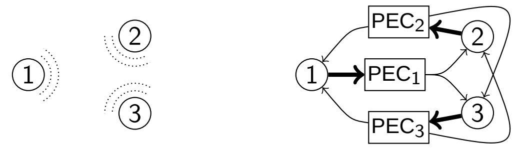
General Capacity Region For The Fully-Connected 3-node Packet Erasure Network

Jaemin Han and Chih-Chun Wang, *Senior Member, IEEE*

***Abstract*—This work studies the capacity region when three nodes $\{1, 2, 3\}$ communicate with each other by sending packets through unreliable wireless medium. For each time slot, with some probabilities a packet sent by node i may be received by both of the other nodes j and k; received only by node j (or node k); or received by neither node. Interference is avoided by enforcing that at most one node can transmit in each time slot. We assume that node i can always reach node j, possibly with the help of the third node k, for any $i \neq j$ pairs (thus the term fully-connected). One notable example of this model is any CSMA-based Wi-Fi network with 3 nodes within the hearing range of each other.**

We consider the most general traffic demands possible in this setting. Namely, there are six private-information flows with rates $(R_{1\to2}, R_{1\to3}, R_{2\to1}, R_{2\to3}, R_{3\to1},\ R_{3\to2})$, respectively, and three common-information flows with rates $(R_{1\to23}, R_{2\to31}, R_{3\to12})$, respectively. We characterize the 9-dimensional Shannon capacity region within a gap that is inversely proportional to the packet size (bits). The gap can be attributed to exchanging reception status (ACK/NACK) and can be further reduced to zero if we allow such feedbacks to be transmitted via a separate control channel. For normal-sized packets, say 12000 bits, our results effectively characterize the capacity region for many important scenarios, e.g., wireless access-point networks with client-to-client cooperative communications, and wireless 2-way relay networks with packet-level coding and processing. Technical contributions of this work include a new converse for many-to-many network communications and a new capacity-approaching scheme based on simple linear network coding operations.

***Index Terms*—Packet Erasure Networks, Packet Erasure Channels, Channel Capacity, Network Coding**



(a) 3 nearby nodes (b) The 3-node Packet Erasure Network

Fig. 1: Illustrations of the 3-node Packet Erasure Network in this work: (a) There are nine co-existing flows possible in general.

I. INTRODUCTION

One of the driving forces that enable high-rate, ubiquitous network communications is the continuous development of Network Information Theory (NIT), which characterizes how much information one can possibly send through a network reliably and thus provides guidance on how to design high-performance (optimal or near optimal) practical network protocols. One notable example in the recent NIT development is the emergence of Linear Network coding (LNC) as a promising technique in modern communication networks. For the single-multicast traffic, it is well known that LNC strictly outperforms non-coding solutions and can achieve the capacity for *error-free networks* [1] and *random erasure networks* [2]. Recent wireless testbeds [3], [4] have also demonstrated that LNC can provide substantial throughput gains over the traditional 802.11 protocols in a practical environment.

Despite the above promising results, our NIT understanding is still nascent for networks with general traffic patterns. When there are only 2 nodes in the network with two co-existing information flows of opposite directions, Shannon [5] provided the first inner and outer bound pair for this simple scenario. The setting of Shannon's work was later generalized under the names of the 3-terminal communication channels [6] and the discrete memoryless network channel [7]. For arbitrary traffic patterns, the simple cut-set outer bound [8, Section 15.10] is often used, which in general is not tight. Despite the continuous development of the cut-set-based bounding techniques, e.g., [9] for the deterministic networks and [10] for general noisy networks, finding the capacity region for networks of general topology and traffic patterns is still an open problem.

There are at least two difficulties when finding the capacity of network communications. Firstly, the information transfer from node A to node B may alter the channel of another transmission. For example, due to the lack of full-duplex hardware, transmission from node B to node A may be completely impossible when node A is sending information to node B. Such a dependence among the point-to-point channels within a network was succinctly characterized by the 2-way model in [5]. Secondly, if there are multiple co-existing flows in a multi-hop network that go in different directions, then each node sometimes has to assume different roles (say, being a sender and/or being a relay) simultaneously. An optimal solution thus needs to balance the roles of each node either through scheduling [11], [12] or through ingenious ways of coding and cooperation [7], [13]. Also see the discussion in [6] for the very detailed case studies for a simple 3-node network. Due to the inherent hardness of the problem, the network capacity region is known only for some scenarios, most of which involve only 1-hop transmissions, say broadcast

This work was supported in part by NSF grants: CCF-0845968, ECCS-1407603, and CCF-1422997. Part of this work was presented in ISIT 2015.

J. Han is with the Next Generation and Standards in Intel, Hillsboro, OR 97124 USA (e-mail: jaemin.han@intel.com).

C.-C. Wang is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: chihw@purdue.edu).

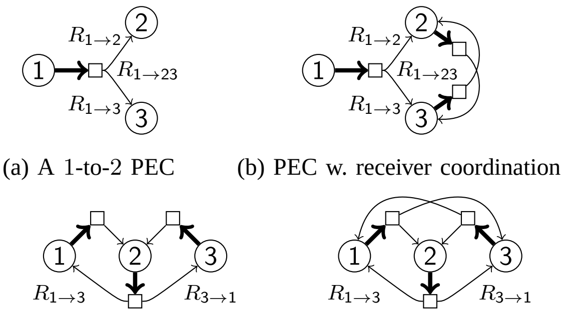
channels or multiple access channels, and/or with all co-existing flows in parallel directions (i.e., flows not forming cycles). As will be seen later, our capacity results consider multi-hop transmission with flows in arbitrary directions.

In this work, we study the 3-node network, Fig. 1(a), with the most general traffic requirements. Namely, there are six co-existing private-information flows with rates $(R_{1\to 2}, R_{1\to 3}, R_{2\to 1}, R_{2\to 3}, R_{3\to 1}, R_{3\to 2})$, respectively, in all possible directions; and there are three co-existing common-information flows with rates $(R_{1\to 23},\ R_{2\to 31}, R_{3\to 12})$, respectively, from a node to the other two nodes. We are interested in characterizing the corresponding 9-dimensional Shannon capacity region. To simplify the analysis, we consider a simple but non-trivial noisy channel model, the random packet erasure network (PEN). That is, each node is associated with its own broadcast packet erasure channel (PEC) such that each node can choose a symbol $X \in \mathbb{F}_q$ from some finite field $\mathbb{F}_q$, transmits the symbol X, and a random subset of the other two nodes will receive the symbol, see Fig. 1(b). The symbol X is sometimes called a packet of size $\log_2(q)$ bits. We assume time-sharing among all three nodes so that interference is fully avoided. In this way, we can concentrate on the topological effects and the broadcast-channel diversity gain within the network.

Specifically, we consider one of the following two scenarios. **Scenario 1**: Motivated by the throughput benefit of the causal packet ACKnowledgment feedback for erasure networks [11], [12], [14]–[24], in this scenario we assume that the reception status is causally available to the entire network after each packet transmission through a separate control channel for free. Such assumption can be justified by the fact that the length of ACK/NACK is 1 bit, much smaller than the size of a regular packet.

Scenario 2: In this scenario we assume that there is no inherent feedback mechanism. Any ACK/NACK signal, if there is any, has to be sent through the regular forward channels along with information messages. As a result, any achievability scheme needs to balance the amount of information and control messages. For example, suppose a particular coding scheme chooses to divide the transmitted packet X into the header and the payload. Then it needs to carefully decide what the content of the control information would be and how many bits the header should have to accommodate the control information. The timeliness of delivering the control messages is also critical since the control information, sent through the forward erasure channel, may get lost as well. Therefore, some critical control information may not arrive in time. Such a setting in Scenario 2 is much closer to practice as it considers the complexity/delay overhead of the coding solution. In Scenario 2, we also assume that the 3-node PEN is *fully-connected*, i.e, node i can always reach node j, possibly with the help of the third node k, for any $i \neq j$ pairs. The formal definition of fully-connectedness is provided in Definition 2 of Section III-B. Note that *the fully-connectedness is assumed only in Scenario 2*. When the causal reception status is available for free (Scenario 1), our results do not need the fully-connectedness assumption.

The main contributions can be summarized as follows. We first characterize the exact 9-dimensional Shannon capacity region for Scenario 1 when the causal ACK/NACK feedbacks are immediately available for free through a separate control channel. For the more practical setting of Scenario 2 where the control messages have to be sent through the forward erasure channel, the capacity for the fully-connected 3-node PEN is then characterized with a gap inversely proportional to $\log_2(q)$. This gap is due to the need of exchanging the reception status (ACK/NACK) within the network. The technical contributions of this work include a new converse for many-to-many network communications and a new capacity-approaching scheme based on simple LNC operations.



(a) A 1-to-2 PEC (b) PEC w. receiver coordination

(c) A Two-way relay PEC (d) A Two-way relay PEC w. opportunistic routing

Fig. 2: Special examples of the 3-node Packet Erasure Network (PEN) considered in this work. The rectangle implies the broadcast packet erasure channel (PEC).

It is worth noting that the considered 3-node PEN contains many important practical and theoretically interesting scenarios as sub-cases. **Example 1:** If we set the broadcast PECs of nodes 2 and 3 to be always erasure (i.e., neither nodes can transmit anything), then Fig. 1(b) collapses to Fig. 2(a), the 2-receiver broadcast PEC scenario. The capacity region $(R_{1\to 2}, R_{1\to 3}, R_{1\to 23})$ derived in our Scenario 1 is identical to the existing results in [14], [18]. **Example 2:** Instead of setting the PECs of nodes 2 and 3 to all erasure, we set $R_{2\to 1}$, $R_{2\to 3}$, $R_{3\to 1}$, $R_{3\to 2}$, $R_{2\to 31}$, $R_{3\to 12}$ to be zeros. Namely, we still allow nodes 2 and 3 to transmit but there is no information message emanating from nodes 2 and 3. In this case, node 2 can potentially be a relay that helps forwarding those node-1 packets destined for node 3 and node 3 can be a relay for flow $1 \to 2$, see Fig. 2(b). This work then characterizes the Shannon capacity[1]$(R_{1\to 2}, R_{1\to 3}, R_{1\to 23})$ of a broadcast PEC with receiver coordination.

Example 3: If we set $R_{1\to 2}$, $R_{2\to 1}$, $R_{2\to 3}$, $R_{3\to 2}$, $R_{1\to 23}$, $R_{2\to 31}$, $R_{3\to 12}$ to be zeros and prohibit any direct communication between nodes 1 and 3, Fig. 1(b) now collapses to Fig. 2(c), in which node 2 is a two-way relay for unicast flows $1 \to 3$ and $3 \to 1$. The results in this work thus characterizes the Shannon capacity region $(R_{1\to 3}, R_{3\to 1})$ of this two-way relay network Fig. 2(c), which is identical to the existing result in [25]. **Example 4:** If we additionally allow direct communication between nodes 1 and 3, Fig. 1(b)

[1]In [11], the LNC capacity of Fig. 2(b) was characterized, but the most general Shannon capacity region was unknown in [11].

now collapses to Fig. 2(d). Namely, when node 1 is sending packets to the relay node 2, the packets might be overheard directly by the destination node 3. If indeed node 3 overhears the communication, then node 1 could inform node 2 opportunistically that there is no need to forward that packet to node 3 anymore. Such a scheme is called *opportunistic routing* and testbed implementation [4] has shown that opportunistic routing can potentially improve the throughput by 20x. The results in this work thus characterize the Shannon capacity region $(R_{1\to 3}, R_{3\to 1})$ of Fig. 2(d), which allows for the possibility of both opportunistic routing and two-way-relay coding. The Shannon capacity region computed by this work again matches the existing result in [12].

In summary, most existing works on packet erasure networks have studied either ≤ 2 co-existing flows [3], [4], [11], [12], [14], [19] or all flows originating from the same node [11], [16], [20], [21], [23], [26], [27]. By characterizing the most general 9-dimensional Shannon capacity region with arbitrary flow directions, this work significantly improves our understanding for communications over the 3-node PEN.

The rest of the paper is organized as follows. Section II formulates the problem. The main results of this work are the general 9-dimensional Shannon capacity region and the corresponding capacity-approaching simple LNC scheme, which are presented in Section III. Section IV applies our results to some important practical scenarios, numerically evaluates the capacity region, and compares them with some existing suboptimal solutions. Section V provides the detailed intuition and a new converse proof for the Shannon capacity outer bound. The details of our simple LNC achievability scheme are provided in Section VI. Finally, Section VII discusses some interference models and Section VIII concludes the paper.

II. Problem Formulation

A. Broadcast Packet Erasure Channels

For any positive integer K, a 1-to-K broadcast packet erasure channel (PEC) is defined as to take an input X from a finite field $\mathbb{F}_q$ with size $q > 0$ and output a K-dimensional vector $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_K)$. We assume that the input is either received perfectly or completely erased, i.e., each output Y_k must be either the input X or an erasure symbol ε, where $Y_k = \varepsilon$ means that the k-th receiver does not correctly receive the input X. As a result, the reception status can be described by a K-dimensional binary vector $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_K)$ where $Z_k = 1$ and ε represents whether the k-th receiver successfully received the input X or not, respectively. Any given PEC can then be described by its distribution of the binary reception status $\mathbf{Z}$.

B. Memoryless 3-node Packet Erasure Network

Consider a network of three nearby nodes labeled as $\{1, 2, 3\}$, see Fig. 1(a). For the ease of exposition, we will use (i, j, k) to represent one of three cyclically shifted tuples of node indices $\{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}$. The 3-node Packet Erasure Network (PEN) is then defined as the collection of three separate 1-to-2 broadcast PECs, each from node i to the other two nodes j and k for all $i \in \{1, 2, 3\}$, see Fig. 1(b).

The channel behaviors of the 3-node PEN can be described by the following definitions. For any time slot t, we use a 6-dimensional *channel output vector* $\mathbf{Z}(t)$ to represent the reception status of the entire network:

$$\begin{aligned}\mathbf{Z}(t) = (Z_{1\to 2}(t), Z_{1\to 3}(t), &Z_{2\to 1}(t), Z_{2\to 3}(t),\\ &Z_{3\to 1}(t), Z_{3\to 2}(t)) \in \{1, \varepsilon\}^6,\end{aligned}$$

where $Z_{i\to h}(t) = 1$ and ε represents whether node h can receive the transmission from node i or not, respectively. We assume that the 3-node PEN is memoryless and stationary,[2] i.e., we allow arbitrary joint distribution for the 6 coordinates of $\mathbf{Z}(t)$ but assume that $\mathbf{Z}(t_1)$ and $\mathbf{Z}(t_2)$ are independently and identically distributed for any $t_1 \neq t_2$. We use $p_{i\to jk} \triangleq \mathsf{Prob}(Z_{i\to j}(t) = 1, Z_{i\to k}(t) = 1)$ to denote the probability that the packet transmitted from node i is successfully received by both nodes j and k; and use $p_{i\to j\overline{k}}$ to denote the probability $\mathsf{Prob}(Z_{i\to j}(t) = 1, Z_{i\to k}(t) = \varepsilon)$ that node-i packet is successfully received by node j but not by node k. Probability $p_{i\to \overline{j}k}$ is defined symmetrically. Define $p_{i\to j\vee k} \triangleq p_{i\to \overline{j}k} + p_{i\to jk} + p_{i\to j\overline{k}}$ as the probability that at least one of nodes j and k receives the packet, and define $p_{i\to j} \triangleq p_{i\to jk} + p_{i\to j\overline{k}}$ (resp. $p_{i\to k} \triangleq p_{i\to jk} + p_{i\to \overline{j}k}$) as the marginal reception probability from node i to node j (resp. node k). We also assume that the random process $\{\mathbf{Z}(t) : \forall t\}$ is independent of any information messages.

Assume synchronized time-slotted transmissions. To model interference, we assume that only one node can successfully transmit at each time slot $t \in \{1, \cdots, n\}$. If two or more nodes happen to transmit in the same time slot, then both transmissions will fail. More specifically, we define the following *scheduling decision* binary variable $\sigma_i(t)$ for any node $i \in \{1, 2, 3\}$. Namely, $\sigma_i(t) = 1$ represents that node i decides to transmit at time t and $\sigma_i(t) = 0$ represents not transmitting. Any transmission is completely destroyed if there are two or more nodes transmitting simultaneously. For example, suppose node i decides to transmit a packet $X_i(t) \in \mathbb{F}_q$ in time t (thus $\sigma_i(t) = 1$). Then, only when $\sigma_j(t) = \sigma_k(t) = 0$ can node i transmit without any interference. Moreover, only when $Z_{i\to h}(t) = 1$ will node $h \neq i$ receive $Y_{i\to h}(t) = X_i(t)$. In all other cases, node h receives an erasure $Y_{i\to h}(t) = \varepsilon$. We summarize this interference and erasure model by the following definition.

$$Y_{i\to h}(t) = \begin{cases} X_i(t) & \text{if } \sigma_i(t) = 1, \sigma_j(t) = \sigma_k(t) = 0, \\ & \quad \text{and } Z_{i\to h}(t) = 1 \\ \varepsilon & \quad \text{otherwise} \end{cases} \tag{1}$$

C. Joint Scheduling and Network Coding Scheme Under Scenario 2

Over the 3-node PEN described above, we consider the following 9-dimensional traffic flows: 6 private-information flows with rates $(R_{1\to 2}, R_{1\to 3}, R_{2\to 1}, R_{2\to 3}, R_{3\to 1}, R_{3\to 2})$, respectively; and 3 common-information flows with rates $(R_{1\to 23}, R_{2\to 31}, R_{3\to 12})$, respectively. Namely, $R_{1\to 23}$ represents the rate of the common-information message from node

[2] The 3-node PEN is a special case of the discrete memoryless network channel [7].

1 to both nodes 2 and 3. We use $\vec{R}_{i*} \triangleq (R_{i\to j}, R_{i\to k}, R_{i\to jk})$ to denote the rates of all three 3 flows originated from node i, for all $i \in \{1,2,3\}$. We use a 9-dimensional rate vector $\vec{R} \triangleq (\vec{R}_{1*}, \vec{R}_{2*}, \vec{R}_{3*})$ to denote the rates of all possible flow directions.

Within a total budget of n time slots, node i would like to send $nR_{i\to h}$ packets (private-information messages), denoted by a row vector $\mathbf{W}_{i\to h}$, to node $h \neq i$, and would like to send $nR_{i\to jk}$ packets (common-information messages), denoted by a row vector $\mathbf{W}_{i\to jk}$, to the other two nodes simultaneously. Namely, the unit of the rate vector $\vec{R}$ is packets per time slot, where each information message packet has $\log_2(q)$ bits and is chosen independently and uniformly randomly from a finite field $\mathbb{F}_q$ with size $q > 1$.

For the ease of exposition, we define $\mathbf{W}_{i*} \triangleq \mathbf{W}_{i\to j} \cup \mathbf{W}_{i\to k} \cup \mathbf{W}_{i\to jk}$ as the collection of all messages originated from node i. Similarly, we define $\mathbf{W}_{*i} \triangleq \mathbf{W}_{j\to i} \cup \mathbf{W}_{j\to ki} \cup \mathbf{W}_{k\to i} \cup \mathbf{W}_{k\to ij}$ as the collection of all messages destined to node i. Sometimes we slightly abuse the above notation and define $\mathbf{W}_{\{i,j\}*} \triangleq \mathbf{W}_{i*} \cup \mathbf{W}_{j*}$ as the collection of messages originated from either node i or node j. Similar "collection-based" notation can also be applied to the received symbols and we can thus define $\mathbf{Y}_{*i}(t) \triangleq \{Y_{j\to i}(t), Y_{k\to i}(t)\}$ and $\mathbf{Y}_{i*}(t) \triangleq \{Y_{i\to j}(t), Y_{i\to k}(t)\}$ as the collection of all symbols received and transmitted by node i during time t, respectively. For simplicity, we also use brackets $[\cdot]_1^t$ to denote the collection from time 1 to t. For example, $[\mathbf{Y}_{*i}, \mathbf{Z}]_1^{t-1}$ is shorthand for the collection $\{Y_{j\to i}(\tau), Y_{k\to i}(\tau), \mathbf{Z}(\tau) : \forall \tau \in \{1, \cdots, t-1\}\}$.

Recall that two scenarios were discussed in Section I. That is, causal ACK/NACK feedback can be transmitted for free in Scenario 1 but has to go through the forward channel when in Scenario 2. We first focus on the detailed formulation under Scenario 2.

