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

Jaemin Han, Chih-Chun Wang; {han83,chihw}@purdue.edu
School of Electrical and Computer Engineering, Purdue University, USA

***Abstract*—This work considers the fully-connected 3-node packet erasure network: For each time slot, with some probabilities a packet sent by any 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 nodes. 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 example of this model is any Wi-Fi network with 3 nodes within the hearing range of each other.**

We consider the most general traffic demands. 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) and can be further reduced to 0 if we allow ACK to be transmitted via a separate control channel. For normal-sized packets, say 12000 bits, our results effectively characterizs the capacity region for many important scenarios, e.g., wireless access-point networks with client-to-client cooperation. Technical contributions of this work include a new converse for many-to-many network communications and a new capacity-approaching simple linear network coding scheme.

I. Introduction

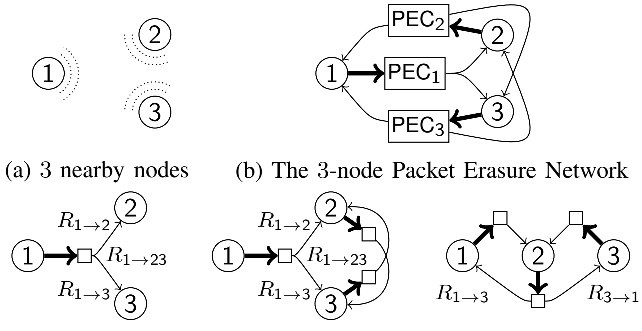
One of the driving forces for high-rate, ubiquitous network communications is Network Information Theory (NIT), which provides guidance on how to design high-performance (optimal or near optimal) practical network protocols. One notable example in the recent NIT development is Linear Network coding (LNC). 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 50-200% throughput gain over the traditional 802.11 protocols.

Despite the above promising results, our NIT understanding is still nascent for networks with general traffic patterns. When there are only 2 nodes with two co-existing information flows of opposite directions, Shannon [5] provided the first inner and outer bounds. 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].

There are at least two difficulties when characterizing 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 lack of full-duplex hardware, transmission from B to A may be impossible when A is sending information to 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 [8] or through ingenious ways of coding and cooperation [7], [9]. 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 channels or multiple access channels, and/or with a small number of co-existing flows in parallel directions (i.e., flows not forming cycles).

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 $(R_{1\to2}, R_{1\to3}, R_{2\to1}, R_{2\to3}, R_{3\to1}, R_{3\to2})$, respectively, in all possible directions; and there are three co-existing common-information flows with rates $(R_{1\to23},\ R_{2\to31}, R_{3\to12})$, 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 X, and a random subset of the other two nodes will receive the packet, 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.

We assume one of the following two scenarios. **Scenario 1**: Motivated by the throughput benefit of the causal packet ACKnowledgment feedback for erasure networks [8], [10]–[13], we assume that the reception status is casually 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**: 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. The timeliness



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

(c) An 1-to-2 PEC (d) PEC w. receiver coordinations (e) A Two-way relay PEC

Fig. 1. Illustrations of 3-node Packet Erasure Network (PEN) and its variants.

of delivering the control messages is also critical since the control information, sent through the erasure channel, may get lost as well. In Scenario 2, we further 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*). Note that *the fully-connectedness is assumed only in Scenario 2*. When the casual reception status is available for free (Scenario 1), our results do not need the fully-connectedness assumption.

Contributions: For Scenario 1, we characterize the exact 9-dimensional Shannon capacity region. For Scenario 2, the capacity region is characterized with a gap inversely proportional to $\log_2(q)$. 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.

It is worth noting that the considered 3-node PEN contains many important practical scenarios as sub-cases. **Example 1:** If we set 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. 1(c), 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 [10]. **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. In this case, node 2 can still potentially help relay the packets destined for node 3 and vice versa, see Fig. 1(d). 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. 1(e) 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. 1(e).

Summary: Most existing works on PENs studied either ≤ 2 co-existing flows [3], [4], [8], [10], [12], [13] or all flows originating from the same node [8], [10], [11], [13]–[15]. By characterizing the most general 9-dimensional Shannon capacity region with arbitrary flow directions, this work significantly improves our understanding for communications over 3-node PENs.

II. Problem Formulation

Throughout this work, we use (i,j,k) to represent one of three cyclically shifted tuples $\{(1,2,3),(2,3,1),(3,1,2)\}$. We define the 9-dimensional rate vector $\vec{R} \triangleq (R_{1\to 2}, R_{1\to 3}, R_{2\to 1}, R_{2\to 3}, R_{3\to 1}, R_{3\to 2}, R_{1\to 23}, R_{2\to 31}, R_{3\to 12})$.

Assume the synchronized time-slotted transmissions. 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. Each uncoded packet is chosen independently and uniformly randomly from a finite field $\mathbb{F}_q$.

For any time slot $t \in \{1, \cdots, n\}$, define the *channel output vector* $\mathbf{Z}(t) \triangleq (Z_{1\to 2}(t), Z_{1\to 3}(t), Z_{2\to 1}(t), Z_{2\to 3}(t), Z_{3\to 1}(t), Z_{3\to 2}(t)) \in \{1, \varepsilon\}^6$, where $Z_{i\to h}(t) = 1$ and ε represents whether node h can receive the transmission from node i or not, respectively. For any node i, the corresponding scheduling decision $\sigma_i(t) = 1$ represents that node i decides to transmit at time t and $\sigma_i(t) = 0$ represents not transmitting. We assume that any transmission is completely destroyed if there are ≥ 2 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 $Z_{i\to h}(t) = 1$ and $\sigma_j(t) = \sigma_k(t) = 0$ will node h receive $Y_{i\to h}(t) = X_i(t)$. In all other cases, node h receives an erasure $Y_{i\to h}(t) = \varepsilon$. To highlight this interference model, we sometimes write

$$Y_{i\to h}(t) = X_i(t) \circ Z_{i\to h}(t) \circ 1_{\{\sigma_i(t)=1, \sigma_j(t)=\sigma_k(t)=0\}}. \quad (1)$$

We further 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 $\mathbf{Z}(t)$ is independently and identically distributed over the time axis t. 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 $X_i(t)$ 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 $X_i(t)$ 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. $p_{i\to k}$) 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.

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 intended 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

[1] In [8], the LNC capacity is characterized instead of the most general Shannon capacity.

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

originated from both nodes i and 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 a 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 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 of 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 [16], 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.

Namely, at every time t, each node decides whether it wants 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 only the control information is used to determine $\sigma_i(t)$ and thus the messages $\mathbf{W}$ cannot be sent through the timing of the scheduling.[4] Once each node decides whether to transmit or not,[5] 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 $\vec{R}$.

[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 12000 bits.

[4]E.g., 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 scheduling depend only on the control information in the header, not the messages in the payload.

[5]If two nodes i and j want to transmit simultaneously, then our channel model (1) automatically leads to full collision and erases both transmissions.

A. 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.

III. Main Results

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[6] 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:

$$\sum_{\forall\, i \in \{1,2,3\}} s^{(i)} \leq 1. \tag{9}$$

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

$$R_{i\to j} + R_{i\to k} + R_{i\to jk} \leq s^{(i)} p_{i\to j\vee k}. \tag{10}$$

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

$$\begin{aligned} R_{j\to i} + R_{j\to ki} + R_{k\to i} + R_{k\to ij} &\leq s^{(j)} p_{j\to i} + s^{(k)} p_{k\to i} \\ &\quad - \left(\frac{p_{j\to i}}{p_{j\to k\vee i}} R_{j\to k} + \frac{p_{k\to i}}{p_{k\to i\vee j}} R_{k\to j} \right), \end{aligned} \tag{11}$$

Proposition 1 considers arbitrary, possibly non-linear ways of designing the scheduling and encoding/decoding functions in (5) to (8), and is derived by entropy-based analysis, which include the results [8], [10] as special cases.

The intuition is 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, (9) holds naturally. (10) 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 (11). The LHS of (11) 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 (11) is the amount of time slots that either node j or node

[6]Proposition 1 is naturally an outer bound for Scenario 2, see Section II-A.

k can communicate with node i. As a result, it resembles a multiple-access cut condition of a typical cut-set argument [17, 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 remaining term on the RHS of (11) quantifies such penalty that is inevitable regardless of what kind of coding schemes being used. The detailed proof of Proposition 1 is relegated to Appendix A.

B. A Linear Network Coding Capacity Achieving Scheme

We first focus on Scenario 2, where 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 3-node PEN. To formalize our discussion, we define

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., $\min(p_{i_1\to i_3}, p_{i_3\to i_2}) > 0$). 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:

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

i.e., $\mathsf{nzmin}\{a, b\}$ 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 $t^{(i)}_{[\mathsf{u}]}$ and $\{t^{(i)}_{[\mathsf{c},l]}\}_{l=1}^4$ 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:

$$\sum_{\forall\, i\in\{1,2,3\}} t^{(i)}_{[\mathsf{u}]} + t^{(i)}_{[\mathsf{c},1]} + t^{(i)}_{[\mathsf{c},2]} + t^{(i)}_{[\mathsf{c},3]} + t^{(i)}_{[\mathsf{c},4]} \leq 1 - t_{\mathsf{FB}}, \quad (12)$$

where t_{FB} is a constant defined as

$$t_{\mathsf{FB}} \triangleq \frac{1}{\log_2(q)} \sum_{\forall i\in\{1,2,3\}} \frac{1}{\mathsf{nzmin}\{p_{i\to j}, p_{i\to k}\}}. \quad (13)$$

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

$$R_{i\to j} + R_{i\to k} + R_{i\to jk} < t^{(i)}_{[\mathsf{u}]} p_{i\to j\vee k}. \quad (14)$$

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

$$\begin{aligned} \Big(R_{i\to j} + R_{i\to jk}\Big)\frac{p_{i\to \overline{j}k}}{p_{i\to j\vee k}} &< \Big(t^{(i)}_{[\mathsf{c},1]} + t^{(i)}_{[\mathsf{c},3]}\Big)\cdot p_{i\to j} \\ &\quad + \Big(t^{(k)}_{[\mathsf{c},2]} + t^{(k)}_{[\mathsf{c},3]}\Big)\cdot p_{k\to j}, \end{aligned} \quad (15)$$

$$\begin{aligned} \Big(R_{i\to k} + R_{i\to jk}\Big)\frac{p_{i\to j\overline{k}}}{p_{i\to j\vee k}} &< \Big(t^{(i)}_{[\mathsf{c},1]} + t^{(i)}_{[\mathsf{c},4]}\Big)\cdot p_{i\to k} + \\ &\quad + \Big(t^{(j)}_{[\mathsf{c},2]} + t^{(j)}_{[\mathsf{c},4]}\Big)\cdot p_{j\to k}. \end{aligned} \quad (16)$$

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 $t^{(i)}_{[\mathsf{u}]}$ and $\{t^{(i)}_{[\mathsf{c},l]}\}_{l=1}^4$ for all $i \in \{1, 2, 3\}$ such that (12), (14) to (16) hold while we set $t_{\mathsf{FB}} = 0$ in (13).

