Two-Flow Capacity Region Of The COPE Principle For Wireless Butterfly Networks with Broadcast Erasure Channels

Wei-Cheng Kuo, Chih-Chun Wang *Member, IEEE*
Center of Wireless Systems and Applications (CWSA)
School of Electrical and Computer Engineering, Purdue University, USA

***Abstract*—This paper characterizes the full capacity region of the *COPE principle* for 2-flow wireless butterfly networks with broadcast packet erasure channels (PECs). The capacity results in this work allow for random overhearing with arbitrary overhearing probabilities, arbitrary scheduling policies, network-wide channel state information (CSI) feedback after each transmission, and potential use of non-linear network codes. An information-theoretic outer bound is derived that takes into account the delayed CSI feedback of the underlying broadcast packet erasure channels. For the achievability, this paper proposes a new class of linear network codes, named as the Space-Based Linear Network Coding (SBLNC), that achieves the capacity outer bound. Further, the proposed outer and inner bounds are later generalized for the setting in which a transmission may be heard by its 2-hop neighbor(s), the so-called *opportunistic routing* scenario. When allowing the possibility of opportunistic routing, the proposed inner and outer bounds do not always meet. Numerical experiments, however, show that the relative gap of the two bounds is less than 0.08% in average. The proposed bounds thus tightly bracket the capacity region even when combining the COPE principle with opportunistic routing.**

I. Introduction

The seminal work by Li *et al.* in 2003 [1] shows that, linear network coding (LNC) achieves the single-flow multicast network capacity. However, most flows are unicast. For the setting of multiple unicast flows, the capacity region of network coding (and/or LNC) remains largely unknown. In 2006, Katti *et al.* proposed a new multiple-unicast LNC protocol, called "COPE" [2], which realizes the network coding gain by exploiting the wireless network diversity created by overhearing packets of other coexisting flows. Take the 5-node wireless butterfly network in Fig. 1(a) for example. Suppose source s_1 would like to send a packet X to destination d_1; source s_2 would like to send a packet Y to d_2; and they share a common relay r. Also suppose that when s_1 (resp. s_2) sends X (resp. Y) to r, destination d_2 (resp. d_1) can overhear packet X (resp. Y). We further assume that after the first two transmissions,

Manuscript received October 29, 2012; revised May 19, 2013; accepted August 2, 2013. This work was supported in parts by NSF grants CCF-0845968 and CNS-0905331. Part of the results was presented in the 2011 IEEE International Symposium on Information Theory.

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: wkuo@purdue.edu; chihw@purdue.edu).

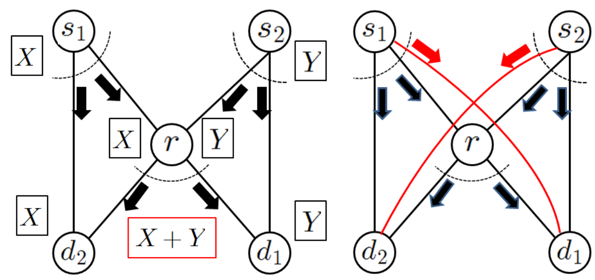


Fig. 1. The illustration of (a) The COPE principle; and (b) The scenario of 2-hop overhearing for which one can apply opportunistic routing to enhance the throughput.

both d_1 and d_2 can use feedback[1] to inform r the overhearing status at d_1 and d_2, respectively. Then instead of transmitting two packets X and Y separately, the relay node r can send the linear combination $[X+Y]$. Each destination d_i can then decode its desired packet by subtracting the overheard packet from the linear combination $[X+Y]$. In [2], it was shown empirically that the throughput improvement of the above COPE principle ranges from 40% to 200% in a multi-hop testbed environment. Throughout this paper, we use the term "COPE principle" to denote the above scenario that involves at least two sessions and periodic reception status feedback. The COPE principle exploits the following two types of coding gain to improve the overall throughput. The first one is "*the side information*" coding gain, for which the destination d_j overhears the transmission from s_i [3], [4]. The second one is "*the feedback coding gain in broadcast channels*," which follows from the unmatched reception status when relay r is broadcasting packets to d_1 and d_2, respectively [5]–[7].

Despite its simple nature, the exact capacity region of the COPE principle remains an open problem even for the simplest case of two coexisting flows. Several attempts have since been made to quantify some suboptimal achievable rate regions of the COPE principle [8]–[17]. One difficulty of deriving the capacity region is due to the use of feedback in the COPE principle. It is shown in [5] that although feedback could strictly enhance the capacity in a multi-unicast environment, the

[1]The use of feedback is critical in the COPE principle. Without feedback, the relay r will not know that sending the linear combination $[X+Y]$ can serve two flows simultaneously.

exact amount of throughput improvement is hard to quantify. Compared with the capacity-based approaches, the results in [4] proposes a queue-based approach for the general wireline and wireless networks while considering both inter-session and intra-session network coding. The proposed queue-based scheme explores the possible inter-session network coding chance by tracking the side information received from the opposite source. However, the results in [4] mainly focus on the side information benefits and on deciding whether to apply inter-session or intra-session network coding. Similarly, [3] circumvents the difficulty of feedback-based analysis by considering a special class of 2-staged coding schemes. Although the results in [3] fully quantify the benefits of message side information [4], [5], [18]–[22] , they capture only partially the feedback benefits, which leads again to a strictly suboptimal achievable rate region.

Recently, [6] and [7] successfully characterized the full capacity region of the 1-hop broadcast packet erasure channel with ≤ 3 coexisting flows. The results in [6] and [7] prompt the possibility of fully characterizing the capacity of the COPE principle over the wireless butterfly network in Fig. 1(a). For comparison, our work focuses on the wireless butterfly network in Fig. 1(a) while [6] and [7] focus on the 1-hop broadcast channel. For the wireless butterfly network in Fig. 1(a), the network designer faces both the *scheduling problem*: which node (out of the two source nodes s_1, s_2, and the relay node r) to transmit at the current time slot, and the *network coding problem*: how to combine the heard/overheard packets and generate the network coded packets. For the 1-hop broadcast channel considered in [6] and [7], there is no scheduling problem since there is only one base station and the base station transmits all the time. As will be seen shortly, for a wireless butterfly network, the feedback/control messages may propagate through the entire network and affect dynamically the scheduling and coding decisions for all three nodes s_1, s_2, and r, which further complicates the analysis.

In this paper, we first characterize the full capacity region of the above COPE principle for 2-flow wireless butterfly networks with broadcast PECs. The setting considered in this work allows for memoryless random overhearing with arbitrary overhearing probabilities, arbitrary scheduling policies[2], network-wide channel state information (CSI) feedback after each transmission, and potential use of non-linear network codes. An information-theoretic outer bound is derived that fully takes into account the delayed CSI feedback of the underlying broadcast packet erasure channels. This paper then proposes a new class of linear network codes, named as the Space-Based Linear Network Coding (SBLNC), for which the throughput can be quantified by solving a linear programming problem. We prove that the SBLNC scheme achieves the capacity region of the COPE principle.

Recently, new opportunistic routing (OpR) protocols, such as the MORE protocol [23]–[25], take advantage of the observation that in Fig. 1(a) the packet sent by source s_i can sometimes be heard directly by its two-hop-away destination

[2]In general, a scheduling policy exchanges control messages between the network nodes and then decides which node to transmit in the current time slot.

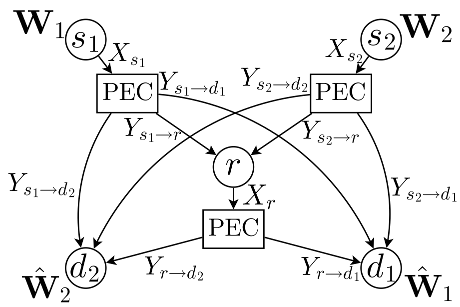


Fig. 2. The 2-flow wireless butterfly network with broadcast PECs.

d_i. (The 2-hop overhearing probability is generally much less than the success probability of the 1-hop direct transmission from s_i to r.) See Fig. 1(b) for illustration. [23] and [25] show that such 2-hop overhearing can also be used to significantly improve the throughput by the concept of *opportunistic routing* without relying on the COPE principle (i.e., no coding across multiple unicast flows). Throughout this paper, we will use "OpR" to denote the scenario in which d_i can directly hear the transmission of s_i with non-zero probability[3] for $i = 1, 2$. An interesting question is thus what is the capacity region when we combine both the COPE principle and opportunistic routing. To that end, the proposed outer and inner bounds are further generalized for the setting in which a transmission may be heard by its 2-hop neighbor(s). In contrast with the case when only the COPE principle is considered, the proposed inner and outer bounds are no longer tight after taking into account the possibility of performing opportunistic routing. Extensive numerical experiments are used to quantify the gap between the two bounds. The results show that for practical settings, the relative gap of the two bounds is less than 0.08% in average. The proposed bounds thus tightly bracket the capacity region even when combining the COPE principle with opportunistic routing.

The rest of the paper is organized as follows. Section II formulates the problem. Section III describes the main result of this paper, the full capacity region of the COPE principle for 2-flow wireless butterfly networks with broadcast PECs. Section IV introduces a new class of LNC, named as the Space-Based Linear Network Code (SBLNC), and provides some examples and motivations for designing the SBLNC code. Section V quantifies the performance of SBLNC and shows that when focusing exclusively on the COPE principle without considering opportunistic routing, SBLNC achieves the capacity. Section VI reports the results of serveral numerical experiments. Section VII concludes the paper.

II. Problem Formulation

A. Memoryless Broadcast Packet Erasure Channels

For any positive integer M, define $[M] \triangleq \{1, \cdots, M\}$. A 1-to-$M$ packet erasure channel takes an input x from a

[3]The concept of "*opportunistic routing*" is different from the term "*intra-session random linear network coding (RLNC)*." For example, one way of realizing the OpR throughput improvement is by RLNC [23]. However, there are other ways of realizing the OpR benefits [26]. Also, when periodic feedback is not available [3], [4], [17], it is shown that even without direct links, one still needs to use intra-session network coding to fully realize the side information benefits of the COPE principle.

finite field $\mathrm{GF}(q)$ and outputs an M-dimensional vector $\mathbf{y} = (y_1, y_2, ..., y_M)$, where $y_i \in \{x, *\}$ for all $i \in [M]$. Here $*$ denotes the erasure symbol. $y_i = *$ means that the i-th receiver does not receive the input x. We also assume that there is no other type of noise, i.e., the received y_i is either x or $*$.

We consider only *memoryless PECs*, i.e., the erasure pattern is independently and identically distributed (i.i.d.) for each channel usage. The characteristics of a memoryless 1-to-M PEC can be fully described by 2^M *successful reception probabilities* $p_{s\to T\overline{[M]\backslash T}}$ indexed by any subset $T \subset [M]$. That is, $p_{s\to T\overline{[M]\backslash T}}$ denote the probability that a packet x sent from source s is heard by and only by the i-th destination for all $i \in T$.

B. The COPE Principle for 2-Flow Wireless Butterfly Networks with Broadcast PECs

The COPE principle for the 2-flow wireless butterfly networks with broadcast PECs is modeled as follows. We consider a 5-node 2-hop relay network with two source-destination pairs (s_1, d_1) and (s_2, d_2) and a common relay r interconnected by three broadcast PECs. See Fig. 2 for illustration. Specifically, source s_i can use a 1-to-3 broadcast PEC to communicate with $\{d_1, d_2, r\}$ for $i = 1, 2$, and relay r can use a 1-to-2 broadcast PEC to communicate with $\{d_1, d_2\}$. To accommodate the discussion of opportunistic routing, we allow s_i to directly communicate with d_i, see Fig. 2. When opportunistic routing is not permitted (as in the case when focusing exclusively on the COPE principle), we simply choose the PEC channel success probabilities $p_{s_i\to\cdot}$ such that the probability that d_i can hear the transmission from s_i is zero.

We assume slotted transmission. Within an overall time budget of n time slots, source s_i would like to convey nR_i packets $\mathbf{W}_i \triangleq (W_{i,1}, \cdots, W_{i,nR_i})$ to destination d_i for all $i \in \{1, 2\}$ where R_i is the rate for flow i. For all $i \in \{1, 2\}$, $j \in [nR_i]$, the information packet $W_{i,j}$ is assumed to be independently and uniformly randomly distributed over $\mathsf{GF}(q)$.

For any time t, we use an 8-dimensional *channel status vector* $\mathbf{Z}(t)$ to represent the channel reception status of the entire network:

$$\begin{aligned}\mathbf{Z}(t) = (&Z_{s_1\to d_1}(t), Z_{s_1\to d_2}(t), Z_{s_1\to r}(t), Z_{s_2\to d_1}(t),\\ &Z_{s_2\to d_2}(t), Z_{s_2\to r}(t), Z_{r\to d_1}(t), Z_{r\to d_2}(t)) \in \{*, 1\}^8\end{aligned}$$

where "$*$" and "1" represent erasure and successful reception, respectively. For example, when s_1 transmits a packet $X_{s_1}(t) \in \mathsf{GF}(q)$ in time t, relay r receives $Y_{s_1\to r}(t) = X_{s_1}(t)$ if $Z_{s_1\to r}(t) = 1$ and receives $Y_{s_1\to r}(t) = *$ if $Z_{s_1\to r}(t) = *$. For simplicity, we use $Y_{s_1\to r}(t) = X_{s_1}(t) \circ Z_{s_1\to r}(t)$ as shorthand.

In this work, we consider the node-exclusive interference model. That is, we allow only one node to be scheduled in each time slot. The scheduling decision at time t is denoted by $\sigma(t)$, which takes value in the set $\{s_1, s_2, r\}$. For example, $\sigma(t) = s_1$ means that node s_1 is scheduled for time slot t. For convenience, when s_1 is not scheduled at time t, we simply set $Y_{s_1\to r}(t) = *$. As a result, the scheduling decision can be incorporated into the following expression of $Y_{s_1\to r}(t)$:

$$Y_{s_1\to r}(t) = X_{s_1}(t) \circ Z_{s_1\to r}(t) \circ 1_{\{\sigma(t)=s_1\}}.$$

Similar notation is used for all other received signals. For example, $Y_{r\to d_2}(t) = X_r(t) \circ Z_{r\to d_2}(t) \circ 1_{\{\sigma(t)=r\}}$ is what d_2 receives from r in time t, where $X_r(t)$ is the packet sent by r in time t.

We assume that the 3 PECs are memoryless and stationary. Namely, we allow arbitrary joint distribution for the 8 coordinates of $\mathbf{Z}(t)$ but assume that $\mathbf{Z}(t)$ is i.i.d. over the time axis t. For example, the individual random variables $Z_{s_2\to d_1}(t)$ and $Z_{s_2\to r}(t)$ may be dependent but the two random vectors $\mathbf{Z}(t_1)$ and $\mathbf{Z}(t_2)$ are independent as long as $t_1 \neq t_2$. We also assume that the random process $\{\mathbf{Z}(t) : \forall t\}$ is independent of the information messages $\mathbf{W}_1$ and $\mathbf{W}_2$.

For simplicity, we use brackets $[\cdot]_1^t$ to denote the collection from time 1 to t. For example, $[\sigma, \mathbf{Z}, Y_{s_1\to d_2}]_1^t \triangleq \{\sigma(\tau), \mathbf{Z}(\tau), Y_{s_1\to d_2}(\tau) : \forall \tau \in \{1, 2, ..., t\}\}$. Also, for any $S \subseteq \{s_1, s_2, r\}$ and $T \subseteq \{r, d_1, d_2\}$, we define

$$\mathbf{Y}_{S\to T}(t) \triangleq \{Y_{s\to d}(t) : \forall s \in S, \forall d \in T\}.$$

For example, $\mathbf{Y}_{\{s_1,r\}\to\{d_1,d_2\}}(t)$ is the collection of $Y_{s_1\to d_1}(t)$, $Y_{s_1\to d_2}(t)$, $Y_{r\to d_1}(t)$, and $Y_{r\to d_2}(t)$.

Given the rate vector (R_1, R_2), a joint scheduling and network coding (NC) scheme is defined by n scheduling decision functions

$$\forall t \in [n],\ \sigma(t) = f_{\sigma,t}([\mathbf{Z}]_1^{t-1}), \tag{1}$$

$3n$ encoding functions at s_1, s_2, and r, respectively: For all $t \in [n]$

$$X_{s_i}(t) = f_{s_i,t}(\mathbf{W}_i, [\mathbf{Z}]_1^{t-1}), \quad \forall i \in \{1, 2\}, \tag{2}$$

$$X_r(t) = f_{r,t}([\mathbf{Y}_{\{s_1,s_2\}\to r}, \mathbf{Z}]_1^{t-1}), \tag{3}$$

and 2 decoding functions at d_1 and d_2, respectively:

$$\hat{\mathbf{W}}_i = f_{d_i}([\mathbf{Y}_{\{s_1,s_2,r\}\to d_i}, \mathbf{Z}]_1^n), \quad \forall i \in \{1, 2\}. \tag{4}$$

By (1), we allow $\sigma(t)$, the scheduling decision at time t, to be a function of the network-wide reception status vectors before time t. By (2), the encoding decision at s_i is a function depending on the information messages and past channel status. Encoding at r depends on what r received in the past and the past channel status vector, see (3). In the end, d_i decodes $\mathbf{W}_i$ based on what d_i has received and the past channel status of the entire network[4]. We allow the encoding and decoding functions $f_{s_i,t}$, $f_{r,t}$, and f_{d_i} to be linear or nonlinear.