Given the rate vector $\vec{R}$, a joint scheduling and network coding scheme is described by $3n$ binary scheduling functions: $\forall\, t \in \{1, \cdots, n\}$ and $\forall\, i \in \{1,2,3\}$,

$$\sigma_i(t) = f^{(t)}_{\mathsf{SCH},\, i}([\mathbf{Y}_{*i}]_1^{t-1}) \tag{2}$$

plus $3n$ encoding functions: $\forall\, t \in \{1, \cdots, n\}$ and $\forall\, i \in \{1,2,3\}$,

$$X_i(t) = f_i^{(t)}(\mathbf{W}_{i*}, [\mathbf{Y}_{*i}]_1^{t-1}), \tag{3}$$

plus 3 decoding functions: $\forall\, i \in \{1,2,3\}$,

$$\hat{\mathbf{W}}_{*i} = g_i(\mathbf{W}_{i*}, [\mathbf{Y}_{*i}]_1^n). \tag{4}$$

To refrain from using the timing-channel[3] techniques [28], we also require the following equality

$$I([\sigma_1, \sigma_2, \sigma_3]_1^n\,;\, \mathbf{W}_{\{1,2,3\}*}) = 0, \tag{5}$$

where $I(\cdot\,;\,\cdot)$ is the mutual information and $\mathbf{W}_{\{1,2,3\}*} \triangleq \mathbf{W}_{1*} \cup \mathbf{W}_{2*} \cup \mathbf{W}_{3*}$ is all the 9-flow information messages as defined earlier.

Intuitively, at every time t, each node decides whether to transmit or not based on what it has received in the past, see (2). Note that the received symbols $[\mathbf{Y}_{*i}]_1^{t-1}$ may contain both the message information and the control information. (5) ensures that the "timing" of the transmission $\sigma_i(t)$ cannot be used to carry[4] the message information. Once each node decides whether to transmit or not, it encodes $X_i(t)$ based on its information messages and what it has received from other nodes in the past, see (3). In the end of time n, each node decodes its desired packets based on its information messages and what it has received, see (4).

We can now define the capacity region.

Definition 1. Fix the distribution of $\mathbf{Z}(t)$ and finite field $\mathbb{F}_q$. A 9-dimensional rate vector $\vec{R}$ is achievable if for any $\epsilon > 0$ there exists a joint scheduling and network code scheme with sufficiently large n such that $\mathsf{Prob}(\hat{\mathbf{W}}_{*i} \neq \mathbf{W}_{*i}) < \epsilon$ for all $i \in \{1,2,3\}$. The capacity region is the closure of all achievable $\vec{R}$.

D. Comparison between Scenarios 1 and 2

The previous formulation focuses on Scenario 2. The difference between Scenarios 1 and 2 is that the former allows the use of causal ACK/NACK feedbacks for free. As a result, for Scenario 1, we simply need to insert the *causal* network-wide channel status information $[\mathbf{Z}]_1^{t-1}$ in the input arguments of (2) and (3), respectively; and insert the *overall* network-wide channels status information $[\mathbf{Z}]_1^n$ in the input argument of (4). The formulation of Scenario 1 thus becomes as follows: $\forall\, t \in \{1, \cdots, n\}$ and $\forall\, i \in \{1,2,3\}$,

$$\sigma_i(t) = \overline{f}^{(t)}_{\mathsf{SCH},\, i}([\mathbf{Y}_{*i}, \mathbf{Z}]_1^{t-1}), \tag{6}$$

$$X_i(t) = \overline{f}_i^{(t)}(\mathbf{W}_{i*}, [\mathbf{Y}_{*i}, \mathbf{Z}]_1^{t-1}), \tag{7}$$

$$\hat{\mathbf{W}}_{*i} = \overline{g}_i(\mathbf{W}_{i*}, [\mathbf{Y}_{*i}, \mathbf{Z}]_1^n), \tag{8}$$

while we still impose no-timing channel information (5). Obviously, with more information to use, the capacity region under Scenario 1 is a superset of that of Scenario 2, which is why we use overlines in the above function descriptions. Following this observation, we will outer bound the (larger) capacity of Scenario 1 and inner bound the (smaller) capacity of Scenario 2 in the subsequent sections.

Without loss of generality, we can further replace the distributed scheduling computation in (6) (each node i computes its own scheduling) by the following centralized scheduling function

$$\sigma(t) = \overline{f}^{(t)}_{\mathsf{SCH}}([\mathbf{Z}]_1^{t-1}) \in \{1,2,3\}, \tag{9}$$

[3]We believe that the use of timing channel techniques will not alter the capacity region much when the packet size is large. One justification is that the rate of the timing channel is at most 3 bits per slot, which is negligible compared to a normal packet size of 12000 bits.

[4]For example, one (not necessarily optimal) way to encode is to divide a packet $X_i(t)$ into the header and the payload. The messages $\mathbf{W}_{i*}$ will be embedded in the payload while the header contains control information such as ACK. If this is indeed the way we encode, then (5) requires that transmit decision depend only on the control information in the header, not the messages in the payload. Note that the control information does not necessarily need to be ACK. For example, a scheme may choose to transmit the current queue lengths instead of ACK and uses only the queue lengths to decide whether a node should transmit or not. If that is the case, the scheme will then put the queue lengths of other nodes in the header of the packets sent to node i, those $[\mathbf{Y}_{*i}]_1^{t-1}$ packets. Node i will decide whether to transmit or not based only on the queue length information it receives. Since the evolution of the queue lengths are independent from the message symbols, the mutual information condition (5) will hold naturally.

that takes the values in the set of three nodes $\{1,2,3\}$. That is, $\sigma(t)=i$ implies that only node i is scheduled to transmit in time t.

To prove why we can replace (6) by (9) without loss of generality, we first introduce the following lemma.

Lemma 1. Without loss of generality, we can replace (6) by the following form:

$$\sigma_i(t)=\overline{f}^{(t)}_{\mathsf{SCH},\,i}([\mathbf{Z}]_1^{t-1}), \tag{10}$$

which is still a binary scheduling function but the input argument $[\mathbf{Y}_{*i}]_1^{t-1}$ in (6) is removed.

The proof of Lemma 1 is relegated to Appendix A. The intuition behind the proof is to show that since the information equality (5) must hold, knowing the past reception status $[\mathbf{Z}]_1^{t-1}$ is sufficient for the scheduling purpose.

Lemma 1 ensures that we can replace the scheduling decision (6) of each individual node i by (10). We then observe that every node i makes its scheduling decision based on the same input argument $[\mathbf{Z}]_1^{t-1}$, which, in Scenario 1, is available to all three nodes for free via a separate control channel. Therefore, it is as if there is a centralized scheduler in Scenario 1 and the centralized scheduler will never induce any scheduling conflict. As a result, we can further replace the individual scheduler (10) by a centralized global scheduling function (9) where $\sigma(t)=i$ implies that node i is the only scheduled node in time t.

In summary, under Scenario 1, the joint network coding and scheduling solution is described by (7), (8), and (9). Here we do not impose (5) anymore since the centralized scheduler (9) satisfies (5) naturally.

III. Main Results

The main results can be summarized as follows. We will first provide the capacity outer bound of the 3-node PEN of Scenario 1 in Section III-A. The capacity-achieving LNC scheme of Scenario 1 and the similar capacity-approaching inner bound of Scenario 2 are provided in Section III-B.

A. Capacity outer bound of 3-node Packet Erasure Network

Proposition 1. For any fixed $\mathbb{F}_q$, a 9-dimensional $\vec{R}$ is achievable under Scenario 1 only if there exist 3 non-negative variables $s^{(i)}$ for all $i\in\{1,2,3\}$ such that jointly they satisfy the following three groups of linear conditions:

• Group 1, termed the *time-sharing condition*, has 1 inequality:

$$\sum_{\forall\, i\in\{1,2,3\}} s^{(i)}\leq 1. \tag{11}$$

• Group 2, termed the *broadcast cut-set condition*, has 3 inequalities: For all $i\in\{1,2,3\}$,

$$R_{i\to j}+R_{i\to k}+R_{i\to jk}\leq s^{(i)}\cdot p_{i\to j\vee k}. \tag{12}$$

• Group 3, termed the 3-*way multiple-access cut-set condition*, has 3 inequalities: For all $i\in\{1,2,3\}$,

$$R_{j\to i}+R_{j\to ki}+R_{k\to i}+R_{k\to ij}\leq s^{(j)}\cdot p_{j\to i}+s^{(k)}\cdot p_{k\to i} \\ -\left(\frac{p_{j\to i}}{p_{j\to k\vee i}}R_{j\to k}+\frac{p_{k\to i}}{p_{k\to i\vee j}}R_{k\to j}\right). \tag{13}$$

Proposition 1 considers arbitrary, possibly non-linear ways of designing the encoding/decoding and scheduling functions in (7), (8), and (9), and is derived by entropy-based analysis. Proposition 1 can also be viewed as strict generalization of the results of the simpler settings [16], [20].

The brief intuitions behind (11) to (13) are as follows. Each variable $s^{(i)}$ counts the expected frequency (normalized over the time budget n) that node i is scheduled for successful transmissions. As a result, (11) holds naturally. (12) is a simple cut-set condition for broadcasting from node i. One main contribution of this work is the derivation of the new 3-way multiple-access outer bound in (13). The LHS of (13) contains all the information destined for node i. The term $s^{(j)}p_{j\to i}+s^{(k)}p_{k\to i}$ on the RHS of (13) is the amount of time slots that either node j or node k can communicate with node i. As a result, it resembles a multiple-access cut condition of a typical cut-set argument [8, Section 15.10]. What is special in our setting is that, since node j may have some private-information for node k and vice versa, sending those private-information has a penalty on the multiple access channel from nodes $\{j,k\}$ to node i. The last term on the RHS of (13) quantifies such penalty that is inevitable regardless of what kind of coding schemes being used. The proof of Proposition 1 and the detailed discussions are relegated to Section V.

Remark: In addition to having a new penalty term on the RHS of (13), the 3-way multiple-access cut-set condition (13) is surprising, not because it upper bounds the *combined information-flow rate* from nodes $\{j,k\}$ entering node i but because, unlike the traditional multiple-access upper bounds, we do not need to upper bound the individual rate from node j (resp. k) to node i.

More specifically, a traditional multi-access channel capacity result will also upper bound the rate $R_{j\to i}+R_{j\to ki}$ by considering the cut from node j to node i (ignoring node k completely). If we follow the above logic and write down naively the "cut condition" from node j to i, then we will have

$$R_{j\to i}+R_{j\to ki}\leq s^{(j)}\cdot p_{j\to i}-\frac{p_{j\to i}}{p_{j\to k\vee i}}R_{j\to k}. \tag{14}$$

where $R_{j\to i}+R_{j\to ki}$ is the rate from nodes j to i, $s^{(j)}\cdot p_{j\to i}$ is the successful time slots, and $\frac{p_{j\to i}}{p_{j\to k\vee i}}R_{j\to k}$ is the penalty term. One might expect that (14) is also a legitimate outer bound. It turns out that (14) is not an outer bound and one can find some LNC solution that contradicts (14).

The reason why (14) is false is as follows. The $\mathbf{W}_{j\to i}$ packets may not necessarily go directly from node j to node i and it is possible that node k can also help relay those packets. As a result, how frequently node k is scheduled can also affect the number of $\mathbf{W}_{j\to i}$ packets that one can hope to deliver from node j to node i. Since (14) does not involve $s^{(k)}$, it does not consider the possibility of node k relaying the packets for node j. In contrast, our outer bound (13) indeed captures such a subtle but critical phenomenon by grouping all $R_{j\to i}$, $R_{k\to i}$, $R_{j\to ki}$, $R_{k\to ij}$, $R_{j\to k}$, and $R_{k\to j}$ as a whole and upper bounds it with the (weighted) sum of scheduling frequencies of nodes j and k.

B. A Capacity Approaching LNC Scheme

Scenario 2 requires the network to be fully-connected, which is defined as follows.

Definition 2. In Scenario 2, we assume the 3-node PEN is *fully-connected* in the sense that the given channel reception probabilities satisfy either $p_{i_1\to i_2} > 0$ or $\min(p_{i_1\to i_3}, p_{i_3\to i_2}) > 0$ for all distinct $i_1, i_2, i_3 \in \{1,2,3\}$.

Namely, node i_1 must be able to communicate with node i_2 either through the direct communication (i.e., $p_{i_1\to i_2} > 0$) or through relaying (i.e., $\min(p_{i_1\to i_3}, p_{i_3\to i_2}) > 0$). Note that in Scenario 2, the control messages have to be sent through the regular forward channel as well. The fully-connectedness assumption guarantees that feedback/control information can be sent successfully from one node to any other node, either directly or through the help of another node.

We also need the following new math operator.

Definition 3. For any 2 non-negative values a and b, the operator $\mathsf{nzmin}\{a,b\}$, standing for non-zero minimum, is defined as:

$$\mathsf{nzmin}\{a,b\} = \begin{cases} \max(a,b) & \text{if } \min(a,b)=0, \\ \min(a,b) & \text{if } \min(a,b)\neq 0. \end{cases}$$

Intuitively, $\mathsf{nzmin}\{a,b\}$ is the minimum of the strictly positive entries.

Proposition 2. For any fixed $\mathbb{F}_q$, a 9-dimensional $\vec{R}$ is LNC-achievable in Scenario 2 if there exist 15 non-negative variables $t^{(i)}_{[\mathrm{u}]}$ and $\{t^{(i)}_{[\mathrm{c},l]}\}_{l=1}^{4}$ for all $i \in \{1,2,3\}$ such that jointly they satisfy the following three groups of linear conditions:

- Group 1, termed the *time-sharing condition*, has 1 inequality:

$$\sum_{\forall i\in\{1,2,3\}} t^{(i)}_{[\mathrm{u}]} + t^{(i)}_{[\mathrm{c},1]} + t^{(i)}_{[\mathrm{c},2]} + t^{(i)}_{[\mathrm{c},3]} + t^{(i)}_{[\mathrm{c},4]} \leq 1 - t_{\mathsf{FB}}, \quad (15)$$

where t_{FB} is a constant defined as

$$t_{\mathsf{FB}} \triangleq \sum_{\forall i\in\{1,2,3\}} \frac{3}{\log_2(q)\cdot \mathsf{nzmin}\{p_{i\to j}, p_{i\to k}\}}. \quad (16)$$

- Group 2 has 3 inequalities: For all $(i,j,k) \in \{(1,2,3), (2,3,1), (3,1,2)\}$,

$$R_{i\to j} + R_{i\to k} + R_{i\to jk} < t^{(i)}_{[\mathrm{u}]}\cdot p_{i\to j\vee k}. \quad (17)$$

- Group 3 has 6 inequalities: For all $(i,j,k) \in \{(1,2,3), (2,3,1), (3,1,2)\}$,

$$\Big(R_{i\to j} + R_{i\to jk}\Big)\frac{p_{i\to \overline{j}k}}{p_{i\to j\vee k}} < \Big(t^{(i)}_{[\mathrm{c},1]} + t^{(i)}_{[\mathrm{c},3]}\Big)\cdot p_{i\to j} + \Big(t^{(k)}_{[\mathrm{c},2]} + t^{(k)}_{[\mathrm{c},3]}\Big)\cdot p_{k\to j}, \quad (18)$$

$$\Big(R_{i\to k} + R_{i\to jk}\Big)\frac{p_{i\to j\overline{k}}}{p_{i\to j\vee k}} < \Big(t^{(i)}_{[\mathrm{c},1]} + t^{(i)}_{[\mathrm{c},4]}\Big)\cdot p_{i\to k} + \Big(t^{(j)}_{[\mathrm{c},2]} + t^{(j)}_{[\mathrm{c},4]}\Big)\cdot p_{j\to k}. \quad (19)$$

Proposition 3. Continue from Proposition 2. If we focus on Scenario 1 instead, then the rate vector $\vec{R}$ is LNC-achievable if there exist 15 non-negative variables $t^{(i)}_{[\mathrm{u}]}$ and $\{t^{(i)}_{[\mathrm{c},l]}\}_{l=1}^{4}$ for all $i \in \{1,2,3\}$ such that (15), (17) to (19) hold while we set $t_{\mathsf{FB}} = 0$ in (16).

In short, the constant term t_{FB} in (16) quantifies the overhead of sending the ACK/NACK feedbacks through the forward erasure channel in Scenario 2 and can be set to 0 in Scenario 1.

Since both the outer bound and the achievable regions can be computed by an LP solver, one can numerically verify that for all possible channel parameters, the rate regions of Propositions 1 and 3 of Scenario 1 always match. We can actually prove this observation by analyzing the underlying linear algebraic structures of the two LP problems.

Proposition 4. The outer bound in Proposition 1 and the closure of the achievable region in Proposition 3 match for all possible channel parameters $\{p_{i\to jk}, p_{i\to \overline{j}k}, p_{i\to j\overline{k}} : \forall (i,j,k)\}$. They thus describe the corresponding 9-dimensional Shannon capacity region under Scenario 1.

From the above discussions, one can see that even for the more practical Scenario 2, in which there is no dedicated feedback control channels, Proposition 2 is indeed capacity-approaching when the 3-node PEN is fully-connected. The gap to the outer bound is inversely proportional to $\log_2(q)$ and diminishes to zero if the packet size $\log_2(q)$ (bits) is large enough. In real life, the actual payload of each packet is roughly 10^4 bits and the gap is thus negligible unless the reception probabilities $p_{i\to j}$ or $p_{i\to k}$ is extremely small.

The proof of Proposition 3, i.e., an achievability scheme for the simpler case of publicly available feedback (Scenario 1), is provided in Section VI. When causal feedback is not freely available (Scenario 2), Proposition 2 needs a scheme that handles when and how to send the control information through the forward erasure channel. Such a scheme is provided and analyzed in Appendix B. This scheme can be viewed as a strict generalization for the simpler scheme in Section VI. The proof of Proposition 4 is relegated to Appendix E.

C. Comments On The Finite Size And The Fully-Connectedness Assumption

We note that when the finite field size $\mathbb{F}_q$ is small, the gap between the outer bound in Proposition 1 and the inner bound in Proposition 2 can still be substantial. We believe that in the small $\mathbb{F}_q$ regime, the inner bound can be further improved. The reason is that, as will be seen in Appendix B, the bound in Proposition 2 is obtained by analyzing a scheme that transmits the feedback information in a very crude way. Namely, it first converts the feedback information into a binary vector and then send the entire vector to all nodes. This provides more than enough information for each node. Although the penalty of such scheme is negligible when q is large, a better design could significantly reduce the amount of information necessary for small q value. For example, when q is extremely small, say $q = 2$, then each packet contains only $\log_2(q) = 1$ bit. Then after sending each packet (each bit), our scheme will then spend a roughly equal amount of time slots to send feedback (also 1 bit) for the 1-bit packet. In the end, roughly half of its resources is on sending feedback information, which is clearly suboptimal since even without feedback we can achieve a substantial throughput rate simply by using traditional MDS codes. The (near-) optimal code design for small q is beyond the scope of this work.

We close this section by discussing some degenerate scenarios and the related fully-connectedness assumption. We first consider Scenario 1, which does not require the fully-connected assumption. It is possible that in Scenario 1, we have $p_{i\to j\vee k}=0$ for some (i,j,k), which implies that (18) and (19) being undefined. However, when $p_{i\to j\vee k}=0$, it is simply impossible to send any messages out of node i. As a result, we can replace the (undefined) (18) and (19) by a hard condition $R_{i\to j}=R_{i\to k}=R_{i\to jk}=0$. Proposition 4 still holds after such a simple revision.

We now consider Scenario 2. We note that Proposition 2 holds only when the network is fully-connected. Actually, when the network is not fully-connected, the denominator of (16) may be zero and (16) becomes undefined. When the network is *not* fully-connected, it is an interesting open problem what the actual capacity region is going to be. Specifically, the outer bound (Proposition 1) still holds even when the network is not fully-connected. However, there are reasons to believe that the outer bound is not tight anymore. For example, suppose $p_{2\to 3\vee 1}=0$, i.e., the PEC from node 2 is completely erasure, there is no dedicated control channel, and any feedback has to be sent through the forward channel, i.e., Scenario 2 but being not fully-connected. In this example, node 2 is completely “in the dark”. Note that being in the dark does not mean that we cannot send messages to node 2. For example, we can use an MDS code to send messages from nodes 1 to node 2. When the MDS code rate is slightly lower than the success probability $p_{1\to 2}$, then node 2 can receive the correct messages with high probability without sending any ACK. However, when node 2 is in the dark, neither node 1 nor node 3 can be made aware of the reception status of node 2. Therefore, the classic network coding techniques in [16] do not apply in this scenario. How to characterize the Shannon capacity region when some node is in the dark is beyond the scope of this work and will be actively investigated in the future.

Remark: The above “asymmetric” feedback scenario is theoretically interesting. In practice, the PEC is usually used to model network communications, for which ACK is often required for any transmission and also necessary for the purpose of channel estimation. Therefore, if $p_{2\to 3\vee 1}=0$ and node 2 is in the dark, then nodes 1 and 3 will give up communicating to node 2 immediately due to the lack of any ACK feedback. The aforementioned MDS code approach will not be used when node 2 cannot acknowledge the transmission in any way.