In short, the constant term t_{FB} in (13) 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. We now characterize the capacity for the 3-node PEN under Scenario 1.

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

The proof of Proposition 4 is by some pure algebraic arguments and thus relegated to Appendix B.

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 q is large enough.

In the following, the sketch of proof for Proposition 3 (Scenario 1) is explained while the detailed construction for Proposition 2 (Scenario 2) is omitted due to space limit.

Sketch of Proposition 3: We only 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 $t^{(i)}_{[\mathsf{u}]}$ and $\{t^{(i)}_{[\mathsf{c},l]}\}_{l=1}^4$ for all $i \in \{1, 2, 3\}$ satisfying Proposition 3.

We propose the following 2-stage scheme. Stage 1: Each node, say node i, has $n(R_{i\to j} + R_{i\to k} + R_{i\to jk})$ unicasts and multicast packets (i.e., $\mathbf{W}_{i*}$) that need to be sent to other nodes j and k. Assume that those packets are ordered in group-wise as $\mathbf{W}_{i\to j}$, $\mathbf{W}_{i\to k}$, and then $\mathbf{W}_{i\to jk}$, and they are indexed by $l = 1$ to $n(R_{i\to j} + R_{i\to k} + R_{i\to jk})$. 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 repeat the same process until each of these $n(R_{1\to 2} + R_{1\to 3} + R_{1\to 23})$ packets is heard by at least one of nodes 2 and 3. By simple analysis, see [13], node 1 can finish the transmission in $nt^{(i)}_{[\mathsf{u}]}$ slots since (14).[7] We repeat this process for nodes 2 and 3, respectively. Stage 1 can be finished in $n(\sum_i t^{(i)}_{[\mathsf{u}]})$ 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, is delivered successfully and thus

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

will not be considered for future operations. 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 $nR_{i\to j}\frac{p_{i\to \bar{j}k}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(k)}_{i\to j}$ packets. Symmetrically, we also have $nR_{i\to k}\frac{p_{i\to j\bar{k}}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(j)}_{i\to k}$ packets that were intended for node k but were overheard only by node j in Stage 1.

Similarly for the common-information packets $\mathbf{W}_{i\to jk}$, each packet was heard by at least one of nodes j and k during Stage 1. Those that have been heard by both nodes j and k, is 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 $nR_{i\to jk}\frac{p_{i\to \bar{j}k}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(k)}_{i\to jk}$ packets. Symmetrically, we also have $nR_{i\to jk}\frac{p_{i\to j\bar{k}}}{p_{i\to j\vee k}}$ number of $\mathbf{W}^{(j)}_{i\to jk}$ packets that were heard only by node j in Stage 1.

Stage 2 is the LNC phase, in which each node i will send a linear combination of overheard packets that it exactly knows as described above. We claim that there are (at least) 4 possible ways of sending LNC packets. That is, for each time t, node i send a linear combination $X_i(t) = [\tilde{W}_j + \tilde{W}_k]$ with 4 possible ways of choosing the individual packets $\tilde{W}_j$ and $\tilde{W}_k$:

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

To further explain this LNC stage, we observe that choice $[\mathsf{c},1]$ is the standard LNC operation for the 2-receiver broadcast channels [10] 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. One can easily prove that transmitting such a linear mixture again benefits both nodes simultaneously.

Since each node i has 4 possible coding choices, we perform coding choice $[\mathsf{c},l]$ for exactly $n\cdot t^{(i)}_{[\mathsf{c},l]}$ times for $l=1$ to 4. Since $\mathbf{W}^{(k)}_{i\to j}\cup\mathbf{W}^{(k)}_{i\to jk}$ participates in coding choices $[\mathsf{c},1]$ and $[\mathsf{c},3]$ of node i and coding choices $[\mathsf{c},2]$ and $[\mathsf{c},3]$ of node k, (15) guarantees 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.[8] Symmetrically, (16) guarantees 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, (12) guarantees that we can finish Stages 1 and 2 in the allotted n time slots. The sketch of the proof is complete. $\square$

[8]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 only, we say herein their new intended destination is node j instead.

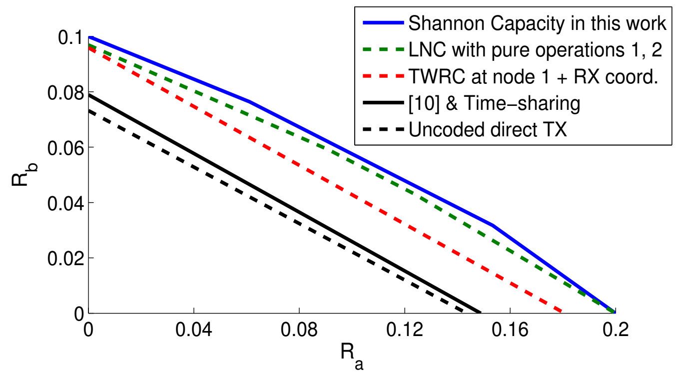


Fig. 2. Comparison of the capacity region with different achievable rates

C. Numerical Evaluation

Consider a 3-node network with marginal channel success probabilities $p_{1\to2}=0.35$, $p_{1\to3}=0.8$, $p_{2\to1}=0.6$, $p_{2\to3}=0.5$, $p_{3\to1}=0.3$, and $p_{3\to2}=0.75$, respectively. To illustrate the 9-dimensional capacity region, we assume that 3 flows are of the same rate $R_{1\to2}=R_{1\to3}=R_{1\to23}=R_a$ and the other 6 flows are of rate $R_{2\to1}=R_{2\to3}=R_{3\to1}=R_{3\to2}=R_{2\to31}=R_{3\to12}=R_b$. Fig. 2 compares the Shannon capacity region of (R_a,R_b) with different achievability schemes.

The smallest achievable rate region is by simply performing uncoded direct transmission. The second achievability scheme combines the broadcast channel LNC in [10] with time-sharing among all three nodes. The third scheme performs two-way relay channel (TWRC) coding in node 1 for those $3\to2$ and $2\to3$ flows while allowing node 2 to relay the packets $\mathbf{W}_{1\to3}$ destined for node 3 and vice versa. The fourth scheme is the largest that can be derived from the existing LNC operations. Namely, we allow all three nodes to perform the broadcast-based LNC and/or TWRC-based LNC operations (coding choices $[\mathsf{c},1]$ and $[\mathsf{c},2]$ in Stage 2) but not the hybrid operations (coding choices $[\mathsf{c},3]$ and $[\mathsf{c},4]$) 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.

IV. Conclusion

This work characterizes 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 quantified for all channel parameters when the casual ACK/NACK feedbacks are immediately available for free through a separate control channel. For the practical setting where the control messages has to be sent through the regular forward channels, the proposed LNC scheme approaches the capacity as the packet size becomes large enough.

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] 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.

[9] 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.

[10] 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.

[11] 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.

[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] 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.

[14] 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.

[15] 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.

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

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

APPENDIX A
PROOF OF PROPOSITION 1

We now provide the proof of Proposition 1. To that end, we prove Proposition 1 based on the following, slightly different, joint scheduling and network coding formulation of Scenario 1: $\forall\, t\in\{1,\cdots,n\}$ and $\forall\, i\in\{1,2,3\}$,

$$\sigma(t) = \overline{\Theta}^{(t)}_{\mathsf{SCH}}([\mathbf{Z}]_1^{t-1}) \in \{1,2,3\}, \tag{17}$$

$$X_i(t) = \overline{f}^{(t)}_i(\mathbf{W}_{i*}, [\mathbf{Y}_{*i}, \mathbf{Z}]_1^{t-1}), \tag{18}$$

$$\hat{\mathbf{W}}_{*i} = \overline{g}_i(\mathbf{W}_{i*}, [\mathbf{Y}_{*i}, \mathbf{Z}]_1^{n}), \tag{19}$$

while we do not impose no-timing-channel condition anymore. Notice that (18) and (19) is equivalent to the original encoding (7) and decoding (8) functions in Scenario 1, respectively. The only difference is that the timing-channel-free scheduling process is described solely by the scheduling decision $\sigma(t)\in\{1,2,3\}$ that takes the values in the set of three nodes and does not depend on any information messages, see (17).[9] The subsequent proofs based on the new descriptions (17) to (19) can be easily applied to the original formulation (5) to (8) of Scenario 1 with very little modifications.