This setting models the scenario in which there is a dedicated, error-free, low-rate control channel that can broadcast the previous network channel status $\mathbf{Z}(t-1)$ causally to all network nodes. The total amount of control information is no larger than 8 bits per time slot, which is much smaller than the actual payload of each packet $\approx 10^4$ bits. As a result, the perfect feedback channel could be implemented by

[4]Since the scheduling decision $\sigma(t)$ is a function of $[\mathbf{Z}]_1^{t-1}$, all the encoding functions in (2) and (3), and the decoding functions in (4) also know implicitly the scheduling decision $\sigma(t)$.

piggybacking[5] on the data packets. The scheduling decision $\sigma(t)$ can be computed centrally (by a central controller) or distributively by each individual node since we allow all nodes to have the knowledge of the reception status of the entire network. This work mainly focuses on the theoretical capacity analysis of the feedback-enabled butterfly networks. Hence we omit the discussion of any practical implementation issues like buffer management and time synchronization, the latter of which is critical when implementing the scheduler $\sigma(t)$.

Definition 1: Fix the distribution of $\mathbf{Z}(t)$ and finite field $\mathsf{GF}(q)$. A rate vector (R_1, R_2) is achievable if for any $\epsilon > 0$, there exists a joint scheduling and NC scheme with sufficiently large n such that

$$\max_{\forall i \in \{1,2\}} \mathsf{Prob}(\mathbf{W}_i \neq \hat{\mathbf{W}}_i) < \epsilon.$$

The capacity region is defined as the closure of all achievable rate vectors (R_1, R_2).

Remark: In (1), the scheduling decision $\sigma(t)$ does not depend on the information messages $\mathbf{W}_i$, which means that we prohibit the use of timing channels [27], [28]. Even when we allow the usage of timing channels, we conjecture that the overall capacity improvement with the timing channel techniques is negligible. A heuristic argument is that each successful packet transmission gives $\log_2(q)$ bits of information while the timing information (to transmit or not) gives roughly 1 bit of information. When focusing on sufficiently large $\mathsf{GF}(q)$, additional gain of timing information is thus likely to be absorbed in our timing-information-free capacity characterization. In our setting, r is the only node that can mix packets from two different data flows. Further relaxation such that s_1 and s_2 can hear each other and perform coding accordingly is beyond the scope of this work.

C. A Useful Notation

In our network model, there are 3 broadcast PECs associated with s_1, s_2, and r, respectively. We sometimes call those PECs the s_i-PEC, $i = 1, 2$, and the r-PEC. Since only one node can be scheduled in each time slot, we can assume that the reception events of each PEC are independent from that of the other PECs. As a result, the distribution of the network-wide channel status vector $\mathbf{Z}(t)$ can be described by the probabilities $p_{s_i \to T\overline{\{r,d_1,d_2\}\backslash T}}$ for all $i \in \{1, 2\}$ and for all $T \subseteq \{r, d_1, d_2\}$, and $p_{r \to U\overline{\{d_1,d_2\}\backslash U}}$ for all $U \subseteq \{d_1, d_2\}$. Totally there are $8 + 8 + 4 = 20$ parameters. By allowing some of the coordinates of $\mathbf{Z}(t)$ to be correlated (i.e., spatially correlated as the correlation is between coordinates, not over the time axis), our setting can also model the scenario in which destinations d_1 and d_2 are situated in the same physical node and thus have perfectly correlated channel success events.

For notational simplicity, we also define the following three *probability* functions $p_{s_i}(\cdot)$, $i = 1, 2$, and $p_r(\cdot)$, one for each of the PECs. The input argument of each function p_s (s being one of $\{s_1, s_2, r\}$) is a collection of the elements in $\{d_1, d_2, r, \overline{d_1}, \overline{d_2}, \overline{r}\}$. The function $p_s(\cdot)$ outputs the probability that the reception event is *compatible* to the specified collection of $\{d_1, d_2, r, \overline{d_1}, \overline{d_2}, \overline{r}\}$. For example,

$$p_{s_1}(d_2\overline{r}) = p_{s_1 \to d_2\overline{d_1 r}} + p_{s_1 \to d_1 d_2 \overline{r}} \tag{5}$$

is the probability that the input of the s_1-PEC is successfully received by d_2 but not by r. Herein, d_1 is a *don't-care* receiver and $p_{s_1}(d_2\overline{r})$ thus sums two joint probabilities together (d_1 receives it or not) as described in (5). Another example is $p_r(d_2) = p_{r \to d_1 d_2} + p_{r \to \overline{d_1} d_2}$, which is the probability that a packet sent by r is heard by d_2. To slightly abuse the notation, we further allow $p_s(\cdot)$ to take multiple input arguments separated by the comma sign ",". With this new notation, $p_s(\cdot)$ then represents the probability that the reception event is compatible to *at least* one of the input arguments. For example,

$$p_{s_1}(d_1\overline{d_2}, r) = p_{s_1 \to d_1\overline{d_2 r}} + p_{s_1 \to d_1 \overline{d_2} r} + p_{s_1 \to d_1 d_2 r} + p_{s_1 \to \overline{d_1} d_2 r} + p_{s_1 \to \overline{d_1 d_2} r}.$$

That is, $p_{s_1}(d_1\overline{d_2}, r)$ represents the probability that $(Z_{s_1 \to d_1}, Z_{s_1 \to d_2}, Z_{s_1 \to r})$ equals one of the following 5 vectors $(1, *, *)$, $(1, *, 1)$, $(1, 1, 1)$, $(*, 1, 1)$, and $(*, *, 1)$. Note that these 5 vectors are compatible to either $d_1\overline{d_2}$ or r or both. Another example of this $p_s(\cdot)$ notation is $p_{s_1}(d_1, d_2, r)$, which represents the probability that a packet sent by s_1 is received by *at least* one of the three nodes d_1, d_2, and r. In the following context, we slightly abuse the notation $p_i(\cdot) \triangleq p_{s_i}(\cdot)$ for $i = 1, 2$ when there is no ambiguity.

TABLE I
THE COMPARISONS TO THE EXISTING WORK [3]

	Features in [3]	Features in this work
Setting	(1) Allow only sequential scheduling, (2) Allow only batch feedback, (3) Arbitrary number of flows.	(1) Allow dynamic scheduling, (2) Allow per-packet feedback, (3) 2 flows.
The outer bound	Side-information-based index coding analysis.	Full feedback and scheduling analysis.
The inner bound	(1) Sequential scheduling, (2) Linear coding functions.	(1) Sequential scheduling, (2) Linear coding functions.
Contribution	Suboptimal achievability scheme for an arbitrary number of flows.	Exact 2-flow capacity region of the COPE principle with per packet scheduling.

D. Comparison to Existing Works and Practical COPE Protocol

In this subsection, we compare the network model of this paper with the existing message-side-information-based results [3] and the practical COPE-based schemes.

There are three major differences between the setting of this work and in [3]. First, the setting of the outer bound in [3] is restricted to considering only the deterministic sequential scheduling policy, which schedules nodes s_1, s_2, and r in a strict order. Namely, s_1 transmits first. After s_1 stops, s_2 can begin to transmit. Only after s_2 stops transmission can r start its own transmission. For comparison, the setting of our outer bound derivation allows for dynamically choosing

[5]Some pipelining may be necessary to mitigate the propagation delay of the feedback control messages.

the schedule $\sigma(t)$ for each time slot t depending on the past reception status $[\mathbf{Z}]_1^{t-1}$, which includes any store-&-forward-based scheduling policies as special cases, such as the back-pressure and the maximal weighted matching schemes (see [25] for references). Our outer bound result thus quantifies the best possible achievable rates with jointly designed scheduling and coding policies.

Secondly, in [3] no feedback is allowed when s_1 and s_2 transmit. More specifically, suppose jointly s_1 and s_2 take $t_{s_1} + t_{s_2}$ time slots to finish transmission. Then only in the beginning of time $(t_{s_1} + t_{s_2} + 1)$ are we allowed to send the channel status $[\mathbf{Z}]_1^{t_{s_1}+t_{s_2}}$ to r. No further feedback is allowed until time n, the end of overall transmission. For comparison, our setting allows constantly broadcasting network-wide channel status $[\mathbf{Z}]_1^{t-1}$ to s_1, s_2, and r, as discussed in Section II-B. This setting thus includes the Automatic Repeat reQuest (ARQ) mechanism as a special case [3], [6]. Broadcasting the control information $[\mathbf{Z}]_1^{t-1}$ to all the network nodes also eliminates the need of estimating/learning the reception status of the neighbors. Thirdly, [3] focuses on an arbitrary number of coexisting flows while this work focuses exclusively on the 2-flow scenario.

In addition to different settings, the outer bound analysis of [3] focuses on the side-information-based index-coding-like analysis, while this work focuses on full feedback and scheduling analysis. Although both the achievability schemes in [3] and in this work are based on linear network coding with sequential scheduling, the latter has to take into account the per-slot causal feedback in the LNC design. The differences between [3] and this work are summarized in Table I.

We also compare the analytical results in this work to the practical COPE implementation in [2]. The COPE protocol in [2] contains three major components: (i) Opportunistic listening: Each destination is in a promiscuous monitoring mode and stores all the overheard packets; (ii) Opportunistic coding: The relay node decides which packets to be coded together opportunistically, based on the overhearing patterns of its neighbors; and (iii) Learning the states of the neighbors: Although in the practical COPE implementation reception reports are periodically sent to advertise the overhearing patterns of the next-hop neighbors of the relay, the relay node still needs to extrapolate the overhearing status of its neighbors since there is always a time lag due to the infrequent periodic feedback.

Our setting closely captures the opportunistic listening component of COPE by modeling the wireless packet transmission as a random broadcast PEC. In (1)–(3), the channel status vector is used to make the coding and scheduling decisions, which captures the opportunistic coding component of COPE. In COPE, the reception reports are broadcast periodically, which is captured by the control information $[\mathbf{Z}]_1^{t-1}$. In sum, our capacity region is a superset of any achievable rates of any COPE-principle-based schemes [2] when focusing on the 2-flow wireless butterfly networks with broadcast PECs in Fig. 1(a) and the node exclusive interference model.

Remark: The setting in Section II-B also includes the wireless erasure 2-way relay channel model (Fig. 3(a) and 3(b)) as a special case. Specifically, if we set the overhearing probabilities: $p_i(d_j) = 1$ for all $i \neq j$, then the capacity region of the setting in Section II-B is also the capacity region of the wireless erasure 2-way relay channel in Fig. 3.

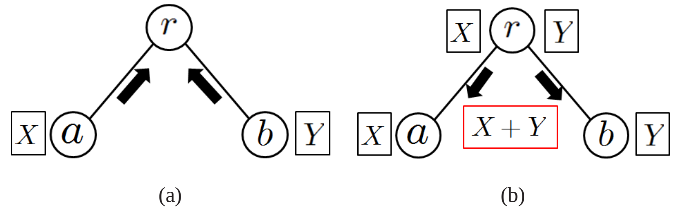


Fig. 3. The illustration of the two–way relay channel for which node a would like to send X to node b and b would like to send Y to a. In (b), the common relay can send a linear combination $[X+Y]$ that benefits both destinations simultaneously.

III. Main Results

In this section, we provide our results based on two cases: The case of considering only the COPE principle and the case of combining COPE with the opportunistic routing technique. The main difference is that for the former setting, we assume that no transmission can be heard by its 2-hop neighbors, i.e., $p_i(d_i) = 0$ for all $i = 1, 2$. For the latter setting, we allow $p_i(d_i)$ to be non-zero.

For the case of using exclusively the COPE principle, the full capacity region has been characterized in Section III-A while for the case of COPE plus opportunistic routing, a pair of outer and inner bounds are provided in Sections III-B and III-C, respectively.

A. The COPE-Principle 2-Flow Wireless Butterfly Network Capacity

Proposition 1: Consider any 2-flow wireless butterfly network with broadcast PECs with $p_i(d_i) = 0$ for all $i = 1, 2$ and consider any finite field $\mathsf{GF}(q)$. The rate pair (R_1, R_2) is in the capacity region if and only if there exist three non-negative time sharing parameters t_{s_1}, t_{s_2} and t_r such that jointly (R_1, R_2) and (t_{s_1}, t_{s_2}, t_r) satisfy

$$t_{s_1} + t_{s_2} + t_r \leq 1 \tag{6}$$
$$\forall i \in \{1, 2\}, R_i \leq t_{s_i} p_i(r) \tag{7}$$
$$\frac{R_1}{p_r(d_1)} + \frac{(R_2 - t_{s_2} p_2(d_1))^+}{p_r(d_1, d_2)} \leq t_r \tag{8}$$
$$\frac{(R_1 - t_{s_1} p_1(d_2))^+}{p_r(d_1, d_2)} + \frac{R_2}{p_r(d_2)} \leq t_r \tag{9}$$

where $(\cdot)^+ \triangleq \max(0, \cdot)$ is the projection to non-negative reals.

The proof of the achievability part of Proposition 1 is relegated to Section V-B and the converse proof is relegated to Appendix A.

The intuition behind (6) to (9) is as follows. (6) is a time sharing bound, which follows from the total time budget being n and the node-exclusive interference model.

Inequality (7) is a simple cut-set bound. That is, the message W_i has to be sent from s_i to the common relay r first. Therefore, the rate is upper bounded by the link capacity from s_i to r.

Inequalities (8) and (9) combine the capacity results on message-side-information [3] and the capacity results on channel output feedback for broadcast channels [6], [7]. A very heuristic, not rigorous explanation of (8) is as follows. $\frac{R_1}{p_r(d_1)}$ represents how many time slots it takes to send all the flow-1 packets to d_1 as if there is no flow-2. $t_{s_2}p_2(d_1)$ characterizes how much flow-2 information can be "overheard" by d_1, and $(R_2 - t_{s_2}p_2(d_1))^+$ thus represents how much flow-2 information that has not been heard by d_1 but still needs to be sent to d_2. Since those flow-2 packets cannot be "coded" together with any flow-1 packets, they need to be sent separately by themselves in addition to the $\frac{R_1}{p_r(d_1)}$ time slots used to send flow-1 packets. In general, it takes $\frac{(R_2 - t_{s_2}p_2(d1))^+}{p_r(d_2)}$ for those packets to arrive at d_2. However, [6] shows that the use of feedback can further reduce the time to $\frac{(R_2 - t_{s_2}p_2(d_1))^+}{p_r(d_1,d_2)}$. We then have (8) since the total transmission time of relay r is nt_r time slots. (9) is symmetric to (8).

B. Capacity Outer Bound for COPE plus OpR

The capacity results in Proposition 1 can be generalized as an outer bound for the case when the destination d_i may overhear directly the transmission of s_i, i.e., $p_i(d_i) > 0$.

Proposition 2: Consider any 2-flow wireless butterfly network with broadcast PECs in Fig. 2 with arbitrary channel characteristics and consider any finite field $\mathsf{GF}(q)$. If a rate vector (R_1, R_2) is achievable, there exist three non-negative scalars t_{s_1}, t_{s_2}, and t_r satisfying

$$t_{s_1} + t_{s_2} + t_r \leq 1 \quad (10)$$

$$\forall i \in \{1,2\}, \quad R_i \leq t_{s_i}p_i(d_i, r) \quad (11)$$

$$\frac{(R_1 - t_{s_1}p_1(d_1))^+}{p_r(d_1)} + \frac{(R_2 - t_{s_2}p_2(d_1,d_2))^+}{p_r(d_1,d_2)} \leq t_r \quad (12)$$

$$\frac{(R_1 - t_{s_1}p_1(d_1,d_2))^+}{p_r(d_1,d_2)} + \frac{(R_2 - t_{s_2}p_2(d_2))^+}{p_r(d_2)} \leq t_r. \quad (13)$$

This outer bound considers channel status feedback, dynamic scheduling, and possibly non-linear encoding functions, and is derived by entropy-based analysis. One major challenge of the outer bound derivation is to incorporate the effects of dynamic scheduling, which was not presented in the existing sequential-scheme-based outer bound analysis [3]. To circumvent this issue, we further analyze the time average of the mutual information over all possible realizations. The detailed proof is relegated to Appendix A.

Remark: One can easily see that when the channel probabilities satisfy $p_i(d_i) = 0$ for all $i = 1, 2$, the outer bound in Proposition 2 collapses to the capacity region in Proposition 1. Proposition 2 is thus a strict generalization of the converse part of Proposition 1.