IV. Special Examples and Numerical Evaluation

In the following, we apply Propositions 1 and 3 to the four special examples discussed in Section I. We also numerically evaluate the 9-dimensional capacity region for some specific channel parameter values.

A. Example 1: The Simplest 1-to-2 Broadcast PEC

Consider the simplest setting of a 1-to-2 broadcast PEC with 2 private-information flows of rates $R_{1\to 2}$ and $R_{1\to 3}$, and 1 common-information flow of rate $R_{1\to 23}$. See Fig. 2(a) for illustration. In this scenario, we assume that only node 1 can transmit and nodes 2 and 3 can only listen and send ACK/NACK feedback after each packet transmission. This simple 1-to-2 broadcast PEC can be viewed as a special example of the general problem by setting $p_{2\to 3\vee 1}=p_{3\to 1\vee 2}=0$, and by hardwiring the unused rates $\{R_{2\to 1},R_{2\to 3},R_{3\to 1},R_{3\to 2}\}$ and $\{R_{2\to 31},R_{3\to 12}\}$ to zeros. One can thus use Proposition 1 to compute the 3-dimensional capacity region $(R_{1\to 2},R_{1\to 3},R_{1\to 23})$ of the 1-to-2 broadcast PEC. More explicitly, by setting $s^{(1)}=1$ and $s^{(2)}=s^{(3)}=0$, (13) with $i=2$ leads to the following (20) and (13) with $i=3$ leads to the following (21):

$$R_{1\to 2}+R_{1\to 23}\leq p_{1\to 2}-\frac{p_{1\to 2}}{p_{1\to 2\vee 3}}R_{1\to 3}, \tag{20}$$

$$R_{1\to 3}+R_{1\to 23}\leq p_{1\to 3}-\frac{p_{1\to 3}}{p_{1\to 2\vee 3}}R_{1\to 2}. \tag{21}$$

As expected, the capacity region $(R_{1\to 2},R_{1\to 3},R_{1\to 23})$ described by (20) and (21) is identical to the existing 1-to-2 broadcast PEC capacity results in [14].

B. Example 2: The 1-to-2 Broadcast PEC With Receiver Coordination

Another special example is the 1-to-2 broadcast PEC with receiver coordination, see Fig. 2(b). In this scenario, node 1 still likes to communicate and send 3 flows to nodes 2 and 3 with rates $(R_{1\to 2},R_{1\to 3},R_{1\to 23})$. However, we allow nodes 2 and 3 to communicate with each other with the constraint that whenever node 2 (or node 3) transmits, node 1 has to remain silent. The communication between nodes 2 and 3 can be used either to relay some overheard packets to the intended destination, or to send carefully designed coded packets that can further enhance the throughput.

Similar to the previous example, such a scenario is a special case of the general problem by setting $p_{2\to 1}=p_{3\to 1}=0$, and by hardwiring $\{R_{2\to 1},R_{2\to 3},R_{3\to 1},R_{3\to 2}\}$ and $\{R_{2\to 31},R_{3\to 12}\}$ to zeros. We can again use Proposition 1 to compute the capacity region $(R_{1\to 2},R_{1\to 3},R_{1\to 23})$ of the 1-to-2 broadcast PEC with receiver coordination:

$$\sum_{\forall\, i\in\{1,2,3\}} s^{(i)}\leq 1, \tag{22}$$

$$R_{1\to 2}+R_{1\to 3}+R_{1\to 23}\leq s^{(1)}\cdot p_{1\to 2\vee 3}, \tag{23}$$

$$R_{1\to 2}+R_{1\to 23}+\frac{p_{1\to 2}}{p_{1\to 2\vee 3}}R_{1\to 3}\leq s^{(3)}\cdot p_{3\to 2}+s^{(1)}\cdot p_{1\to 2}, \tag{24}$$

$$R_{1\to 3}+R_{1\to 23}+\frac{p_{1\to 3}}{p_{1\to 2\vee 3}}R_{1\to 2}\leq s^{(1)}\cdot p_{1\to 3}+s^{(2)}\cdot p_{2\to 3}, \tag{25}$$

where (22) follows from (11); (23) follows from (12); and (24) and (25) follow from (13).

Compared to the existing work [11], our results have characterized the more general Shannon capacity region instead of linear capacity region while also considering the possibility of co-existing common-information rate $R_{1\to 23}$.

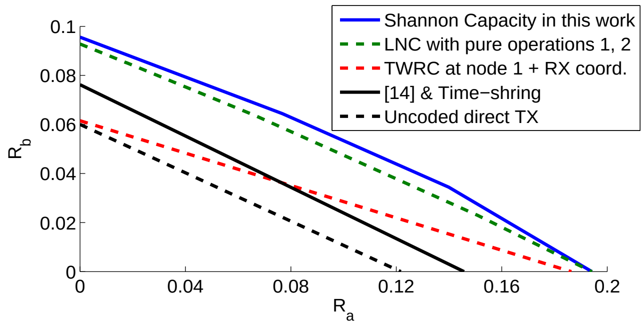


Fig. 3: Comparison of the capacity region with different achievable rates

C. Example 3: The Two-way Relay PEC

Another example is the two-way relay PEC as described in Fig. 2(c). Namely, nodes 1 and 3 want to communicate with each other with rates $(R_{1\to 3}, R_{3\to 1})$, respectively. The communication must be achieved via a relaying node 2. Such a scenario is a special case of the general problem by simply hardwiring $\{R_{1\to 2}, R_{2\to 1}, R_{2\to 3}, R_{3\to 2}\}$ and $\{R_{1\to 23}, R_{2\to 31}, R_{3\to 12}\}$ to zeros. We can again use Proposition 1 to compute the capacity region $(R_{1\to 3}, R_{3\to 1})$:

$$\sum_{\forall i\in\{1,2,3\}} s^{(i)} \leq 1, \tag{26}$$

$$R_{1\to 3} \leq s^{(1)} \cdot p_{1\to 2}, \qquad R_{3\to 1} \leq s^{(3)} \cdot p_{3\to 2}, \tag{27}$$

$$R_{1\to 3} \leq s^{(2)} \cdot p_{2\to 3}, \qquad R_{3\to 1} \leq s^{(2)} \cdot p_{2\to 1}, \tag{28}$$

where (26) and (27) follow from (11) and (12), respectively, and (28) follows from (13). One can easily verify that the capacity region described by (26) to (28) matches the existing results in [25].

D. Example 4: The Two-way Relay PEC with Opportunistic Routing

For the same setting as in **Example 3** but allowing the direct communications between node 1 and node 3, see Fig. 2(d), we can also use Proposition 1 to compute the two-way relay PEC capacity region $(R_{1\to 3}, R_{3\to 1})$ with opportunistic routing:

$$\sum_{\forall i\in\{1,2,3\}} s^{(i)} \leq 1, \tag{29}$$

$$R_{1\to 3} \leq s^{(1)} p_{1\to 2\vee 3}, \quad R_{3\to 1} \leq s^{(3)} p_{3\to 1\vee 2}, \tag{30}$$

$$R_{1\to 3} \leq s^{(1)} p_{1\to 3} + s^{(2)} p_{2\to 3}, \tag{31}$$

$$R_{3\to 1} \leq s^{(2)} p_{2\to 1} + s^{(3)} p_{3\to 1}. \tag{32}$$

One can verify that the capacity region described by (29) to (32) matches the existing results in [12].

E. Numerical Evaluation

Consider a 3-node network with marginal channel success probabilities $p_{1\to 2} = 0.35$, $p_{1\to 3} = 0.8$, $p_{2\to 1} = 0.6$, $p_{2\to 3} = 0.5$, $p_{3\to 1} = 0.3$, and $p_{3\to 2} = 0.75$, respectively, and we assume that all the erasure events are independent. That is, $p_{i\to j\vee k} = 1 - (1 - p_{i\to j})(1 - p_{i\to k})$. To illustrate the 9-dimensional capacity region, we further assume that the following 3 flows are of the same rate $R_{1\to 2} = R_{1\to 3} = R_{1\to 23} = R_a$ and the other 6 flows are of rate $R_{2\to 1} = R_{2\to 3} = R_{3\to 1} = R_{3\to 2} = R_{2\to 31} = R_{3\to 12} = R_b$. We will use Proposition 1 to find the largest R_a and R_b value for this example scenario.

Fig. 3 compares the Shannon capacity region of (R_a, R_b) with different achievability schemes. The smallest rate region is achieved by simply performing uncoded direct transmission. The second achievability scheme combines the broadcast channel LNC in [14] with time-sharing among all three nodes. The third scheme performs two-way relay channel (TWRC) coding in node 1 for those $3 \to 2$ and $2 \to 3$ flows while allowing node 2 to relay the node 1's packets destined for node 3 and vice versa. The fourth scheme is derived from our achievability scheme in the proof of Proposition 3 except when we impose the restriction that the scheme can only use LNC choices that were known previously. Namely, we allow all three nodes to perform the broadcast-based LNC and/or TWRC-based LNC operations (coding choices $[\mathrm{c}, 1]$ and $[\mathrm{c}, 2]$ in Stage 2) but not the hybrid operations (coding choices $[\mathrm{c}, 3]$ and $[\mathrm{c}, 4]$) proposed in this work. One can see that the result is strictly suboptimal. It shows that the proposed hybrid operations are critical for achieving the Shannon capacity in Propositions 1 and 3. The detailed rate region description of each sub-optimal achievability scheme is described in Appendix C.

V. Proof of Proposition 1

We now prove the capacity outer bound in Proposition 1. Given any reception probabilities and any $\epsilon > 0$, consider a joint network coding and scheduling scheme (7), (8), and (9) that can send 9 flows with rates $\vec{R}$ in n time slots with the overall error probability no larger than ϵ. Based on the given scheme, define $s^{(i)}$ as the normalized expected number of time slots for which node i is scheduled. That is,

$$s^{(i)} \triangleq \frac{1}{n}\mathbb{E}\left\{\sum_{t=1}^{n} 1_{\{\sigma(t)=i\}}\right\}, \tag{33}$$

where $1_{\{\cdot\}}$ is the indicator function. By the above definition, the computed scheduling frequencies $\{s^{(1)}, s^{(2)}, s^{(3)}\}$ must satisfy the time-sharing condition (11).

We will now prove (12) and (13) of Proposition 1. To that end, we assume that the logarithm of the mutual information and the entropy is of base q, the order of the underlying finite field $\mathbb{F}_q$. For the case when the logarithm of the entropy is base-2, we will distinguish it by using $H_2(\cdot)$.

A. Proof of the broadcast cut-set condition (12)

The inequality (12) can be proven by proving the following two inequalities separately:

$$I(\mathbf{W}_{i*}\,;\, [\mathbf{Y}_{*j}, \mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{j,k\}*}, [\mathbf{Z}]_1^n) \geq n\left(R_{i\to j} + R_{i\to k} + R_{i\to jk} - 2\epsilon - \frac{H_2(2\epsilon)}{n\log_2 q}\right), \tag{12A}$$

$$I(\mathbf{W}_{i*}\,;\, [\mathbf{Y}_{*j}, \mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{j,k\}*}, [\mathbf{Z}]_1^n) \leq n s^{(i)} p_{i\to j\vee k}. \tag{12B}$$

Intuitively, (12A) follows from the Fano's inequality and (12B) follows from a simple cut condition. By choosing $\epsilon \to 0$, we have proven (12).

Firstly, (12A) can be derived as follows:

$$I(\mathbf{W}_{i*}\,;\,[\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n\,|\,\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^n)$$
$$= I(\mathbf{W}_{i*}\,;\,\mathbf{W}_{\{j,k\}*},[\mathbf{Y}_{*j},\mathbf{Y}_{*k},\mathbf{Z}]_1^n) \tag{34}$$
$$\geq n(R_{i\to j}+R_{i\to k}+R_{i\to jk})(1-2\epsilon)-\frac{H_2(2\epsilon)}{\log_2 q}, \tag{35}$$

where (34) follows from the definition of mutual information and the fact that $\mathbf{W}_{i*}$, $\mathbf{W}_{\{j,k\}*}$, and $[\mathbf{Z}]_1^n$ are independent of each other. To derive (35), we observe that the messages $\mathbf{W}_{i*}$ can be decoded from $[\mathbf{Y}_{*j},\mathbf{Y}_{*k},\mathbf{Z}]_1^n$ and $\mathbf{W}_{\{j,k\}*}\triangleq\mathbf{W}_{j*}\cup\mathbf{W}_{k*}$, see (8) for nodes j and k, with error probability being at most 2ϵ by the union bound. As a result, by Fano's inequality, we have (35).

Secondly, (12B) can be derived as follows:

$$I(\mathbf{W}_{i*}\,;\,[\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n\,|\,\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^n)$$
$$\leq H([\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n\,|\,\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^n) \tag{36}$$
$$= \sum_{t=1}^n H(\mathbf{Y}_{*j}(t),\mathbf{Y}_{*k}(t)\,|\,[\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^{t-1},\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^t) \tag{37}$$
$$= \sum_{t=1}^n H(\mathbf{Y}_{*j}(t),\mathbf{Y}_{*k}(t)\,|\,[\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^{t-1},\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^t,X_j(t),X_k(t)) \tag{38}$$
$$= \sum_{t=1}^n H(Y_{i\to j}(t),Y_{i\to k}(t)\,|\,[\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^{t-1},\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^t,X_j(t),X_k(t)) \tag{39}$$
$$\leq \sum_{t=1}^n \mathbb{E}\Big\{1_{\{\sigma(t)=i\}}\circ 1_{\{Z_{i\to j}(t)=1 \text{ or } Z_{i\to k}(t)=1\}}\Big\} \tag{40}$$
$$= \sum_{t=1}^n \mathbb{E}\Big\{1_{\{\sigma(t)=i\}}\Big\}\mathbb{E}\Big\{1_{\{Z_{i\to j}(t)=1 \text{ or } Z_{i\to k}(t)=1\}}\Big\} \tag{41}$$
$$= p_{i\to j\vee k}\mathbb{E}\left\{\sum_{t=1}^n 1_{\{\sigma(t)=i\}}\right\} = ns^{(i)}p_{i\to j\vee k}, \tag{42}$$

where (36) follows from the definition of mutual information; (37) follows from the chain rule and from the fact that the future channel outputs $[\mathbf{Z}]_{t+1}^n$ are independent of $\mathbf{Y}_{*j}(t),\mathbf{Y}_{*k}(t)$; (38) follows from the fact that the transmitted symbol $X_j(t)$ (resp. $X_k(t)$) is a function of the past received symbols $[\mathbf{Y}_{*j}]_1^{t-1}$ (resp. $[\mathbf{Y}_{*k}]_1^{t-1}$), the information messages $\mathbf{W}_{j*}$ (resp. $\mathbf{W}_{k*}$), and the past channel outputs $[\mathbf{Z}]_1^{t-1}$, see (7); (39) follows from the fact that the received symbol $Y_{k\to j}(t)$ in $\mathbf{Y}_{*j}(t)$ (resp. $Y_{j\to k}(t)$ in $\mathbf{Y}_{*k}(t)$) can be uniquely computed from the values of the current input $X_k(t)$ (resp. $X_j(t)$), the current channel output $\mathbf{Z}(t)$, and the current scheduling decision $\sigma(t)$, which depends only on the past channel outputs $[\mathbf{Z}]_1^{t-1}$, see (9); (40) follows from that only when $\sigma(t)=i$ with $Z_{i\to j}(t)=1$ or $Z_{i\to k}(t)=1$, we will have a non-zero value of the entropy and it is upper bounded by 1 since the base of the logarithm is q; (41) follows from the fact that since the scheduling decision $\sigma(t)$ depends only on the past channel outputs $[\mathbf{Z}]_1^{t-1}$, see (9), the random variables $\sigma(t)$ and $\mathbf{Z}(t)$ are independent; and (42) follows from the definition (33).

B. Proof of the 3-way multiple-access cut-set condition (13)

We now prove (13) by proving the following two inequalities:

$$I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n) \geq n\bigg(R_{j\to i}+R_{k\to i}+R_{j\to ki}+R_{k\to ij}+\frac{p_{j\to i}}{p_{j\to k\vee i}}R_{j\to k} + \frac{p_{k\to i}}{p_{k\to i\vee j}}R_{k\to j}-6\epsilon-\frac{3H_2(\epsilon)}{n\log_2 q}\bigg), \tag{13A}$$
$$I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n) \leq n(s^{(j)}p_{j\to i}+s^{(k)}p_{k\to i}). \tag{13B}$$

Intuitively, (13B) follows a simple cut condition. By choosing $\epsilon\to 0$, we have proven (13).

We now provide the detailed derivation of (13A) and (13B). Firstly, the inequality (13B) can be derived in a similar way as (12B). Specifically, we have

$$I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n)$$
$$\leq H([\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n) \tag{43}$$
$$= \sum_{t=1}^n H(Y_{j\to i}(t),Y_{k\to i}(t)\,|\,[\mathbf{Y}_{*i}]_1^{t-1},\mathbf{W}_{i*},[\mathbf{Z}]_1^t), \tag{44}$$
$$\leq n(s^{(j)}p_{j\to i}+s^{(k)}p_{k\to i}), \tag{45}$$

where (43) follows from the definition of mutual information; (44) follows from the chain rule and the fact that the future channel outputs $[\mathbf{Z}]_{t+1}^n$ are independent of $Y_{j\to i}(t),Y_{k\to i}(t)$; and (45) follows from similar arguments as used in (40) to (42).

We now prove (13A). For the ease of exposition, we only prove for the case when the node indices are fixed to $(i,j,k)=(1,2,3)$. Then (13A) becomes

$$I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}]_1^n\,|\,\mathbf{W}_{1*},[\mathbf{Z}]_1^n) \geq n\bigg(R_{2\to 1}+R_{3\to 1}+R_{2\to 31}+R_{3\to 12}+\frac{p_{2\to 1}}{p_{2\to 3\vee 1}}R_{2\to 3} + \frac{p_{3\to 1}}{p_{3\to 1\vee 2}}R_{3\to 2}-6\epsilon-\frac{3H_2(\epsilon)}{n\log_2 q}\bigg).$$

The cases of other node indices $(i,j,k)\in\{(2,3,1),(3,1,2)\}$ can be proven by symmetry.

Consider the following claims, whose proofs are relegated to Appendix D.