Given the reception probabilities, consider a joint scheduling and network code scheme (17) to (19) that achieves $\vec{R}$ within the total time budget n. Let us choose $s^{(i)}$ as the normalized expected number of time slots for which node i is scheduled in this scheme. That is,

$$\begin{aligned} s^{(i)} &\triangleq \frac{1}{n}\mathbb{E}\left\{\sum_{t=1}^{n} 1_{\{\sigma(t)=i\}}\right\}, \\ &= \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{\forall [\mathbf{z}]_1^{t-1} \text{ s.t. } \overline{\Theta}^{(t)}_{\mathsf{SCH}}([\mathbf{z}]_1^{t-1})=i} \mathsf{Prob}([\mathbf{Z}]_1^{t-1}=[\mathbf{z}]_1^{t-1})\right), \end{aligned} \tag{20}$$

where $1_{\{\cdot\}}$ is the indicator function and (20) follows from the scheduling description (17) of the network code. Since $\sigma(t)\in\{1,2,3\}$ is decided based on the past CSI $[\mathbf{Z}]_1^{t-1}$, it is averaged over all possible past channel outputs realizations $[\mathbf{z}]_1^{t-1}\in\{1,\varepsilon\}^{6(t-1)}$. The computed scheduling frequencies $\{s^{(1)},s^{(2)},s^{(3)}\}$ for all three nodes must satisfy the time-sharing condition (9).

In the subsequent proofs, we will derive two conditions (10) and (11) of Proposition 1, respectively. To that end, we use the information-theoretic mutual information $I(\cdot\,;\cdot)$ and the entropy $H(\cdot)$. We also assume that the logarithm for the mutual information and the entropy is taken with respect to the base q from a finite field $\mathbb{F}_q$. For the case when the logarithm of the entropy is based on 2, we clearly distinguish it by using $H_2(\cdot)$.

The following lemma is useful in deriving the proof of Proposition 1.

Lemma 1: In Scenario 1, fix a time slot $t\in\{1,\cdots,n\}$. Then, knowing all the messages $\mathbf{W}_{\{1,2,3\}*}$ and a past channel output realization $[\mathbf{z}]_1^{t-1}\in[\mathbf{Z}]_1^{t-1}$ 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 1. 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 (18). As a result, $\{X_1(1),X_2(1),X_3(1)\}$ can be uniquely determined given $\mathbf{W}_{\{1,2,3\}*}$.

[9] Originally, the *message-independent* information of *which node is scheduled to transmit without any interference at time t* had to be determined jointly through each node's binary variable (whether to transmit or not) (6); no-timing-channel condition (5); and the interference model (1). Although such modeling has a good practical value of describing the scheduling process in an implementation-friendly manner, it has a little value when deriving a capacity outer bound, since a central genie can freely choose the interference-free transmissions over the network. The new scheduling description (17) succinctly captures the necessary timing-channel-free scheduling process for the simpler entropy-based analysis.

Assume that the statement of Lemma 1 is true until time $t-1$ such that $[X_1, X_2, X_3]_1^{t-1}$ and $[\mathbf{Y}_{1*}, \mathbf{Y}_{2*}, \mathbf{Y}_{3*}]_1^{t-2}$ are uniquely determined by $\mathbf{W}_{\{1,2,3\}*}$ and the past channel output realizations $[\mathbf{z}]_1^{t-2}$ up to time $t-2$. Then given the uniquely determined input symbols $\{X_1(t-1), X_2(t-1), X_3(t-1)\}$, all the received symbols $\{\mathbf{Y}_{1*}(t-1), \mathbf{Y}_{2*}(t-1), \mathbf{Y}_{3*}(t-1)\}$ can also be uniquely determined as well once we know the channel output realization $\mathbf{z}(t-1)$ at time $t-1$. As a result, knowing $\mathbf{W}_{\{1,2,3\}*}$ and $[\mathbf{z}]_1^{t-1}$ can uniquely decide $[\mathbf{Y}_{1*}, \mathbf{Y}_{2*}, \mathbf{Y}_{3*}]_1^{t-1}$ up to $t-1$. Then by the encoding descriptions (18), the input symbols $\{X_1(t), X_2(t), X_3(t)\}$ at time t can be uniquely determined as well. The proof of Lemma 1 is thus complete. □

A-1. The broadcasting cut-set condition (10)

The broadcasting cut-set condition (10) can be easily deriven based on the following mutual information term:

$$I(\mathbf{W}_{i*}\,;\,[\mathbf{Y}_{*j}, \mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{j,k\}*}, [\mathbf{Z}]_1^n). \tag{21}$$

That is, we will prove that

$$n\begin{pmatrix} R_{i\to j} + R_{i\to k} + R_{i\to jk} \\ -\,2\epsilon_n - \dfrac{H_2(2\epsilon_n)}{n\log_2 q} \end{pmatrix} \overset{\text{(10A)}}{\leq} (21) \overset{\text{(10B)}}{\leq} n s^{(i)} p_{i\to j\vee k},$$

for any feasible scheme given n. Therefore, any sequence of feasible schemes with $\epsilon_n \to 0$ as $n \to \infty$ must satisfy (10). The detailed derivations of (10A) and (10B) are relegated to Appendix C.

A-2. The 3-way multiple-access cut-set condition (11)

The 3-way multiple-access cut-set condition (11) can be deriven based on the following mutual information term:

$$I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n \,|\, \mathbf{W}_{i*}, [\mathbf{Z}]_1^n). \tag{22}$$

That is, we will prove that

$$n\begin{pmatrix} R_{j\to i} + R_{k\to i} \\ +\,R_{j\to ki} + R_{k\to ij} \\ +\,\dfrac{p_{j\to i}}{p_{j\to k\vee i}} R_{j\to k} \\ +\,\dfrac{p_{k\to i}}{p_{k\to i\vee j}} R_{k\to j} \\ -\,6\epsilon_n - \dfrac{3H_2(\epsilon_n)}{n\log_2 q} \end{pmatrix} \overset{\text{(11A)}}{\leq} (22) \overset{\text{(11B)}}{\leq} n\begin{pmatrix} s^{(j)} p_{j\to i} \\ + \\ s^{(k)} p_{k\to i} \end{pmatrix}.$$

for any feasible scheme given n. Therefore, any sequence of feasible schemes with $\epsilon_n \to 0$ as $n \to \infty$ must satisfy the cut-set outer bound (11). The detailed derivations of (11A) and (11B) are relegated to Appendix D.

APPENDIX B
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.

Then, we now introduce the following Lemma.

Lemma 2: 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 $t_{[\mathrm{u}]}^{(i)}$ and $\{t_{[\mathrm{c},\,l]}^{(i)}\}_{l=1}^4$ for all $i \in \{1,2,3\}$ such that jointly satisfy three groups of linear conditions in Proposition 3 (when replacing all strict inequality $<$ by $\leq$).

One can clearly see that Lemma 2 imply 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.

The proof of Lemma 2: 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 $t_{[\mathrm{u}]}^{(i)} \triangleq \frac{R_{i\to j} + R_{i\to k} + R_{i\to jk}}{p_{i\to j\vee k}}$ which is non-negative by definition, and then construct 3 non-negative values $\{\tilde{s}^{(i)}\}$ as follows: For all $i \in \{1,2,3\}$,

$$\tilde{s}^{(i)} \triangleq s^{(i)} - \frac{R_{i\to j} + R_{i\to k} + R_{i\to jk}}{p_{i\to j\vee k}}. \tag{23}$$

Notice that since the given values $\{s^{(i)}\}$ satisfy the broadcasting cut-set condition (10) in Proposition 1, the newly constructed values $\{\tilde{s}^{(i)}\}$ must be non-negative. Then, we can rewrite the 3-way multiple-access cut-set condition (11) in Proposition 1 as follows: For all $i \in \{1,2,3\}$,

$$\Big(R_{j\to i} + R_{j\to ki}\Big)\frac{p_{j\to k\bar{i}}}{p_{j\to k\vee i}} + \Big(R_{k\to i} + R_{k\to ij}\Big)\frac{p_{k\to \bar{i}j}}{p_{k\to i\vee j}} \leq \tilde{s}^{(j)}\cdot p_{j\to i} + \tilde{s}^{(k)}\cdot p_{k\to i}. \tag{24}$$

Notice that there are three equations in the form of (24), each for $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$. Then for each (i,j,k), we construct the four non-negative values $\{\tilde{s}_{i+}^{(j)}, \tilde{s}_{i-}^{(j)}, \tilde{s}_{i+}^{(k)}, \tilde{s}_{i-}^{(k)}\}$ such that they jointly satisfy

$$\tilde{s}_{i+}^{(j)} + \tilde{s}_{i-}^{(j)} = \tilde{s}^{(j)}, \tag{25}$$

$$\tilde{s}_{i+}^{(k)} + \tilde{s}_{i-}^{(k)} = \tilde{s}^{(k)}, \tag{26}$$

so that we can split (24) into two halves as follows:

$$\Big(R_{j\to i} + R_{j\to ki}\Big)\frac{p_{j\to k\bar{i}}}{p_{j\to k\vee i}} \leq \tilde{s}_{i+}^{(j)}\cdot p_{j\to i} + \tilde{s}_{i+}^{(k)}\cdot p_{k\to i}, \tag{27}$$

$$\Big(R_{k\to i} + R_{k\to ij}\Big)\frac{p_{k\to \bar{i}j}}{p_{k\to i\vee j}} \leq \tilde{s}_{i-}^{(j)}\cdot p_{j\to i} + \tilde{s}_{i-}^{(k)}\cdot p_{k\to i}. \tag{28}$$

Note that finding such four non-negative values $\{\tilde{s}_{i+}^{(j)}, \tilde{s}_{i-}^{(j)}, \tilde{s}_{i+}^{(k)}, \tilde{s}_{i-}^{(k)}\}$ satisfying (25) to (28) can be always done. Thus from the given three values $\{\tilde{s}^{(i)}\}$, we have constructed in total 12 non-negative values $\{\tilde{s}_{i+}^{(j)}, \tilde{s}_{i-}^{(j)}, \tilde{s}_{i+}^{(k)}, \tilde{s}_{i-}^{(k)}\}$ for all $(i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\}$. Since the node indices are cyclically decided, one can easily see that, for each $i \in \{1,2,3\}$ and the corresponding $\tilde{s}^{(i)}$-value, we have four associated values $\{\tilde{s}_{j+}^{(i)}, \tilde{s}_{j-}^{(i)}, \tilde{s}_{k+}^{(i)}, \tilde{s}_{k-}^{(i)}\}$ where they jointly satisfy

$$\tilde{s}_{j+}^{(i)} + \tilde{s}_{j-}^{(i)} = \tilde{s}_{k+}^{(i)} + \tilde{s}_{k-}^{(i)} = \tilde{s}^{(i)}. \tag{29}$$

We now prove the following claim.