C. Capacity Inner Bound for COPE plus OpR

An inner bound for the general case of $p_i(d_i) \geq 0$ is described as follows.

Proposition 3: A rate vector (R_1, R_2) is achievable by a linear network code if there exist 3 non-negative variables t_{s_1}, t_{s_2}, t_r, 10 non-negative variables, $\omega_{s_i}^k$, where $i \in \{1,2\}$ and $k \in \{0,1,2,3,4\}$, 4 non-negative variables $\omega_{r,\mathrm{N}}^k$, $\omega_{r,\mathrm{C}}^k$ for $k = 1, 2$, such that jointly the 17 variables[6] and (R_1, R_2) satisfy the following four groups of inequalities:

Group 1 has 5 inequalities, named as *the time budget constraints*.

$$\forall i = 1,2, \quad \sum_{k=0}^{4} \omega_{s_i}^k \leq t_{s_i} \quad (14)$$

$$\forall i = 1,2, \quad \omega_{r,\mathrm{N}}^1 + \omega_{r,\mathrm{N}}^2 + \omega_{r,\mathrm{C}}^i \leq t_r \quad (15)$$

$$t_{s_1} + t_{s_2} + t_r < 1 \quad (16)$$

Group 2 has 12 inequalities, named as *the packet conservation laws at the source nodes*. Consider any $i, j \in \{1,2\}$ satisfying $i \neq j$. For each (i,j) pair (out of the two choices $(1,2)$ and $(2,1)$), we have the following 6 inequalities.

$$\omega_{s_i}^0 p_i(d_i, d_j, r) \leq R_i \quad (17)$$

$$\omega_{s_i}^1 p_i(d_i, r) \leq \omega_{s_i}^0 p_i(d_j \overline{d_i r}) \quad (18)$$

$$\omega_{s_i}^1 p_i(d_i, d_j) \leq \omega_{s_i}^0 p_i(r \overline{d_i d_j}) \quad (19)$$

$$\omega_{s_i}^2 p_i(d_i, r) \leq \omega_{s_i}^0 p_i(d_j \overline{d_i r}) - \omega_{s_i}^1 p_i(d_i, r) \quad (20)$$

$$\omega_{s_i}^3 p_i(d_i, d_j) \leq \omega_{s_i}^0 p_i(r \overline{d_i d_j}) - \omega_{s_i}^1 p_i(d_j, d_i r) \quad (21)$$

$$\begin{aligned}\omega_{s_i}^4 p_i(d_i) \leq\ & \omega_{s_i}^0 p_i(d_j r \overline{d_i}) \\ & + \omega_{s_i}^1 (p_i(d_j) + p_i(r) - p_i(d_i d_j r)) \\ & + \omega_{s_i}^2 p_i(r \overline{d_i}) + \omega_{s_i}^3 p_i(d_j \overline{d_i})\end{aligned} \quad (22)$$

Group 3 has 4 inequalities, named as *the packet conservation laws at the relay node*. For each (i,j) pair with $i \neq j$, we have the following 2 inequalities.

$$\begin{aligned}\omega_{r,\mathrm{N}}^i p_r(d_i, d_j) \leq\ & \omega_{s_i}^0 p_i(r \overline{d_i d_j}) - \omega_{s_i}^1 p_i(d_j, d_i r) \\ & - \omega_{s_i}^3 p_i(d_i, d_j)\end{aligned} \quad (23)$$

$$\begin{aligned}\omega_{r,\mathrm{C}}^i p_r(d_i) \leq\ & \omega_{s_i}^0 p_i(d_j r \overline{d_i}) \\ & + \omega_{s_i}^1 (p_i(d_j) + p_i(r) - p_i(d_i d_j r)) \\ & + \omega_{s_i}^2 p_i(r \overline{d_i}) + \omega_{s_i}^3 p_i(d_j \overline{d_i}) \\ & - \omega_{s_i}^4 p_i(d_i) + \omega_{r,\mathrm{N}}^i p_r(d_j \overline{d_i})\end{aligned} \quad (24)$$

Group 4 has 2 inequalities, named as *the decodability conditions*. Consider $i = 1, 2$. For each i, we have the following inequality.

$$\left(\sum_{k=0}^{4} \omega_{s_i}^k\right) p_i(d_i) + \left(\omega_{r,\mathrm{N}}^i + \omega_{r,\mathrm{C}}^i\right) p_r(d_i) \geq R_i \quad (25)$$

The inner bound inequalities are based on the SBLNC scheme constructed in Section IV. A heuristic but not rigorous explanation is as follows. The time budget constraints (14)–(16) describe the fact that each time slot can be assigned to one of the transmitting nodes s_1, s_2, or r and the overall normalized time budget is one. The conservation laws (17)–(24) correspond to the fact that to select any one of the SBLNC policies, the corresponding *coding set* of the policy must be non-empty. The non-emptiness of the coding set can

[6] In the achieving algorithm in Section V, the t variables correspond to the numbers of time slots that each of the sources and the relay is used; and the ω variables correspond to the numbers of time slots each policy is used.

be described by (17)–(24), which determine the size of the coding sets. The decodability conditions describe how many packets need to be received at each destination before they can decode the desired messages.

Proposition 3 will be proved by explicit construction of an achievability scheme based on the SBLNC scheme described in the next section. The detailed proof of Proposition 3 is relegated to Section V-A.

IV. A Space-Based Linear Network Code(SBLNC) Construction

In the existing network coding scheme works (e.g. [1], [3]), designing the encoding coefficients is always a challenging task. With per-time-slot causal feedback, the sources (including the relay) can continuously track what has been received by the destinations and the best network coding strategy might evolve overtime, which further exacerbates the problem of designing the right coding coefficients. In this section, we introduce a new class of network coding scheme named as the "Space-Based Linear Network Code (SBLNC)" scheme that significantly simplifies the design of the coding coefficients. The SBLNC scheme will later be used to prove the capacity inner bound in Proposition 3.

A. Linear-Space-Based Definitions

We first provide some basic definitions that will be used when describing an SBLNC scheme.

For $i = 1, 2$, a *flow-i coding vector* $\mathbf{v}^{(i)}$ is an nR_i-dimensional row vector with each coordinate being a scalar in $\mathsf{GF}(q)$. Any linear combination of the message symbols $W_{i,1}$ to W_{i,nR_i} can thus be represented by $\mathbf{v}^{(i)}\mathbf{W}_i^{\mathrm{T}}$ where $\mathbf{W}_i^{\mathrm{T}}$ is the transpose of $\mathbf{W}_i$. We use the superscript "(i)" to emphasize that we are focusing on a flow-i vector.

We define the *flow-i message space* by $\Omega_i \triangleq (\mathsf{GF}(q))^{nR_i}$, an nR_i-dimensional linear space. In the following, we define the following 6 *knowledge spaces* S_r, S_{d_1}, S_{d_2}, T_r, T_{d_1}, and T_{d_2} for the 5-node relay network in Fig. 2.

The knowledge spaces S_r, S_{d_2}, S_{d_1} are linear subspaces of Ω_1 and represent the knowledge about the flow-1 packets at nodes r, d_2, and d_1, respectively. Symmetrically, the knowledge spaces T_r, T_{d_1}, and T_{d_2} are linear subspaces of Ω_2 and represent the knowledge about the flow-2 packets at nodes r, d_1, and d_2, respectively. In the following, we discuss the detailed construction of S_r, S_{d_2}, and S_{d_1} and the construction of T_r to T_{d_2} follows symmetrically[7].

- In the end of any time t, $S_r(t) \subset \Omega_1$ is the *linear span* of a group of $\mathbf{v}^{(1)}$ vectors, denoted by $\mathbf{V}^{(1)}_{s_1\to r}$. The group $\mathbf{V}^{(1)}_{s_1\to r}$ contains the $\mathbf{v}^{(1)}$ vectors sent by s_1 during time 1 to t and have been received successfully by r. Throughout the paper, we use the convention that the linear span of an empty set is a set containing the zero vector, i.e., $\mathsf{span}\{\emptyset\} = \{0\}$. For example, if r has not yet received any packet from s_1, then by convention $S_r(t) = \{0\}$.
- In the end of time t, $S_{d_2}(t) \subset \Omega_1$ is the linear span of two groups of $\mathbf{v}^{(1)}$ vectors, denoted by $\mathbf{V}^{(1)}_{s_1\to d_2}$ and $\mathbf{V}^{(1)}_{\mathrm{N},r\to d_2}$. The first group $\mathbf{V}^{(1)}_{s_1\to d_2}$ contains the $\mathbf{v}^{(1)}$ vectors corresponding to the packets sent by s_1 during time 1 to t and have been received successfully by d_2. The second group $\mathbf{V}^{(1)}_{\mathrm{N},r\to d_2}$ contains the $\mathbf{v}^{(1)}$ vectors corresponding to the packets sent by r during time 1 to t that *are not mixed with any other flow-2 packets*. The letter "N" in the subscript stands for Not-inter-flow-coded transmission.
- In the end of time t, $S_{d_1}(t) \subset \Omega_1$ is the linear span of three groups of $\mathbf{v}^{(1)}$ vectors, denoted by $\mathbf{V}^{(1)}_{s_1\to d_1}$, $\mathbf{V}^{(1)}_{\mathrm{N},r\to d_1}$, and $\mathbf{V}^{(1)}_{\mathrm{C},r\to d_1}$. The first group $\mathbf{V}^{(1)}_{s_1\to d_1}$ contains the $\mathbf{v}^{(1)}$ vectors corresponding to the packets sent by s_1 during time 1 to t and have been received successfully by d_1. The second group $\mathbf{V}^{(1)}_{\mathrm{N},r\to d_1}$ contains the $\mathbf{v}^{(1)}$ vectors corresponding to the packets sent by r during time 1 to t that *are not mixed with any other flow-2 packets*. The third group $\mathbf{V}^{(1)}_{\mathrm{C},r\to d_1}$ contains the $\mathbf{v}^{(1)}$ vectors *that can be decoded from the inter-flow coded packets*[8] *sent by r during time 1 to t*. The letter "C" in the subscript stands for inter-flow-Coded transmission.

In sum, we use S and T to distinguish whether we are focusing on flow-1 or flow-2 packets, respectively, and we use the subscripts to describe the node of interest. One can easily see that these six knowledge spaces evolve over time since each node may receive more and more packets that can be used to obtain/decode new information. We use the following example to illustrate the definitions of S_r to T_{d_2}.

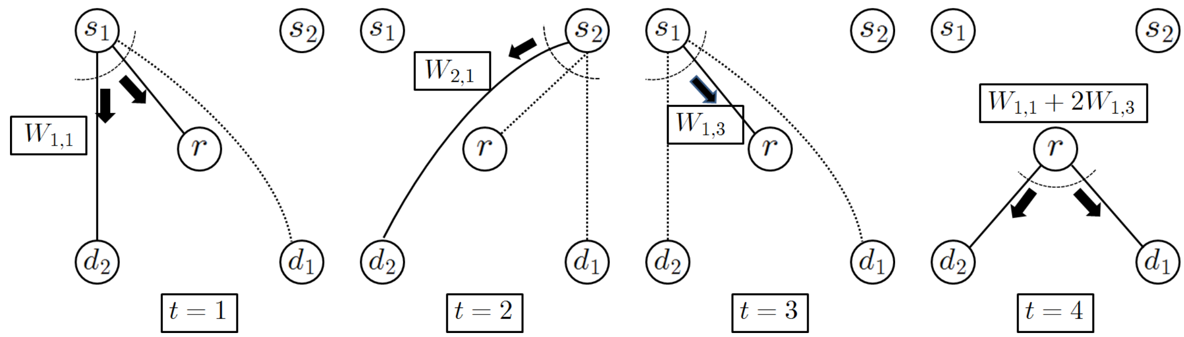
Example 1: Consider $\mathsf{GF}(3)$ and $nR_1 = 3$ and $nR_2 = 2$. That is, flow-1 contains 3 message symbols $W_{1,1}$ to $W_{1,3}$ and flow-2 contains 2 message symbols $W_{2,1}$ and $W_{2,2}$. Ω_1 and Ω_2 are thus 3-dimensional and 2-dimensional linear spaces in $\mathsf{GF}(3)$, respectively. Consider the first four time slots $t = 1$ to 4 for our discussion.

When $t = 1$, suppose that s_1 is scheduled; an uncoded flow-1 message symbol $W_{1,1}$ is transmitted; and the packet is heard by and only by d_2 and r. See Fig. 4(a) for illustration, for which we use the solid lines to represent that d_2 and r have received the packet. We use the dashed line to denote that d_1 does not receive the packet. When $t = 2$, suppose that s_2 is scheduled; an uncoded flow-2 message symbol $W_{2,1}$ is transmitted; and the packet is heard by and only by d_2, see Fig. 4(b). When $t = 3$, suppose that s_1 is scheduled; an uncoded flow-1 symbol $W_{1,3}$ is transmitted; and the packet is heard by and only by r. When $t = 4$, suppose that r is scheduled; r sends a linear combination $[W_{1,1} + 2W_{1,3}]$ of the two flow-1 packets it has received thus far; and the packet $[W_{1,1} + 2W_{1,3}]$ is heard by both d_1 and d_2.

We now describe the six knowledge spaces S_r to T_{d_2} in the end of $t = 4$. By Figs. 4(a) and 4(d), d_2 has received two flow-1 packets $W_{1,1}$ and $[W_{1,1} + 2W_{1,3}]$, one from s_1 and one from r. Therefore, by the end of $t = 4$, the flow-1 knowledge space at d_2 becomes $S_{d_2}(4) = \mathsf{span}\{(1, 0, 0), (1, 0, 2)\}$. Also, neither r nor d_1 has received any flow-2 packets by the end

[7]The construction of T_{d_1}(resp. T_{d_2}) follows the construction of S_{d_2}(resp. S_{d_1}).

[8]When the relay r sends a linear combination of both flow-1 and -2 packets.



(a) (b) (c) (d)

Fig. 4. The illustration of the coding procedure in Example 1. We use a solid line to represent that the corresponding receiver has successfully received the packet and use a dotted line to represent erasure.

TABLE II
THE RESULTING KNOWLEDGE SPACES AT THE END OF EXAMPLE 1

Flow-1		Flow-2	
$S_{d_1}(4)$	$\mathsf{span}\{(1,0,2)\}$	$T_{d_1}(4)$	$\{(0,0)\}$
$S_{d_2}(4)$	$\mathsf{span}\{(1,0,0),(1,0,2)\}$	$T_{d_2}(4)$	$\mathsf{span}\{(1,0)\}$
$S_r(4)$	$\mathsf{span}\{(1,0,0),(0,0,1)\}$	$T_r(4)$	$\{(0,0)\}$

of $t=4$. Therefore, T_r and T_{d_1}, the flow-2 knowledge spaces at r and d_1, respectively, contain only the zero element. The other knowledge spaces S_{d_1}, S_r, and T_{d_2} in the end of $t=4$ can be derived similarly and they are summarized in Table II.

The above definitions also lead to the following self-explanatory lemma.

Lemma 1: The two destinations d_1 and d_2 can decode the desired message symbols $\mathbf{W}_1$ and $\mathbf{W}_2$, respectively, if and only if by the end of time n

$$S_{d_1}(n)=\Omega_1 \text{ and } T_{d_2}(n)=\Omega_2.$$

For simplicity, we use $S_i(t)$ and $T_i(t)$ to denote the knowledge space $S_{d_i}(t)$ and $T_{d_i}(t)$ for $i=1,2$. We also omit the input argument "(t)" if the time index is clear from the context. To conclude this subsection, we introduce the notation of the sum space: $(A\oplus B)\triangleq \mathsf{span}\{\mathbf{v}:\forall \mathbf{v}\in A\cup B\}$. Notice that $A\oplus B$ and $A\cup B$ are different. For example, suppose we consider a 2-dimensional linear space with $\mathsf{GF}(3)$ with two linear subspaces $A=\mathsf{span}\{(1,0)\}$ and $B=\mathsf{span}\{(1,1)\}$. Then $A\cup B=\{(0,0),(1,0),(2,0),(1,1),(2,2)\}$ is not a linear subspace anymore, but $A\oplus B=\mathsf{span}\{(1,0),(1,1)\}=\{(0,0),(1,0),(2,0),(1,1),(2,2),(2,1),(1,2),(0,1),(0,2)\}$ is a linear subspace. By simple algebra, we have the following lemma.