Claim 1. The following is true:

$$I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}]_1^n\,|\,\mathbf{W}_{1*},[\mathbf{Z}]_1^n) = \sum_{t=1}^n I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t)). \tag{46}$$

Claim 2. Define

$$\mathbf{W}_{\overline{2\to 3}} \triangleq \mathbf{W}_{\{1,3\}*}\cup\mathbf{W}_{2\to 1}\cup\mathbf{W}_{2\to 31}, \tag{47}$$

That is, $\mathbf{W}_{\overline{2\to 3}}$ is the collection of all the 9-flow information messages except $\mathbf{W}_{2\to 3}$. This is why we use the overline in the subscript. Symmetrically, define

$$\mathbf{W}_{\overline{3\to 2}} \triangleq \mathbf{W}_{\{1,2\}*}\cup\mathbf{W}_{3\to 1}\cup\mathbf{W}_{3\to 12}. \tag{48}$$

Then, the following is true: $\forall\, t \in \{1, \cdots, n\}$,

$$\begin{aligned}
&I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t))\\
&\geq I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t))\\
&\quad+\frac{p_{2\to1}}{p_{2\to3\vee1}}I(\mathbf{W}_{2\to3}\,;\,\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{t-1},\mathbf{W}_{\overline{2\to3}},\mathbf{Z}(t))\\
&\quad+\frac{p_{3\to1}}{p_{3\to1\vee2}}I(\mathbf{W}_{3\to2}\,;\,\mathbf{Y}_{3*}(t)\,|\,[\mathbf{Y}_{3*},\mathbf{Z}]_1^{t-1},\mathbf{W}_{\overline{3\to2}},\mathbf{Z}(t)).
\end{aligned}\tag{49}$$

Claim 3. The followings are true:

$$\begin{aligned}
\sum_{t=1}^{n}&I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t))\\
&\qquad\qquad= I(\mathbf{W}_{*1}\,;\,\mathbf{W}_{1*},[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n}),
\end{aligned}\tag{50}$$

$$\begin{aligned}
\sum_{t=1}^{n}&I(\mathbf{W}_{2\to3}\,;\,\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{t-1},\mathbf{W}_{\overline{2\to3}},\mathbf{Z}(t))\\
&\qquad\qquad\geq I(\mathbf{W}_{2\to3}\,;\,\mathbf{W}_{3*},[\mathbf{Y}_{*3},\mathbf{Z}]_1^{n}),
\end{aligned}\tag{51}$$

$$\begin{aligned}
\sum_{t=1}^{n}&I(\mathbf{W}_{3\to2}\,;\,\mathbf{Y}_{3*}(t)\,|\,[\mathbf{Y}_{3*},\mathbf{Z}]_1^{t-1},\mathbf{W}_{\overline{3\to2}},\mathbf{Z}(t))\\
&\qquad\qquad\geq I(\mathbf{W}_{3\to2}\,;\,\mathbf{W}_{2*},[\mathbf{Y}_{*2},\mathbf{Z}]_1^{n}).
\end{aligned}\tag{52}$$

By the above Claims 1 to 3 we have

$$\begin{aligned}
&I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}]_1^n\,|\,\mathbf{W}_{1*},[\mathbf{Z}]_1^n)\\
&\geq I(\mathbf{W}_{*1}\,;\,\mathbf{W}_{1*},[\mathbf{Y}_{*1},\mathbf{Z}]_1^n)\\
&\quad+\frac{p_{2\to1}}{p_{2\to3\vee1}}I(\mathbf{W}_{2\to3}\,;\,\mathbf{W}_{3*},[\mathbf{Y}_{*3},\mathbf{Z}]_1^n)\\
&\quad+\frac{p_{3\to1}}{p_{3\to1\vee2}}I(\mathbf{W}_{3\to2}\,;\,\mathbf{W}_{2*},[\mathbf{Y}_{*2},\mathbf{Z}]_1^n), \qquad (53)\\
&\geq n(R_{2\to1}+R_{3\to1}+R_{2\to31}+R_{3\to12})(1-\epsilon)\\
&\quad-\frac{H_2(\epsilon)}{\log_2 q}+\frac{p_{2\to1}}{p_{2\to3\vee1}}\left(nR_{2\to3}(1-\epsilon)-\frac{H_2(\epsilon)}{\log_2 q}\right)\\
&\quad+\frac{p_{3\to1}}{p_{3\to1\vee2}}\left(nR_{3\to2}(1-\epsilon)-\frac{H_2(\epsilon)}{\log_2 q}\right) \qquad (54)
\end{aligned}$$

where (53) follows from jointly combining (46) to (52); and (54) follows from applying Fano's inequality to each individual term. Since we can choose ϵ arbitrarily, by letting $\epsilon \to 0$, we have proven (13A).

Remark: As discussed in Section III-A, (13) is inspired by the *multiple-access channel* (MAC) cut-set bound. When considering the MAC, one usually focuses on all *incoming* traffic entering node i, i.e., $R_{j\to i}$, $R_{j\to ki}$, $R_{k\to i}$, and $R_{k\to ij}$, and thus might be interested in quantifying/bounding the following mutual information term:

$$\begin{aligned}
I(\,&\mathbf{W}_{j\to i},\mathbf{W}_{j\to ki},\mathbf{W}_{k\to i},\mathbf{W}_{k\to ij}\,;\,[\mathbf{Y}_{*i}]_1^n\\
&\qquad|\,\mathbf{W}_{i*},\mathbf{W}_{j\to k},\mathbf{W}_{k\to j},[\mathbf{Z}]_1^n).
\end{aligned}\tag{55}$$

Unfortunately, (55) does not take into the fact that node j has some private information that need to be delivered to node k (those $\mathbf{W}_{j\to k}$ packets) and vice versa. Due to such an observation, we quantify the mutual information term $I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n)$ instead of (55). Comparing $I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n)$ and (55), we can use the chain rule to show that

$$\begin{aligned}
(55) = {}&I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n)\\
&-I(\mathbf{W}_{j\to k},\mathbf{W}_{k\to j}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n),
\end{aligned}$$

and the difference $I(\mathbf{W}_{j\to k},\mathbf{W}_{k\to j}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n)$ can be viewed as the amount of the private information $\mathbf{W}_{j\to k}$ and $\mathbf{W}_{k\to j}$ that has been "leaked" to the other node i. In some broad sense, (13) (or equivalent (13A)) characterizes a new lower bound on the information leakage

$$\begin{aligned}
I(\mathbf{W}_{j\to k},&\mathbf{W}_{k\to j}\,;\,[\mathbf{Y}_{*i}]_1^n\,|\,\mathbf{W}_{i*},[\mathbf{Z}]_1^n)\\
&\geq\frac{p_{j\to i}}{p_{j\to k\vee i}}R_{j\to k}+\frac{p_{k\to i}}{p_{k\to i\vee j}}R_{k\to j}.
\end{aligned}$$

This is why in our discussion right after Proposition 1 we referred to the term $\frac{p_{j\to i}}{p_{j\to k\vee i}}R_{j\to k}+\frac{p_{k\to i}}{p_{k\to i\vee j}}R_{k\to j}$ as the penalty for sending those private-information. Note that similar information leakage arguments have been used in other channel models, e.g., the wireless deterministic channels [29].

VI. Proof of Proposition 3

We provide the so-called *first-order analysis* for the achievability of a LNC solution.

We assume that all nodes know the channel reception probabilities, the total time budget n, and the rate vector $\vec{R}$ they want to achieve in the beginning of time 0. As a result, each node can compute the same 15 non-negative values $t^{(i)}_{[\mathrm{u}]}$ and $\{t^{(i)}_{[\mathrm{c},l]}\}_{l=1}^4$ for all $i \in \{1,2,3\}$ satisfying Proposition 3.

Our construction consists of 2 stages. Stage 1: Each node, say node i, has $n(R_{i\to j}+R_{i\to k}+R_{i\to jk})$ unicast and multicast packets (i.e., $\mathbf{W}_{i*}$) that need to be sent to other nodes j and k. Assume that those packets are grouped together and indexed as $l = 1$ to $n(R_{i\to j}+R_{i\to k}+R_{i\to jk})$. That is, the packet indices $l = 1$ to $nR_{i\to j}$ correspond to $\mathbf{W}_{i\to j}$ packets, the packet indices $l = nR_{i\to j}+1$ to $n(R_{i\to j}+R_{i\to k})$ correspond to $\mathbf{W}_{i\to k}$ packets, and so forth. Then in the beginning of time 1, node 1 chooses the first packet (index 1) and repeatedly sends it uncodedly until at least one of nodes 2 and 3 receives it. Whether it is received or not can be known causally by network-wide feedbacks $\mathbf{Z}(t-1)$. Then node 1 picks the next indexed packet and repeats the same process until each of these $n(R_{1\to2}+R_{1\to3}+R_{1\to23})$ packets is heard by at least one of nodes 2 and 3. By simple analysis, see [21], node 1 can finish the transmission in $nt^{(i)}_{[\mathrm{u}]}$ slots since (17).[5] We repeat this process for nodes 2 and 3, respectively. Note that once node 1 has finished transmitting all its own packets $\mathbf{W}_{1*}$, node 2 can immediately take over and start transmitting its own packets $\mathbf{W}_{2*}$ because node 2 knows the value of $n(R_{1\to2}+R_{1\to3}+R_{1\to23})$ and from the instant, error-free, network-wide feedback, node 2 can count in the end of each time slot how many packets node 1 finished transmission. By the same reason, node 3 can immediately take over after node 2 has finished. Stage 1 can be finished in $n(\sum_i t^{(i)}_{[\mathrm{u}]})$ slots.

After Stage 1, the status of all packets is summarized as follows. Each of $\mathbf{W}_{i\to j}$ packets is heard by at least one of nodes j and k. Those that have already been heard by node j, the intended destination, are delivered successfully and thus will not be considered for future operations (Stage 2). We denote those $\mathbf{W}_{i\to j}$ packets that are overheard by node k only (not by node j) as $\mathbf{W}^{(k)}_{i\to j}$. In average, there are

[5] By the law of large numbers, we can ignore the randomness of the events and treat them as deterministic when n is sufficiently large.

$nR_{i\to j}\frac{p_{i\to \overline{j}k}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(k)}_{i\to j}$ packets. Since the causal feedback is available to all network nodes (not only node i), by letting all three nodes perform some simple bookkeeping, any one of the three network nodes (not only node i) is aware of the indices of all the $\mathbf{W}^{(k)}_{i\to j}$ packets. We denote the corresponding index set by $\mathbf{I}^{(k)}_{i\to j}$. Symmetrically, we also have $nR_{i\to k}\frac{p_{i\to j\overline{k}}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(j)}_{i\to k}$ packets that was intended for node k but was overheard only by node j in Stage 1, and all three nodes can individually create the corresponding index set $\mathbf{I}^{(j)}_{i\to k}$.

Similarly for the common-information packets $\mathbf{W}_{i\to jk}$, each packet was heard by at least one of nodes j and k in Stage 1. Those that have been heard by both nodes j and k, are delivered successfully and thus will not be considered in Stage 2. We similarly denote those $\mathbf{W}_{i\to jk}$ packets that are heard by node k only (not by node j) as $\mathbf{W}^{(k)}_{i\to jk}$. In average, there are $nR_{i\to jk}\frac{p_{i\to \overline{j}k}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(k)}_{i\to jk}$ packets. Symmetrically, we also have $nR_{i\to jk}\frac{p_{i\to j\overline{k}}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(j)}_{i\to jk}$ packets that were heard only by node j in Stage 1. The corresponding index sets are denoted by $\mathbf{I}^{(k)}_{i\to jk}$ and $\mathbf{I}^{(j)}_{i\to jk}$, respectively, and they can be individually created by all three nodes through simple bookkeeping.

In summary, all three nodes individually know all 12 index sets $\{\mathbf{I}^{(k)}_{i\to j},\mathbf{I}^{(k)}_{i\to jk},\mathbf{I}^{(j)}_{i\to k},\mathbf{I}^{(j)}_{i\to jk} : \forall(i,j,k)\}$ after Stage 1. In addition, each node i knows the content of its own packets $\mathbf{W}_{i\to j}$, $\mathbf{W}_{i\to k}$, and $\mathbf{W}_{i\to j}k$, and the content of what it has received from other nodes $(\mathbf{W}^{(i)}_{j\to k},\mathbf{W}^{(i)}_{j\to ki},\mathbf{W}^{(i)}_{k\to j},\mathbf{W}^{(j)}_{k\to ij})$ during Stage 1.

Stage 2 is the LNC phase, in which each node i will send a linear combination of the overheard packets. That is, for each time t, node i sends a linear combination $X_i(t)=[\tilde{W}_j+\tilde{W}_k]$ with 4 possible ways of choosing the constituent packets $\tilde{W}_j$ and $\tilde{W}_k$, which are detailed as follows.

$$\begin{aligned}
&[\mathsf{c},1]: && \tilde{W}_j\in\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk} \text{ and } \tilde{W}_k\in\mathbf{W}^{(j)}_{i\to k}\cup\mathbf{W}^{(j)}_{i\to jk},\\
&[\mathsf{c},2]: && \tilde{W}_j\in\mathbf{W}^{(i)}_{k\to j}\cup\mathbf{W}^{(i)}_{k\to ij} \text{ and } \tilde{W}_k\in\mathbf{W}^{(i)}_{j\to k}\cup\mathbf{W}^{(i)}_{j\to ki},\\
&[\mathsf{c},3]: && \tilde{W}_j\in\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk} \text{ and } \tilde{W}_k\in\mathbf{W}^{(i)}_{j\to k}\cup\mathbf{W}^{(i)}_{j\to ki},\\
&[\mathsf{c},4]: && \tilde{W}_j\in\mathbf{W}^{(i)}_{k\to j}\cup\mathbf{W}^{(i)}_{k\to ij} \text{ and } \tilde{W}_k\in\mathbf{W}^{(j)}_{i\to k}\cup\mathbf{W}^{(j)}_{i\to jk}.
\end{aligned}$$

To explain the intuition behind the 4 coding choices $[\mathsf{c},1]$ to $[\mathsf{c},4]$, we observe that choice $[\mathsf{c},1]$ is the standard LNC operation for the 2-receiver broadcast channels [14] since node i sends a linear sum that benefits both nodes j and k simultaneously, i.e., the sum of two packets, each overheard by an undesired receiver. Choice $[\mathsf{c},2]$ is the standard LNC operation for the 2-way relay channels, since node i, as a relay for the 2-way traffic from $j\to k$ and from $k\to j$, respectively, mixes the packets from two opposite directions and sends their linear sum.

Choices $[\mathsf{c},3]$ and $[\mathsf{c},4]$ are the new "hybrid" cases that are proposed in this work, for which we can mix part of the broadcast traffic and part of the 2-way traffic. We argue that transmitting such a linear mixture again benefits both nodes simultaneously. For example, suppose that coding choice $[\mathsf{c},3]$ is used, and the linear sum $[\tilde{W}_j+\tilde{W}_k]$ is received by node j. Since $\tilde{W}_k$ is a function of all packets originated from node j, node j can compute $\tilde{W}_k$ by itself and then subtract it from the linear sum and derive the desired packet $\tilde{W}_j$. Similarly, if node k receives the linear sum, since it has overheard all packets in $\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk}$, it can subtract $\tilde{W}_j$ and decode its desired $\tilde{W}_k$. The argument for coding choice $[\mathsf{c},4]$ is symmetric.

We now explain in details how to implement the above 4 coding choices for each of the time slots in Stage 2. The best way to explain the implementation is to temporarily view the overheard packets as being stored in a big queue. Namely, in the beginning of Stage 2, all the packets in $\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk}$ are put into a big queue. Similarly, all the packets in $\mathbf{W}^{(j)}_{i\to k}\cup\mathbf{W}^{(j)}_{i\to jk}$, $\mathbf{W}^{(i)}_{k\to j}\cup\mathbf{W}^{(i)}_{k\to ij}$, and $\mathbf{W}^{(i)}_{j\to k}\cup\mathbf{W}^{(i)}_{j\to ki}$ are put into 3 big queues as well, one queue for each set of packets respectively. Then coding choice $[\mathsf{c},1]$ means that node i takes the head-of-line packet from the queue of $\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk}$, and combines it with the head-of-line packet from the queue of $\mathbf{W}^{(j)}_{i\to k}\cup\mathbf{W}^{(j)}_{i\to jk}$. Coding choices $[\mathsf{c},2]$ to $[\mathsf{c},4]$ can be interpreted similarly by combining the head-of-line packets from different queues.

Since each node i has 4 possible coding choices, we perform coding choice $[\mathsf{c},l]$ for exactly $nt^{(i)}_{[\mathsf{c},l]}$ times sequentially for $l=1$ to 4. After sending the 4 coding choices for a combined total of $n\big(t^{(i)}_{[\mathsf{c},1]}+t^{(i)}_{[\mathsf{c},2]}+t^{(i)}_{[\mathsf{c},3]}+t^{(i)}_{[\mathsf{c},4]}\big)$ time slots for node i, we set $i=i+1$ and repeat the same process until all three nodes have finished transmission. Totally, Stage 2 takes $\sum_{i\in\{1,2,3\}} n\big(t^{(i)}_{[\mathsf{c},1]}+t^{(i)}_{[\mathsf{c},2]}+t^{(i)}_{[\mathsf{c},3]}+t^{(i)}_{[\mathsf{c},4]}\big)$ time slots. We now describe how to manage the "queues" within each node during transmission.

Suppose that node i is performing the coding choice $[\mathsf{c},1]$ and chooses two head-of-line packets $\tilde{W}_j\in\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk}$ and $\tilde{W}_k\in\mathbf{W}^{(j)}_{i\to k}\cup\mathbf{W}^{(j)}_{i\to jk}$ from the individual queues, respectively. If the linear combination $[\tilde{W}_j+\tilde{W}_k]$ is received by node j, then node j will decode the desired $\tilde{W}_j$ by subtracting the overheard packet $\tilde{W}_k$. As a result, we remove the successfully delivered packet $\tilde{W}_j$ from the queue of $\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk}$. Similarly, if the combination $[\tilde{W}_j+\tilde{W}_k]$ is received by node k, then node k can decode the desired packet $\tilde{W}_k$ and we remove $\tilde{W}_k$ from the corresponding queue of $\mathbf{W}^{(j)}_{i\to k}\cup\mathbf{W}^{(j)}_{i\to jk}$. If any one of the two queues is empty, say the queue of $\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk}$ is empty during coding choice $[\mathsf{c},1]$, then we simply set $\tilde{W}_j=0$. Namely, in such a degenerate case we choose to send an uncoded packet $[0+\tilde{W}_k]$ instead of a linear combination $[\tilde{W}_j+\tilde{W}_k]$. If both queues are empty, then we simply send a 0 packet. The same queue management is applied to coding choices $[\mathsf{c},2]$ to $[\mathsf{c},4]$ as well.

Note that the above process requires very detailed bookkeeping at each node. Namely, node j (resp. node k) needs to know the head-of-line packet $\tilde{W}_k$ (resp. $\tilde{W}_j$) while node i is executing Stage 2, so that it can subtract the overheard from the linear combination $[\tilde{W}_j+\tilde{W}_k]$ when received. This is possible since in the beginning of Stage 2, each node knows all 12 index sets: $\{\mathbf{I}^{(k)}_{i\to j},\mathbf{I}^{(k)}_{i\to jk},\mathbf{I}^{(j)}_{i\to k},\mathbf{I}^{(j)}_{i\to jk} : \forall(i,j,k)\}$. Since the reception status $[\mathbf{Z}]_1^{t-1}$ is available to all nodes for free, through detailed bookkeeping and a proper counting mechanism over the index sets, each node (not only node i but also nodes j and k) can successfully trace the status of the queues when node i is executing Stage 2. In this way each node maintains a synchronized view of the queue status of the other nodes and can thus identify the head-of-line packets that constitute the linear combination.

Another important point worth emphasizing is that the queues cannot be replenished during Stage 2. Namely, if a packet is removed from the queue in one coding operation, then it will be removed from the synchronized queues at all three nodes and will not participate in any future coding operations. For example, the packets in $\mathbf{W}^{(k)}_{i\to j} \cup \mathbf{W}^{(k)}_{i\to jk}$ will participate in coding choice $[\mathrm{c}, 1]$ of node i, but they can also participate in coding choice $[\mathrm{c}, 3]$ of node i, and coding choices $[\mathrm{c}, 2]$ and $[\mathrm{c}, 3]$ of node k. If a packet in $\mathbf{W}^{(k)}_{i\to j} \cup \mathbf{W}^{(k)}_{i\to jk}$ is successfully delivered through coding choice $[\mathrm{c}, 1]$ of node i, then it will be removed from the queue and will not participate in any subsequent coding choices $[\mathrm{c}, 3]$ of node i, and coding choices $[\mathrm{c}, 2]$ and $[\mathrm{c}, 3]$ of node k in the future time slots. Again, this LNC design is possible since each node maintains a synchronized view of the queue status of the other nodes with the help of the causal feedback $[\mathbf{Z}]_1^{t-1}$.

Since $\mathbf{W}^{(k)}_{i\to j} \cup \mathbf{W}^{(k)}_{i\to jk}$ participates in coding choices $[\mathrm{c}, 1]$ and $[\mathrm{c}, 3]$ of node i and coding choices $[\mathrm{c}, 2]$ and $[\mathrm{c}, 3]$ of node k, (18) guarantees that the queue of $\mathbf{W}^{(k)}_{i\to j} \cup \mathbf{W}^{(k)}_{i\to jk}$ will be empty in the end of Stage 2, which means that we can finish sending all $\mathbf{W}^{(k)}_{i\to j} \cup \mathbf{W}^{(k)}_{i\to jk}$ packets and they will all successfully arrive at node j, the intended destination.[6] Symmetrically, (19) guarantees that the queue of $\mathbf{W}^{(j)}_{i\to k} \cup \mathbf{W}^{(j)}_{i\to jk}$ will be empty, which means that we can finish sending all $\mathbf{W}^{(j)}_{i\to k} \cup \mathbf{W}^{(j)}_{i\to jk}$ packets to their intended destination node k in the end of Stage 2. Finally, (15) guarantees that we can finish Stages 1 and 2 in the allotted n time slots. The proof of Proposition 3 is complete.

Remark: An astute reader might notice that it would be a strict waste of resource if a node transmits $X_i(t) = 0$ packet, which happens when both queues are empty during Stage 2. The reason why we include such a "wasteful" operation is that the scheme described herein has to operate under any arbitrary t-values satisfying (15)–(19). The underlying t-values may not be optimal and sometimes they can *overly allocate* time slots. Therefore, we include the *idle operation* (i.e., sending $X_i(t) = 0$) in our scheme description so that we can properly consume the overly allocated time slots. One can actually prove that if the optimal t-values derived in Proposition 4 are used, then the proposed scheme will never be idle or equivalently will not send such a fixed $X_i(t) = 0$.