Claim: Given the above four non-negative values of $\{\tilde{s}^{(i)}_{j+}, \tilde{s}^{(i)}_{j-}, \tilde{s}^{(i)}_{k+}, \tilde{s}^{(i)}_{k-}\}$ and the corresponding non-negative $\tilde{s}^{(i)}$-value, we can always find another four non-negative values $t^{(i)}_{[c,1]}$, $t^{(i)}_{[c,2]}$, $t^{(i)}_{[c,3]}$, and $t^{(i)}_{[c,4]}$ such that

$$t^{(i)}_{[c,2]} + t^{(i)}_{[c,4]} = \tilde{s}^{(i)}_{j+}, \tag{30}$$
$$t^{(i)}_{[c,1]} + t^{(i)}_{[c,3]} = \tilde{s}^{(i)}_{j-}, \tag{31}$$
$$t^{(i)}_{[c,1]} + t^{(i)}_{[c,4]} = \tilde{s}^{(i)}_{k+}, \tag{32}$$
$$t^{(i)}_{[c,2]} + t^{(i)}_{[c,3]} = \tilde{s}^{(i)}_{k-}, \tag{33}$$
$$\text{and } t^{(i)}_{[c,1]} + t^{(i)}_{[c,2]} + t^{(i)}_{[c,3]} + t^{(i)}_{[c,4]} = \tilde{s}^{(i)}. \tag{34}$$

Proof of Claim: Since the given values $\{\tilde{s}^{(i)}_{j+}, \tilde{s}^{(i)}_{j-}, \tilde{s}^{(i)}_{k+}, \tilde{s}^{(i)}_{k-}\}$ satisfy (29), consider the following two cases depending on the values of $\tilde{s}^{(i)}_{j-}$ and $\tilde{s}^{(i)}_{k+}$.

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

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

Since $t^{(i)}_{[c,4]} = 0$, one can easily check that both (30) and (32) are satisfied by the above construction. Eq. (31) also holds because $t^{(i)}_{[c,1]} + t^{(i)}_{[c,3]} = \tilde{s}^{(i)} - \tilde{s}^{(i)}_{j+}$ and by (29), this is equal to $\tilde{s}^{(i)}_{j-}$. Similarly, $t^{(i)}_{[c,2]} + t^{(i)}_{[c,3]} = \tilde{s}^{(i)} - \tilde{s}^{(i)}_{k+}$ and by (29), this is equal to $\tilde{s}^{(i)}_{k-}$. Therefore, (33) also holds from the above construction. Moreover, one can easily see that (34) holds as well by adding them up.

We now argue that the constructed $\{t^{(i)}_{[c,1]}, t^{(i)}_{[c,2]}, t^{(i)}_{[c,3]}, t^{(i)}_{[c,4]}\}$ are all non-negative values. Obviously, $t^{(i)}_{[c,1]}$, $t^{(i)}_{[c,2]}$, and $t^{(i)}_{[c,4]}$ are non-negative by the construction and the previous discussions. The third value $t^{(i)}_{[c,3]}$ is non-negative as well because $\tilde{s}^{(i)} - \tilde{s}^{(i)}_{k+} - \tilde{s}^{(i)}_{j+} = \tilde{s}^{(i)}_{j-} - \tilde{s}^{(i)}_{k+}$ by (29) and we assumed $\tilde{s}^{(i)}_{j-} \geq \tilde{s}^{(i)}_{k+}$ in the beginning.

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

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

Following similar arguments as in *Case* 1, one can prove that (30) to (34) hold by the above construction. Moreover, the constructed $\{t^{(i)}_{[c,1]}, t^{(i)}_{[c,2]}, t^{(i)}_{[c,3]}, t^{(i)}_{[c,4]}\}$ values are all non-negative because, for the example of $t^{(i)}_{[c,4]}$, we have $\tilde{s}^{(i)} - \tilde{s}^{(i)}_{j-} - \tilde{s}^{(i)}_{k-} = \tilde{s}^{(i)}_{k+} - \tilde{s}^{(i)}_{j-}$ by (29) and we assumed $\tilde{s}^{(i)}_{j-} < \tilde{s}^{(i)}_{k+}$ in the beginning.

Since the above two cases disjointly cover the entire scenario, the claim is thus proven. □

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

First note that the above $\{t^{(i)}\}$-values satisfy the time-sharing condition (12) of Proposition 3 since for each i, we have $t^{(i)}_{[u]} + t^{(i)}_{[c,1]} + t^{(i)}_{[c,2]} + t^{(i)}_{[c,3]} + t^{(i)}_{[c,4]} = \frac{R_{i\to j}+R_{i\to k}+R_{i\to jk}}{p_{i\to j\vee k}} + \tilde{s}^{(i)}$ by the chosen $t^{(i)}_{[u]}$ and (34), and by the definition (23) of $\tilde{s}^{(i)}$, this is equal to $s^{(i)}$. Since the given values $s^{(i)}$ for all $i \in \{1,2,3\}$ satisfy the time-sharing condition (9) of Proposition 1, the time-sharing condition of Proposition 3 must hold as well.

Moreover, the second condition (14) of Proposition 3 obviously holds by the chosen $t^{(i)}_{[u]} \triangleq \frac{R_{i\to j}+R_{i\to k}+R_{i\to jk}}{p_{i\to j\vee k}}$. Thus we only need to show that the third conditions (15) and (16) of Proposition 3, when $<$ is replaced by $\leq$, hold by the constructed values of $\{t^{(i)}_{[c,1]}, t^{(i)}_{[c,2]}, t^{(i)}_{[c,3]}, t^{(i)}_{[c,4]}\}$ for all $i \in \{1,2,3\}$. However for clarity, here we only show for the case when $(i,j,k) = (1,2,3)$ and other cases can be followed symmetrically. In other words, we prove that the following equations hold:

$$\begin{aligned}\Big(R_{1\to 2} + R_{1\to 23}\Big)\frac{p_{1\to \overline{2}3}}{p_{1\to 2\vee 3}} \leq \Big(t^{(1)}_{[c,1]} + t^{(1)}_{[c,3]}\Big)\cdot p_{1\to 2} \\ + \Big(t^{(3)}_{[c,2]} + t^{(3)}_{[c,3]}\Big)\cdot p_{3\to 2},\end{aligned} \tag{35}$$

$$\begin{aligned}\Big(R_{1\to 3} + R_{1\to 23}\Big)\frac{p_{1\to 2\overline{3}}}{p_{1\to 2\vee 3}} \leq \Big(t^{(1)}_{[c,1]} + t^{(1)}_{[c,4]}\Big)\cdot p_{1\to 3} \\ + \Big(t^{(2)}_{[c,2]} + t^{(2)}_{[c,4]}\Big)\cdot p_{2\to 3},\end{aligned} \tag{36}$$

where (35) and (36) corresponds to the third conditions (15) and (16) of Proposition 3, respectively, when $(i,j,k) = (1,2,3)$ and $<$ is replaced by $\leq$.

To that end, we use (27), (28), and the above claim. We first rewrite (27) and (28) as follows:

$$\Big(R_{1\to 2} + R_{1\to 23}\Big)\frac{p_{1\to \overline{2}3}}{p_{1\to 2\vee 3}} \leq \tilde{s}^{(3)}_{2-}\cdot p_{3\to 2} + \tilde{s}^{(1)}_{2-}\cdot p_{1\to 2}, \tag{37}$$
$$\Big(R_{1\to 3} + R_{1\to 23}\Big)\frac{p_{1\to 2\overline{3}}}{p_{1\to 2\vee 3}} \leq \tilde{s}^{(1)}_{3+}\cdot p_{1\to 3} + \tilde{s}^{(2)}_{3+}\cdot p_{2\to 3}, \tag{38}$$

where (37) is of (28) when $(i,j,k) = (2,3,1)$ and (38) is of (27) when $(i,j,k) = (3,1,2)$. Note that the four values $\tilde{s}^{(3)}_{2-}$, $\tilde{s}^{(1)}_{2-}$, $\tilde{s}^{(1)}_{3+}$, and $\tilde{s}^{(2)}_{3+}$ are involved in the above equations (37) to (38). But if we follow the results of the above claim, then we have

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

where the first is of (33) when $(i,j,k) = (3,1,2)$; the second is of (31) when $(i,j,k) = (1,2,3)$; the third is of (32) when

$(i,j,k) = (1,2,3)$; and the last is of (30) when $(i,j,k) = (2,3,1)$.

Plugging the above substitutions into (37) and (38), then one can easily see that (37) and (38) are equivalent to (35) and (36), respectively.