Lemma 2: For any two linear subspaces A and B in Ω, the following equality always holds.

$$\mathsf{dim}(A\oplus B)=\mathsf{dim}(A)+\mathsf{dim}(B)-\mathsf{dim}(A\cap B).$$

B. An Instance of The SBLNC Schemes

In the following, we will introduce a new class of network codes, named as the Space-Based Linear Network Code (SBLNC). An SBLNC scheme contains a finite number of *policies*. Each policy Γ contains a linear subspace $A^{(\Gamma)}$, named as *the inclusion space/set*, and a finite collection of linear subspaces $B_l^{(\Gamma)}$ for $l=1$ to $L^{(\Gamma)}$, named as *the exclusion spaces/sets*. For each time slot t, the SBLNC chooses one of the specified policies and uses it to generate the coded packet. For example, say node s is scheduled for transmission and we decide to choose a policy Γ for encoding. Then s will first choose arbitrarily a coding vector $\mathbf{v}^{(i)}$ from the set $A^{(\Gamma)}\backslash\left(\bigcup_{l=1}^{L^{(\Gamma)}}B_l^{(\Gamma)}\right)$, and then transmit a linearly encoded packet $X=\mathbf{v}^{(i)}\mathbf{W}_i^{\mathrm{T}}$. That is, the coding vector must be in the inclusion set $A^{(\Gamma)}$ but not in any of the exclusion sets $B_l^{(\Gamma)}$. Obviously, a policy can be used/chosen only when the corresponding set $A^{(\Gamma)}\backslash\left(\bigcup_{l=1}^{L^{(\Gamma)}}B_l^{(\Gamma)}\right)$ is non-empty. For notational simplicity, we say a policy is *feasible* if the corresponding $A^{(\Gamma)}\backslash\left(\bigcup_{l=1}^{L^{(\Gamma)}}B_l^{(\Gamma)}\right)$ is non-empty.

For illustration, consider the following policy for node s_1, named as Policy $\Gamma_{s_1,0}$. When Policy $\Gamma_{s_1,0}$ is used/chosen, we let source node s_1 choose arbitrarily a coding vector $\mathbf{v}^{(1)}$ from $\Omega_1\backslash(S_1\oplus S_2\oplus S_r)$ and send the corresponding coded packet $X_{s_1}=\mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}}$. That is, the inclusion set is $A^{(\Gamma_{s_1,0})}=\Omega_1$ and the exclusion set is $B_1^{(\Gamma_{s_1,0})}=S_1\oplus S_2\oplus S_r$.

Continue the example in Section IV-A for which the knowledge spaces are summarized in Table II. In the beginning of $t=5$ (or equivalently in the end of $t=4$), we have $A^{(\Gamma_{s_1,0})}=\Omega_1$ and $B^{(\Gamma_{s_1,0})}=S_1\oplus S_2\oplus S_r=\{(a,0,c):\forall a,c\in\mathsf{GF}(q)\}$. As a result, if we choose Policy $\Gamma_{s_1,0}$ for $t=5$, any coding vectors of the form (a,b,c) with $b\neq 0$ are in the set $\Omega_1\backslash(S_1\oplus S_2\oplus S_r)$. There are totally 18 such vectors since $\mathsf{GF}(3)$ is used. Source s_1 can then choose arbitrarily from any one of the 18 vectors and send $X=aW_{1,1}+bW_{1,2}+cW_{1,3}$ in time $t=5$.

In the following, we define 13 policies that will be used in the proof of the achievability part of Propositions 1 and 3.

There are 5 policies governing the coding operations at source s_1, which are named as Policy $\Gamma_{s_1,j}$ for $j=0$ to 4. When Policy $\Gamma_{s_1,j}$ is used, s_1 sends $X_{s_1}=\mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}}$ for some $\mathbf{v}^{(1)}$. That is, source s_1 only mixes/encodes flow-1 packets together. In the following, we describe how to choose the vector $\mathbf{v}^{(1)}$ for each individual policy.

§ *Policy* $\Gamma_{s_1,0}$*:* Choose $\mathbf{v}^{(1)}$ arbitrarily from

$$\Omega_1\backslash(S_1\oplus S_2\oplus S_r). \tag{26}$$

§ *Policy* $\Gamma_{s_1,1}$*:* Choose $\mathbf{v}^{(1)}$ arbitrarily from

$$(S_2 \oplus S_r)\backslash \left((S_1 \oplus S_r) \cup (S_1 \oplus S_2)\right). \tag{27}$$

§ *Policy* $\Gamma_{s_1,2}$*:* Choose $\mathbf{v}^{(1)}$ arbitrarily from

$$S_2\backslash(S_1 \oplus S_r). \tag{28}$$

§ *Policy* $\Gamma_{s_1,3}$*:* Choose $\mathbf{v}^{(1)}$ arbitrarily from

$$S_r\backslash \left(S_1 \oplus (S_2 \cap S_r)\right). \tag{29}$$

§ *Policy* $\Gamma_{s_1,4}$*:* Choose $\mathbf{v}^{(1)}$ arbitrarily from

$$(S_2 \cap S_r)\backslash S_1. \tag{30}$$

Policy $\Gamma_{s_2,j}$, $j = 0$ to 4 are symmetric versions of *Policy* $\Gamma_{s_1,j}$ that concern source s_2 and mix/encode flow-2 packets instead. More explicitly, source s_2 sends $X_{s_2} = \mathbf{v}^{(2)}\mathbf{W}_2^{\mathrm{T}}$ for which the coding vector $\mathbf{v}^{(2)}$ is chosen according to the following specification.

§ *Policy* $\Gamma_{s_2,0}$*:* Choose $\mathbf{v}^{(2)}$ arbitrarily from

$$\Omega_2\backslash(T_1 \oplus T_2 \oplus T_r). \tag{31}$$

§ *Policy* $\Gamma_{s_2,1}$*:* Choose $\mathbf{v}^{(2)}$ arbitrarily from

$$(T_1 \oplus T_r)\backslash \left((T_2 \oplus T_r) \cup (T_1 \oplus T_2)\right). \tag{32}$$

§ *Policy* $\Gamma_{s_2,2}$*:* Choose $\mathbf{v}^{(2)}$ arbitrarily from

$$T_1\backslash(T_2 \oplus T_r). \tag{33}$$

§ *Policy* $\Gamma_{s_2,3}$*:* Choose $\mathbf{v}^{(2)}$ arbitrarily from

$$T_r\backslash \left(T_2 \oplus (T_1 \cap T_r)\right). \tag{34}$$

§ *Policy* $\Gamma_{s_2,4}$*:* Choose $\mathbf{v}^{(2)}$ arbitrarily from

$$(T_1 \cap T_r)\backslash T_2. \tag{35}$$

There are 3 policies $\Gamma_{r,j}$, $j = 1, 2, 3$, governing the coding operations at the relay r, which are described as follows.

§ *Policy* $\Gamma_{r,1}$*:* The relay r chooses arbitrarily a vector $\mathbf{v}^{(1)}$ from

$$S_r\backslash \left((S_r \cap S_2) \oplus S_1\right) \tag{36}$$

and sends an intra-flow-coded flow-1 packet $X_r = \mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}}$.

§ *Policy* $\Gamma_{r,2}$*:* The relay r chooses arbitrarily a vector $\mathbf{v}^{(2)}$ from

$$T_r\backslash \left((T_r \cap T_1) \oplus T_2\right) \tag{37}$$

and sends an intra-flow-coded flow-2 packet $X_r = \mathbf{v}^{(2)}\mathbf{W}_2^{\mathrm{T}}$.

§ *Policy* $\Gamma_{r,3}$ is for the relay node r to send an interflow-coded packet $X_r = \mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}} + \mathbf{v}^{(2)}\mathbf{W}_2^{\mathrm{T}}$, with $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ chosen as follows: If $(S_2 \cap S_r)\backslash S_1$ is non-empty, choose $\mathbf{v}^{(1)}$ arbitrarily from

$$(S_2 \cap S_r)\backslash S_1, \tag{38}$$

otherwise choose $\mathbf{v}^{(1)} = \mathbf{0}$, a zero vector. If $(T_1 \cap T_r)\backslash T_2$ is non-empty, choose $\mathbf{v}^{(2)}$ arbitrarily from

$$(T_1 \cap T_r)\backslash T_2, \tag{39}$$

otherwise choose $\mathbf{v}^{(2)} = \mathbf{0}$.

Continue from Example 1 in Section IV-A with the knowledge spaces in the end of $t = 4$ described in Table II. Consider Policy $\Gamma_{s_1,3}$ as defined in (29). Since $S_2 \cap S_r = S_r$ in the end of $t = 4$, we have $S_r\backslash \left(S_1 \oplus (S_2 \cap S_r)\right) \subseteq S_r\backslash(S_2 \cap S_r) = \emptyset$ being an empty set. Thus, in contrast with the fact that Policy $\Gamma_{s_1,0}$ is feasible in the beginning of $t = 5$ as shown in our previous discussion, Policy $\Gamma_{s_1,3}$ is infeasible in the beginning of $t = 5$.

One can repeat the above analysis and verify that out of all 13 policies, only 4 of them are feasible in the beginning of $t = 5$, which are $\Gamma_{s_1,0}$, $\Gamma_{s_1,4}$, $\Gamma_{s_2,0}$, and $\Gamma_{r,3}$. The network code designer can thus apply one of the four policies in $t = 5$.

Suppose the network designer chooses policy $\Gamma_{s_1,0}$ for $t = 5$ and sends a flow-1 coded packet with the coding vector being $\mathbf{v}^{(1)} = (2, 1, 0)$. Also suppose that the packet is received by r but by neither d_1 nor d_2. Then in the end of time $t = 5$, the knowledge space S_r evolves from the original $\mathsf{span}\{(1, 0, 0), (0, 0, 1)\}$ to the new space $\mathsf{span}\{(1, 0, 0), (0, 0, 1), (2, 1, 0)\}$. We now notice that the Policy $\Gamma_{s_1,0}$ is no longer feasible since with the new S_r, the exclusion space of $\Gamma_{s_1,0}$ becomes $S_1 \oplus S_2 \oplus S_r = \mathsf{span}\{(1, 0, 0), (0, 0, 1), (2, 1, 0)\}$ and $\Omega_1\backslash \left(S_1 \oplus S_2 \oplus S_r\right)$ is now empty. On the other hand, the new S_r also lets some previously infeasible policies become feasible. For example, consider Policy $\Gamma_{s_1,3}$. With the new S_r, we have $S_r = \mathsf{span}\{(1, 0, 0), (0, 0, 1), (2, 1, 0)\}$ and $S_1 \oplus (S_2 \cap S_r) = \mathsf{span}\{(1, 0, 0), (1, 0, 2)\}$. Therefore, $S_r\backslash \left(S_1 \oplus (S_2 \cap S_r)\right) \neq \emptyset$. Policy $\Gamma_{s_1,3}$ is thus feasible and can be used for transmission in $t = 6$. With similar analysis, one can verify that in the beginning of $t = 6$, we have 5 feasible policies: $\Gamma_{s_1,3}$, $\Gamma_{s_1,4}$, $\Gamma_{s_2,0}$, $\Gamma_{r,1}$, and $\Gamma_{r,3}$. This example shows that due to the evolution of the knowledge spaces over time, each coding policy may become feasible or infeasible depending on the reception status until the present. Focusing on coding policies frees us from designing the value of each coordinate of the coding vector $\mathbf{v}^{(1)}$ (resp. $\mathbf{v}^{(2)}$). Instead, we only need to choose $\mathbf{v}^{(1)}$ (resp. $\mathbf{v}^{(2)}$) from one of the policies that are currently feasible, which significantly simplifies the corresponding analysis.

C. The Intuition Behind the Proposed SBLNC Policies

In Section V, we will prove that the proposed SBLNC scheme can achieve the capacity in Proposition 1 and the inner bound in Proposition 3 when we carefully decide which of the 13 policies to apply for each time instant. We conclude Section IV by discussing the intuition behind the proposed 13 policies. We first consider the relay policies $\Gamma_{r,1}$ to $\Gamma_{r,3}$ due to its conceptual simplicity. We then discuss the source policies $\Gamma_{s_i,0}$ to $\Gamma_{s_i,4}$.

1) The Relay Policies: We first notice that for all relay policies $\Gamma_{r,1}$, $\Gamma_{r,2}$, and $\Gamma_{r,3}$, the corresponding inclusion space is either a subspace of S_r or a subspace of T_r. The reason is that for node r to send a coded packet, the encoded packet must already be in S_r or T_r, the knowledge spaces of r. As a result, the transmitted vector $\mathbf{v}^{(1)}$ (or $\mathbf{v}^{(2)}$) must be drawn from a subset of S_r (or T_r).

It is clear that *a good network code should try to serve two flows simultaneously in order to maximize the throughput.* We now focus on Policy $\Gamma_{r,3}$. First notice that by (39), $\mathbf{v}^{(2)}$ is drawn from $(T_1 \cap T_r)$. This means that the value of $\mathbf{v}^{(2)}\mathbf{W}_2^{\mathrm{T}}$ is already known by destination d_1 since $\mathbf{v}^{(2)}$ is in the flow-2 knowledge space T_1 at d_1. Hence whenever d_1 receives the packet $X_r(t) = \mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}} + \mathbf{v}^{(2)}\mathbf{W}_2^{\mathrm{T}}$, it can extract its desired information and recover $\mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}}$ by substracting $\mathbf{v}^{(2)}\mathbf{W}_2^{\mathrm{T}}$. We then note that Policy $\Gamma_{r,3}$ ensures that whenever (38) is not empty the selected $\mathbf{v}^{(1)}$ is not in S_1, the flow-1 knowledge space at d_1. Hence upon the reception of such a coded packet, $\mathsf{dim}(S_1)$ will increase by one. By Lemma 1 destination d_1 is one step closer to fully decode its desired message $\mathbf{W}_1$. Symmetrically, by (38) d_2 has already known the value of $\mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}}$ and thus d_2 can compute the value of $\mathbf{v}^{(2)}\mathbf{W}_2^{\mathrm{T}}$ upon the reception of the inter-flow coded packet generated by Policy $\Gamma_{r,3}$. Since $\mathbf{v}^{(2)}$ is not in T_2, d_2 can decode one extra linear combination of flow-2 packets. Policy $\Gamma_{r,3}$ thus serves both d_1 and d_2 simultaneously.

Although Policy $\Gamma_{r,3}$ can serve both destinations simultaneously, there is a limit on how much information can be sent by $\Gamma_{r,3}$. That is, if we use only Policy $\Gamma_{r,3}$ and nothing else, the information that can be received by d_1 through Policy $\Gamma_{r,3}$ is at most $(S_r \cap S_2)$ since all $\mathbf{v}^{(1)}$ are drawn from $(S_r \cap S_2)$. The largest flow-1 knowledge space that d_1 can possibly attain is thus $S_1 \oplus (S_r \cap S_2)$, where S_1 represents the flow-1 information that d_1 has accumulated by overhearing the transmission directly from its two-hop neighbor s_1, and $(S_r \cap S_2)$ represents the information that can be conveyed by $\Gamma_{r,3}$. Note that it is possible that S_r is not a subspace of $S_1 \oplus (S_r \cap S_2)$, which means that relay r still possesses some flow-1 information that cannot be conveyed to d_1 by $\Gamma_{r,3}$ alone. $\Gamma_{r,1}$ is devised to address this problem. That is, the $\mathbf{v}^{(1)}$ vector chosen from (36) is (i) from the knowledge space of r, and (ii) not in $S_1 \oplus (S_r \cap S_2)$, the largest flow-1 knowledge space that d_1 can attain when using exclusively Policy $\Gamma_{r,3}$. Such $\mathbf{v}^{(1)}$ vector thus represents an information packet that is complementary to the inter-flow-coded Policy $\Gamma_{r,3}$.

2) The Source Policies: Here without loss of generality we focus on source-1 policies.

Before explaining the source policies (26)–(30), we first discuss several network coding goals for the transmission of s_1. The highest priority is to enlarge S_1 and S_2 such that $(S_1 \oplus S_2) = \Omega_1$ at the end of the source-1 transmission since $s_1 \to r \to d_1$ and $s_1 \to d_1$ are the only two routes from s_1 to d_1. By the cut-set bound, we must achieve $(S_1 \oplus S_r) = \Omega_1$ at the end of the source-1 transmission otherwise d_1 cannot decode the complete flow-1 messages. Another priority is to maximize $\mathsf{dim}(S_1 \oplus (S_2 \cap S_r))$. As discussed in the previous paragraph, the largest amount of information that can be transmitted to d_1 through inter-session coded messages is $\mathsf{dim}(S_1 \oplus (S_2 \cap S_r))$. Therefore, it is important to maximize the inter-session coding benefits by maximizing $\mathsf{dim}(S_1 \oplus (S_2 \cap S_r))$ during the transmission of s_1 so that the relay r can harvest the largest amount of inter-session coding benefits during Policy $\Gamma_{r,3}$. Another priority is to ensure that destination d_1 overhears directly from s_1 as much information as possible since those information heard directly by d_1 does not need to be sent by relay r anymore. This is equivalent to maximize $\mathsf{dim}(S_1)$.