VII. Discussions on Channel Interference Models

In this paper, the 3-node packet erasure problem is formulated based on the full collision model. Namely, whenever two or more nodes transmit, both transmissions always collide regardless whether erasure happens. Essentially, this full-collision model necessitates time-sharing among all three nodes. One main reason behind this model is to characterize the throughput benefits of broadcast diversity and find the largest attainable capacity of a 3-node PEN without worrying about the effects of interference.

For future work, there are a couple of ways that can generalize the above full-collision model. One is called the partial collision model. Namely, when two nodes transmit, then they collide only if both are non-erasure. If only one is non-erasure, then the receiver still hears the non-erased one successfully. Another interference model is the additive interference model. Namely, when both nodes transmit and both are non-erasure, we add the transmitted files X_1 and X_2 in the finite field. This is the most complicated one and similar to the settings used in some existing works [30]–[34]. There currently are no known capacity results on the general 3-node networks with 9 co-existing flows.

VIII. Conclusion

This work studies the capacity of the 3-node packet erasure network, when the most general 9-dimensional private- and common-information traffics are considered. The Shannon capacity has been exactly characterized for all channel parameters when the causal ACK/NACK feedbacks are causally available for free through a separate control channel. For the practical setting where the control messages have to be sent through the regular forward channels, under the assumption of the network being *fully-connected*, the capacity region is bracketed by a pair of upper and lower bounds, the gap of which is inversely proportional to the packet size (in bits). The capacity region is thus effectively characterized when the packet size becomes sufficiently large. Technical contributions of this work include a new converse for many-to-many network communications and a new capacity-approaching scheme based on simple linear network coding operations.

Appendix A
Proof of Lemma 1

We prove this by induction. When $t = 1$, then (6) and (10) are equivalent by definition. Suppose (6) and (10) are equivalent for $t = 1$ to $t_0 - 1$. We now consider $t = t_0$. By Lemma 2 in Appendix D, $[\mathbf{Y}_{*i}]_1^{t_0-1}$ can be uniquely computed by the values of $\mathbf{W}_{\{1,2,3\}*}$ and $[\mathbf{Z}]_1^{t_0-1}$. As a result, we can rewrite (6) by

$$\sigma_i(t_0) = \overline{f}^{(t_0)}_{\mathsf{SCH},\,i}(\mathbf{W}_{\{1,2,3\}*}, [\mathbf{Z}]_1^{t_0-1}). \tag{56}$$

Then due to the information equality (5), there is no dependence between $\sigma_i(t_0)$ and $\mathbf{W}_{\{1,2,3\}*}$. As a result, we can further remove $\mathbf{W}_{\{1,2,3\}*}$ from the input arguments in (56), which leads to (10). By induction, the proof of Lemma 1 is thus complete.

Appendix B
Proof of Proposition 2

We provide the first-order analysis for the achievability scheme.

Suppose that all the network nodes share the same parameters before initiating, see the discussion in Section VI. Also

[6] Those $\mathbf{W}^{(k)}_{i\to jk}$ packets are the common-information packets that are intended for both nodes j and k. However, since our definition of $\mathbf{W}^{(k)}_{i\to jk}$ counts only those that have already been received by node k, we say herein their new intended destination is node j instead.

assume that a common random seed is available to all nodes in advance.

We similarly follow the 2-stage scheme used in the proof of Proposition 3. The difference is that we need to revise the scheme in Scenario 1 to take into account the assumption of Scenario 2 that any feedback information needs to be sent over the regular channels as well. Our scheme closely mimics the scheme in Section VI but now uses some form of random linear network coding (RLNC), which allows us to circumvent the need of instant causal feedback (after each transmission) and can thus use "batch feedback" that reports the reception status with delay. With a common random seed available to all three nodes, the RLNC operations of one node can be "simulated" in the other nodes as well. This allows the same kind of "bookkeeping" as used in the proof of Proposition 3. Since bookkeeping may be computationally expensive, in practice, network code designers can place the coding vectors used by the RLNC in the header of the packets, which circumvents the need of bookkeeping. However, putting the coding vectors in the header reduces the data rate. As a result, to minimize the loss of capacity, we opt to use bookkeeping instead of the traditional practice of putting the coding vectors in the header of the packet.

We now explain the main RLNC process for each stage. In each stage, we assume that nodes will sequentially transmit following the order of the node indices $\{1, 2, 3\}$. We now define the following three constants for each node $i \in \{1, 2, 3\}$ that can facilitate our discussion:

$$\eta_1^{(i)} \triangleq \frac{R_{i\to j}}{R_{i\to j} + R_{i\to k} + R_{i\to jk}}, \tag{57}$$

$$\eta_2^{(i)} \triangleq \frac{R_{i\to k}}{R_{i\to j} + R_{i\to k} + R_{i\to jk}}, \tag{58}$$

$$\eta_3^{(i)} \triangleq \frac{R_{i\to jk}}{R_{i\to j} + R_{i\to k} + R_{i\to jk}}. \tag{59}$$

Obviously, $\eta_1^{(i)}+\eta_2^{(i)}+\eta_3^{(i)} = 1$ for any $i \in \{1, 2, 3\}$. Totally, there are 9 such constants. Note that each of the network node can compute all 9 constants since $\vec{R}$ is available to all nodes. Without loss of generality, we can also assume

$$\begin{aligned} &t_{[\mathrm{u}]}^{(i)} \left(\eta_1^{(i)} + \eta_3^{(i)}\right) \cdot p_{i\to \overline{j}k} \\ &\quad < \left(t_{[\mathrm{c},1]}^{(i)} + t_{[\mathrm{c},3]}^{(i)}\right) \cdot p_{i\to j} + \left(t_{[\mathrm{c},2]}^{(k)} + t_{[\mathrm{c},3]}^{(k)}\right) \cdot p_{k\to j}, \end{aligned} \tag{60}$$

$$\begin{aligned} &t_{[\mathrm{u}]}^{(i)} \left(\eta_2^{(i)} + \eta_3^{(i)}\right) \cdot p_{i\to j\overline{k}} \\ &\quad < \left(t_{[\mathrm{c},1]}^{(i)} + t_{[\mathrm{c},4]}^{(i)}\right) \cdot p_{i\to k} + \left(t_{[\mathrm{c},2]}^{(j)} + t_{[\mathrm{c},4]}^{(j)}\right) \cdot p_{j\to k}. \end{aligned} \tag{61}$$

The reason is that we can always set $t_{[\mathrm{u}]}^{(i)}$ to be arbitrarily close to $\frac{R_{i\to j}+R_{i\to k}+R_{i\to jk}}{p_{i\to j\vee k}}$ but still larger than $\frac{R_{i\to j}+R_{i\to k}+R_{i\to jk}}{p_{i\to j\vee k}}$ without violating any of the inequalities (15) and (17). As a result, $t_{[\mathrm{u}]}^{(i)}\eta_1^{(i)}$ can be made arbitrarily close to $\frac{R_{i\to j}}{p_{i\to j\vee k}}$ and $t_{[\mathrm{u}]}^{(i)}\eta_3^{(i)}$ arbitrarily close to $\frac{R_{i\to jk}}{p_{i\to j\vee k}}$. By (18), we thus have (60). Similarly, since $t_{[\mathrm{u}]}^{(i)}\eta_2^{(i)}$ and $t_{[\mathrm{u}]}^{(i)}\eta_3^{(i)}$ can be made arbitrarily close to $\frac{R_{i\to k}}{p_{i\to j\vee k}}$ and $\frac{R_{i\to jk}}{p_{i\to j\vee k}}$, respectively, (19) implies (61).

Description of Stage 1: Each node i performs the following RLNC operations exactly for $nt_{[\mathrm{u}]}^{(i)}$ number of time slots. Specifically, consider the first $\eta_1^{(i)}$ portion of the allotted $nt_{[\mathrm{u}]}^{(i)}$ times. In each of those $nt_{[\mathrm{u}]}^{(i)}\eta_1^{(i)}$ time slots, node i chooses a $1 \times (nR_{i\to j})$ random encoding row vector $\mathbf{c}_t \in \mathbb{F}_q^{nR_{i\to j}}$ independently and uniformly randomly and transmits $X_i(t)$ by $X_i(t) = \mathbf{c}_t\mathbf{W}_{i\to j}^\top$. We now consider the second $\eta_2^{(i)}$ portion of the $nt_{[\mathrm{u}]}^{(i)}$ time slots. In each of those $nt_{[\mathrm{u}]}^{(i)}\eta_2^{(i)}$ time slots, node i chooses a $1 \times (nR_{i\to k})$ random coding vector $\mathbf{c}_t$ independently and uniformly randomly and transmits $X_i(t) = \mathbf{c}_t\mathbf{W}_{i\to k}^\top$. Finally, consider the last $\eta_3^{(i)}$ portion of the $nt_{[\mathrm{u}]}^{(i)}$ time slots. In each of those $nt_{[\mathrm{u}]}^{(i)}\eta_3^{(i)}$ time slots, node i chooses a $1 \times (nR_{i\to jk})$ vector $\mathbf{c}_t$ independently and uniformly randomly and transmits $X_i(t) = \mathbf{c}_t\mathbf{W}_{i\to jk}^\top$. Namely for the allotted $nt_{[\mathrm{u}]}^{(i)}$ time slots, node i sequentially transmits some random mixture of the packets $\mathbf{W}_{i\to j}$, $\mathbf{W}_{i\to k}$, and $\mathbf{W}_{i\to jk}$ over the fixed fractions $\eta_1^{(i)}$, $\eta_2^{(i)}$, and $\eta_3^{(i)}$ of times, respectively, and does not care whether the transmitted packet is correctly received or not. Stage 1 can be finished in exactly $n(\sum_i t_{[\mathrm{u}]}^{(i)})$ slots.

Note that when node i computes the coding vectors $\mathbf{c}_t$, the other nodes j and k can also "simulate" the computation and thus know the $\mathbf{c}_t$ vector used by node i. As a result, if node j receives a coded packet $X_i(t) = \mathbf{c}_t\mathbf{W}_{i\to jk}^\top$ during the third fraction of node i's transmission, node j knows the $\mathbf{c}_t$ vector used for encoding.

New Packet Regrouping After Stage 1: After Stage 1, we put some of those $\{\mathbf{c}_t\mathbf{W}_{i\to j}\}$ packets that were sent during the first fraction of Stage 1, totally there are $nt_{[\mathrm{u}]}^{(i)}\eta_1^{(i)}$ such packets, into two disjoint groups. Specifically, we use $\{\mathbf{c}_t\mathbf{W}_{i\to j}\}_{\overline{j}k}$ to denote those packets $\{\mathbf{c}_t\mathbf{W}_{i\to j}\}$ that are heard only by node k and not by node j; and we use $\{\mathbf{c}_t\mathbf{W}_{i\to j}\}_j$ to denote those packets that are heard by node j (may or may not be heard by node k). In average, there are $nt_{[\mathrm{u}]}^{(i)}\eta_1^{(i)}p_{i\to\overline{j}k}$ number of $\{\mathbf{c}_t\mathbf{W}_{i\to j}\}_{\overline{j}k}$ packets and $nt_{[\mathrm{u}]}^{(i)}\eta_1^{(i)}p_{i\to j}$ number of $\{\mathbf{c}_t\mathbf{W}_{i\to j}\}_j$ packets.

Symmetrically, we put some of those $nt_{[\mathrm{u}]}^{(i)}\eta_2^{(i)}$ packets sent during the second fraction of Stage 1 into two disjoint groups. That is, $\{\mathbf{c}_t\mathbf{W}_{i\to k}\}_{j\overline{k}}$ and $\{\mathbf{c}_t\mathbf{W}_{i\to k}\}_k$ denote those packets that are heard by node j only, and by node k (may or may not be heard by node j), respectively. The size of each group, in average, is $nt_{[\mathrm{u}]}^{(i)}\eta_2^{(i)}p_{i\to j\overline{k}}$ and $nt_{[\mathrm{u}]}^{(i)}\eta_2^{(i)}p_{i\to k}$, respectively.

Finally, among the $nt_{[\mathrm{u}]}^{(i)}\eta_3^{(i)}$ number of the packets $\{\mathbf{c}_t\mathbf{W}_{i\to jk}\}$ sent in the third fraction $\eta_3^{(i)}$, we place them into 4 different groups but this time the groups are not necessarily disjoint. Specifically, we use $\{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_{\overline{j}k}$ and $\{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_j$ to denote, respectively, the packets that are received by node k only (not by node j) and by node j (regardless whether node k receives them). We use $\{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_{j\overline{k}}$ and $\{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_k$ to denote, respectively, those packets that are heard by node j only (not by node k) and by node k (regardless whether node j receives them). The first two groups of packets are disjoint and the last two groups of packets are disjoint. But there may be overlap between $\{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_j$ and $\{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_k$. In average, the sizes of these four groups are $nt_{[\mathrm{u}]}^{(i)}\eta_3^{(i)}p_{i\to\overline{j}k}$,

$nt_{[\mathsf{u}]}^{(i)}\eta_3^{(i)}p_{i\to j}$, $nt_{[\mathsf{u}]}^{(i)}\eta_3^{(i)}p_{i\to j\overline{k}}$, and $nt_{[\mathsf{u}]}^{(i)}\eta_3^{(i)}p_{i\to k}$, respectively.

For ease of description, we further put some of the groups of the packets into super groups. Specifically, for all $(i,j,k)\in\{(1,2,3),\,(2,3,1),(3,1,2)\}$,

$$\tilde{\mathbf{W}}_{i\to j}^{(k)} \triangleq \{\mathbf{c}_t\mathbf{W}_{i\to j}\}_{\overline{j}k} \cup \{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_{\overline{j}k}, \tag{62}$$

$$\tilde{\mathbf{W}}_{i\to k}^{(j)} \triangleq \{\mathbf{c}_t\mathbf{W}_{i\to k}\}_{j\overline{k}} \cup \{\mathbf{c}_t\mathbf{W}_{i\to jk}\}_{j\overline{k}}. \tag{63}$$

In total, there are 6 such $\tilde{\mathbf{W}}$-groups by definition and their sizes, in average, are

$$|\tilde{\mathbf{W}}_{i\to j}^{(k)}| = nt_{[\mathsf{u}]}^{(i)}\left(\eta_1^{(i)}+\eta_3^{(i)}\right)p_{i\to\overline{j}k}, \tag{64}$$

$$|\tilde{\mathbf{W}}_{i\to k}^{(j)}| = nt_{[\mathsf{u}]}^{(i)}\left(\eta_2^{(i)}+\eta_3^{(i)}\right)p_{i\to j\overline{k}}. \tag{65}$$

Description of The Feedback Stage: Thus far, the above re-grouping of the packets can be made only when one has the full knowledge of the reception status. However, right after Stage 1, no node has received any feedback yet and it is thus impossible to perform the packet regrouping as described previously. After Stage 1, we thus perform the following feedback stage so that after the feedback stage, all nodes can share a synchronized view about which packets are in which groups. Again, we emphasize that the goal of the feedback stage is to convey the reception status ACK/NACK. We never send any actual coded/uncoded messages (the payload) during the feedback stage.

Specifically, during Stage 1, each node i has been on the listening side for a total duration of $n\left(t_{[\mathsf{u}]}^{(j)}+t_{[\mathsf{u}]}^{(k)}\right)$ number of time slots. As a result, each node i can record whether it received a packet or not during those time slots and generate a single file of $n\left(t_{[\mathsf{u}]}^{(j)}+t_{[\mathsf{u}]}^{(k)}\right)$ bits. Then node i would like to deliver this file to both nodes j and k. It can be achieved by the following two-step approach. Step 1: Node i converts the file into $\left\lceil\frac{n\left(t_{[\mathsf{u}]}^{(j)}+t_{[\mathsf{u}]}^{(k)}\right)}{\log_2 q}\right\rceil$ number of packets. Then it uses an MDS code or a rate-less code to *broadcast* those packets for totally

$$\frac{n\left(t_{[\mathsf{u}]}^{(j)}+t_{[\mathsf{u}]}^{(k)}\right)}{\log_2(q)\cdot\mathsf{nzmin}\{p_{i\to j},p_{i\to k}\}} \tag{66}$$

number of time slots. As a result, if $\min(p_{i\to j},p_{i\to k})>0$, then $\mathsf{nzmin}\{p_{i\to j},p_{i\to k}\}=\min(p_{i\to j},p_{i\to k})$ and both nodes j and k can recover the file. The feedback transmission for node i is thus complete.

However, it is possible that $p_{i\to j}=0$ (or $p_{i\to k}=0$). In this case, $\mathsf{nzmin}\{p_{i\to j},p_{i\to k}\}=p_{i\to k}$ and only node k can recover the file of node i. In this case, we let node k help relay the file to node j, which will take additionally

$$\frac{n\left(t_{[\mathsf{u}]}^{(j)}+t_{[\mathsf{u}]}^{(k)}\right)}{\log_2(q)\cdot p_{k\to j}} \tag{67}$$

number[7] of time slots. The feedback stage of node i finishes after node k helps relay the file of node i. Note that the number of time slots used for the file of node i during the feedback stage can be upper bounded by

$$\frac{n}{\log_2(q)\cdot\mathsf{nzmin}\{p_{i\to j},p_{i\to k}\}} + \frac{n}{\log_2(q)\cdot\mathsf{nzmin}\{p_{j\to k},p_{j\to i}\}} + \frac{n}{\log_2(q)\cdot\mathsf{nzmin}\{p_{k\to i},p_{k\to j}\}}$$

where the first term $\frac{n}{\log_2(q)\cdot\mathsf{nzmin}\{p_{i\to j},p_{i\to k}\}}$ upper bounds (66) since $t_{[\mathsf{u}]}^{(j)}+t_{[\mathsf{u}]}^{(k)}\le 1$; and the summation $\frac{n}{\log_2(q)\cdot\mathsf{nzmin}\{p_{j\to k},p_{j\to i}\}}+\frac{n}{\log_2(q)\cdot\mathsf{nzmin}\{p_{k\to i},p_{k\to j}\}}$ upper bounds (67) regardless whether $p_{i\to j}=0$ or $p_{i\to k}=0$.

Since the feedback stage has to be executed for all three nodes, the total number of time slots of the feedback stage is upper bounded by nt_{FB}, as defined in (16).

After the feedback stage, every node will know the reception status of all other nodes during Stage 1. All three nodes can thus share a synchronized view of the packet reception status and the packet regrouping, as discussed in (62) and (63). In particular, each node i exactly knows

- The contents and size of the RLNC packet groups $(\tilde{\mathbf{W}}_{i\to j}^{(k)},\tilde{\mathbf{W}}_{i\to k}^{(j)})$. The content of the packets in each group is known since those are the messages originated from node i.
- The contents and size of the RLNC packet groups $(\tilde{\mathbf{W}}_{j\to k}^{(i)},\tilde{\mathbf{W}}_{k\to j}^{(i)})$. The content of the packets in each group is known since those are the packets overheard by node i.
- The sizes of $|\tilde{\mathbf{W}}_{j\to i}^{(k)}|$ and $|\tilde{\mathbf{W}}_{k\to i}^{(j)}|$, which are obtained by the feedback it has received from nodes j and k.
- The content of all packets in $(\{\mathbf{c}_t\mathbf{W}_{j\to i}\}_i$, $\{\mathbf{c}_t\mathbf{W}_{j\to ki}\}_i$, $\{\mathbf{c}_t\mathbf{W}_{k\to i}\}_i$, $\{\mathbf{c}_t\mathbf{W}_{k\to ij}\}_i)$ are known by node i since it has received those packets during Stage 1. Note that these are the packets that have already been delivered to their target destination, which is node i. In comparison, the $(\tilde{\mathbf{W}}_{j\to k}^{(i)},\tilde{\mathbf{W}}_{k\to j}^{(i)})$ in the second bullet are those packets destined for either node j or k but is *overheard* by i.
- The random coding vectors $\{\mathbf{c}_t\}$ for all RLNC packets sent during Stage 1. This is due to that all three nodes compute the coding vectors based on a common random seed.