In sum, 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 $\{t^{(i)}_{[\mathrm{u}]}, t^{(i)}_{[\mathrm{c},1]}, t^{(i)}_{[\mathrm{c},2]}, t^{(i)}_{[\mathrm{c},3]}, t^{(i)}_{[\mathrm{c},4]}\}$ for all $i \in \{1,2,3\}$ and have proven that they jointly satisfy the linear conditions of the closure version of Proposition 3. The proof of Lemma 2 is thus complete. □

Appendix C
Proofs of (10A) and (10B) for the Broadcasting Cut-Set Condition (10)

Firstly, (10A) can be derived as follows:

$$
\begin{aligned}
&(21) \triangleq I(\mathbf{W}_{i*}\,;\,[\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^n) \\
&= I(\mathbf{W}_{i*}\,;\,\mathbf{W}_{\{j,k\}*},[\mathbf{Y}_{*j},\mathbf{Y}_{*k},\mathbf{Z}]_1^n) && (39)\\
&\geq n(R_{i\to j}+R_{i\to k}+R_{i\to jk})(1-2\epsilon_n) - \frac{H_2(2\epsilon_n)}{\log_2 q}, && (40)
\end{aligned}
$$

where (39) 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 with each other; and (40) follows from Fano's inequality. Note that for any feasible network code, the messages $\mathbf{W}_{i*}$ can be decoded from $[\mathbf{Y}_{*j},\mathbf{Y}_{*k},\mathbf{Z}]_1^n$ and $\mathbf{W}_{\{j,k\}*} \triangleq \mathbf{W}_{j*}\cup\mathbf{W}_{k*}$ at nodes j and k, see (19), with error at most $2\epsilon_n > 0$ given n.

Secondly, (10B) can be derived as follows:

$$
\begin{aligned}
&(21) \triangleq I(\mathbf{W}_{i*}\,;\,[\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^n) \\
&= H([\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^n) \\
&\quad - H([\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{i,j,k\}*},[\mathbf{Z}]_1^n) && (41)\\
&= H([\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^n \,|\, \mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^n) && (42)\\
&= \sum_{t=1}^n H(\mathbf{Y}_{*j}(t),\mathbf{Y}_{*k}(t) \,| \\
&\qquad [\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^{t-1},\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^t) && (43)\\
&= \sum_{t=1}^n H(\mathbf{Y}_{*j}(t),\mathbf{Y}_{*k}(t) \,| \\
&\qquad [\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^{t-1},\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^t, X_j(t), X_k(t)) && (44)\\
&= \sum_{t=1}^n H(Y_{i\to j}(t),Y_{i\to k}(t) \,| \\
&\qquad [\mathbf{Y}_{*j},\mathbf{Y}_{*k}]_1^{t-1},\mathbf{W}_{\{j,k\}*},[\mathbf{Z}]_1^t, X_j(t), X_k(t)) && (45)\\
&\leq \sum_{t=1}^n \mathbb{E}\Big\{1_{\{\sigma(t)=i\}} \circ 1_{\{Z_{i\to j}(t)=1 \text{ or } Z_{i\to k}(t)=1\}}\Big\} && (46)\\
&= \sum_{t=1}^n \mathbb{E}\Big\{1_{\{\sigma(t)=i\}}\Big\}\mathbb{E}\Big\{1_{\{Z_{i\to j}(t)=1 \text{ or } Z_{i\to k}(t)=1\}}\Big\} && (47)\\
&= p_{i\to j\vee k}\mathbb{E}\left\{\sum_{t=1}^n 1_{\{\sigma(t)=i\}}\right\} = ns^{(i)}p_{i\to j\vee k}, && (48)
\end{aligned}
$$

where (41) follows from the definition of mutual information; (42) follows from Lemma 1 that any received symbols are functions of all the messages and the channel outputs; (43) follows from the chain rule and from the casuality such that the received symbols at time t do not depend on the future channel realizations $[\mathbf{Z}]_{t+1}^n$; (44) 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 (18); (45) 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 $\mathbf{Y}_{*k}(t)$) depends only on 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 (17); (46) 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; (47) follows from the fact that $\mathbf{Z}(t)$ is independently and identically distributed over the time axis. Since the scheduling decision $\sigma(t)$ depends only on the past channel outputs $[\mathbf{Z}]_1^{t-1}$, see (17), two indicator random variables are independent with each other; and (48) follows from the fact that $\mathbf{Z}(t)$ is independently and identically distributed over the time axis, the linearity of expectation, and the definition (20) given a scheme of interest.

Appendix D
Proofs of (11A) and (11B) for the 3-Way Multiple-Access Cut-Set Condition (11)

Firstly, (11B) can be derived as similar to (10B) in the following way:

$$
\begin{aligned}
&(22) \triangleq I(\mathbf{W}_{\{j,k\}*}\,;\,[\mathbf{Y}_{*i}]_1^n \,|\, \mathbf{W}_{i*},[\mathbf{Z}]_1^n) \\
&= H([\mathbf{Y}_{*i}]_1^n \,|\, \mathbf{W}_{i*},[\mathbf{Z}]_1^n) - H([\mathbf{Y}_{*i}]_1^n \,|\, \mathbf{W}_{\{i,j,k\}*},[\mathbf{Z}]_1^n) && (49)\\
&= H([\mathbf{Y}_{*i}]_1^n \,|\, \mathbf{W}_{i*},[\mathbf{Z}]_1^n) && (50)\\
&= \sum_{t=1}^n H(Y_{j\to i}(t),Y_{k\to i}(t) \,|\, [\mathbf{Y}_{*i}]_1^{t-1},\mathbf{W}_{i*},[\mathbf{Z}]_1^t), && (51)\\
&\leq n(s^{(j)}p_{j\to i}+s^{(k)}p_{k\to i}), && (52)
\end{aligned}
$$

where (49) follows from the definition of mutual information; (50) follows from Lemma 1 that any received symbols are functions of all the messages and the channel outputs; (51) follows from the chain rule and from the casuality such that the received symbols at time t do not depend on the future channel realizations $[\mathbf{Z}]_{t+1}^n$; and (52) follows from the similar discussions as in (46) to (48) for the proof of (10B) that the entropy is upper-bounded by 1 only when either $\sigma(t)=j$ with $Z_{j\to i}(t)=1$ or $\sigma(t)=k$ with $Z_{k\to i}(t)=1$, and further follows by the independently and identically distributed $\mathbf{Z}(t)$ over the time axis, the linearity of expectation, and the definition (20) given a scheme of interest.

We now show the detailed derivation of (11A). For the ease of exposition, we prove for the case when the node indices are fixed to $(i,j,k)=(1,2,3)$. Then (22) reduces to

$$I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}]_1^n \,|\, \mathbf{W}_{1*},[\mathbf{Z}]_1^n). \tag{53}$$

Namely, we will show that the following holds for any feasible scheme given n:

$$n\begin{pmatrix} R_{2\to 1}+R_{3\to 1}\\ +R_{2\to 31}+R_{3\to 12}\\ +\dfrac{p_{2\to 1}}{p_{2\to 3\vee 1}}R_{2\to 3}\\ +\dfrac{p_{3\to 1}}{p_{3\to 1\vee 2}}R_{3\to 2}\\ -6\epsilon_n-\dfrac{3H_2(\epsilon_n)}{n\log_2 q}\end{pmatrix}\overset{\text{(11A)}}{\leq}(53),$$

and the cases of other node indices $(i,j,k)\in\{(2,3,1),(3,1,2)\}$ can be followed symmetrically.

To that end, let us denote W_2 and W_3 as shorthand notations for the following collections of information messages:

$$\begin{aligned} W_2 &\triangleq \mathbf{W}_{1*}\cup\mathbf{W}_{*1}\cup\mathbf{W}_{3\to 2}\\ &=\mathbf{W}_{\{1,3\}*}\cup\mathbf{W}_{2\to 1}\cup\mathbf{W}_{2\to 31}, \end{aligned}\tag{54}$$

$$\begin{aligned} W_3 &\triangleq \mathbf{W}_{1*}\cup\mathbf{W}_{*1}\cup\mathbf{W}_{2\to 3}\\ &=\mathbf{W}_{\{1,2\}*}\cup\mathbf{W}_{3\to 1}\cup\mathbf{W}_{3\to 12}. \end{aligned}\tag{55}$$

That is, W_2 (resp. W_3) is a collection of all the 9-flow information messages except $\mathbf{W}_{2\to 3}$ (resp. $\mathbf{W}_{3\to 2}$). In other words, $W_2\cup\mathbf{W}_{2\to 3}$ and $W_3\cup\mathbf{W}_{3\to 2}$ implies all the 9-flow information messages $\mathbf{W}_{\{1,2,3\}*}$ in the 3-node network, respectively.

Consider the following claims, of which their proofs are relegated to the subsequent subsections.

Claim 1: The following is true:

$$(53)=\sum_{t=1}^{n} I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t)).\tag{56}$$

Claim 2: The following is true: $\forall\, t\in\{1,\cdots,n\}$,

$$\begin{aligned} &I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t))\\ &\geq I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t))\\ &\quad+\frac{p_{2\to 1}}{p_{2\to 3\vee 1}}I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{t-1},W_2,\mathbf{Z}(t))\\ &\quad+\frac{p_{3\to 1}}{p_{3\to 1\vee 2}}I(\mathbf{W}_{3\to 2}\,;\,\mathbf{Y}_{3*}(t)\,|\,[\mathbf{Y}_{3*},\mathbf{Z}]_1^{t-1},W_3,\mathbf{Z}(t)). \end{aligned}\tag{57}$$

Claim 3: The followings are true:

$$\begin{aligned} &\sum_{t=1}^{n} I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{t-1},\mathbf{W}_{1*},\mathbf{Z}(t))\\ &\qquad\qquad=I(\mathbf{W}_{*1}\,;\,\mathbf{W}_{1*},[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n}), \end{aligned}\tag{58}$$

$$\begin{aligned} &\sum_{t=1}^{n} I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{t-1},W_2,\mathbf{Z}(t))\\ &\qquad\qquad\geq I(\mathbf{W}_{2\to 3}\,;\,\mathbf{W}_{3*},[\mathbf{Y}_{*3},\mathbf{Z}]_1^{n}), \end{aligned}\tag{59}$$

$$\begin{aligned} &\sum_{t=1}^{n} I(\mathbf{W}_{3\to 2}\,;\,\mathbf{Y}_{3*}(t)\,|\,[\mathbf{Y}_{3*},\mathbf{Z}]_1^{t-1},W_3,\mathbf{Z}(t))\\ &\qquad\qquad\geq I(\mathbf{W}_{3\to 2}\,;\,\mathbf{W}_{2*},[\mathbf{Y}_{*2},\mathbf{Z}]_1^{n}). \end{aligned}\tag{60}$$