With the above three goals in mind, we now discuss the intuition of each source policy. Policy $\Gamma_{s_1,j}$, $j = 0, 1, 2$, are designed to maximize $\mathsf{dim}(S_1 \oplus S_r)$ since $S_1 \oplus S_r$ is a subset of the exclusion sets for these policies. Therefore, every time one of d_1 and r receives a packet encoded by Policy $\Gamma_{s_1,j}$, $j = 0, 1, 2$, the term $\mathsf{dim}(S_1 \oplus S_r)$ will increase by one. Policy $\Gamma_{s_1,j}$, $j = 0$ to 3, are designed to maximize $\mathsf{dim}(S_1 \oplus (S_2 \cap S_r))$ since $S_1 \oplus (S_2 \cap S_r)$ is a subset of the exclusion sets for all these policies. Policies $\Gamma_{s_1,j}$, $j = 0$ to 4, are designed to maximize $\mathsf{dim}(S_1)$ since S_1 is a subset of the exclusion sets for all these policies. As one can see that the five s_1-policies aim at simultaneously achieving the three goals.

To explain the heuristics why we choose the specific inclusion and exclusion sets for these policies, we use Policy $\Gamma_{s_1,1}$ as an example. To that end, we notice that by Lemma 2, we have

$$\begin{aligned}
&\mathsf{dim}(S_1 \oplus (S_2 \cap S_r)) \\
=&\mathsf{dim}(S_1) + \mathsf{dim}(S_2 \cap S_r) - \mathsf{dim}(S_1 \cap S_2 \cap S_r) \\
=&\mathsf{dim}(S_1) + \mathsf{dim}(S_2) + \mathsf{dim}(S_r) - \mathsf{dim}(S_2 \oplus S_r) \\
&- \mathsf{dim}(S_1 \cap S_2 \cap S_r). \qquad (40)
\end{aligned}$$

Since $S_2 \oplus S_r$ is the inclusion set for Policy $\Gamma_{s_1,1}$; and S_1, S_2, and S_r are subsets of the exclusion sets for Policy $\Gamma_{s_1,1}$, every time any one of the $\{d_1, d_2, r\}$ nodes receives a policy-$\Gamma_{s_1,1}$ packet, at least one of the three terms $\mathsf{dim}(S_1)$, $\mathsf{dim}(S_2)$, and $\mathsf{dim}(S_r)$ will increase and the term $\mathsf{dim}(S_2 \oplus S_r)$ remains unchanged. Assuming that $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ does not change too much[9], Policy $\Gamma_{s_1,1}$ increases $\mathsf{dim}(S_1 \oplus (S_2 \cap S_r))$ in a very efficient way since all three positive terms in (40) can have a good chance of increase while one negative term in (40) remains constant and the other negative term in (40) only increases slightly.

To summarize, the proposed source-1 policies aim at simultaneously maximizing $\mathsf{dim}(S_1 \oplus S_r)$, $\mathsf{dim}(S_1 \oplus (S_2 \cap S_r))$, and $\mathsf{dim}(S_1)$. The specific design of the inclusion and exclusions sets is to maximize the above three different terms in an efficient way. Detailed description about how the three terms increase is relegated to the throughput analysis in Section V-A.

V. Capacity Approaching Coding Scheme

In this section, we will first prove the capacity inner bound *Proposition 3* for the 2-flow wireless butterfly network setting considering both the COPE principle and opportunistic routing. We will then prove that the inner bound coincides with the capacity characterization in Proposition 1 when considering only the COPE principle.

A. Achieving The Inner Bound of Proposition 3

We prove Proposition 3 by properly scheduling the 13 policies described in Section IV-B.

Consider any t_{s_1}, t_{s_2}, t_r, $\omega_{s_i}^k$, $i \in \{1, 2\}$ and $k \in \{0, 1, 2, 3, 4\}$, $\omega_{r,\mathrm{N}}^k$, and $\omega_{r,\mathrm{C}}^k$, $k = 1, 2$, satisfying the inequalities (14) to (25) in Proposition 3. For any $\epsilon > 0$, we can always construct another set of t' and ω' variables such that the new

[9] $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ usually changes only slightly since it is relatively difficult for *all nodes* to acquire the same common information.

t' and ω' variables satisfy (14) to (24) with strict inequality, and satisfy the following inequality

$$\left(\sum_{k=0}^{4}\omega_{s_i}^k\right)p_i(d_i)+\left(\omega_{r,\mathrm{N}}^i+\omega_{r,\mathrm{C}}^i\right)p_r(d_i)>R_i-\epsilon \tag{41}$$

instead of (25). Based on the above observation, we will assume that the given t_{s_1}, t_{s_2}, t_r, $\omega_{s_i}^k$, $\omega_{r,\mathrm{N}}^k$, and $\omega_{r,\mathrm{C}}^k$ satisfy (14) to (24) and (41) with strict inequality. In the following, we will construct an SBLNC solution such that the scheme "properly terminates" within the allocated n time slots with close-to-one probability and after the SBLNC scheme stops, each d_i has received at least $n(R_i-\epsilon)$ number of its desired information packets.

We construct the SBLNC scheme as follows. We first schedule the s_1-policies sequentially from $\Gamma_{s_1,0}$ to $\Gamma_{s_1,4}$. Each policy $\Gamma_{s_1,k}$ lasts for $n\cdot\omega_{s_1}^k$ time slots. After finishing $\Gamma_{s_1,k}$ we move on to Policy $\Gamma_{s_1,k+1}$ until finishing all 5 s_1-policies. After finishing the s_1-policies, we move on to the s_2-policies. Again, we choose the s_2-policies sequentially from $\Gamma_{s_2,0}$ to $\Gamma_{s_2,4}$ and each policy lasts for $n\cdot\omega_{s_2}^k$ time slots. After the s_2-policies, we schedule the r-policies sequentially from $\Gamma_{r,1}$ to $\Gamma_{r,3}$. Policies $\Gamma_{r,1}$ and $\Gamma_{r,2}$ last for $n\cdot\omega_{r,\mathrm{N}}^1$ and $n\cdot\omega_{r,\mathrm{N}}^2$ time slots, respectively. Policy $\Gamma_{r,3}$ lasts for $n\cdot\max\{\omega_{r,\mathrm{C}}^1,\omega_{r,\mathrm{C}}^2\}$ time slots. Feedback is critical for the SBLNC scheme as it is used to decide the evolution of the knowledge spaces S_1, S_2,..., T_r, which in turn decides the sets in (26)–(39).

To prove the correctness of the above construction, we need to show that the following two statements hold with close-to-one probability: (i) During each time slot, it is always possible to construct the desired coding vectors $\mathbf{v}^{(1)}$ (or $\mathbf{v}^{(2)}$). That is, we never schedule an infeasible policy throughout the operation; (ii) Destination d_i can decode $n(R_i-\epsilon)$ of the desired information packets when the scheme terminates[10]. In addition to the above two statements, we will also prove that (iii) during the first $n\cdot\omega_{r,\mathrm{C}}^1$ (resp. $n\cdot\omega_{r,\mathrm{C}}^2$) time slots of scheduling $\Gamma_{r,3}$, the computed flow-1 vector $\mathbf{v}^{(1)}$ (resp. flow-2 vector $\mathbf{v}^{(2)}$) is not zero with close-to-one probability.

We first prove (ii) while assuming both (i) and (iii) are true. We notice that all the exclusion spaces of policies $\Gamma_{s_1,0}$ to $\Gamma_{s_1,4}$, and $\Gamma_{r,1}$ contain S_1 as a subset. As a result, all those packets carry some new flow-1 information that has not yet been received by d_1. If d_1 receives any of those packets, the dimension of S_1 will increase by one. Similarly, during the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$, the computed $\mathbf{v}^{(1)}$ vector does not belong to S_1, see (38). As a result, if d_1 receives any of those packets, the dimension of S_1 will increase by one. From the above reasoning, the expected value of $\mathsf{dim}(S_1)$ in the end of the SBLNC scheme must satisfy

$$\begin{aligned}\mathsf{E}\{\mathsf{dim}(S_1)\}=&p_1(d_1)\left(\sum_{k=0}^{4}n\omega_{s_1}^k\right)\\&+p_r(d_1)(n\omega_{r,\mathrm{N}}^1+n\omega_{r,\mathrm{C}}^1) \qquad (42)\\>&n(R_1-\epsilon) \qquad (43)\end{aligned}$$

[10]Since we focus on the t and ω variables satisfying inequality (14)–(25), we can use inequality (14)–(16) to show that we can finish transmission within the allocated n time slots.

where the right-hand side of (42) quantifies the expected number of packets received by d_1 during Policies $\Gamma_{s_1,0}$ to $\Gamma_{s_1,4}$, $\Gamma_{r,1}$, and the first $n\omega_{r,\mathrm{C}}^1$ time slots of $\Gamma_{r,3}$. (43) follows from (41). By the law of large number, $\mathsf{dim}(S_1)>n(R_1-\epsilon)$ with close-to-one probability when n is sufficiently large. The above inequality ensures that d_1 can decode $n(R_1-\epsilon)$ of the flow-1 information packets at the end of the SBLNC scheme. By symmetry, d_2 can also decode $n(R_2-\epsilon)$ of the flow-2 packets $\mathbf{W}_2$ in the end of time $t=n$. What remains to be shown is to prove that (i) and (iii) hold with close-to-one probability.

Next we prove (i) and (iii) by the first order analysis that assumes sufficiently large n. We first consider Policy $\Gamma_{s_1,0}$. For any time t, $\Gamma_{s_1,0}$ is a feasible policy if (26) is non-empty. Suppose we have $q\geq 2$, then (26) being non-empty is equivalent to the following condition.

$$\begin{aligned}&\mathsf{dim}(\Omega_1)-\mathsf{dim}(\Omega_1\cap(S_1\oplus S_2\oplus S_r))\\&=\mathsf{dim}(\Omega_1)-\mathsf{dim}(S_1\oplus S_2\oplus S_r)>0.\end{aligned}\tag{44}$$

The reason is as follows. Let us temporarily define $A=\Omega_1$ and $B=\Omega_1\cap(S_1\oplus S_2\oplus S_r)$. Then we have

$$\begin{aligned}&A\backslash B\neq\emptyset\Leftrightarrow|A|-|B|>0\\&\Leftrightarrow q^{\mathsf{dim}(A)}-q^{\mathsf{dim}(B)}>0\Leftrightarrow\mathsf{dim}(A)-\mathsf{dim}(B)>0.\end{aligned}$$

We then note that $\mathsf{dim}(\Omega_1)=nR_1$ is a constant and does not change over time. Also note that $\mathsf{dim}(S_1\oplus S_2\oplus S_r)$ increases monotonically over time since a node accumulates more "knowledge" over time. As a result, if we can prove that (44) holds in the end of the duration of (executing) Policy $\Gamma_{s_1,0}$, then throughout the entire duration of $\Gamma_{s_1,0}$, we can always find some $\mathbf{v}^{(1)}$ belong to (26).

To that end, we notice that when we choose $\Gamma_{s_1,0}$ as our coding policy, the coding vector $\mathbf{v}^{(1)}$ is chosen from (26). Since $\mathbf{v}^{(1)}$ does not belong to the exclusion space $S_1\oplus S_2\oplus S_r$, $\mathsf{dim}(S_1\oplus S_2\oplus S_r)$ *increases by one if and only if at least one of* d_1, d_2, *and* r *receives the transmitted packet* $X_{s_1}=\mathbf{v}^{(1)}\mathbf{W}_1^{\mathrm{T}}$. Also note that in the beginning of Policy $\Gamma_{s_1,0}$, $\mathsf{dim}(S_1\oplus S_2\oplus S_r)=0$. As a result, in the end of the duration of $\Gamma_{s_1,0}$, we have

$$\begin{aligned}&\mathsf{E}\{\mathsf{dim}(S_1\oplus S_2\oplus S_r)\}\\&=0+n\cdot\omega_{s_1}^0\cdot p_1(d_1,d_2,r) \qquad (45)\\&<nR_1=\mathsf{dim}(\Omega_1), \qquad (46)\end{aligned}$$

where (45) follows from quantifying the expected number of time slots (out of totally $n\omega_{s_1}^0$ time slots) in which at least one of d_1, d_2, and r receives it. (46) follows from (17).

By the law of large numbers, (46) implies that (44) holds in the end of the duration of $\Gamma_{s_1,0}$ with close-to-one probability. As a result, with close-to-one probability Policy $\Gamma_{s_1,0}$ remains feasible during the assigned duration of $n\cdot\omega_{s_1}^0$ time slots.

We now consider Policy $\Gamma_{s_1,1}$. For any time t, $\Gamma_{s_1,1}$ is feasible if (27) is non-empty, which is equivalent to having

$$\begin{aligned}
&q^{\mathsf{dim}(S_2\oplus S_r)} = |S_2 \oplus S_r| \\
&\quad > |(S_2 \oplus S_r) \cap ((S_1 \oplus S_r) \cup (S_1 \oplus S_2))| \\
&\quad = |((S_2 \oplus S_r) \cap (S_1 \oplus S_r)) \cup ((S_2 \oplus S_r) \cap (S_1 \oplus S_2))| \\
&\Leftrightarrow \mathsf{dim}(S_2 \oplus S_r) \\
&\quad > \max\{\mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_r)), \\
&\qquad\qquad \mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_2))\}
\end{aligned} \tag{47}$$

where "$\Leftrightarrow$" holds assuming the underlying finite field $\mathsf{GF}(q)$ satisfying $q \geq 2$. The reason is as follows. Let us temporarily define $A = S_2 \oplus S_r$, $B = S_1 \oplus S_r$, and $C = S_1 \oplus S_2$. Then we have

$$A\backslash(B \cup C) \neq \emptyset \Leftrightarrow |A| - |(A \cap B) \cup (A \cap C)| > 0. \tag{48}$$

Also, we have the following inequality

$$\begin{aligned}
&q^{\max\{\mathsf{dim}(A\cap B),\mathsf{dim}(A\cap C)\}} \leq |(A \cap B) \cup (A \cap C)| \\
&\leq q^{\mathsf{dim}(A\cap B)} + q^{\mathsf{dim}(A\cap C)} - 1
\end{aligned} \tag{49}$$

where the last inequality follows from the idea of the union bound and the observation that the all-zero vector is always in both $A \cap B$ and $A \cap C$. Based on (49), we then have

$$(48) \Leftrightarrow \mathsf{dim}(A) > \max\{\mathsf{dim}(A \cap B), \mathsf{dim}(A \cap C)\},$$

which leads to (47).

When we choose Policy $\Gamma_{s_1,1}$ as our coding policy, the coding vector $\mathbf{v}^{(1)}$ is chosen from (27). Therefore, $\mathbf{v}^{(1)}$ must belong to the inclusion space $S_2 \oplus S_r$, which implies that no matter how many nodes in $\{d_1, d_2, r\}$ receive the packet, $\mathsf{dim}(S_2 \oplus S_r)$ remains the same. Also note that similar to the case of $\Gamma_{s_1,0}$, $\mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_r))$ and $\mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_2))$ increase monotonically over time. As a result, if we can prove that (47) holds in the end of the duration of Policy $\Gamma_{s_1,1}$, then throughout the entire duration of $\Gamma_{s_1,1}$, we can always find some $\mathbf{v}^{(1)}$ belong to (27). The remaining task is thus to quantify the three different terms $\mathsf{dim}(S_2 \oplus S_r)$, $\mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_r))$, and $\mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_2))$ at the end of (the duration of) $\Gamma_{s_1,1}$. All the following discussions hold with close-to-one probability when focusing on the first order analysis of n.