Description of Stage 2: We describe the LNC operations of node i only and the operations for other nodes follow symmetrically. Similar to Stage 2 of Proposition 3, each node i will perform 4 different types of LNC operations and each operation will last for $nt_{[\mathsf{c},1]}^{(i)}$ to $nt_{[\mathsf{c},4]}^{(i)}$, respectively. For each time slot of the first coding operations (out of totally $nt_{[\mathsf{c},1]}^{(i)}$ time slots), we let node i choose two coding vectors $\mathbf{c}_{t;j}$ and $\mathbf{c}_{t;k}$ independently and uniformly randomly, where $\mathbf{c}_{t;j}$ is a $1\times|\tilde{\mathbf{W}}_{i\to j}^{(k)}|$ random row vector and $\mathbf{c}_{t;k}$ is a $1\times|\tilde{\mathbf{W}}_{i\to k}^{(j)}|$ random row vector. Then we let node i send a linear combination

$$[\mathsf{c},1]:\quad X_i(t)=\tilde{\mathbf{W}}_{i\to j}^{(k)}\mathbf{c}_{t;j}^\top+\tilde{\mathbf{W}}_{i\to k}^{(j)}\mathbf{c}_{t;k}^\top. \tag{68}$$

For the next time slot, another pair of $\mathbf{c}_{t;j}$ and $\mathbf{c}_{t;k}$ coding vectors are randomly chosen and used to encode $X_i(t)$

[7] If $p_{i\to j}=0$, then by our fully-connectedness assumption, $p_{k\to j}>0$.

according to (68). Repeat the above operations until the time-budget $nt^{(i)}_{[\mathsf{c},1]}$ is used up. Then we move on and encode the next coding type $[\mathsf{c},l], l=2,3,4$

$$[\mathsf{c},2]: \quad X_i(t) = \tilde{\mathbf{W}}^{(i)}_{k\to j}\mathbf{c}^\top_{t;j} + \tilde{\mathbf{W}}^{(i)}_{j\to k}\mathbf{c}^\top_{t;k}, \tag{69}$$

$$[\mathsf{c},3]: \quad X_i(t) = \tilde{\mathbf{W}}^{(k)}_{i\to j}\mathbf{c}^\top_{t;j} + \tilde{\mathbf{W}}^{(i)}_{j\to k}\mathbf{c}^\top_{t;k}, \tag{70}$$

$$[\mathsf{c},4]: \quad X_i(t) = \tilde{\mathbf{W}}^{(i)}_{k\to j}\mathbf{c}^\top_{t;j} + \tilde{\mathbf{W}}^{(j)}_{i\to k}\mathbf{c}^\top_{t;k}. \tag{71}$$

Each coding type $[\mathsf{c},l]$ will last for $nt^{(i)}_{[\mathsf{c},l]}$ time slots and the coding vectors $\mathbf{c}_{t;j}$ and $\mathbf{c}_{t;k}$ are chosen independently and uniformly randomly with the properly selected dimension. For example of the coding choice $[\mathsf{c},3]$, the randomly chosen $\mathbf{c}_{t;j}$ is a $1\times|\tilde{\mathbf{W}}^{(k)}_{i\to j}|$ row vector and the randomly chosen $\mathbf{c}_{t;k}$ is a $1\times|\tilde{\mathbf{W}}^{(i)}_{j\to k}|$ row vector.

Stage 2 is completed after all three nodes have finished sending their corresponding 4 coding types. The description of the proposed scheme is complete. (There is no need to have the second feedback stage.)

Analysis of the scheme: The total amount of time to finish the transmission is upper bounded by

$$n\Bigg(\Big(\sum_{\forall\, i\in\{1,2,3\}} t^{(i)}_{[\mathsf{u}]}\Big) + t_{\mathsf{FB}} + \Big(\sum_{\forall\, i\in\{1,2,3\}} t^{(i)}_{[\mathsf{c},1]} + t^{(i)}_{[\mathsf{c},2]} + t^{(i)}_{[\mathsf{c},3]} + t^{(i)}_{[\mathsf{c},4]}\Big)\Bigg).$$

By (15), we can thus finish all the transmissions within the total time budget of n time slots.

We now argue that after finishing transmission, all nodes can decode their desired packets. To that end, we focus only on node 1. The discussions of nodes 2 and 3 can be made by symmetry.

During Stage 2, consider the transmission of node 3. Node 3 has 4 possible coding choices. In each coding choice, it randomly mixes from two groups of packets. For example, in coding choice $[\mathsf{c},1]$, node 3 mixes $\tilde{\mathbf{W}}^{(2)}_{3\to1}$ and $\tilde{\mathbf{W}}^{(1)}_{3\to2}$, see (68) when $(i,j,k)=(3,1,2)$. Since the content of any packets in $\tilde{\mathbf{W}}^{(1)}_{3\to2}$ is known to node 1, see the discussion in the end of the feedback stage, node 1, upon the reception of any $[\mathsf{c},1]$ packet transmitted by node 3, can subtract the term $\tilde{\mathbf{W}}^{(1)}_{3\to2}\mathbf{c}^\top_{t;2}$ from the received packet. Therefore, it is as if node 1 has received a packet of the form

$$\tilde{\mathbf{W}}^{(2)}_{3\to1}\mathbf{c}^\top_{t;1} \tag{72}$$

without the corruption term $\tilde{\mathbf{W}}^{(1)}_{3\to2}\mathbf{c}^\top_{t;2}$. Similarly, when node 3 performs coding choice $[\mathsf{c},3]$, again, node 1 will receive coded packets of the form (72) after subtracting those $\tilde{\mathbf{W}}^{(3)}_{1\to2}\mathbf{c}^\top_{t;2}$ packets of its own, see (70) when $(i,j,k)=(3,1,2)$. Also, during node 2 performing coding choices $[\mathsf{c},2]$ and $[\mathsf{c},3]$, node 1 can again receive coded packets of the form (72) after subtracting those known packets (either of the form $\tilde{\mathbf{W}}^{(2)}_{1\to3}\mathbf{c}^\top_{t;3}$ or of the form $\tilde{\mathbf{W}}^{(1)}_{2\to3}\mathbf{c}^\top_{t;3}$), see (69) and (70) when $(i,j,k)=(2,3,1)$.

Since $\tilde{\mathbf{W}}^{(2)}_{3\to1}$ participates in coding choices $[\mathsf{c},1]$ and $[\mathsf{c},3]$ of node 3 and coding choices $[\mathsf{c},2]$ and $[\mathsf{c},3]$ of node 2, node 1 will receive $n\left(t^{(3)}_{[\mathsf{c},1]}+t^{(3)}_{[\mathsf{c},3]}\right)\cdot p_{3\to1}+n\left(t^{(2)}_{[\mathsf{c},2]}+t^{(2)}_{[\mathsf{c},3]}\right)\cdot p_{2\to1}$ number of packets of the form (72). Note that the number of $\tilde{\mathbf{W}}^{(2)}_{3\to1}$ packets, in average, has been computed in (64). By (60), the number of linear combinations (72) received by node 1 is larger than the number of $\tilde{\mathbf{W}}^{(2)}_{3\to1}$ packets to be begin with. As a result, node 1 is guaranteed to decode $\tilde{\mathbf{W}}^{(2)}_{3\to1}$ correctly with close-to-one probability when the finite-field size q is sufficiently large enough.

Recall that by definition (62), $\tilde{\mathbf{W}}^{(2)}_{3\to1}=\{\mathbf{c}_t\mathbf{W}_{3\to1}\}_{\overline{1}2}\cup\{\mathbf{c}_t\mathbf{W}_{3\to12}\}_{\overline{1}2}$. We now observe that node 1 has also received all the RLNC packets of $(\{\mathbf{c}_t\mathbf{W}_{3\to1}\}_1, \{\mathbf{c}_t\mathbf{W}_{3\to12}\}_1)$ during Stage 1. As a result, in the end of Stage 2, node 1 has correctly received $nt^{(3)}_{[\mathsf{u}]}\eta^{(3)}_1 p_{3\to1\vee2}$ number of packets of the form $\{\mathbf{c}_t\mathbf{W}_{3\to1}\}$ and $nt^{(3)}_{[\mathsf{u}]}\eta^{(3)}_3 p_{3\to1\vee2}$ number of packets of the form $\{\mathbf{c}_t\mathbf{W}_{3\to12}\}$. Note that we only have $nR_{3\to1}$ of $\mathbf{W}_{3\to1}$ packets and $nR_{3\to12}$ of $\mathbf{W}_{3\to12}$ packets to begin with. Since by definition $t^{(3)}_{[\mathsf{u}]}$ is strictly larger than $\frac{R_{3\to1}+R_{3\to2}+R_{3\to12}}{p_{3\to1\vee2}}$, and also by definitions (57) and (59), the number of linear combinations received by node 1 is larger than the number of uncoded message symbols $\mathbf{W}_{3\to1}$ and $\mathbf{W}_{3\to12}$. As a result, node 1 is guaranteed to decode $\mathbf{W}_{3\to1}$ and $\mathbf{W}_{3\to12}$ correctly with close-to-one probability when the finite field size q is sufficiently large enough.

By symmetric arguments, with close-to-one probability node 1 can also decode $\tilde{\mathbf{W}}^{(3)}_{2\to1}$ in the end of Stage 2 and later combines $\tilde{\mathbf{W}}^{(3)}_{2\to1}$ with the packets $(\{\mathbf{c}_t\mathbf{W}_{2\to1}\}_1, \{\mathbf{c}_t\mathbf{W}_{2\to31}\}_1)$ it has received in Stage 1 to decode message symbols $\mathbf{W}_{2\to1}$ and $\mathbf{W}_{2\to31}$. Symmetric arguments can be used to shown that nodes 2 and 3 can also decode their desired messages. The proof of Proposition 2 is thus complete.

Remark: The arguments of letting the finite field size approach infinity is to ensure that the simple RLNC construction leads to legitimate MDS codes. When the finite field size is small, say $q=2$, we can use the fact that for any fixed $\mathbb{F}_q$ we can always construct a code with dimension k that is *nearly MDS* in the sense that as long as we receive $k+O(\sqrt{k})$ number of encoded packets we can reconstruct the original file. Since we focus only on the normalized throughput, such a near-MDS code is sufficient for our achievability construction.

APPENDIX C
DETAILED DESCRIPTION OF ACHIEVABILITY SCHEMES IN FIG. 3

In the following, we describe the 9-dimensional rate regions of each suboptimal achievability scheme used for the numerical evaluation in Section IV-E.

• **LNC with pure operations 1, 2:** The rate regions can be described by Proposition 3 with the variables $t^{(i)}_{[\mathsf{c},3]}$ and $t^{(i)}_{[\mathsf{c},4]}$ hardwired to 0 for all $i\in\{1,2,3\}$.

• **TWRC at node** 1 **and RX coord.:** This scheme performs two-way relay channel (TWRC) coding only at node 1 for those $3\to2$ and $2\to3$ flows while allowing both nodes 2 and 3 to relay the node 1's packets destined to each other. That is, node 2 can relay $\mathbf{W}_{1\to3}$ and $\mathbf{W}_{1\to23}$ for node 3, and node 3 can relay $\mathbf{W}_{1\to2}$ and $\mathbf{W}_{1\to23}$ for node 2. The corresponding rate regions can be described as follows:

$$\sum_{\forall i\in\{1,2,3\}} t^{(i)}_{[\mathrm{u}]} + t^{(i)}_{[\mathrm{c}]} \leq 1, \tag{73}$$

$$\frac{R_{1\to 2}+R_{1\to 3}+R_{1\to 23}}{p_{1\to 2\vee 3}} < t^{(1)}_{[\mathrm{u}]}, \tag{74}$$

$$\frac{R_{2\to 1}}{p_{2\to 1}} + \frac{R_{2\to 3}}{p_{2\to 3\vee 1}} + \frac{R_{2\to 31}}{p_{2\to 1}} + \frac{R_{2\to 31}}{p_{2\to 3}} < t^{(2)}_{[\mathrm{u}]}, \tag{75}$$

$$\frac{R_{3\to 1}}{p_{3\to 1}} + \frac{R_{3\to 2}}{p_{3\to 1\vee 2}} + \frac{R_{3\to 12}}{p_{3\to 1}} + \frac{R_{3\to 12}}{p_{3\to 2}} < t^{(3)}_{[\mathrm{u}]}, \tag{76}$$

$$R_{2\to 3}\frac{p_{2\to\overline{3}1}}{p_{2\to 3\vee 1}} < t^{(1)}_{[\mathrm{c}]}\cdot p_{1\to 3}, \tag{77}$$

$$R_{3\to 2}\frac{p_{3\to 1\overline{2}}}{p_{3\to 1\vee 2}} < t^{(1)}_{[\mathrm{c}]}\cdot p_{1\to 2}, \tag{78}$$

$$\Big(R_{1\to 3}+R_{1\to 23}\Big)\frac{p_{1\to 2\overline{3}}}{p_{1\to 2\vee 3}} < t^{(2)}_{[\mathrm{c}]}\cdot p_{2\to 3}, \tag{79}$$

$$\Big(R_{1\to 2}+R_{1\to 23}\Big)\frac{p_{1\to \overline{2}3}}{p_{1\to 2\vee 3}} < t^{(3)}_{[\mathrm{c}]}\cdot p_{3\to 2}. \tag{80}$$

Namely, each node i has two variables $t^{(i)}_{[\mathrm{u}]}$ and $t^{(i)}_{[\mathrm{c}]}$ for the respective stages, see (73). During Stage 1, node 1 repeatedly transmits its packets uncodedly until at least one of nodes 2 and 3 receives it. This stage can be finished within $nt^{(1)}_{[\mathrm{u}]}$ time slots, see (74). For node 2, we send all $\mathbf{W}_{2\to 1}$ and all $\mathbf{W}_{2\to 31}$ messages directly to node 1 and send all $\mathbf{W}_{2\to 31}$ directly to node 3; but we send all $\mathbf{W}_{2\to 3}$ messages uncodedly until at least one of the nodes 1 and 3 receives it. Such an *uncoded stage* can be finished in $nt^{(2)}_{[\mathrm{u}]}$ time slots, see (75). Node 3's uncoded stage is symmetric to that of node 2.

Eq. (77) to (78) allow node 1 to perform Two-Way-Relay coding over the $3\to 2$ and $2\to 3$ packets overheard at node 1. (79) allows node 2 to relay those packets it has overheard from node 1 to the desired destination node 3. (80) is symmetric to (79).

• **[14] & Time-sharing:** The rate regions can be described by Proposition 3 with the variables $t^{(i)}_{[\mathrm{c},2]}$, $t^{(i)}_{[\mathrm{c},3]}$, and $t^{(i)}_{[\mathrm{c},4]}=0$ hardwired to 0 for all $i\in\{1,2,3\}$. Namely, we only allow, as in [14], the broadcast channel LNC of coding choice $[\mathrm{c},1]$ during the Stage 2.

• **Uncoded direct TX:** This scheme does not perform any coding operation when transmitting, and just uncodedly transmits packets one by one until the desired receivers receive it. The rate region of this primitive scheme can be described by

$$\sum_{\forall i\in\{1,2,3\}} \frac{R_{i\to j}}{p_{i\to j}} + \frac{R_{i\to k}}{p_{i\to k}} + \frac{R_{i\to jk}}{p_{i\to j}} + \frac{R_{i\to jk}}{p_{i\to k}} \leq 1.$$

APPENDIX D
PROOFS OF THREE CLAIMS IN SECTION V-B

Before we prove Claims 1 to 3, we prove the following useful lemmas.

Lemma 2. Consider Scenario 1 and any fixed $t\in\{1,\cdots,n\}$. Then, knowing all the messages $\mathbf{W}_{\{1,2,3\}*}$ and the past channel outputs $[\mathbf{Z}]_1^{t-1}$ can uniquely decide $[X_1,X_2,X_3]_1^t$ and $[\mathbf{Y}_{1*},\mathbf{Y}_{2*},\mathbf{Y}_{3*}]_1^{t-1}$. Namely, $[X_1,X_2,X_3]_1^t$ and $[\mathbf{Y}_{1*},\mathbf{Y}_{2*},\mathbf{Y}_{3*}]_1^{t-1}$ are functions of the random variables $\{\mathbf{W}_{\{1,2,3\}*},[\mathbf{Z}]_1^{t-1}\}$ for any time $t\in\{1,\cdots,n\}$.

Proof of Lemma 2: The proof follows from the induction on time t. When $t=1$, each node i encodes the input symbol $X_i(1)$ purely based on its information messages $\mathbf{W}_{i*}$, see (7). As a result, $\{X_1(1),X_2(1),X_3(1)\}$ can be uniquely determined by $\mathbf{W}_{\{1,2,3\}*}$. Lemma 2 thus holds for $t=1$.

Suppose that the statement of Lemma 2 is true until time $t=t_0-1$. Consider $t=t_0$. By induction, $[X_1,X_2,X_3]_1^{t_0-1}$ can be uniquely decided by $\mathbf{W}_{\{1,2,3\}*}$ and $[\mathbf{Z}]_1^{t_0-2}$. Since $[\mathbf{Y}_{1*},\mathbf{Y}_{2*},\mathbf{Y}_{3*}]_1^{t_0-1}$ is a function of $[X_1,X_2,X_3]_1^{t_0-1}$ and $[\mathbf{Z}]_1^{t_0-1}$, we know that $[\mathbf{Y}_{1*},\mathbf{Y}_{2*},\mathbf{Y}_{3*}]_1^{t_0-1}$ can be uniquely decided by $\mathbf{W}_{\{1,2,3\}*}$ and $[\mathbf{Z}]_1^{t_0-1}$. Then by the encoding functions in (7), the input symbols $\{X_1(t_0),X_2(t_0),X_3(t_0)\}$ at time $t=t_0$ can be uniquely determined as well. The proof of Lemma 2 is thus complete. □

Lemma 3. Consider Scenario 1 and any fixed time slot $t\in\{1,\cdots,n\}$. Then, knowing the messages $\mathbf{W}_{\{1,3\}*}$, the received symbols $[\mathbf{Y}_{2*}]_1^{t-1}$, and the past channel outputs $[\mathbf{Z}]_1^{t-1}$ can uniquely decide $[X_1,X_3]_1^t$. Namely, $[X_1,X_3]_1^t$ is a function of the random variables $\{\mathbf{W}_{\{1,3\}*},[\mathbf{Y}_{2*}]_1^{t-1},[\mathbf{Z}]_1^{t-1}\}$ for any time $t\in\{1,\cdots,n\}$.

Proof of Lemma 3: Similar to Lemma 2, the proof follows from induction on time t. When $t=1$, in the beginning of time slot 1, $X_1(1)$ (resp. $X_3(1)$) is encoded purely based on the message $\mathbf{W}_{1*}$ (resp. $\mathbf{W}_{3*}$), see (7). As a result, $\{X_1(1),X_3(1)\}$ can be uniquely determined by $\mathbf{W}_{\{1,3\}*}$.

Assume that the statement of Lemma 3 is true until time $t=t_0-1$. By induction, $[X_1,X_3]_1^{t_0-1}$ can be uniquely determined by $\{\mathbf{W}_{\{1,3\}*},[\mathbf{Y}_{2*}]_1^{t_0-2},[\mathbf{Z}]_1^{t_0-2}\}$. Now consider time $t=t_0$. Compared to time $t=t_0-1$, we know additionally $Y_{2\to 1}(t_0-1)$, $Y_{2\to 3}(t_0-1)$, and $\mathbf{Z}(t_0-1)$. Since we already knew $[X_3]_1^{t_0-1}$, the received symbols $[Y_{3\to 1}]_1^{t_0-1}$ can be uniquely determined from the given $[\mathbf{Z}]_1^{t_0-1}$. Jointly with the known messages $\mathbf{W}_{1*}$, the received symbols $[Y_{2\to 1}]_1^{t_0-1}$, and $[\mathbf{Z}]_1^{t_0-1}$, we can also uniquely determine $X_1(t_0)$, see the encoding function of node 1 in (7). The proof regarding to $X_3(t_0)$ can be done by symmetry. The proof of Lemma 3 is thus complete. □

D-1. Proof of Claim 1

The equality (46) in Claim 1 can be proven as follows. Notice that

$$\begin{aligned}
&I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}]_1^n \,|\, \mathbf{W}_{1*},[\mathbf{Z}]_1^n)\\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) - I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) && (81)\\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) && (82)\\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{1*})\\
&\quad + I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Z}(n) \,|\, [\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1},\mathbf{W}_{1*})\\
&\quad + I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(n) \,|\, [\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1},\mathbf{W}_{1*},\mathbf{Z}(n)) && (83)\\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{1*})\\
&\quad + I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(n) \,|\, [\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1},\mathbf{W}_{1*},\mathbf{Z}(n)), && (84)
\end{aligned}$$

where (81) follows from the chain rule; (82) follows from the fact that $\mathbf{W}_{\{2,3\}*}$, $\mathbf{W}_{1*}$ and $[\mathbf{Z}]_1^n$ are independent with each other; (83) follows from the chain rule; and (84) can be obtained by showing that the second term of (83) is zero. The reason is that by our problem formulation, $\mathbf{Z}(n)$ is independent of $\mathbf{W}_{\{2,3\}*}$, $[\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}$, and $\mathbf{W}_{1*}$.

By iteratively applying the equalities (83) to (84), we have proven Claim 1.