Then, one can easily see that by the above Claims 1 to 3 we have (11A):

$$\begin{aligned} (53) &\triangleq I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}]_1^n\,|\,\mathbf{W}_{1*},[\mathbf{Z}]_1^n)\\ &\geq I(\mathbf{W}_{*1}\,;\,\mathbf{W}_{1*},[\mathbf{Y}_{*1},\mathbf{Z}]_1^n)\\ &\quad+\frac{p_{2\to 1}}{p_{2\to 3\vee 1}}I(\mathbf{W}_{2\to 3}\,;\,\mathbf{W}_{3*},[\mathbf{Y}_{*3},\mathbf{Z}]_1^n)\\ &\quad+\frac{p_{3\to 1}}{p_{3\to 1\vee 2}}I(\mathbf{W}_{3\to 2}\,;\,\mathbf{W}_{2*},[\mathbf{Y}_{*2},\mathbf{Z}]_1^n), \qquad (61)\\ &\geq n(R_{2\to 1}+R_{3\to 1}+R_{2\to 31}+R_{3\to 12})(1-\epsilon_n)\\ &\quad-\frac{H_2(\epsilon_n)}{\log_2 q}+\frac{p_{2\to 1}}{p_{2\to 3\vee 1}}\left(nR_{2\to 3}(1-\epsilon_n)-\frac{H_2(\epsilon_n)}{\log_2 q}\right)\\ &\quad+\frac{p_{3\to 1}}{p_{3\to 1\vee 2}}\left(nR_{3\to 2}(1-\epsilon_n)-\frac{H_2(\epsilon_n)}{\log_2 q}\right), \qquad (62)\\ &\geq nR_{2\to 1}+nR_{3\to 1}+nR_{2\to 31}+nR_{3\to 12}\\ &\quad+n\frac{p_{2\to 1}}{p_{2\to 3\vee 1}}R_{2\to 3}+n\frac{p_{3\to 1}}{p_{3\to 1\vee 2}}R_{3\to 2}\\ &\quad-6n\epsilon_n-3\frac{H_2(\epsilon_n)}{\log_2 q}, \qquad (63) \end{aligned}$$

where (61) follows from jointly combining (56) to (60) in the above Claims 1 to 3; (62) follows from Fano's inequality; and (63) follows from the fact that any rate value is less than or equal to 1.

We now provide the proofs of Claims 1 to 3 in the subsequent subsections. To that end, the following lemmas are also useful.

Lemma 3: Let U_1 and U_2 be some random variables that are independent with each other, and $h_{11}(\cdot)$, $h_{12}(\cdot)$ and $h_2(\cdot)$ be some arbitrary function. Then, $h_{11}(U_1)\to h_{12}(U_1)\to h_2(U_2)$ forms a Markov chain. Namely, they satisfy

$$\begin{aligned} &I(h_{11}(U_1)\,;\,h_2(U_2)\,|\,h_{12}(U_1))=0,\\ &H(h_{11}(U_1)\,|\,h_{12}(U_1),h_2(U_2))=H(h_{11}(U_1)\,|\,h_{12}(U_1)). \end{aligned}$$

Proof of Lemma 3: The proof is omitted due to the simplicity. □

Lemma 4: In Scenario 1, fix a 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 a past channel output realization $[\mathbf{z}]_1^{t-1}\in[\mathbf{Z}]_1^{t-1}$ can uniquely decide $[X_1,X_3]_1^t$. Namely, $[X_1,X_3]_1^t$ are 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 4: As similar to Lemma 1, the proof follows from the 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 (18). As a result, $\{X_1(1),X_3(1)\}$ can be uniquely determined from the given $\mathbf{W}_{\{1,3\}*}$.

Assume that the statement of Lemma 4 is true until time $t-1$ such that $[X_1,X_3]_1^{t-1}$ can be uniquely determined by $\{\mathbf{W}_{\{1,3\}*},[\mathbf{Y}_{2*}]_1^{t-2},[\mathbf{z}]_1^{t-2}\}$. Now suppose that we also know $Y_{2\to 1}(t-1)$, $Y_{2\to 3}(t-1)$, and $\mathbf{z}(t-1)$. Since we already knew $[X_3]_1^{t-1}$, the received symbols $[Y_{3\to 1}]_1^{t-1}$ can be uniquely determined from the given $[\mathbf{z}]_1^{t-1}$. Jointly with the known messages $\mathbf{W}_{1*}$, the received symbols $[Y_{2\to 1}]_1^{t-1}$, and $[\mathbf{z}]_1^{t-1}$, we can also uniquely determine $X_1(t)$ by following the node 1's encoding description, see (18). Again from the given $[\mathbf{z}]_1^{t-1}$, the known input symbols $[X_1]_1^{t-1}$ uniquely determine $[Y_{1\to 3}]_1^{t-1}$. Then jointly with the already-known messages $\mathbf{W}_{3*}$, the received symbols $[Y_{2\to 3}]_1^{t-1}$, and $[\mathbf{z}]_1^{t-1}$, we can further uniquely determine $X_3(t)$ by following the node 3's encoding description, see (18). The proof of Lemma 4 is thus complete. □

Remark: The symmetrical version of Lemma 4 also holds when we swap the node indices 2 and 3. That is, for any time $t \in \{1, \cdots, n\}$, knowing the messages $\mathbf{W}_{\{1,2\}*}$, the received symbols $[\mathbf{Y}_{3*}]_1^{t-1}$, and a past channel output realization $[\mathbf{z}]_1^{t-1}$ can uniquely decide $[X_1, X_2]_1^t$. Namely, $[X_1, X_2]_1^t$ are function of the random variables $\{\mathbf{W}_{\{1,2\}*}, [\mathbf{Y}_{3*}]_1^{t-1}, [\mathbf{Z}]_1^{t-1}\}$ for any time $t \in \{1, \cdots, n\}$.

D-1. Proof of Claim 1

The equality (56) in Claim 1 can be proven as follows. Notice that

$$
\begin{aligned}
&(53) \triangleq I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}]_1^n \,|\, \mathbf{W}_{1*}, [\mathbf{Z}]_1^n) \\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}, \mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) - I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) \quad (64)\\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}, \mathbf{Z}]_1^n \,|\, \mathbf{W}_{1*}) \quad (65)\\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{1*}) \\
&\quad + I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Z}(n) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{1*}) \\
&\quad + I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(n) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{1*}, \mathbf{Z}(n)) \quad (66)\\
&= I(\mathbf{W}_{\{2,3\}*}\,;\,[\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1} \,|\, \mathbf{W}_{1*}) \\
&\quad + I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(n) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{n-1}, \mathbf{W}_{1*}, \mathbf{Z}(n)), \quad (67)
\end{aligned}
$$

where (64) follows from the chain rule; (65) follows from the fact that $\mathbf{W}_{\{2,3\}*}$, $\mathbf{W}_{1*}$ and $[\mathbf{Z}]_1^n$ are independent with each other; (66) follows from the chain rule; and (67) follows from Lemma 3 that the second term becomes zero when we choose $U_1 \triangleq \{\mathbf{W}_{\{1,2,3\}*}, [\mathbf{Z}]_1^{n-1}\}$ and $U_2 \triangleq \mathbf{Z}(n)$ that are independent with each other, because by Lemma 1 the received symbols $[\mathbf{Y}_{*1}]_1^{n-1}$ are basically a function of U_1.

Iteratively applying the equalities (66) to (67), we have the result. The proof of Claim 1 is thus complete.

D-2. Proof of Claim 2

Given a scheme of interest, fix a time slot $t \in \{1, \cdots, n\}$. The inequality (57) in Claim 2 can be proven as follows.

Since the scheduling decision $\sigma(t)$ is a function of the past CSI $[\mathbf{Z}]_1^{t-1}$, see (17), we consider the mutual information terms in (57) based on a past channel output realization $[\mathbf{z}]_1^{t-1} \in [\mathbf{Z}]_1^{t-1}$. To that end, let us define

$$\mathsf{term}_0^{[\mathbf{z}]} \triangleq I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z} \rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)), \quad (68)$$

$$\mathsf{term}_1^{[\mathbf{z}]} \triangleq I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z} \rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)), \quad (69)$$

$$\mathsf{term}_2^{[\mathbf{z}]} \triangleq I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}]_1^{t-1}, \langle \mathbf{z} \rangle, W_2, \mathbf{Z}(t)), \quad (70)$$

$$\mathsf{term}_3^{[\mathbf{z}]} \triangleq I(\mathbf{W}_{3\to 2}\,;\,\mathbf{Y}_{3*}(t) \,|\, [\mathbf{Y}_{3*}]_1^{t-1}, \langle \mathbf{z} \rangle, W_3, \mathbf{Z}(t)), \quad (71)$$

where we denote $\langle \mathbf{z} \rangle \triangleq \{[\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}\}$ as shorthand for the event of a specific channel output realizations $[\mathbf{z}]_1^{t-1}$ in the past. By the definition of mutual information, we have

$$
\begin{aligned}
&I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{t-1}, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\
&\qquad = \sum_{\forall [\mathbf{z}]_1^{t-1}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}) \cdot \mathsf{term}_0^{[\mathbf{z}]}, \quad (72)\\
&I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(t) \,|\, [\mathbf{Y}_{*1}, \mathbf{Z}]_1^{t-1}, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\
&\qquad = \sum_{\forall [\mathbf{z}]_1^{t-1}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}) \cdot \mathsf{term}_1^{[\mathbf{z}]}, \quad (73)\\
&I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t) \,|\, [\mathbf{Y}_{2*}, \mathbf{Z}]_1^{t-1}, W_2, \mathbf{Z}(t)) \\
&\qquad = \sum_{\forall [\mathbf{z}]_1^{t-1}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}) \cdot \mathsf{term}_2^{[\mathbf{z}]}, \quad (74)\\
&I(\mathbf{W}_{3\to 2}\,;\,\mathbf{Y}_{3*}(t) \,|\, [\mathbf{Y}_{3*}, \mathbf{Z}]_1^{t-1}, W_3, \mathbf{Z}(t)) \\
&\qquad = \sum_{\forall [\mathbf{z}]_1^{t-1}} \mathsf{Prob}([\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}) \cdot \mathsf{term}_3^{[\mathbf{z}]}. \quad (75)
\end{aligned}
$$