We will first decide the value of $\mathsf{dim}(S_2 \oplus S_r)$. We know that $\mathsf{dim}(S_2 \oplus S_r)$ remains the same during Policy $\Gamma_{s_1,1}$. Therefore, the value of $\mathsf{dim}(S_2 \oplus S_r)$ is decided by how much it increases during $\Gamma_{s_1,0}$. Since any $\mathbf{v}^{(1)}$ in Policy $\Gamma_{s_1,0}$ does not belong to $S_2 \oplus S_r$ (see (26)), every time one of d_2 and r receives a packet of $\Gamma_{s_1,0}$, $\mathsf{dim}(S_2 \oplus S_r)$ will increase by one. As a result, in the end of $\Gamma_{s_1,1}$ we have

$$\mathsf{E}\{\mathsf{dim}(S_2 \oplus S_r)\} = n\omega_{s_1}^0 p_1(d_2, r) + n\omega_{s_1}^1 \cdot 0. \tag{50}$$

We now consider the first term $\mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_r))$ in the max operation in (47). By Lemma 2, we can rewrite $\mathsf{dim}((S_1 \oplus S_r) \cap (S_2 \oplus S_r))$ by

$$\begin{aligned}
&\mathsf{dim}((S_1 \oplus S_r) \cap (S_2 \oplus S_r)) \\
&= \mathsf{dim}(S_2 \oplus S_r) + \mathsf{dim}(S_1 \oplus S_r) - \mathsf{dim}(S_1 \oplus S_2 \oplus S_r).
\end{aligned} \tag{51}$$

The value of $\mathsf{dim}(S_2 \oplus S_r)$ is quantified in (50). Since any $\mathbf{v}^{(1)}$ in Policy $\Gamma_{s_1,0}$ does not belong to $S_1 \oplus S_r$ (see (26)) and any $\mathbf{v}^{(1)}$ in Policy $\Gamma_{s_1,1}$ does not belong to $S_1 \oplus S_r$ either (see (27)), every time one of d_1 and r receives a packet of $\Gamma_{s_1,0}$ or $\Gamma_{s_1,1}$, $\mathsf{dim}(S_1 \oplus S_r)$ will increase by one. In the end of $\Gamma_{s_1,1}$ we thus have

$$\mathsf{E}\{\mathsf{dim}(S_1 \oplus S_r)\} = n\omega_{s_1}^0 \cdot p_1(d_1, r) + n\omega_{s_1}^1 \cdot p_1(d_1, r). \tag{52}$$

Similarly, since any $\mathbf{v}^{(1)}$ in Policy $\Gamma_{s_1,0}$ does not belong to $S_1 \oplus S_2 \oplus S_r$ (see (26)) and any $\mathbf{v}^{(1)}$ in Policy $\Gamma_{s_1,1}$ belongs to $S_1 \oplus S_2 \oplus S_r$ (see (27)), every time one of d_1, d_2, and r receives a packet of $\Gamma_{s_1,0}$, $\mathsf{dim}(S_1 \oplus S_2 \oplus S_r)$ will increase by one. In the end of $\Gamma_{s_1,1}$ we thus have

$$\mathsf{E}\{\mathsf{dim}(S_1 \oplus S_2 \oplus S_r)\} = n\omega_{s_1}^0 p_1(d_1, d_2, r) + n\omega_{s_1}^1 \cdot 0. \tag{53}$$

By (50), (51), (52), and (53), we can verify that (18) implies that $\mathsf{dim}(S_2 \oplus S_r) > \mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_r))$ in the end of Policy $\Gamma_{s_1,1}$. By swapping the roles of d_2 and r, symmetric arguments can be used to prove that (19) implies $\mathsf{dim}(S_2 \oplus S_r) > \mathsf{dim}((S_2 \oplus S_r) \cap (S_1 \oplus S_2))$ in the end of Policy $\Gamma_{s_1,1}$. Therefore, $\Gamma_{s_1,1}$ is feasible throughout its duration of $n\omega_{s_1}^1$ time slots.

Similar dimension-comparison arguments can be used to complete the proof of (i) and (iii). The remaining derivation repeats similar steps described above, and is relegated to Appendix B. The proof of Proposition 3 is thus complete.

B. Capacity of The COPE Principle On 2-Flow Wireless Butterfly Networks with Broadcast PECs

In this subsection we will prove that the capacity outer bound in Proposition 2 and the capacity inner bound in Proposition 3 coincide when destination d_i cannot directly hear from source s_i for $i = 1, 2$, which prohibits the use of opportunistic routing. Proposition 1 thus describes the exact capacity region of the COPE principle on 2-flow wireless butterfly networks with broadcast PECs. Recall that in Proposition 2, the outer bound considers channel status feedback, dynamic scheduling, and possibly non-linear encoding functions. Meanwhile, the proposed achieving scheme in Section V-A uses only linear encoding functions and a sequential-order scheduling. By proving that the outer and inner bounds match, we have shown that the proposed linear encoding and sequential scheduling scheme in Section V-A is as good as any other non-linear encoding and dynamic scheduling scheme when the use of opportunistic routing is prohibited.

To complete the proof, we note that when $p_i(d_i) = 0$, for $i = 1, 2$, (17)–(25) of the inner bound in Proposition 3 is reduced to the following forms:[11]

$$\omega_{s_i}^0 p_i(d_j, r) \le R_i, \tag{54}$$
$$\omega_{s_i}^1 p_i(r) \le \omega_{s_i}^0 p_i(d_j\overline{r}), \tag{55}$$
$$\omega_{s_i}^1 p_i(d_j) \le \omega_{s_i}^0 p_i(r\overline{d_j}), \tag{56}$$
$$\omega_{s_i}^2 p_i(r) \le \omega_{s_i}^0 p_i(d_j\overline{r}) - \omega_{s_i}^1 p_i(r), \tag{57}$$
$$\omega_{s_i}^3 p_i(d_j) \le \omega_{s_i}^0 p_i(r\overline{d_j}) - \omega_{s_i}^1 p_i(d_j), \tag{58}$$
$$\omega_{r,\mathrm{N}}^i p_r(d_i, d_j) \le \omega_{s_i}^0 p_i(r\overline{d_j}) - \omega_{s_i}^1 p_i(d_j) - \omega_{s_i}^3 p_i(d_j), \tag{59}$$
$$\omega_{r,\mathrm{C}}^i p_r(d_i) \le \omega_{s_i}^0 p_i(d_j r) + \omega_{s_i}^1 (p_i(d_j) + p_i(r)) + \omega_{s_i}^2 p_i(r) + \omega_{s_i}^3 p_i(d_j) + \omega_{r,\mathrm{N}}^i p_r(d_j\overline{d_i}). \tag{60}$$

and (25) becomes

$$p_r(d_i)(\omega_{r,\mathrm{N}}^i + \omega_{r,\mathrm{C}}^i) \ge R_i. \tag{61}$$

The following lemma proves the tightness of the bounds when there is no 2-hop overhearing, i.e., $p_i(d_i) = 0$ for $i = 1, 2$.

Lemma 3: For any 5-tuple $(R_1, R_2, t_1, t_2, t_r)$ satisfying the capacity outer bound (6)–(9), we can always find 14 accompanying variables $\omega_{s_i}^j, \omega_{r,\mathrm{N}}^i, \omega_{r,\mathrm{C}}^i$ for $i = 1, 2$ and $j = 0, 1, 2, 3, 4$, such that jointly the $5 + 14 = 19$ variables satisfy (14), (15), (54) to (61).

Proof: Given any $(R_1, R_2, t_{s_1}, t_{s_2}, t_r)$ satisfying (6)–(9), we construct $\{\omega_{s_i}^j, \omega_{r,\mathrm{N}}^i, \omega_{r,\mathrm{C}}^i : i \in \{1, 2\}, j \in \{0, 1, 2, 3, 4\}\}$ in the following way. For each pair $(i, j) = (1, 2)$ or $(2, 1)$, we define

$$\omega_{s_i}^0 = \frac{R_i}{p_i(d_j, r)}, \tag{62}$$
$$\omega_{s_i}^1 = R_i \left(\min\left\{ \frac{1}{p_i(r)}, \frac{1}{p_i(d_j)} \right\} - \frac{1}{p_i(d_j, r)} \right), \tag{63}$$
$$\omega_{s_i}^2 = R_i \left(\frac{1}{p_i(r)} - \frac{1}{p_i(d_j)} \right)^+, \tag{64}$$
$$\omega_{s_i}^3 = \min\left\{ R_i \left(\frac{1}{p_i(d_j)} - \frac{1}{p_i(r)} \right)^+, t_{s_i} - \frac{R_i}{p_i(r)} \right\}, \tag{65}$$
$$\omega_{s_i}^4 = 0, \tag{66}$$
$$\omega_{r,\mathrm{N}}^i = \frac{(R_i - t_{s_i} p_i(d_j))^+}{p_r(d_i, d_j)}, \tag{67}$$
$$\omega_{r,\mathrm{C}}^i = \frac{R_i}{p_r(d_i)} - \frac{(R_i - t_{s_i} p_i(d_j))^+}{p_r(d_i, d_j)}. \tag{68}$$

One can verify that the above assignment $\{R_1, R_2, t_{s_1}, t_{s_2}, t_r, \omega_{s_i}^j, \omega_{r,\mathrm{N}}^i, \omega_{r,\mathrm{C}}^i : i \in \{1, 2\}, j \in \{0, 1, 2, 3, 4\}\}$ is always non-negative and satisfies (14), (15), (54) to (61). The detailed verification is relegated to Appendix C. The proof of Lemma 3 is thus complete. ■

VI. Numerical Results

In this section, we apply the capacity results to some numerically generated scenarios so that we can explicitly quantify the throughput/capacity improvement of the SBLNC scheme under the COPE principle with and without opportunistic routing. The detailed simulation setting is described as follows.

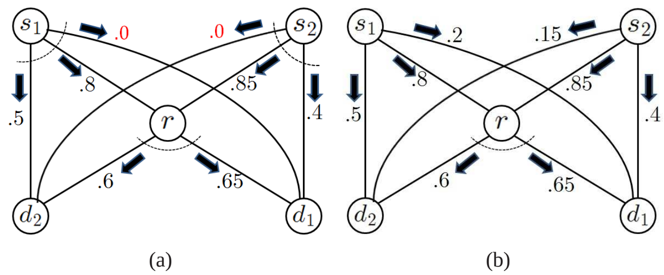


Fig. 5. An instance of the 2-flow wireless butterfly network with broadcast PECs with the success probabilities being indicated next to the corresponding arrows: (a) The COPE principle only; and (b) The COPE principle plus opportunistic routing. We also assume that the success events between different node pairs are independent.

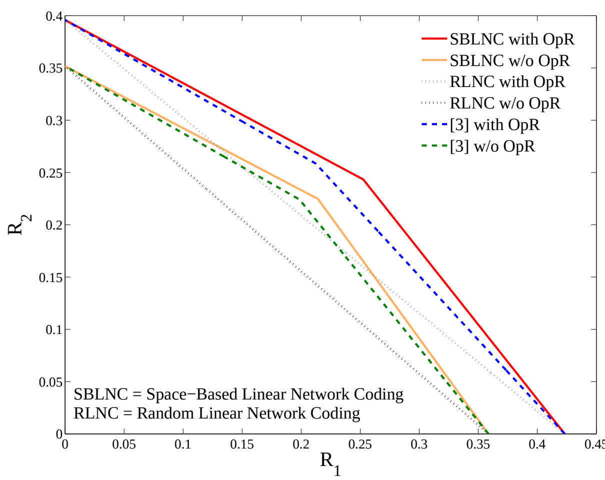


Fig. 6. The achievable regions of the scenario in Fig. 5(a) and Fig. 5(b). The solid lines indicate the achievable regions of the SBLNC scheme under the scenarios of COPE only (the orange line) or COPE plus OpR (the red line). The dash lines indicate the achievable regions of the existing result [3] under the scenarios of COPE only (the green line) or COPE plus OpR (the blue line). The dotted lines indicate the achievable regions of intra-session network coding only (random linear network coding) under the scenarios of OpR (the gray line) or not (the black line).

Consider one particular channel parameter assignment of the 2-flow wireless butterfly network with broadcast PECs. Fig. 5(a) describes the transmission *success probability* between each node pair as the number next to the corresponding arrow without opportunistic routing. And Fig. 5(b) describes the same set of channel parameter assignment, except for that now we allow OpR. We also assume that the success events between different node pairs are independent. For example, when allowing opportunistic routing in Fig. 5(b), the probability that a packet sent by s_1 is heard by d_1 is $p_1(d_1) = .2$ and the probability that a packet sent by r is received by d_2 is $p_r(d_2) = .6$. We then compute 6 different achievable regions and plot them in Fig. 6.

The solid lines in Fig. 6 represent the achievable regions[12] of

[11] Inequality (22) becomes trivial since the left-hand side of (22) becomes zero and the right-hand side of (22) is always non-negative.

[12] Our main results provide a pair of outer and inner bounds for this capacity region. Since the gap between the inner and outer bounds is negligible (with relative gap less than 0.08%), we plot only the inner bound (the achievable rate) in Fig. 6.

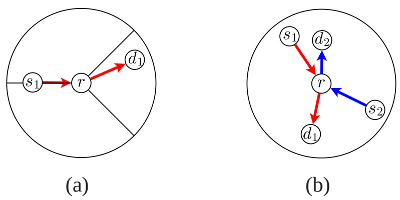


Fig. 7. (a) The relative location of (s_i, d_i). (b) Topology of two (s_i, d_i) pairs.

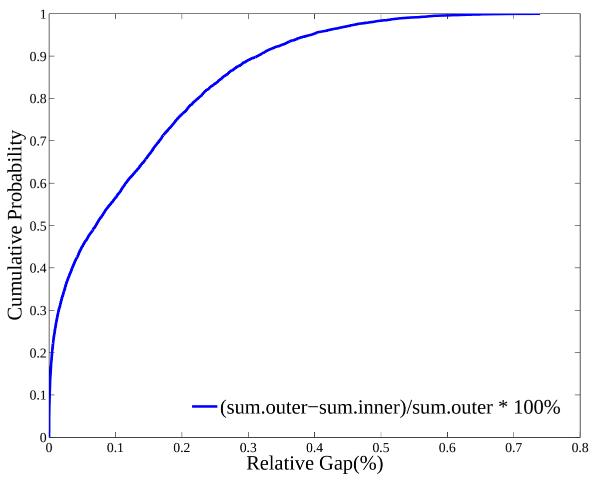


Fig. 8. The cumulative distribution of the relative gap between the outer and the inner bounds with proportional fairness. The outer and the inner bounds are described in Propositions 2 and 3, respectively.

the SBLNC scheme with and without OpR, respectively. The dash lines represent the achievable regions of the existing side-information-based feedback-free result [3] with and without OpR. The dotted lines represent the achievable regions of intra-session network coding while performing time-sharing between different sessions. When OpR is allowed, the dotted line corresponds to the intra-session RLNC discussed in [29]. When OpR is not allowed, the dotted line is the scheduling-based capacity (stability) of the store-&-forward solutions.

As can be seen, when there is only one flow in the network (say $R_2 = 0$), then OpR (the gray dotted line) is optimal as was first established in [29]. However, when there are two coexisting flows (when both R_1 and $R_2 > 0$), the COPE principle without OpR (the orange solid line) can sometimes outperform OpR due to the stronger overhearing between $s_2 \to d_1$ and $s_1 \to d_2$, $p_2(d_1) = 0.4$ and $p_1(d_2) = 0.5$, than the two-hop direct overhearing from $s_1 \to d_1$ and $s_2 \to d_2$, $p_1(d_1) = 0.2$ and $p_2(d_2) = 0.15$. On the other hand, the throughput can be further enhanced by the proposed joint COPE and OpR solution (the red solid line). Last but not least, the proposed SBLNC scheme (the solid lines) always outperforms all the existing schemes since it achieves/approaches the capacity region.

We are also interested in quantifying the average throughput benefits of COPE and OpR in a randomly placed network. To generate a typical XOR-in-the-air scenario, we first place the relay node in the center of a unit circle. Then we randomly place four nodes (s_1, s_2, d_1, d_2) inside the unit circle. To simulate the need of the relay for each session pair, we force the placement of each pair to be in the opposite 90 degree area. That is, d_i must be located in the opposite 90 degree area of s_i's location for $i = 1, 2$. See Fig. 7(a) for illustration. Fig. 7(b) illustrates one realization of our random node placement.

We use the Euclidean distance D between any two nodes to decide the overhearing probability when a packet is transmitted. More explicitly, we use the Rayleigh model

$$\mathsf{Prob}(\mathsf{success}) = \int_{T^*}^{\infty} \frac{2x}{\gamma} e^{-\frac{x^2}{\gamma}} dx \quad \text{where } \gamma \triangleq \frac{1}{(4\pi)^2 D^\alpha},$$

where α is the path loss factor, and T^* is the decodable SNR threshold. To reflect the packet delivery ratio measured in practical environments, we choose $\alpha = 2.5$ and $T^* = 0.006$ so that the overhearing probability for a 1-hop neighbor is around 0.7–0.8 while overhearing probability for a 2-hop neighbor is around 0.2–0.3. If no direct overhearing is allowed, we simply hardwire the probability that d_i overhears s_i to be zero. We again assume that the success events between different node pairs are independent.

We consider three different fairness requirements: (a) No fairness requirement; (b) Proportional fairness; and (c) Min-cut-based fairness requirement. When there is no fairness constraint, we use a linear programming solver to find the largest sum rate $R_1 + R_2$ that satisfies the capacity outer bound in Proposition 2, which is denoted by $R_{\text{sum.outer}}$. Similarly, we find the largest sum rate $R_1 + R_2$ that satisfies the capacity inner bound in Proposition 3 and denote it by $R_{\text{sum.inner}}$. After computing the sum rates $R_{\text{sum.outer}}$ and $R_{\text{sum.inner}}$, we repeat the above experiment with different randomly chosen node placements for 10000 times. For the setting of (b) proportional fairness, we replace the sum rate objective function $R_1 + R_2$ by the logarithmic objective function $\log(R_1) + \log(R_2)$. We again compute the $R_{\text{sum.outer}}$ and $R_{\text{sum.inner}}$ using the new objective function. For the setting of (c) min-cut-based requirement, we impose an additional constraint $R_i = \beta \min\left(p_i(d_i, r), p_i(d_i) + p_r(d_i)\right)$ for $i = 1, 2$ with a common β, which enforces the individual rate R_i being proportional to the min-cut value from s_i to d_i assuming no other sessions are transmitting and s_i and r are scheduled with the same frequency. The results are summarized in Table III and Fig. 8.

TABLE III
AVERAGE SUM-RATES OVER 10000 RANDOM NODE PLACEMENTS.