D-2. Proof of Claim 2

For any fixed deterministic channel realization $[\mathbf{z}]_1^{t-1}$, we will consider the mutual information terms in (49), conditioning on the event $[\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}$. For notational simplicity, we use $\vec{\mathbf{z}}$ to denote the deterministic channel realization $[\mathbf{z}]_1^{t-1}$ of interest and use $\langle\vec{\mathbf{z}}\rangle \triangleq \{[\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}\}$ to denote the corresponding event.

For any fixed deterministic $\vec{\mathbf{z}}$ and fixed time instant t, we define

$$\mathsf{term}_0^{[\vec{\mathbf{z}}]} \triangleq I(\mathbf{W}_{\{2,3\}*}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)), \tag{85}$$

$$\mathsf{term}_1^{[\vec{\mathbf{z}}]} \triangleq I(\mathbf{W}_{*1}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)), \tag{86}$$

$$\mathsf{term}_2^{[\vec{\mathbf{z}}]} \triangleq I(\mathbf{W}_{2\to 3}\,;\, \mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to 3}}, \mathbf{Z}(t)), \tag{87}$$

$$\mathsf{term}_3^{[\vec{\mathbf{z}}]} \triangleq I(\mathbf{W}_{3\to 2}\,;\, \mathbf{Y}_{3*}(t) \,|\, [\mathbf{Y}_{3*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{3\to 2}}, \mathbf{Z}(t)). \tag{88}$$

By the definition of mutual information, we have

$$I(\mathbf{W}_{\{2,3\}*}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{t-1}, \mathbf{W}_{1*}, \mathbf{Z}(t)) = \sum_{\forall \vec{\mathbf{z}}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = \vec{\mathbf{z}}) \cdot \mathsf{term}_0^{[\vec{\mathbf{z}}]}, \tag{89}$$

$$I(\mathbf{W}_{*1}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{t-1}, \mathbf{W}_{1*}, \mathbf{Z}(t)) = \sum_{\forall \vec{\mathbf{z}}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = \vec{\mathbf{z}}) \cdot \mathsf{term}_1^{[\vec{\mathbf{z}}]}, \tag{90}$$

$$I(\mathbf{W}_{2\to 3}\,;\, \mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^{t-1}, \mathbf{W}_{\overline{2\to 3}}, \mathbf{Z}(t)) = \sum_{\forall \vec{\mathbf{z}}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = \vec{\mathbf{z}}) \cdot \mathsf{term}_2^{[\vec{\mathbf{z}}]}, \tag{91}$$

$$I(\mathbf{W}_{3\to 2}\,;\, \mathbf{Y}_{3*}(t) \,|\, [\mathbf{Y}_{3*}, \mathbf{Z}]_1^{t-1}, \mathbf{W}_{\overline{3\to 2}}, \mathbf{Z}(t)) = \sum_{\forall \vec{\mathbf{z}}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = \vec{\mathbf{z}}) \cdot \mathsf{term}_3^{[\vec{\mathbf{z}}]}. \tag{92}$$

Comparing (49) and equalities (89) to (92), it is clear that we only need to prove that for all $\vec{\mathbf{z}}$, the following inequality holds:

$$\mathsf{term}_0^{[\vec{\mathbf{z}}]} \geq \mathsf{term}_1^{[\vec{\mathbf{z}}]} + \frac{p_{2\to 1}}{p_{2\to 3\vee 1}}\mathsf{term}_2^{[\vec{\mathbf{z}}]} + \frac{p_{3\to 1}}{p_{3\to 1\vee 2}}\mathsf{term}_3^{[\vec{\mathbf{z}}]}. \tag{93}$$

To prove (93), we first partition all the past channel status realizations $\vec{\mathbf{z}}$ into three disjoint sets, depending on the value of the scheduling decision $\sigma(t)$, see (9). That is, for all $i \in \{1,2,3\}$,

$$\mathcal{Z}_i \triangleq \{\, \vec{\mathbf{z}} : \sigma(t) = i \,\}.$$

This partition can be done uniquely since the scheduling decision $\sigma(t)$ is a function of the past channel status $[\mathbf{Z}]_1^{t-1}$.

We now prove (93) depending on to which $\mathcal{Z}_i$ the realization vector $\vec{\mathbf{z}}$ belong. Specifically, we will prove the following:

- For all $\vec{\mathbf{z}} \in \mathcal{Z}_1$, we have

$$\mathsf{term}_0^{[\vec{\mathbf{z}}]} = 0, \tag{94}$$

$$\mathsf{term}_1^{[\vec{\mathbf{z}}]} = 0, \tag{95}$$

$$\mathsf{term}_2^{[\vec{\mathbf{z}}]} = 0, \tag{96}$$

$$\mathsf{term}_3^{[\vec{\mathbf{z}}]} = 0. \tag{97}$$

- For all $\vec{\mathbf{z}} \in \mathcal{Z}_2$, we have

$$\mathsf{term}_0^{[\vec{\mathbf{z}}]} \geq \mathsf{term}_1^{[\vec{\mathbf{z}}]} + \frac{p_{2\to 1}}{p_{2\to 3\vee 1}}\mathsf{term}_2^{[\vec{\mathbf{z}}]}, \tag{98}$$

$$\mathsf{term}_3^{[\vec{\mathbf{z}}]} = 0. \tag{99}$$

- For all $\vec{\mathbf{z}} \in \mathcal{Z}_3$, we have

$$\mathsf{term}_0^{[\vec{\mathbf{z}}]} \geq \mathsf{term}_1^{[\vec{\mathbf{z}}]} + \frac{p_{3\to 1}}{p_{3\to 1\vee 2}}\mathsf{term}_3^{[\vec{\mathbf{z}}]}, \tag{100}$$

$$\mathsf{term}_2^{[\vec{\mathbf{z}}]} = 0. \tag{101}$$

Then, one can see that (94) to (101) jointly imply that (93) holds for all the past channel output realizations $\vec{\mathbf{z}}$.

Consider the first case in which $\vec{\mathbf{z}} \in \mathcal{Z}_1$. (94) is true because

$$\begin{aligned}
&\mathsf{term}_0^{[\vec{\mathbf{z}}]} \triangleq I(\mathbf{W}_{\{2,3\}*}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\
&\leq H(\mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle) && (102) \\
&= 0, && (103)
\end{aligned}$$

where (102) follows from the definition of mutual information, non-negativity of entropy, and the fact that conditioning reduces entropy; and (103) follows from that, when the scheduling decision is $\sigma(t) = 1$, the received symbols at node 1, i.e., $\mathbf{Y}_{*1}(t)$, are always erasure.

Similarly applying the above arguments, one can prove that (95) to (97) are true as well when $\vec{\mathbf{z}} \in \mathcal{Z}_1$. The first case is thus proven.

Consider the second case in which $\vec{\mathbf{z}} \in \mathcal{Z}_2$. By the same argument as used in proving (94) to (97), we can easily prove (99). We now prove (98). Then notice that

$$\begin{aligned}
&\mathsf{term}_0^{[\vec{\mathbf{z}}]} \triangleq I(\mathbf{W}_{\{2,3\}*}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\
&= I(\mathbf{W}_{*1}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\
&\quad + I(\mathbf{W}_{3\to 2}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{1*}, \mathbf{W}_{*1}, \mathbf{Z}(t)) \\
&\quad + I(\mathbf{W}_{2\to 3}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to 3}}, \mathbf{Z}(t)) && (104) \\
&\geq \mathsf{term}_1^{[\vec{\mathbf{z}}]} \\
&\quad + I(\mathbf{W}_{2\to 3}\,;\, \mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to 3}}, \mathbf{Z}(t)) && (105) \\
&= \mathsf{term}_1^{[\vec{\mathbf{z}}]} \\
&\quad + I(\mathbf{W}_{2\to 3}\,;\, Y_{2\to 1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to 3}}, \mathbf{Z}(t)), && (106)
\end{aligned}$$

where (104) follows from the chain rule and the fact that $\mathbf{W}_{1*} \cup \mathbf{W}_{*1} \cup \mathbf{W}_{3\to 2}$ contains all 9-flow messages except for $\mathbf{W}_{2\to 3}$, which, by definition (47), equals $\mathbf{W}_{\overline{2\to 3}}$. (105) follows from the definition (86) and the non-negativity of mutual information. (106) follows from that when $\vec{\mathbf{z}} \in \mathcal{Z}_2$, the received symbol $Y_{3\to 1}(t) \subset \mathbf{Y}_{*1}(t)$ is always erasure.

The second term in the RHS of (106) satisfies

$$\begin{aligned}
&I(\mathbf{W}_{2\to 3}\,;\, Y_{2\to 1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to 3}}, \mathbf{Z}(t)) \\
&= \frac{p_{2\to 1}}{p_{2\to 3\vee 1}} I(\mathbf{W}_{2\to 3}\,;\, \mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to 3}}, \mathbf{Z}(t)).
\end{aligned} \tag{107}$$

Proof of (107): For the ease of exposition, let us denote $\mathbf{V} \triangleq \{[\mathbf{Y}_{*1}]_1^{t-1}, \mathbf{W}_{\overline{2\to3}}\}$. Rewriting (107), we thus need to prove

$$\begin{aligned} &I(\mathbf{W}_{2\to3} \,;\, Y_{2\to1}(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle, \mathbf{Z}(t)) \\ &= \frac{p_{2\to1}}{p_{2\to3\vee1}} I(\mathbf{W}_{2\to3} \,;\, \mathbf{Y}_{2*}(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle, \mathbf{Z}(t)). \end{aligned} \tag{108}$$

Since $\vec{\mathbf{z}} \in \mathcal{Z}_2$, we have $Y_{2\to1}(t) = X_2(t) \circ Z_{2\to1}(t)$. Since $Z_{2\to1}(t)$ is independent of $\mathbf{W}_{2\to3}$, $X_2(t)$, $\mathbf{V}$, and the random event $\langle\vec{\mathbf{z}}\rangle$, we thus have

$$\begin{aligned} &I(\mathbf{W}_{2\to3} \,;\, Y_{2\to1}(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle, \mathbf{Z}(t)) \\ &= \mathsf{Prob}(Z_{2\to1}(t) = 1) \cdot I(\mathbf{W}_{2\to3} \,;\, X_2(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle) \\ &= p_{2\to1} \cdot I(\mathbf{W}_{2\to3} \,;\, X_2(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle). \end{aligned} \tag{109}$$

By similar arguments, we can also prove that

$$\begin{aligned} &I(\mathbf{W}_{2\to3} \,;\, \mathbf{Y}_{2*}(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle, \mathbf{Z}(t)) \\ &= \mathsf{Prob}(\{Z_{2\to1}(t) = 1\} \cup \{Z_{2\to3}(t) = 1\}) \\ &\qquad\qquad \cdot I(\mathbf{W}_{2\to3} \,;\, X_2(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle) \\ &= p_{2\to3\vee1} \cdot I(\mathbf{W}_{2\to3} \,;\, X_2(t) \,|\, \mathbf{V}, \langle\vec{\mathbf{z}}\rangle). \end{aligned} \tag{110}$$

Equalities (109) and (110) jointly imply (108), which completes the proof of (107). $\square$

Then we observe that the mutual information term on the RHS of (107) also satisfies

$$\begin{aligned} &I(\mathbf{W}_{2\to3} \,;\, \mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(t)) \\ &= H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{W}_{2\to3}, \mathbf{Z}(t)) \quad (111) \\ &= H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{W}_{2\to3}, \mathbf{Z}(t)) \quad (112) \\ &\geq H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}, \mathbf{Y}_{3*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{W}_{2\to3}, \mathbf{Z}(t)) \quad (113) \\ &= H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{W}_{2\to3}, \mathbf{Z}(t)) \quad (114) \\ &= I(\mathbf{W}_{2\to3} \,;\, \mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}]_1^{t-1}, \langle\vec{\mathbf{z}}\rangle, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(t)) \quad (115) \\ &= \mathsf{term}_2^{[\vec{\mathbf{z}}]}, \quad (116) \end{aligned}$$

where (111) follows from the definition of mutual information; (112) follows from that (i) $\mathbf{W}_{\overline{2\to3}} \cup \mathbf{W}_{2\to3}$ contains all the 9-flow information messages $\mathbf{W}_{\{1,2,3\}*}$, and (ii) by Lemma 2, both $[\mathbf{Y}_{*1}]_1^{t-1}$ and $[\mathbf{Y}_{2*}]_1^{t-1}$ can be uniquely computed once we know all the messages $\mathbf{W}_{\{1,2,3\}*} = \mathbf{W}_{\overline{2\to3}} \cup \mathbf{W}_{2\to3}$ and the past channel realizations $\vec{\mathbf{z}} = [\mathbf{z}]_1^{t-1}$. Therefore, the conditional entropy remains identical even when we substitute $[\mathbf{Y}_{*1}]_1^{t-1}$ by $[\mathbf{Y}_{2*}]_1^{t-1}$; (113) follows from the fact that conditioning reduces entropy; (114) follows from Lemma 3 that knowing the messages $\{\mathbf{W}_{1*}, \mathbf{W}_{3*}\} \subset \mathbf{W}_{\overline{2\to3}}$, the received symbols $[\mathbf{Y}_{2*}]_1^{t-1}$, and the past channel realizations $\vec{\mathbf{z}} = [\mathbf{z}]_1^{t-1}$ can uniquely decide $[X_3]_1^t$, and thus also the received symbols $[\mathbf{Y}_{3*}]_1^{t-1}$ (since $[\mathbf{z}]_1^{t-1}$ is known). As a result, removing $[\mathbf{Y}_{3*}]_1^{t-1}$ in the first term of (113) will not change the conditional entropy; (115) follows from the definition of mutual information; and (116) follows from the definition (87).

Jointly (106), (107), and (116) imply (98).

The third case, $\vec{\mathbf{z}} \in \mathcal{Z}_3$, is symmetric to the case of $\vec{\mathbf{z}} \in \mathcal{Z}_2$. The proof of Claim 2 is thus complete.

D-3. *Proof of Claim 3*

We provide the proofs for the (in)equalities (50) to (52) in Claim 3. We first show the proof for (50).

Proof of (50): Note that

$$\begin{aligned} &I(\mathbf{W}_{*1} \,;\, \mathbf{W}_{1*}, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^n) \\ &= I(\mathbf{W}_{*1} \,;\, \mathbf{W}_{1*}) + I(\mathbf{W}_{*1} \,;\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) \quad (117) \\ &= I(\mathbf{W}_{*1} \,;\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) \quad (118) \\ &= I(\mathbf{W}_{*1} \,;\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{1*}) \\ &\quad + I(\mathbf{W}_{*1} \,;\, \mathbf{Z}(n) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{1*}) \\ &\quad + I(\mathbf{W}_{*1} \,;\, \mathbf{Y}_{*1}(n) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{1*}, \mathbf{Z}(n)) \quad (119) \\ &= I(\mathbf{W}_{*1} \,;\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{1*}) \\ &\quad + I(\mathbf{W}_{*1} \,;\, \mathbf{Y}_{*1}(n) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{1*}, \mathbf{Z}(n)), \quad (120) \end{aligned}$$

where (117) follows from the chain rule; (118) follows from the fact the messages $\mathbf{W}_{*1}$ and $\mathbf{W}_{1*}$ are independent with each other; (119) follows from the chain rule; and (120) can be obtained by showing that the second term of (119) is zero. The reason is because $\mathbf{Z}(n)$ is independent of $\mathbf{W}_{*1}$, $[\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}$, and $\mathbf{W}_{1*}$. By iteratively applying the equalities (119) to (120) for $t = n-1$ back to $t = 1$, the result (50) follows. $\square$

Secondly, we prove (51). The proof of (52) can be derived symmetrically by swapping the node indices 2 and 3.

Proof of (51): Note that

$$\begin{aligned} &I(\mathbf{W}_{2\to3} \,;\, \mathbf{W}_{3*}, [\mathbf{Y}_{*3}, \mathbf{Z}]_1^n) \\ &\leq I(\mathbf{W}_{2\to3} \,;\, \mathbf{W}_{\{1,3\}*}, \mathbf{W}_{2\to1}, \mathbf{W}_{2\to31}, [\mathbf{Y}_{2*}, Y_{1\to3}, \mathbf{Z}]_1^n) \quad (121) \\ &= I(\mathbf{W}_{2\to3} \,;\, \mathbf{W}_{\{1,3\}*}, \mathbf{W}_{2\to1}, \mathbf{W}_{2\to31}, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^n) \quad (122) \\ &= I(\mathbf{W}_{2\to3} \,;\, \mathbf{W}_{\overline{2\to3}}) + I(\mathbf{W}_{2\to3} \,;\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^n \,|\, \mathbf{W}_{\overline{2\to3}}), \quad (123) \\ &= I(\mathbf{W}_{2\to3} \,;\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^n \,|\, \mathbf{W}_{\overline{2\to3}}), \quad (124) \\ &= I(\mathbf{W}_{2\to3} \,;\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{\overline{2\to3}}) \\ &\quad + I(\mathbf{W}_{2\to3} \,;\, \mathbf{Z}(n) \,|\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{\overline{2\to3}}) \\ &\quad + I(\mathbf{W}_{2\to3} \,;\, \mathbf{Y}_{2*}(n) \,|\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(n)) \quad (125) \\ &= I(\mathbf{W}_{2\to3} \,;\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{\overline{2\to3}}) \\ &\quad + I(\mathbf{W}_{2\to3} \,;\, \mathbf{Y}_{2*}(n) \,|\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{\overline{2\to3}}, \mathbf{Z}(n)), \quad (126) \end{aligned}$$

where (121) follows from the fact that adding the observations $\mathbf{W}_{1*}$, $\mathbf{W}_{2\to1}$, $\mathbf{W}_{2\to31}$, and $[Y_{2\to1}]_1^n$ increases the mutual information; (122) follows from Lemma 3 that $[X_1]_1^n$ is a function of $\mathbf{W}_{\{1,3\}*}$, $[\mathbf{Y}_{2*}]_1^{n-1}$, and $[\mathbf{Z}]_1^{n-1}$, which in turn implies that $[Y_{1\to3}]_1^n$ is a function of $\mathbf{W}_{\{1,3\}*}$, $[\mathbf{Y}_{2*}]_1^{n-1}$, and $[\mathbf{Z}]_1^n$ since $[Y_{1\to3}]_1^n$ is a function of $[X_1]_1^n$ and $[\mathbf{Z}]_1^n$. As a result, removing $[Y_{1\to3}]_1^n$ does not decrease the mutual information; (123) follows from the chain rule and the definition of $\mathbf{W}_{\overline{2\to3}}$ in (47); (124) follows from the fact the messages $\mathbf{W}_{2\to3}$ and $\mathbf{W}_{\overline{2\to3}}$ are independent of each other; (125) follows from the chain rule; and (126) follows from the second term of (125) being zero, since $\mathbf{Z}(n)$ is independent of $\mathbf{W}_{2\to3}$, $[\mathbf{Y}_{2*}, \mathbf{Z}]_1^{n-1}$, and $\mathbf{W}_{\overline{2\to3}}$. By iteratively applying the equalities (125) to (126), the inequality (51) follows. $\square$

APPENDIX E
PROOF OF PROPOSITION 4

Without loss of generality, we assume that $p_{i\to j} > 0$ and $p_{i\to k} > 0$ for all $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$ since the case that any one of them is zero can be viewed as a limiting scenario and the polytope of the capacity outer bound in Proposition 1 is continuous with respect to the channel success probability parameters.

We first introduce the following Lemma.

Lemma 4. Given any $\vec{R}$ and the associated 3 non-negative values $\{s^{(i)}\}$ that satisfy Proposition 1, we can always find 15 non-negative values $t^{(i)}_{[\mathrm{u}]}$ and $\{t^{(i)}_{[\mathrm{c},l]}\}_{l=1}^4$ for all $i \in \{1,2,3\}$ such that jointly they satisfy the groups of linear conditions in Proposition 3 (when replacing all strict inequality $<$ by $\leq$).

One can clearly see that Lemma 4 implies that the capacity outer bound in Proposition 1 matches the closure of the inner bound in Proposition 3. The proof of Proposition 4 is thus complete.

Proof of Lemma 4: Given $\vec{R}$ and the reception probabilities, consider 3 non-negative values $\{s^{(i)}\}$ that jointly satisfy the linear conditions of Proposition 1.