Then, we only need to prove that for all the past channel output realizations $[\mathbf{z}]_1^{t-1} \in [\mathbf{Z}]_1^{t-1}$, the following equation holds:

$$\mathsf{term}_0^{[\mathbf{z}]} \geq \mathsf{term}_1^{[\mathbf{z}]} + \frac{p_{2\to 1}}{p_{2\to 3\vee 1}} \mathsf{term}_2^{[\mathbf{z}]} + \frac{p_{3\to 1}}{p_{3\to 1\vee 2}} \mathsf{term}_3^{[\mathbf{z}]}, \quad (76)$$

and the reason is because (57), for any t, simply follows by the definitions (72) to (75) when we multiply the probability $\mathsf{Prob}([\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1})$ on both sides of (76), and then sum over all the past channel output realizations $[\mathbf{z}]_1^{t-1} \in [\mathbf{Z}]_1^{t-1}$.

To that end, given a scheme of interest, we divide all the past channel output realizations $[\mathbf{z}]_1^{t-1} \in [\mathbf{Z}]_1^{t-1}$ into the following disjoint groups based on the scheduling decision $\sigma(t)$, see (17), i.e., the node index scheduled at time t: For all $a \in \{1, 2, 3\}$,

$$\mathcal{Z}_{\sigma(t)=a} \triangleq \{\, \forall\, [\mathbf{z}]_1^{t-1} : \sigma(t) = a\}.$$

Specifically, we will prove the following:

- For all $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=1}$, we have

$$
\begin{aligned}
\mathsf{term}_0^{[\mathbf{z}]} &= 0, \quad (77)\\
\mathsf{term}_1^{[\mathbf{z}]} &= 0, \quad (78)\\
\mathsf{term}_2^{[\mathbf{z}]} &= 0, \quad (79)\\
\mathsf{term}_3^{[\mathbf{z}]} &= 0. \quad (80)
\end{aligned}
$$

- For all $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=2}$, we have

$$
\begin{aligned}
\mathsf{term}_0^{[\mathbf{z}]} &\geq \mathsf{term}_1^{[\mathbf{z}]} + \frac{p_{2\to 1}}{p_{2\to 3\vee 1}} \mathsf{term}_2^{[\mathbf{z}]}, \quad (81)\\
\mathsf{term}_3^{[\mathbf{z}]} &= 0. \quad (82)
\end{aligned}
$$

- For all $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=3}$, we have

$$
\begin{aligned}
\mathsf{term}_0^{[\mathbf{z}]} &\geq \mathsf{term}_1^{[\mathbf{z}]} + \frac{p_{3\to 1}}{p_{3\to 1\vee 2}} \mathsf{term}_3^{[\mathbf{z}]}, \quad (83)\\
\mathsf{term}_2^{[\mathbf{z}]} &= 0. \quad (84)
\end{aligned}
$$

Then, one can easily see that (77) to (84) jointly imply that (76) holds for all the past channel output realizations $[\mathbf{z}]_1^{t-1} \in [\mathbf{Z}]_1^{t-1}$.

First note that when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=1}$, (77) is true because

$$\begin{aligned} \mathsf{term}_0^{[\mathbf{z}]} &\triangleq I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\ &\leq H(\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle) \qquad (85) \\ &= 0, \qquad (86) \end{aligned}$$

where (85) follows from the definition of mutual information, non-negativity of entropy, and the fact that conditioning reduces entropy; and (86) 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 erasures.

Similarly applying the above arguments, one can prove that (78) to (80) are true as well when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=1}$. Moreover, one can also prove that (82) and (84) holds when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=2}$ and when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=3}$, respectively, by the similar arguments.

We now prove (81) and (83). To that end, we only show the proof for (81) of which is the case when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=2}$. Since (83) is about when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=3}$, the proof can be followed symmetrically by swapping the node indices 2 and 3. First, notice that

$$\begin{aligned} \mathsf{term}_0^{[\mathbf{z}]} &\triangleq I(\mathbf{W}_{\{2,3\}*}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\ &= I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, \mathbf{W}_{1*}, \mathbf{Z}(t)) \\ &\quad + I(\mathbf{W}_{3\to 2}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, \mathbf{W}_{1*}, \mathbf{W}_{*1}, \mathbf{Z}(t)) \\ &\quad + I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, W_2, \mathbf{Z}(t)) \qquad (87) \\ &\geq \mathsf{term}_1^{[\mathbf{z}]} + I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{*1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, W_2, \mathbf{Z}(t)) \qquad (88) \\ &= \mathsf{term}_1^{[\mathbf{z}]} + I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, W_2, \mathbf{Z}(t)) \qquad (89) \end{aligned}$$

where (87) follows from the chain rule and the definition W_2, see (54); (88) follows from the definition (69) and the non-negativity of mutual information; and (89) follows from that when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=2}$, the received symbol $Y_{3\to 1}(t) \subset \mathbf{Y}_{*1}(t)$ is fixed to an erasure.

We now prove that the second term in the right-hand side of (89) further satisfies

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, W_2, \mathbf{Z}(t)) \\ &= \frac{p_{2\to 1}}{p_{2\to 3\vee 1}} I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1}, \langle \mathbf{z}\rangle, W_2, \mathbf{Z}(t)), \qquad (90) \end{aligned}$$

and the proof is as follows.

Proof of (90): For the ease of exposition, let us denote $\mathbf{V} \triangleq \{[\mathbf{Y}_{*1}]_1^{t-1}, W_2\}$. Rewriting (90), then we have

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \mathbf{Z}(t)) \\ &= \frac{p_{2\to 1}}{p_{2\to 3\vee 1}} I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \mathbf{Z}(t)). \qquad (91) \end{aligned}$$

Then, note that

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \mathbf{Z}(t)) \\ &= I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \mathbf{Z}_{2*}(t)) \qquad (92) \\ &= \sum_{\forall\,\mathbf{z}_{2*}\in\{1,\varepsilon\}^2} \mathsf{Prob}(\{\mathbf{Z}_{2*}(t)=\mathbf{z}_{2*}\}\,|\,\{[\mathbf{Z}]_1^{t-1}=[\mathbf{z}]_1^{t-1}\}) \\ &\qquad \cdot I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\mathbf{z}_{2*}\}) \qquad (93) \\ &= \sum_{\forall\,\mathbf{z}_{2*}\in\{1,\varepsilon\}^2} \mathsf{Prob}(\{\mathbf{Z}_{2*}(t)=\mathbf{z}_{2*}\}) \\ &\qquad \cdot I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\mathbf{z}_{2*}\}) \qquad (94) \\ &= p_{2\to 31}\cdot I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\{1,1\}\}) \\ &\quad + p_{2\to \overline{3}1}\cdot I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\{\varepsilon,1\}\}) \qquad (95) \\ &= p_{2\to 31}\cdot I(\mathbf{W}_{2\to 3}\,;\,X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\{1,1\}\}) \\ &\quad + p_{2\to \overline{3}1}\cdot I(\mathbf{W}_{2\to 3}\,;\,X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\{\varepsilon,1\}\}), \qquad (96) \end{aligned}$$

where (92) follows from the definition $\mathbf{Z}_{2*}(t) = (Z_{2\to 3}(t), Z_{2\to 1}(t))$ and from that when $[\mathbf{z}]_1^{t-1} \in \mathcal{Z}_{\sigma(t)=2}$ meaning that node 2 is scheduled at time t, the channel outputs $Z_{1\to 2}(t)$, $Z_{1\to 3}(t)$, $Z_{3\to 1}(t)$, and $Z_{3\to 2}(t)$ are deterministically fixed to erasures; (93) follows from the definition of conditional mutual information when conditioned on the event $\langle \mathbf{z}\rangle \triangleq \{[\mathbf{Z}]_1^{t-1} = [\mathbf{z}]_1^{t-1}\}$; (94) follows from that $[\mathbf{Z}]_1^{t-1}$ and $\mathbf{Z}(t)$ are independent with each other; (95) follows from that when conditioned on the event of $\mathbf{Z}_{2*}(t) \triangleq (Z_{2\to 3}(t), Z_{2\to 1}(t))$ where $Z_{2\to 1}(t) = \varepsilon$, the received symbol $Y_{2\to 1}(t)$ is always fixed to an erasure, and thus the corresponding mutual information becomes zero; and (96) follows from that when conditioned on the event of $\mathbf{Z}_{2*}(t) \triangleq (Z_{2\to 3}(t), Z_{2\to 1}(t))$ where $Z_{2\to 1}(t) = 1$, the received symbol $Y_{2\to 1}(t)$ is always the same as the input packet $X_2(t)$ transmitted from node 2.