Fairness Constraints	OpR	SBLNC	[3]	RLNC
No	allowed	.6599/.6594	.6472	.6180
	not allowed	.4820	.4779	.4116
Proportional	allowed	.6294/.6286	.6101	.5484
	not allowed	.4775	.4726	.3854
Min-cut	allowed	.6031/.6026	.5892	.5406
	not allowed	.4671	.4626	.3856

Table III lists the sum-rate averaged over 10000 simulations. When opportunistic routing is allowed ($p_i(d_i) > 0$), then the inner and outer bounds do not always meet. Therefore, for the entries with SBLNC and OpR, the number on the left is the average of $R_{\text{sum.outer}}$ while the number on the right is the average of $R_{\text{sum.inner}}$. When there is no OpR ($p_i(d_i) = 0$), as was proven Section V-B, the sum-rate outer and inner bounds always coincide and hence only one number is shown in each corresponding entry. The achievable rates in [3] for different scenarios are also listed for comparison. The capacity of pure routing and pure OpR is provided in [29] and we also compute the capacity of the intra-session-coding-only RLNC-based achievability scheme in [29] for comparison. In Table III, we first note that in terms of the averaged throughput, the difference between the outer and the inner bounds is around 0.08%. Among all 10000 instances, the largest absolute difference is with $R_{\text{sum.outer}} = 0.6409$ and $R_{\text{sum.inner}} = 0.6375$. The proposed bounds thus effectively bracket the capacity when combining the COPE and the OpR principles. With the proportional fairness constraint, the proposed SBLNC scheme that combines the benefits of COPE and OpR provides roughly 14.6% throughput improvement when compared to the scheme using only intra-session RLNC and OpR without the inter-session COPE-based network coding operations. Again assuming proportional fairness, if we compare the SBLNC scheme that combines COPE and OpR (sum rate = 0.6286) with the benchmark store-&-forward scheme without inter-session coding and without OpR (sum rate = 0.3854), it shows that COPE and OpR can provide in average 63% throughput improvement over the traditional TCP/IP solutions. We also observe that the throughput improvement of COPE is greater when there is some kind of fairness constraints. The intuition is that inter-session network coding helps resolve the bottleneck experienced by the *weaker* user since the weaker user's packets can now be coded together with the stronger user's packets.

Fig. 8 focuses on the relative gap per experiment when allowing for both COPE and OpR with proportional fairness. Specifically, considering the setting of proportional fairness, we compute the relative gap per each experiment, $(R_{\text{sum.outer}} - R_{\text{sum.inner}})/R_{\text{sum.outer}}$, and then plot the cumulative distribution function (cdf) for the relative gaps. We can see that with more than 80% of the experiments, the relative gap between the outer and inner bounds is smaller than 0.25%.

VII. Conclusion

This paper has introduced a new network coding architecture, named as the "Space-Based Linear Network Code (SBLNC)." The SBLNC scheme has been used to find the exact capacity region of the COPE principle 2-flow wireless butterfly networks with broadcast PECs. The result has also been extended to bracket the capacity when combining the COPE principle and the concept of opportunistic routing. Numerical results show that the proposed outer and inner bounds effectively quantify the capacity for almost all practical scenarios.

Appendix

A. Proof of Proposition 2

For any joint scheduling and NC scheme, we choose t_{s_i} (resp. t_r) as the normalized *expected number of time slots for which s_i (resp. r) is scheduled*. Namely,

$$t_{s_i} \triangleq \frac{1}{n}\mathsf{E}\left\{\sum_{\tau=1}^{n} 1_{\{\sigma(\tau)=s_i\}}\right\} \text{ and } t_r \triangleq \frac{1}{n}\mathsf{E}\left\{\sum_{\tau=1}^{n} 1_{\{\sigma(\tau)=r\}}\right\}.$$

By definition, t_{s_1}, t_{s_2}, and t_r must satisfy (10).

In the subsequent proofs, the logarithm is taken with base q. We prove (11) first. To that end, we notice that

$$\begin{aligned}
&I(\mathbf{W}_1;\hat{\mathbf{W}}_1) \le I(\mathbf{W}_1;[\mathbf{Y}_{\{s_1,s_2,r\}\to d_1},\mathbf{Z}]_1^n) && (69)\\
&= I(\mathbf{W}_1;[\mathbf{Z}]_1^n) + I(\mathbf{W}_1;[\mathbf{Y}_{\{s_1,s_2,r\}\to d_1}]_1^n|[\mathbf{Z}]_1^n) && (70)\\
&\le I(\mathbf{W}_1;[\mathbf{Y}_{\{s_1,s_2\}\to\{d_1,r\}}]_1^n|[\mathbf{Z}]_1^n) && (71)\\
&= I(\mathbf{W}_1;[\mathbf{Y}_{s_1\to\{d_1,r\}}]_1^n|[\mathbf{Z}]_1^n) && (72)\\
&\le H([\mathbf{Y}_{s_1\to\{d_1,r\}}]_1^n|[\mathbf{Z}]_1^n)\\
&= \sum_{t=1}^{n} H(\mathbf{Y}_{s_1\to\{d_1,r\}}(t)|[\mathbf{Z}]_1^n,[\mathbf{Y}_{s_1\to\{d_1,r\}}]_1^{t-1}) && (73)\\
&\le \sum_{t=1}^{n} \mathsf{E}\left\{1_{\{Z_{s_1\to d_1}(t)=1 \text{ or } Z_{s_1\to r}(t)=1\}} \circ 1_{\{\sigma(t)=s_1\}}\right\} && (74)\\
&= n t_{s_1} p_1(d_1,r) && (75)
\end{aligned}$$

where (69) follow from (4); (70) follows from the chain rule; (71) follows from (3), the data processing inequality, and the fact that $\mathbf{Z}$ is independent of $\mathbf{W}_1$; (72) follows from that conditioning on $\mathbf{Z}$ (and σ since σ is a function of $\mathbf{Z}$) $\mathbf{Y}_{s_2\to\{d_1,r\}}$ is a deterministic function of $\mathbf{W}_2$ and is thus independent of $\mathbf{W}_1$; (73) follows from the chain rule; (74) follows from that only when $1_{\{Z_{s_1\to d_1}(t)=1 \text{ or } Z_{s_1\to r}(t)=1\}} = 1$ and $1_{\{\sigma(t)=s_1\}} = 1$ will we have a non-zero entropy value $H(\mathbf{Y}_{s_1\to\{d_1,r\}}(t)|[\mathbf{Z}]_1^n,[\mathbf{Y}_{s_1\to\{d_1,r\}}]_1^{t-1})$, and when $H(\mathbf{Y}_{s_1\to\{d_1,r\}}(t)|[\mathbf{Z}]_1^n,[\mathbf{Y}_{s_1\to\{d_1,r\}}]_1^{t-1}) > 0$, it is upper bounded by 1 since the base of the logarithm is q; (75) follows from Wald's lemma.

On the other hand, Fano's inequality gives us

$$I(\mathbf{W}_1;\hat{\mathbf{W}}_1) \ge nR_1(1-\epsilon) - H(\epsilon). \tag{76}$$

Combining (75) and (76), we have

$$R_1(1-\epsilon) - \frac{H(\epsilon)}{n} \le t_{s_1} p_1(d_1,r). \tag{77}$$

Letting $\epsilon \to 0$, (77) implies (11) for the case of $i = 1$. With symmetric arguments, we can derive (11) for $i = 2$.

We prove (13) by similar techniques as used in [5], [30]. Specifically, we create a new network from the original network by adding an auxiliary pipe that sends all information available at d_2 directly to d_1. Later we will show that even with the additional information, the achievable rates R_1 and R_2 are still upper bounded by (13). As a result, the achievable R_1 and R_2 for the original network must satisfy (13) as well. (12) is a symmetric version of (13).

With the additional information at d_1, the decoding function (see (4)) at d_1 for the new network becomes

$$\hat{\mathbf{W}}_1 = f_{d_1}([\mathbf{Y}_{\{s_1,s_2,r\}\to\{d_1,d_2\}},\mathbf{Z}]_1^n). \tag{78}$$

For any $t \in [n]$, define

$$U(t) \triangleq (\mathbf{W}_2,[\mathbf{Y}_{\{s_1,s_2,r\}\to\{d_1,d_2\}},\mathbf{Z}]_1^{t-1}). \tag{79}$$

We then have

$$
\begin{aligned}
nR_1 &= H(\mathbf{W}_1|\mathbf{W}_2) \\
&\le I(\mathbf{W}_1;\hat{\mathbf{W}}_1|\mathbf{W}_2) + n\epsilon_1 && (80)\\
&\le I(\mathbf{W}_1;[\mathbf{Y}_{\{s_1,s_2,r\}\to\{d_1,d_2\}},\mathbf{Z}]_1^n|\mathbf{W}_2) + n\epsilon_1 && (81)\\
&= I(\mathbf{W}_1;[\mathbf{Z}]_1^n|\mathbf{W}_2) \\
&\quad + I(\mathbf{W}_1;[\mathbf{Y}_{\{s_1,s_2,r\}\to\{d_1,d_2\}}]_1^n|\mathbf{W}_2,[\mathbf{Z}]_1^n) + n\epsilon_1 && (82)\\
&= \sum_{t=1}^{n} I(\mathbf{W}_1;\mathbf{Y}_{\{s_1,s_2,r\}\to\{d_1,d_2\}}(t) \\
&\qquad |\mathbf{W}_2,[\mathbf{Z}]_1^n,[\mathbf{Y}_{\{s_1,s_2,r\}\to\{d_1,d_2\}}]_1^{t-1}) + n\epsilon_1 && (83)\\
&= n\epsilon_1 + \sum_{t=1}^{n}\big(I(\mathbf{W}_1;\mathbf{Y}_{r\to\{d_1,d_2\}}(t)|U(t),[\mathbf{Z}]_1^n) \\
&\quad + I(\mathbf{W}_1;\mathbf{Y}_{s_1\to\{d_1,d_2\}}(t)|U(t),\mathbf{Y}_{r\to\{d_1,d_2\}}(t),[\mathbf{Z}]_1^n) \\
&\quad + I(\mathbf{W}_1;\mathbf{Y}_{s_2\to\{d_1,d_2\}}(t) \\
&\qquad |U(t),\mathbf{Y}_{\{r,s_1\}\to\{d_1,d_2\}}(t),[\mathbf{Z}]_1^n)\big) && (84)\\
&\le n\epsilon_1 + \left(\sum_{t=1}^{n} I(\mathbf{W}_1;\mathbf{Y}_{r\to\{d_1,d_2\}}(t)|U(t),[\mathbf{Z}]_1^n)\right) \\
&\quad + nt_{s_1}p_1(d_1,d_2) + 0 && (85)\\
&\le n\epsilon_1 + nt_{s_1}p_1(d_1,d_2) \\
&\quad + \sum_{t=1}^{n} I(X_r(t);\mathbf{Y}_{r\to\{d_1,d_2\}}(t)|U(t),[\mathbf{Z}]_1^n), && (86)
\end{aligned}
$$

where (80) follows from Fano's inequality where ϵ_1 goes to 0 when $\epsilon \to 0$; (81) follows from the data processing inequality and (78); (82), (83), and (84) follow from the chain rule and the fact that the distribution of $\mathbf{Z}$ is independent of $\mathbf{W}_1$ and $\mathbf{W}_2$; (85) follows from the observation that the second term of the summation can be upper bounded by Wald's lemma (similar to (75)) and $\mathbf{Y}_{s_2\to\{d_1,d_2\}}(t)$ is independent of $\mathbf{W}_1$ given $\mathbf{Z}$ (similar to (72)); and (86) follows from the data processing inequality.

To continue, we define the time sharing random variable $Q_t \in \{1,2,...,n\}$ with $\mathsf{Prob}(Q_t = i) = \frac{1}{n}$ for all $i \in \{1,2,...,n\}$ and Q_t being independent of $[\mathbf{Z}]_1^n$, $\mathbf{W}_1$, and $\mathbf{W}_2$. Since the mutual information is always non-negative, we can rewrite (86) as

$$
\begin{aligned}
&(R_1 - t_{s_1}p_1(d_1,d_2) - \epsilon_1)^+ \\
&\le \sum_{t=1}^{n}\frac{1}{n} I(X_r(t);\mathbf{Y}_{r\to\{d_1,d_2\}}(t)|U(t),[\mathbf{Z}]_1^n) \\
&\le \sum_{t=1}^{n}\frac{1}{n} H(\mathbf{Y}_{r\to\{d_1,d_2\}}(t)|U(t),[\mathbf{Z}]_1^n) && (87)\\
&= \sum_{q_t=1}^{n} \mathsf{Prob}(Q_t = q_t) \\
&\quad \cdot H(\mathbf{Y}_{r\to\{d_1,d_2\}}(q_t)|U(q_t),[\mathbf{Z}]_1^{q_t},Q_t = q_t) && (88)
\end{aligned}
$$

where (87) follows from the definition of the mutual information; (88) follows from replacing the time index t by the time sharing random variable Q_t and the distribution of $U(q_t)$ and $\mathbf{Y}_{r\to\{d_1,d_2\}}(q_t)$ does not depend on the future channel realization $[\mathbf{Z}]_{q_t+1}^n$.

We define three binary random variables $\Theta_\sigma \triangleq 1_{\{\sigma(Q_t)=r\}}$, $\Theta_{Z_1} \triangleq 1_{\{Z_{r\to d_1}(Q_t)=1\}}$, and $\Theta_{Z_2} \triangleq 1_{\{Z_{r\to d_2}(Q_t)=1\}}$, which are functions of Q_t and $[\mathbf{Z}]_1^{Q_t}$. Then we can rewrite (88) as the following.

$$
\begin{aligned}
&(R_1 - t_{s_1}p_1(d_1,d_2) - \epsilon_1)^+ \\
&\le \sum_{q_t=1}^{n}\frac{1}{n} H(\mathbf{Y}_{r\to\{d_1,d_2\}}(q_t) \\
&\qquad |U(q_t),[\mathbf{Z}]_1^{q_t},Q_t = q_t,\Theta_\sigma,\Theta_{Z_1},\Theta_{Z_2}) && (89)\\
&= \sum_{q_t=1}^{n}\frac{1}{n} \\
&\quad \sum_{\substack{\forall u,[z]_1^{q_t},\\ \theta_\sigma,\theta_{Z_1},\theta_{Z_2}}} p_{U(q_t),[\mathbf{Z}]_1^{q_t},\Theta_\sigma,\Theta_{Z_1},\Theta_{Z_2}}(u,[z]_1^{q_t},\theta_\sigma,\theta_{Z_1},\theta_{Z_2}) \\
&\quad \cdot H(\mathbf{Y}_{r\to\{d_1,d_2\}}(q_t)|U(q_t) = u,[\mathbf{Z}]_1^{q_t} = [z]_1^{q_t}, \\
&\qquad \Theta_\sigma = \theta_\sigma,\Theta_{Z_1} = \theta_{Z_1},\Theta_{Z_2} = \theta_{Z_2}) && (90)\\
&= \sum_{q_t=1}^{n}\frac{1}{n} \sum_{\substack{\forall u,[z]_1^{q_t},\theta_{Z_1},\theta_{Z_2}\\ \text{s.t. } \max\{\theta_{Z_1},\theta_{Z_2}\}=1}} p(u,[z]_1^{q_t},1,\theta_{Z_1},\theta_{Z_2}) \\
&\quad \cdot H(X_r(q_t)|U(q_t) = u,[\mathbf{Z}]_1^{q_t} = [z]_1^{q_t}, \\
&\qquad \Theta_\sigma = 1,\Theta_{Z_1} = \theta_{Z_1},\Theta_{Z_2} = \theta_{Z_2}) && (91)
\end{aligned}
$$

where (89) follows from the fact that Θ's are functions of Q_t and $[\mathbf{Z}]_1^{Q_t}$; (90) follows from the definition of the conditional entropy; and (91) follows from the fact that $Y_{r\to\{d_1,d_2\}}(q_t)$ is not erasure only if $\sigma(q_t) = r$ and at least one of $Z_{r\to d_1}$ and $Z_{r\to d_2}$ equals to one and furthermore $Y_{r\to\{d_1,d_2\}}(q_t) = X_r(q_t)$ under such a condition, where we use $p(u,[z]_1^{q_t},1,\theta_{Z_1},\theta_{Z_2})$ as the shorthand of $p_{U(q_t),[\mathbf{Z}]_1^{q_t},\Theta_\sigma,\Theta_{Z_1},\Theta_{Z_2}}(u,[z]_1^{q_t},1,\theta_{Z_1},\theta_{Z_2})$.