We first choose $t^{(i)}_{[\mathrm{u}]} \triangleq \frac{R_{i\to j}+R_{i\to k}+R_{i\to jk}}{p_{i\to j\vee k}}$ which is non-negative by definition. Then define $\tilde{s}^{(i)} \triangleq s^{(i)} - t^{(i)}_{[\mathrm{u}]}$ for all $i \in \{1,2,3\}$. By (12) in Proposition 1, the newly constructed values $\{\tilde{s}^{(i)}\}$ must be non-negative. Then, we can rewrite (13) in Proposition 1 as follows: For all $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$, we have

$$\Big(R_{j\to i}+R_{j\to ki}\Big)\frac{p_{j\to k\bar{i}}}{p_{j\to k\vee i}}+\Big(R_{k\to i}+R_{k\to ij}\Big)\frac{p_{k\to \bar{i}j}}{p_{k\to i\vee j}} \leq \tilde{s}^{(j)}\cdot p_{j\to i}+\tilde{s}^{(k)}\cdot p_{k\to i}.$$

For each tuple (i,j,k), define a constant α_{ijk} as follows:

$$\alpha_{ijk} = \frac{\Big(R_{j\to i}+R_{j\to ki}\Big)\frac{p_{j\to k\bar{i}}}{p_{j\to k\vee i}}}{\Big(R_{j\to i}+R_{j\to ki}\Big)\frac{p_{j\to k\bar{i}}}{p_{j\to k\vee i}}+\Big(R_{k\to i}+R_{k\to ij}\Big)\frac{p_{k\to \bar{i}j}}{p_{k\to i\vee j}}}.$$

For each tuple (i,j,k), we will use α_{ijk}, $\tilde{s}^{(j)}$ and $\tilde{s}^{(k)}$ to define/compute 4 more variables.

$$\begin{aligned}
\tilde{s}^{(j)}_{ijk,+} &= \alpha_{ijk}\cdot \tilde{s}^{(j)},\\
\tilde{s}^{(k)}_{ijk,+} &= \alpha_{ijk}\cdot \tilde{s}^{(k)},\\
\tilde{s}^{(j)}_{ijk,-} &= (1-\alpha_{ijk})\cdot \tilde{s}^{(j)},\\
\tilde{s}^{(k)}_{ijk,-} &= (1-\alpha_{ijk})\cdot \tilde{s}^{(k)}.
\end{aligned}$$

By the above construction, we quickly have

$$\tilde{s}^{(j)}_{ijk,+}+\tilde{s}^{(j)}_{ijk,-} = \tilde{s}^{(j)}, \tag{127}$$
$$\tilde{s}^{(k)}_{ijk,+}+\tilde{s}^{(k)}_{ijk,-} = \tilde{s}^{(k)}, \tag{128}$$

and

$$\Big(R_{j\to i}+R_{j\to ki}\Big)\frac{p_{j\to k\bar{i}}}{p_{j\to k\vee i}} \leq \tilde{s}^{(j)}_{ijk,+}\cdot p_{j\to i}+\tilde{s}^{(k)}_{ijk,+}\cdot p_{k\to i}, \tag{129}$$

$$\Big(R_{k\to i}+R_{k\to ij}\Big)\frac{p_{k\to \bar{i}j}}{p_{k\to i\vee j}} \leq \tilde{s}^{(j)}_{ijk,-}\cdot p_{j\to i}+\tilde{s}^{(k)}_{ijk,-}\cdot p_{k\to i}, \tag{130}$$

for every cyclically shifted (i,j,k) tuple. Totally, we have 3 variables of the form $\tilde{s}^{(i)}$ and 12 variables of the forms $\tilde{s}^{(j)}_{ijk,+}$, $\tilde{s}^{(j)}_{ijk,-}$, $\tilde{s}^{(k)}_{ijk,+}$, and $\tilde{s}^{(k)}_{ijk,-}$. Since each $\tilde{s}^{(i)}$ may participate in more than one "splitting operations (127) and (128)", we thus have that for all $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$,

$$\tilde{s}^{(i)}_{jki,+}+\tilde{s}^{(i)}_{jki,-} = \tilde{s}^{(i)}_{kij,+}+\tilde{s}^{(i)}_{kij,-} = \tilde{s}^{(i)}. \tag{131}$$

The following claim allows us to convert the $\tilde{s}^{(i)}_{jki,+}$, $\tilde{s}^{(i)}_{jki,-}$, $\tilde{s}^{(i)}_{kij,+}$, and $\tilde{s}^{(i)}_{kij,-}$ values to the targeted $t^{(i)}_{[\mathrm{c},1]}$ to $t^{(i)}_{[\mathrm{c},4]}$ values.

Claim: For any cyclically shifted (i,j,k) tuple, given the above four values of $\{\tilde{s}^{(i)}_{jki,+},\tilde{s}^{(i)}_{jki,-},\tilde{s}^{(i)}_{kij,+},\tilde{s}^{(i)}_{kij,-}\}$ and the value of $\tilde{s}^{(i)}$, we can always find another four non-negative values $t^{(i)}_{[\mathrm{c},1]}$, $t^{(i)}_{[\mathrm{c},2]}$, $t^{(i)}_{[\mathrm{c},3]}$, and $t^{(i)}_{[\mathrm{c},4]}$ such that

$$t^{(i)}_{[\mathrm{c},2]}+t^{(i)}_{[\mathrm{c},4]} = \tilde{s}^{(i)}_{jki,+}, \tag{132}$$
$$t^{(i)}_{[\mathrm{c},1]}+t^{(i)}_{[\mathrm{c},3]} = \tilde{s}^{(i)}_{jki,-}, \tag{133}$$
$$t^{(i)}_{[\mathrm{c},1]}+t^{(i)}_{[\mathrm{c},4]} = \tilde{s}^{(i)}_{kij,+}, \tag{134}$$
$$t^{(i)}_{[\mathrm{c},2]}+t^{(i)}_{[\mathrm{c},3]} = \tilde{s}^{(i)}_{kij,-}, \tag{135}$$
$$\text{and } t^{(i)}_{[\mathrm{c},1]}+t^{(i)}_{[\mathrm{c},2]}+t^{(i)}_{[\mathrm{c},3]}+t^{(i)}_{[\mathrm{c},4]} = \tilde{s}^{(i)}. \tag{136}$$

Proof of Claim: Since the given values $\{\tilde{s}^{(i)}_{jki,+},\tilde{s}^{(i)}_{jki,-},\tilde{s}^{(i)}_{kij,+},\tilde{s}^{(i)}_{kij,-}\}$ satisfy (131), consider the following two cases depending on the order of the two values $\tilde{s}^{(i)}_{jki,-}$ and $\tilde{s}^{(i)}_{kij,+}$.

Case 1: $\tilde{s}^{(i)}_{jki,-} \geq \tilde{s}^{(i)}_{kij,+}$. We then construct four values $t^{(i)}_{[\mathrm{c},1]}$, $t^{(i)}_{[\mathrm{c},2]}$, $t^{(i)}_{[\mathrm{c},3]}$, and $t^{(i)}_{[\mathrm{c},4]}$ in the following way:

$$\begin{aligned}
t^{(i)}_{[\mathrm{c},1]} &= \tilde{s}^{(i)}_{kij,+},\\
t^{(i)}_{[\mathrm{c},2]} &= \tilde{s}^{(i)}_{jki,+},\\
t^{(i)}_{[\mathrm{c},3]} &= \tilde{s}^{(i)}_{kij,-}-\tilde{s}^{(i)}_{jki,+},\\
t^{(i)}_{[\mathrm{c},4]} &= 0.
\end{aligned}$$

The above construction clearly gives non-negative $t^{(i)}_{[\mathrm{c},1]}$ to $t^{(i)}_{[\mathrm{c},4]}$ values. One can easily verify that the above construction satisfies all the equalities (132) to (136). For example, by our construction $t^{(i)}_{[\mathrm{c},2]}+t^{(i)}_{[\mathrm{c},3]} = \tilde{s}^{(i)}_{jki,+}+\tilde{s}^{(i)}_{kij,-}-\tilde{s}^{(i)}_{jki,+} = \tilde{s}^{(i)}_{kij,-}$, which satisfies (135).

Case 2: $\tilde{s}^{(i)}_{jki,-} < \tilde{s}^{(i)}_{kij,+}$. We then construct four non-negative values $t^{(i)}_{[\mathrm{c},1]}$, $t^{(i)}_{[\mathrm{c},2]}$, $t^{(i)}_{[\mathrm{c},3]}$, and $t^{(i)}_{[\mathrm{c},4]}$ in the following way:

$$\begin{aligned}
t^{(i)}_{[\mathrm{c},1]} &= \tilde{s}^{(i)}_{jki,-},\\
t^{(i)}_{[\mathrm{c},2]} &= \tilde{s}^{(i)}_{kij,-},\\
t^{(i)}_{[\mathrm{c},3]} &= 0,\\
t^{(i)}_{[\mathrm{c},4]} &= \tilde{s}^{(i)}_{jki,+}-\tilde{s}^{(i)}_{kij,-}.
\end{aligned}$$

Again, the above construction leads to non-negative $t^{(i)}_{[\mathrm{c},1]}$ to $t^{(i)}_{[\mathrm{c},4]}$ values that satisfy (132) to (136). Since the above two cases cover all possible scenarios, the claim is thus proven. □

Using the above claim, we now prove that the constructed values $\{t^{(i)}_{[\mathrm{c},1]},t^{(i)}_{[\mathrm{c},2]},t^{(i)}_{[\mathrm{c},3]},t^{(i)}_{[\mathrm{c},4]}\}$ for all $i \in \{1,2,3\}$ together with the previously chosen $t^{(i)}_{[\mathrm{u}]} \triangleq \frac{R_{i\to j}+R_{i\to k}+R_{i\to jk}}{p_{i\to j\vee k}}$ satisfy the linear conditions of Proposition 3 (when $<$ being replaced by $\leq$).

To that end, we first notice that

$$\begin{aligned} & t_{[\mathrm{u}]}^{(i)}+t_{[\mathrm{c},1]}^{(i)}+t_{[\mathrm{c},2]}^{(i)}+t_{[\mathrm{c},3]}^{(i)}+t_{[\mathrm{c},4]}^{(i)} \\ & \quad=\frac{R_{i\to j}+R_{i\to k}+R_{i\to jk}}{p_{i\to j\vee k}}+\tilde{s}^{(i)} \\ & \quad=s^{(i)}, \end{aligned}$$

where the first equality follows from the definition of $t_{[\mathrm{u}]}^{(i)}$ and (136); and the second equality follows from the definition of $\tilde{s}^{(i)}$. Since the given values $s^{(i)}$ for all $i\in\{1,2,3\}$ satisfy the time-sharing condition (11) of Proposition 1, the time-sharing condition (15) of Proposition 3 must hold as well.

Moreover, the second condition (17) of Proposition 3 obviously holds by the definition of $t_{[\mathrm{u}]}^{(i)}$. In the following, we prove (18) and (19) for the case when $(i,j,k)=(1,2,3)$ and other cases can be proven symmetrically. In other words, we will prove the following equalities:

$$\begin{aligned} \left(R_{1\to 2}+R_{1\to 23}\right)\frac{p_{1\to\overline{2}3}}{p_{1\to 2\vee 3}} &\leq \left(t_{[\mathrm{c},1]}^{(1)}+t_{[\mathrm{c},3]}^{(1)}\right)\cdot p_{1\to 2} \\ &\quad+\left(t_{[\mathrm{c},2]}^{(3)}+t_{[\mathrm{c},3]}^{(3)}\right)\cdot p_{3\to 2}, \end{aligned} \tag{137}$$

$$\begin{aligned} \left(R_{1\to 3}+R_{1\to 23}\right)\frac{p_{1\to 2\overline{3}}}{p_{1\to 2\vee 3}} &\leq \left(t_{[\mathrm{c},1]}^{(1)}+t_{[\mathrm{c},4]}^{(1)}\right)\cdot p_{1\to 3} \\ &\quad+\left(t_{[\mathrm{c},2]}^{(2)}+t_{[\mathrm{c},4]}^{(2)}\right)\cdot p_{2\to 3}. \end{aligned} \tag{138}$$

By (133) and (135), we have

$$\begin{aligned} \left(t_{[\mathrm{c},1]}^{(1)}+t_{[\mathrm{c},3]}^{(1)}\right)\cdot p_{1\to 2}+\left(t_{[\mathrm{c},2]}^{(3)}+t_{[\mathrm{c},3]}^{(3)}\right)\cdot p_{3\to 2} \\ =\tilde{s}_{231,-}^{(1)}\cdot p_{1\to 2}+\tilde{s}_{231,-}^{(3)}\cdot p_{3\to 2}. \end{aligned}$$

As a result, by (130) with the (i,j,k) substituted by $(2,3,1)$, we have proven (137). Similarly, by (134) and (132), we have

$$\begin{aligned} \left(t_{[\mathrm{c},1]}^{(1)}+t_{[\mathrm{c},4]}^{(1)}\right)\cdot p_{1\to 3}+\left(t_{[\mathrm{c},2]}^{(2)}+t_{[\mathrm{c},4]}^{(2)}\right)\cdot p_{2\to 3} \\ =\tilde{s}_{312,+}^{(1)}\cdot p_{1\to 3}+\tilde{s}_{312,+}^{(2)}\cdot p_{2\to 3}. \end{aligned}$$

As a result, by (129) with (i,j,k) substituted by $(3,1,2)$, we have proven (138).

In summary, from the given values $\{s^{(i)}\}$ for all $i\in\{1,2,3\}$ satisfying the linear conditions of Proposition 1, we have constructed 15 non-negative values $\{t_{[\mathrm{u}]}^{(i)},t_{[\mathrm{c},1]}^{(i)},t_{[\mathrm{c},2]}^{(i)},t_{[\mathrm{c},3]}^{(i)},t_{[\mathrm{c},4]}^{(i)}\}$ for all $i\in\{1,2,3\}$ such that they jointly satisfy the linear inequalities of Proposition 3. The proof of Lemma 4 is thus complete. □

REFERENCES

[1] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[2] A. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 789–804, March 2006.

[3] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," in *Proc. ACM SIGCOMM*, Pisa, Italy, September 2006, pp. 243–254.

[4] D. Koutsonikolas, C.-C. Wang, and Y. Hu, "Efficient network coding based opportunistic routing through cumulative coded acknowledgment," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1368–1381, 2011.

[5] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob*, vol. 1, 1961, pp. 611–644.

[6] E. C. V. der Meulen, "Three-terminal communication channels," in *Adv. Appl. Probab.*,, vol. 3, 1971, pp. 120–154.

[7] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 4–21, January 2003.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons Inc., 2006.

[9] I. Shomorony and S. Avestimehr, "A generalized cut-set bound for deterministic multi-flow networks and its applications," in *Proc. IEEE Int'l Symp. Inform. Theory.*, Honolulu, HI, USA, Jun. 29-Jul. 4 2014, pp. 271–275.

[10] S. Kamath and Y.-H. Kim, "Chop and roll: Improving the cutset bound," in *Proc. 52st Annual Allerton Conf. on Comm., Contr., and Computing.*, Monticello, Illinois, USA, October 2014.

[11] C.-C. Wang, "Linear network coding capacity for broadcast erasure channels with feedback, receiver coordination, and arbitrary security requirement," in *Proc. IEEE Int'l Symp. Inform. Theory.*, Istanbul, Turkey, July 2013, pp. 2900–2904.

[12] W.-C. Kuo and C.-C. Wang, "Two-flow capacity region of the cope principle for two-hop wireless erasure networks," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7553–7575, November 2013.

[13] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[14] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback - capacity and algorithms," in *Proc. 5th Workshop on Network Coding, Theory, and Applications (NetCod)*, Lausanne, Switzerland, June 2009, pp. 54–61.

[15] W.-C. Kuo and C.-C. Wang, "On the capacity of 2-user 1-hop relay erasure networks - the union of feedback, scheduling, opportunistic routing, and network coding," in *Proc. IEEE Int'l Symp. Inform. Theory.*, St. Peterburg, Russia, July 2011, pp. 1337–1341.

[16] C.-C. Wang, "On the capacity of 1-to-K broadcast packet erasure channels with channel output feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 931–956, February 2012.

[17] ——, "On the capacity of wireless 1-hop intersession network coding - a broadcast packet erasure channel approach," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 957–988, February 2012.

[18] C.-C. Wang and D. J. Love, "Linear network coding capacity region of 2-receiver mimo broadcast packet erasure channels with feedback," in *Proc. IEEE Int'l Symp. Inform. Theory.*, Boston, MA, USA, July 2012, pp. 2062–2066.

[19] C. C. Wang, "Capacity region of two symmetric nearby erasure channels with channel state feedback," in *Proc. IEEE Information Theory Workshop.*, Lausanne, Switzerland, September 2012, pp. 352–356.

[20] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback - capacity and algorithms," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5779–5804, September 2013.

[21] C.-C. Wang and J. Han, "The capacity region of 2-receiver multiple-input broadcast packet erasure channels with channel output feedback," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5597–5626, Sept. 2014.

[22] W.-C. Kuo and C.-C. Wang, "Robust and optimal opportunistic scheduling for downlink 2-flow inter-session network coding," in *Proc. 33rd IEEE Conf. Comput. Commun. (INFOCOM)*, Toronto, Canada, May 2014, pp. 655–663.

[23] A. Papadopoulos and L. Georgiadis, "Multiuser broadcast erasure channel with feedback and side information and related index coding results," in *Proc. 52st Annual Allerton Conf. on Comm., Contr., and Computing.*, Monticello, Illinois, USA, October 2014.

[24] J. Han and C.-C. Wang, "General capacity region for the fully-connected 3-node packet erasure network," in *Proc. IEEE Int'l Symp. Inform. Theory.*, Hong Kong, June 2015, pp. 2648–2652.

[25] S. M. S. T. Yazdi, S. A. Savari, and G. Kramer, "Network coding in node-constrained line and star networks," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4452–4468, July 2011.

[26] A. Dana and B. Hassibi, "The capacity region of multiple input erasure broadcast channels," in *Proc. IEEE Int'l Symp. Inform. Theory.*, Adelaide, Australia, September 2005, pp. 2315–2319.

[27] L. Czap, V. M. Prabhakaran, S. Diggavi, and C. Fragouli, “Broadcasting private messages securely,” in *Proc. IEEE Int’l Symp. Inform. Theory.*, Cambridge, MA, USA, July 2012, pp. 428–432.

[28] V. Anantharam and S. Verdu, “Bits through queues,” *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 4–18, January 1996.

[29] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, “Binary fading interference channel with no csit,” in *Proc. IEEE Int’l Symp. Inform. Theory.*, Honolulu, HI, USA, Jun. 29-Jul. 4 2014, pp. 666–670.

[30] S. Bhadra, P. Gupta, and S. Shakkottai, “On network coding for inteference networks,” in *Proc. IEEE Int’l Symp. Inform. Theory.*, Seattle, Washington, USA, July 2006, p. 207211.

[31] S. Avestimehr, S. Diggavi, and D. Tse, “A deterministic approach to wireless relay network,” in *Proc. 45th Annual Allerton Conf. on Comm., Contr., and Computing.*, Monticello, Illinois, USA, September 2007.

[32] S. Ray, M. Mdard, and J. Abounadi, “Random coding in noise-free multiple access networks over finite fields,” in *Proc. IEEE GLOBECOM.*, San Francisco, California, USA, December 2003, p. 18981902.

[33] B. Smith and S. Vishwanath, “Unicast transmission over multiple access erasure networks: Capacity and duality,” in *Proc. IEEE Information Theory Workshop.*, Tahoe City, California, USA, September 2007, pp. 331–336.

[34] Y. Zhu and D. Guo, “On the capacity region of fading z-interference channels without csit,” in *Proc. IEEE Int’l Symp. Inform. Theory.*, Austin, Texas, USA, June 2010, p. 350354.

Jaemin Han received the B.S. and M.S. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, South Korea, in 2007 and 2009, respectively, and the Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, Indiana, in 2016. He worked in Intel’s Mobile Communication Group from August 2012 to January 2013, and Communication and Device Group from April to August 2015, both as a graduate intern. He is currently a wireless standards engineer in Intel. His research interests are the graph-theoretic and information-theoretic characterizations of network coding and scheduling algorithms.

Chih-Chun Wang is currently an Associate Professor of the School of Electrical and Computer Engineering of Purdue University. He received the B.E. degree in E.E. from National Taiwan University, Taipei, Taiwan in 1999, the M.S. degree in E.E., the Ph.D. degree in E.E. from Princeton University in 2002 and 2005, respectively. He worked in Comtrend Corporation, Taipei, Taiwan, as a design engineer in 2000 and spent the summer of 2004 with Flarion Technologies, New Jersey. In 2005, he held a post-doctoral researcher position in the Department of Electrical Engineering of Princeton University. He joined Purdue University as an Assistant Professor in 2006, and became an Associate Professor in 2012. He is currently a senior member of IEEE and has been an associate editor of IEEE Transactions on Information Theory since 2014 and the technical co-chair of the 2017 IEEE Information Theory Workshop. His current research interests are in the delay-constrained information theory and network coding. Other research interests of his fall in the general areas of networking, optimal control, information theory, detection theory, and coding theory.

Dr. Wang received the National Science Foundation Faculty Early Career Development (CAREER) Award in 2009.