Then we can further remove the event of $\mathbf{Z}_{2*}(t)$ in the conditioning part of each mutual information term in (96). Namely, for any realization $\mathbf{z}_{2*} \in \{1,\varepsilon\}^2$, we have

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\mathbf{z}_{2*}\}) \\ &= H(X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\mathbf{z}_{2*}\}) \\ &\quad - H(X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \{\mathbf{Z}_{2*}(t)=\mathbf{z}_{2*}\}, \mathbf{W}_{2\to 3}) \qquad (97) \\ &= H(X_2(t)\,|\,\mathbf{V}) - H(X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \mathbf{W}_{2\to 3}) \qquad (98) \\ &= I(\mathbf{W}_{2\to 3}\,;\,X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle), \qquad (99) \end{aligned}$$

where (97) follows from the definition of mutual information; (98) follows from Lemma 3 when we choose $U_1 \triangleq \{\mathbf{W}_{\{1,2,3\}*}, [\mathbf{Z}]_1^{t-1}\}$ and $U_2 \triangleq \mathbf{Z}(t)$ that are independent with each other, because by Lemma 1 we know that $X_2(t)$, $\mathbf{V} \triangleq \{[\mathbf{Y}_{*1}]_1^{t-1}, W_2\}$, and $\mathbf{W}_{2\to 3}$ are functions of U_1, respectively; and (99) again follows from the definition of mutual information.

As a result, jointly combining (96) and (99), we thus have proven that

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,Y_{2\to 1}(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle, \mathbf{Z}(t)) \\ &\quad = p_{2\to 1}\cdot I(\mathbf{W}_{2\to 3}\,;\,X_2(t)\,|\,\mathbf{V}, \langle \mathbf{z}\rangle). \qquad (100) \end{aligned}$$

Similarly applying the above discussions to the mutual information term $I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,\mathbf{V},\langle\mathbf{z}\rangle,\mathbf{Z}(t))$ on the right-hand side of (91), we also have

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,\mathbf{V},\langle\mathbf{z}\rangle,\mathbf{Z}(t)) \\ &\quad= p_{2\to 3\vee 1}\cdot I(\mathbf{W}_{2\to 3}\,;\,X_2(t)\,|\,\mathbf{V},\langle\mathbf{z}\rangle). \end{aligned} \tag{101}$$

Manipulating over the common mutual information term $I(\mathbf{W}_{2\to 3}\,;\,X_2(t)\,|\,\mathbf{V},\langle\mathbf{z}\rangle)$ on both (100) and (101), then the result of (91) follows. The proof of (90) is complete. □

Then the mutual information term on the right-hand side of (90) further reduces to

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{Z}(t)) \\ &= H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{W}_{2\to 3},\mathbf{Z}(t)) && (102)\\ &= H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{*1}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{W}_{2\to 3},\mathbf{Z}(t)) && (103)\\ &\geq H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*},\mathbf{Y}_{3*}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{W}_{2\to 3},\mathbf{Z}(t)) && (104)\\ &= H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{Z}(t)) \\ &\quad - H(\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{W}_{2\to 3},\mathbf{Z}(t)) && (105)\\ &= I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(t)\,|\,[\mathbf{Y}_{2*}]_1^{t-1},\langle\mathbf{z}\rangle,W_2,\mathbf{Z}(t)) && (106)\\ &= \mathsf{term}_2^{[\mathbf{z}]}, && (107) \end{aligned}$$

where (102) follows from the definition of mutual information; (103) follows from that $W_2\cup\mathbf{W}_{2\to 3}$ is all the 9-flow information messages $\mathbf{W}_{\{1,2,3\}*}$ and thus follows from Lemma 1 that all the received symbols up to time $t-1$ can be uniquely determined given $\mathbf{W}_{\{1,2,3\}*}$ and a channel output realization $\langle\mathbf{z}\rangle\triangleq\{[\mathbf{Z}]_1^{t-1}=[\mathbf{z}]_1^{t-1}\}$ up to time $t-1$; (104) follows from the fact that conditioning reduces entropy; (105) follows from Lemma 4 that the messages $\{\mathbf{W}_{1*},\mathbf{W}_{3*}\}\subset W_2$, the received symbols $[\mathbf{Y}_{2*}]_1^{t-1}$, and a past channel output realization $[\mathbf{z}]_1^{t-1}$ can uniquely decide $[X_3]_1^t$, and thus by the given $[\mathbf{z}]_1^{t-1}$, the received symbols $[\mathbf{Y}_{3*}]_1^{t-1}$ are uniquely determined as well; (106) follows from the definition of mutual information; and (107) follows from the definition (70).

As a result, (81) follows by combining (89), (90), and (107). The proof of Claim 2 is thus complete.

D-3. Proof of Claim 3

We provide the proofs for the (in)equalities (58) to (60) in Claim 3. All the proofs can be followed by the iterative process as similar to the proof of (56) in Claim 1. We first show the proof for (58).

Proof of (58): Note that

$$\begin{aligned} &I(\mathbf{W}_{*1}\,;\,\mathbf{W}_{1*},[\mathbf{Y}_{*1},\mathbf{Z}]_1^n) \\ &= I(\mathbf{W}_{*1}\,;\,\mathbf{W}_{1*})+I(\mathbf{W}_{*1}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^n\,|\,\mathbf{W}_{1*}) && (108)\\ &= I(\mathbf{W}_{*1}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^n\,|\,\mathbf{W}_{1*}) && (109)\\ &= I(\mathbf{W}_{*1}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1}\,|\,\mathbf{W}_{1*}) \\ &\quad + I(\mathbf{W}_{*1}\,;\,\mathbf{Z}(n)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1},\mathbf{W}_{1*}) \\ &\quad + I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(n)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1},\mathbf{W}_{1*},\mathbf{Z}(n)) && (110)\\ &= I(\mathbf{W}_{*1}\,;\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1}\,|\,\mathbf{W}_{1*}) \\ &\quad + I(\mathbf{W}_{*1}\,;\,\mathbf{Y}_{*1}(n)\,|\,[\mathbf{Y}_{*1},\mathbf{Z}]_1^{n-1},\mathbf{W}_{1*},\mathbf{Z}(n)), && (111) \end{aligned}$$

where (108) follows from the chain rule; (109) follows from the fact the messages $\mathbf{W}_{*1}$ and $\mathbf{W}_{1*}$ are independent with each other; (110) follows from the chain rule; and (111) follows by Lemma 3 that the second term becomes zero when we choose $U_1\triangleq\{\mathbf{W}_{\{1,2,3\}*},[\mathbf{Z}]_1^{n-1}\}$ and $U_2\triangleq\mathbf{Z}(n)$ that are independent with each other, because by Lemma 1 the received symbols $[\mathbf{Y}_{*1}]_1^{n-1}$ are basically a function of U_1. As a result, iteratively applying the equalities (110) to (111), the result (58) follows. □

Secondly, we prove (59) and (60). To that end, we only show the proof for (59) and then the proof of (60) can be followed symmetrically by swapping the node indices 2 and 3.

Proof of (59): Note that

$$\begin{aligned} &I(\mathbf{W}_{2\to 3}\,;\,\mathbf{W}_{3*},[\mathbf{Y}_{*3},\mathbf{Z}]_1^n) \\ &\leq I(\mathbf{W}_{2\to 3}\,;\,\mathbf{W}_{\{1,3\}*},\mathbf{W}_{2\to 1},\mathbf{W}_{2\to 31},[\mathbf{Y}_{2*},Y_{1\to 3},\mathbf{Z}]_1^n) && (112)\\ &= I(\mathbf{W}_{2\to 3}\,;\,\mathbf{W}_{\{1,3\}*},\mathbf{W}_{2\to 1},\mathbf{W}_{2\to 31},[\mathbf{Y}_{2*},\mathbf{Z}]_1^n) && (113)\\ &= I(\mathbf{W}_{2\to 3}\,;\,W_2)+I(\mathbf{W}_{2\to 3}\,;\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^n\,|\,W_2), && (114)\\ &= I(\mathbf{W}_{2\to 3}\,;\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^n\,|\,W_2), && (115)\\ &= I(\mathbf{W}_{2\to 3}\,;\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{n-1}\,|\,W_2) \\ &\quad + I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Z}(n)\,|\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{n-1},W_2) \\ &\quad + I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(n)\,|\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{n-1},W_2,\mathbf{Z}(n)) && (116)\\ &= I(\mathbf{W}_{2\to 3}\,;\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{n-1}\,|\,W_2) \\ &\quad + I(\mathbf{W}_{2\to 3}\,;\,\mathbf{Y}_{2*}(n)\,|\,[\mathbf{Y}_{2*},\mathbf{Z}]_1^{n-1},W_2,\mathbf{Z}(n)), && (117) \end{aligned}$$

where (112) follows from the fact that adding the observations $\mathbf{W}_{1*}$, $\mathbf{W}_{2\to 1}$, $\mathbf{W}_{2\to 31}$, and $[Y_{2\to 1}]_1^n$ increases the mutual information; (113) follows from Lemma 4 that the input symbols $[X_1]_1^n$ from node 1 are function of the messages $\mathbf{W}_{\{1,3\}*}$, the received symbols $[\mathbf{Y}_{2*}]_1^{n-1}$, and the channel outputs $[\mathbf{Z}]_1^{n-1}$, and thus follows from the fact that the received symbols $[Y_{1\to 3}]_1^n$ are function of $[X_1]_1^n$ and $[\mathbf{Z}]_1^n$. Namely, $[Y_{1\to 3}]_1^n$ are function of $\{\mathbf{W}_{\{1,3\}*},[\mathbf{Y}_{2*}]_1^{n-1},[\mathbf{Z}]_1^n\}$ and thus removing the observation $[Y_{1\to 3}]_1^n$ does not decrease the mutual information; (114) follows from the chain rule and the definition (54); (115) follows from the fact the messages $\mathbf{W}_{2\to 3}$ and W_2 are independent with each other; (116) follows from the chain rule; and (117) follows from Lemma 3 that the second term becomes zero when we choose $U_1\triangleq\{\mathbf{W}_{\{1,2,3\}*},[\mathbf{Z}]_1^{n-1}\}$ and

$U_2 \triangleq \mathbf{Z}(n)$ that are independent with each other, because by Lemma 1 the received symbols $[\mathbf{Y}_{2*}]_1^{n-1}$ are basically a function of U_1. As a result, iteratively applying the equalities (116) to (117), the result (59) follows. $\square$