e^{-t_j}$ is the non-erasure probability of slot $j \in \mathcal{K}\cup\mathcal{S}$ at destination D, and $t_j \triangleq \lambda_{jD}/\Gamma$ if $j \in \mathcal{K}$ and $t_j \triangleq \lambda_{RD}/\Gamma$ if $j \in \mathcal{S}$. As a result, (23) holds if

$$\frac{1}{1 - o_{RD}} \prod_{j=1}^{K+S} (1 - o_j) \geq \frac{1}{o_{RD}} \sum_{x=2}^{K+S} \Bigg(\sum_{y=1}^{\min(x,S)} y\Pr[L_{\text{II}} = y]\Pr[L_{\text{I}} = x - y]\Bigg). \quad (25)$$

When Γ approaches zero, the LHS of the above inequality tends to one. We claim that the RHS tends to zero. Note that x represents the total number of non-erasure slots in both phases. It is not difficult to see that when $\Gamma \to 0$, the term when $x = 2$ dominates other terms in the outer summation. When $x = 2$, there are at most two terms left: $\frac{1}{o_{RD}}\Pr[L_{\text{II}} = 1]\Pr[L_{\text{I}} = 1]$ and $\frac{2}{o_{RD}}\Pr[L_{\text{II}} = 2]\Pr[L_{\text{I}} = 0]$. These two terms both go to zero because the probability that having two non-erasure slots is an order of magnitude lower than the probability of a non-erasure slot in the second phase.

Therefore, $G_{\text{coded}} \leq G_{\text{uncoded}}$ when Γ is small enough. The statement of this theorem is then established by invoking Theorem 3. ■

REFERENCES

[1] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc. 2000 IEEE Int. Symp. on Inf. Theory*, pp. 40.

[2] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[3] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, "Wireless broadcast using network coding," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 914–925, Feb. 2009.

[4] P. Fan, C. Zhi, C. Wei, and K. B. Letaief, "Reliable relay assisted wireless multicast using network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 749–762, June 2009.

[5] Y. Liu and C. W. Sung, "A cross-layer design of network coded retransmissions in wireless relay channels," in *Proc. 2011 Wireless Advanced*, pp. 201-206.

[6] H. D. T. Nguyen, L.-N. Tran, and E.-K. Hong, "On transmission efficiency for wireless broadcast using network coding and fountain codes," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 569–571, May 2011.

[7] Y. Chen, S. Kishore, and J. Li, "Wireless diversity through network coding," in *Proc. 2006 IEEE Wireless Commun. and Netw. Conf.*, pp. 1681–1686.

[8] C. Peng, Q. Zhang, M. Zhao, Y. Yao, and W. Jia, "On the performance analysis of network-coded cooperation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3090–3097, Aug. 2008.

[9] H. Topakkaya and Z. Wang, "Wireless network code design and performance analysis using diversity-multiplexing tradeoff," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 488–496, Feb. 2011.

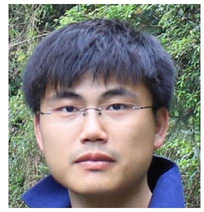
[10] C. Wang, M. Xiao, and M. Skoglund, "Diversity-multiplexing tradeoff analysis of coded multi-user relay networks," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1995–2005, July 2011.

[11] S. B. Wicker and V. K. Bhargava, *Reed-Solomon Codes and their Applications*. IEEE Press, 1994.

[12] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, *et al.*, "A random linear network coding approach to multicust," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[13] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. 2003 Annual Allerton Conf. on Commun., Control, and Comput.*

[14] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.



Mingjun Dai received the B.Eng and M.Phil degrees in control theory and engineering from the Harbin Institute of Technology, Harbin, China, in 2004 and 2006, respectively. He received the Ph.D degree in electronic engineering from City University of Hong Kong in 2012. He joined the faculty at Shenzhen University, Shenzhen, China, in 2012, and is now a lecturer with the College of Information Engineering. His research interests include cooperative relay network, network coding design, and optimization of wireless networks.

Ho Yuet Kwan received the Ph.D. degree in information engineering from the Chinese University of Hong Kong in 2004. His research interests include network coding and communication theory.

Chi Wan Sung (M'98) received the B.Eng, M.Phil, and Ph.D. degrees in information engineering from the Chinese University of Hong Kong in 1993, 1995, and 1998, respectively. He joined the faculty at City University of Hong Kong in 2000, and is now Associate Professor with the Department of Electronic Engineering. He has been an editor of the ETRI Journal since 2009. His research interests include wireless communications, network coding, and algorithms and complexity.