We can further simplify (91) by the following steps. We first note that conditioning on $U(q_t) = u$, $[\mathbf{Z}]_1^{q_t-1} = [z]_1^{q_t-1}$, and $\Theta_\sigma = 1$, the random variable $X_r(q_t)$ is independent of $\mathbf{Z}(q_t)$, Θ_{Z_1}, and Θ_{Z_2}. Notice that $[\mathbf{Z}]_1^{q_t-1}$ is a subset of $U(q_t)$. Therefore, we have

$$
\begin{aligned}
&H(X_r(q_t)|U(q_t) = u,[\mathbf{Z}]_1^{q_t} = [z]_1^{q_t},\Theta_\sigma = 1,\Theta_{Z_1} = \theta_{Z_1}, \\
&\qquad \Theta_{Z_2} = \theta_{Z_2}) \\
&= H(X_r(q_t)|U(q_t) = u,\Theta_\sigma = 1). && (92)
\end{aligned}
$$

Also the joint probability can be rewritten as

$$
\begin{aligned}
&\sum_{\substack{\forall u,[z]_1^{q_t},\theta_{Z_1},\theta_{Z_2}\\ \text{s.t. } \max\{\theta_{Z_1},\theta_{Z_2}\}=1}} p_{U(q_t),[\mathbf{Z}]_1^{q_t},\Theta_\sigma,\Theta_{Z_1},\Theta_{Z_2}}(u,[z]_1^{q_t},1,\theta_{Z_1},\theta_{Z_2}) \\
&= \sum_{\forall u} p_{U(q_t),\Theta_\sigma}(u,1) \\
&\qquad \sum_{\substack{\forall z,\theta_{Z_1},\theta_{Z_2}\\ \text{s.t. } \max\{\theta_{Z_1},\theta_{Z_2}\}=1}} p_{\mathbf{Z}(q_t),\Theta_{Z_1},\Theta_{Z_2}|U(q_t),\Theta_\sigma}(z,\theta_{Z_1},\theta_{Z_2}|u,1) \\
& && (93)\\
&= \left(\sum_{\forall u} p_{U(q_t),\Theta_\sigma}(u,1)\right)\cdot p_r(d_1,d_2), && (94)
\end{aligned}
$$

where (93) follows from the basic probability definition, and (94) follows from the assumption that the channel is memoryless.

(92) and (94) help us rewrite (91) as

$$
\begin{aligned}
(91) =& t_r \cdot p_r(d_1, d_2) \\
& \cdot \frac{\sum_{q_t=1}^{n} \frac{1}{n} \sum_{\forall u} p(u,1) \cdot H(X_r(q_t)|u,1)}{t_r},
\end{aligned} \tag{95}
$$

where $p(u,1)$ and $H(X_r(q_t)|u,1)$ are the shorthand for $p_{U(q_t),\Theta_\sigma}(u,1)$ and $H(X_r(q_t)|U(q_t)=u,\Theta_\sigma=1)$, respectively.

We now focus on flow 2. By Fano's inequality, for some $\epsilon_2 > 0$ that goes to 0 as $\epsilon \to 0$, with similar steps as in (80)–(86), we can also show that

$$
\begin{aligned}
& nR_2 = H(\mathbf{W}_2) \\
\le & I(\mathbf{W}_2; [\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}, \mathbf{Z}]_1^n) + n\epsilon_2 \\
= & I(\mathbf{W}_2; [\mathbf{Z}]_1^n) + I(\mathbf{W}_2; [\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}]_1^n | [\mathbf{Z}]_1^n) + n\epsilon_2 \quad (96) \\
= & \sum_{t=1}^{n} I(\mathbf{W}_2; \mathbf{Y}_{\{s_1,s_2,r\}\to d_2}(t) | [\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}]_1^{t-1}, [\mathbf{Z}]_1^n) \\
& \quad + n\epsilon_2 \quad (97) \\
= & n\epsilon_2 + \sum_{t=1}^{n} \big(I(\mathbf{W}_2; Y_{r\to d_2}(t) | [\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}]_1^{t-1}, [\mathbf{Z}]_1^n) \\
& + I(\mathbf{W}_2; Y_{s_2\to d_2}(t) | [\mathbf{Y}_{\{s_1,s_2\}\to d_2}]_1^{t-1}, [\mathbf{Y}_{r\to d_2}]_1^t, [\mathbf{Z}]_1^n) \\
& + I(\mathbf{W}_2; Y_{s_1\to d_2}(t) | [\mathbf{Y}_{s_1\to d_2}]_1^{t-1}, [\mathbf{Y}_{\{s_2,r\}\to d_2}]_1^t, [\mathbf{Z}]_1^n) \big) \quad (98) \\
\le & n\epsilon_2 + \sum_{t=1}^{n} I(\mathbf{W}_2; Y_{r\to d_2}(t) | [\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}]_1^{t-1}, [\mathbf{Z}]_1^n) \\
& + n t_{s_2} p_2(d_2) + 0 \quad (99)
\end{aligned}
$$

where (96), (97), and (98) follows from the chain rule and the independence between $\mathbf{W}_2$ and $[\mathbf{Z}]_1^n$; and (99) follows from similar derivation as in (85). We then have

$$
\begin{aligned}
(99) = & n\epsilon_2 + n t_{s_2} p_2(d_2) \\
& + \sum_{t=1}^{n} \big(H(Y_{r\to d_2}(t) | [\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}]_1^{t-1}, [\mathbf{Z}]_1^n) \\
& - H(Y_{r\to d_2}(t) | \mathbf{W}_2, [\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}]_1^{t-1}, [\mathbf{Z}]_1^n) \big) \quad (100) \\
\le & n\epsilon_2 + n t_{s_2} p_2(d_2) \\
& + \sum_{t=1}^{n} \left(H(Y_{r\to d_2}(t) | [\mathbf{Z}]_1^n) - H(Y_{r\to d_2}(t) | U(t), [\mathbf{Z}]_1^n) \right) \\
& \quad (101) \\
= & n\epsilon_2 + n t_{s_2} p_2(d_2) + \sum_{t=1}^{n} I(U(t); Y_{r\to d_2}(t) | [\mathbf{Z}]_1^n), \quad (102)
\end{aligned}
$$

where (100) and (102) follows from the definition of the mutual information; and (101) follows from the fact that *conditioning does not increase the entropy* and $[\mathbf{Y}_{\{s_1,s_2,r\}\to d_2}]_1^{t-1}$ is a subset of $U(t)$. Since the mutual information is always non-negative, we now have

$$
\begin{aligned}
& (R_2 - t_{s_2} p_{s_2}(d_2) - \epsilon_2)^+ \\
\le & \frac{1}{n} \sum_{t=1}^{n} I(U(t); Y_{r\to d_2}(t) | [\sigma, \mathbf{Z}]_1^n) \\
= & \sum_{q_t=1}^{n} \mathsf{Prob}(Q_t = q_t) \cdot I(U(q_t); Y_{r\to d_2}(q_t) | [\mathbf{Z}]_1^{q_t}, Q_t = q_t) \\
& \quad (103) \\
= & \sum_{q_t=1}^{n} \frac{1}{n} \cdot H(Y_{r\to d_2}(q_t) | [\mathbf{Z}]_1^{q_t}, Q_t = q_t) \\
& - \sum_{q_t=1}^{n} \frac{1}{n} \cdot H(Y_{r\to d_2}(q_t) | U(q_t), [\mathbf{Z}]_1^{q_t}, Q_t = q_t), \quad (104)
\end{aligned}
$$

where (103) follows from the definition of the conditional mutual information and the fact that the distribution of $U(q_t)$ and $Y_{r\to d_2}(q_t)$ does not depend on the future channel realization $[\mathbf{Z}]_{q_t+1}^n$; and (104) follows from the definition of the mutual information.We now discuss the first summation in (104)

$$
\begin{aligned}
& \sum_{q_t=1}^{n} \frac{1}{n} \cdot H(Y_{r\to d_2}(q_t) | [\mathbf{Z}]_1^{q_t}, Q_t = q_t) \\
= & \sum_{q_t=1}^{n} \frac{1}{n} \cdot H(Y_{r\to d_2}(q_t) | [\mathbf{Z}]_1^{q_t}, Q_t = q_t, \Theta_\sigma, \Theta_{Z_2}) \quad (105) \\
= & \sum_{q_t=1}^{n} \frac{1}{n} \sum_{\substack{\forall [z]_1^{q_t}, \\ \theta_\sigma, \theta_{Z_2}}} p_{[\mathbf{Z}]_1^{q_t}, \Theta_\sigma, \Theta_{Z_2}}([z]_1^{q_t}, \theta_\sigma, \theta_{Z_2}) \\
& \cdot H(Y_{r\to d_2}(q_t) | [\mathbf{Z}]_1^{q_t} = [z]_1^{q_t}, \Theta_\sigma = \theta_\sigma, \Theta_{Z_2} = \theta_{Z_2}) \\
& \quad (106) \\
= & \sum_{q_t=1}^{n} \frac{1}{n} \sum_{\forall [z]_1^{q_t}} p_{[\mathbf{Z}]_1^{q_t}, \Theta_\sigma, \Theta_{Z_2}}([z]_1^{q_t}, 1, 1) \\
& \cdot H(X_r(q_t) | [\mathbf{Z}]_1^{q_t} = [z]_1^{q_t}, \Theta_\sigma = 1, \Theta_{Z_2} = 1) \quad (107)
\end{aligned}
$$

where (105) follows from the fact that Θ's are functions of Q_t and $[\mathbf{Z}]_1^{Q_t}$; (106) follows from the definition of the conditional entropy; and (107) follows from the fact that $Y_{r\to d_2}(q_t)$ is not erasure only if $\sigma(q_t) = r$ and $Z_{r\to d_2}$ equals to one. Furthermore $Y_{r\to d_2}(q_t) = X_r(q_t)$ under such a condition.

We can further simplify (107) by the following steps. We first note that conditioning on $[\mathbf{Z}]_1^{q_t-1} = [z]_1^{q_t-1}$ and $\Theta_\sigma = 1$, the random variable $X_r(q_t)$ is independent of $\mathbf{Z}(q_t)$ and Θ_{Z_2}. Therefore, we have

$$
\begin{aligned}
& H(X_r(q_t) | [\mathbf{Z}]_1^{q_t} = [z]_1^{q_t}, \Theta_\sigma = 1, \Theta_{Z_2} = 1) \\
& = H(X_r(q_t) | [\mathbf{Z}]_1^{q_t-1} = [z]_1^{q_t-1}, \Theta_\sigma = 1). \quad (108)
\end{aligned}
$$

Also the joint probability can be rewritten as

$$\sum_{\forall [z]_1^{q_t}} p_{[\mathbf{Z}]_1^{q_t},\Theta_\sigma,\Theta_{Z_2}}([z]_1^{q_t},1,1)$$
$$= \sum_{\forall [z]_1^{q_t-1}} p_{[\mathbf{Z}]_1^{q_t-1},\Theta_\sigma}([z]_1^{q_t-1},1) \sum_{\forall z} p_{\mathbf{Z}(q_t),\Theta_{Z_2}|[\mathbf{Z}]_1^{q_t-1},\Theta_\sigma}(z,1|[z]_1^{q_t-1},1) \tag{109}$$
$$= \left(\sum_{\forall [z]_1^{q_t-1}} p_{[\mathbf{Z}]_1^{q_t-1},\Theta_\sigma}([z]_1^{q_t-1},1)\right)\cdot p_r(d_2). \tag{110}$$

where (109) follows from the basic probability definition, and (110) follows from the assumption that the channel is memoryless.

(108) and (110) help us rewrite (107) as

$$(107) = t_r\cdot p_r(d_2) \cdot \frac{\sum_{q_t=1}^n \frac{1}{n}\sum_{\forall [z]_1^{q_t-1}} p([z]_1^{q_t-1},1)\cdot H(X_r(q_t)|[z]_1^{q_t-1},1)}{t_r} \tag{111}$$

where $p([z]_1^{q_t-1},1)$ and $H(X_r(q_t)|[z]_1^{q_t-1},1)$ are the shorthand for $p_{[\mathbf{Z}]_1^{q_t-1},\Theta_\sigma}([z]_1^{q_t-1},1)$ and $H(X_r(q_t)|[\mathbf{Z}]_1^{q_t-1} = [z]_1^{q_t-1},\Theta_\sigma=1)$, respectively.

Similarly, for the second summation in (104),

$$\sum_{q_t=1}^n \frac{1}{n}\cdot H(Y_{r\to d_2}(q_t)|U(q_t),[\mathbf{Z}]_1^{q_t},Q_t=q_t)$$
$$= \sum_{q_t=1}^n \frac{1}{n}\cdot H(Y_{r\to d_2}(q_t)|U(q_t),[\mathbf{Z}]_1^{q_t},Q_t=q_t,\Theta_\sigma,\Theta_{Z_2}) \tag{112}$$
$$= \sum_{q_t=1}^n \frac{1}{n} \sum_{\substack{\forall u,[z]_1^{q_t},\\ \theta_\sigma,\theta_{Z_2}}} p_{U(q_t),[\mathbf{Z}]_1^{q_t},\Theta_\sigma,\Theta_{Z_2}}(u,[z]_1^{q_t},\theta_\sigma,\theta_{Z_2}) \cdot H(Y_{r\to d_2}(q_t)|U(q_t)=u,[\mathbf{Z}]_1^{q_t}=[z]_1^{q_t}, \Theta_\sigma=\theta_\sigma,\Theta_{Z_2}=\theta_{Z_2}) \tag{113}$$
$$= \sum_{q_t=1}^n \frac{1}{n} \sum_{\forall u,[z]_1^{q_t}} p_{U(q_t),[\mathbf{Z}]_1^{q_t},\Theta_\sigma,\Theta_{Z_2}}(u,[z]_1^{q_t},1,1) \cdot H(X_r(q_t)|U(q_t)=u,[\mathbf{Z}]_1^{q_t}=[z]_1^{q_t},\Theta_\sigma=1,\Theta_{Z_2}=1) \tag{114}$$

where (112) follows from the fact that Θ's are functions of Q and $[\mathbf{Z}]_1^{Q_t}$; (113) follows from the definition of the conditional entropy; and (114) follows from the fact that $Y_{r\to d_2}(q_t)$ is not erasure only if $\sigma(q_t)=r$ and $Z_{r\to d_2}$ equals to one. Furthermore $Y_{r\to d_2}(q_t)=X_r(q_t)$ under such a condition.

We can further simplify (114) by the following steps. We first note that conditioning on $U(q_t)=u$, $[\mathbf{Z}]_1^{q_t-1}=[z]_1^{q_t-1}$, and $\Theta_\sigma=1$, the random variable $X_r(q_t)$ is independent of $\mathbf{Z}(q_t)$ and Θ_{Z_2}. Notice that $[\mathbf{Z}]_1^{q_t-1}$ is a subset of $U(q_t)$. Therefore, we have

$$H(X_r(q_t)|U(q_t)=u,[\mathbf{Z}]_1^{q_t}=[z]_1^{q_t},\Theta_\sigma=1,\Theta_{Z_2}=1) = H(X_r(q_t)|U(q_t)=u,\Theta_\sigma=1). \tag{115}$$

Also the joint probability can be rewritten as

$$\sum_{\forall u,[z]_1^{q_t}} p_{U(q_t),[\mathbf{Z}]_1^{q_t},\Theta_\sigma,\Theta_{Z_2}}(u,[z]_1^{q_t},1,1)$$
$$= \sum_{\forall u} p_{U(q_t),\Theta_\sigma}(u,1) \sum_{\forall z} p_{\mathbf{Z}(q_t),\Theta_{Z_2}|U(q_t),\Theta_\sigma}(z,1|u,1) \tag{116}$$
$$= \left(\sum_{\forall u} p_{U(q_t),\Theta_\sigma}(u,1)\right)\cdot p_r(d_2). \tag{117}$$

where (116) follows from the basic probability definition, and (117) follows from the assumption that the channel is memoryless.

(115) and (117) help us rewrite (114) as

$$(107) = t_r\cdot p_r(d_2) \cdot \frac{\sum_{q_t=1}^n \frac{1}{n}\sum_{\forall u} p(u,1)\cdot H(X_r(q_t)|u,1)}{t_r} \tag{118}$$

where $p(u,1)$ and $H(X_r(q_t)|u,1)$ are the shorthand for $p_{U(q_t),\Theta_\sigma}(u,1)$ and $H(X_r(q_t)|U(q_t)=u,\Theta_\sigma=1)$, respectively.

Combining (111) and (118), we can rewrite (104) in the following form.

$$(R_2 - t_{s_2}p_{s_2}(d_2)-\epsilon_2)^+ \le t_r\cdot p_r(d_2) \cdot \left(\frac{\sum_{q_t=1}^n \frac{1}{n}\sum_{\forall [z]_1^{q_t-1}} p([z]_1^{q_t-1},1)\cdot H(X_r(q_t)|[z]_1^{q_t-1},1)}{t_r} - \frac{\sum_{q_t=1}^n \frac{1}{n}\sum_{\forall u} p(u,1)\cdot H(X_r(q)|u,1)}{t_r}\right). \tag{119}$$

Summing up $\frac{(95)}{p_r(d_1,d_2)}$ and $\frac{(119)}{p_r(d_2)}$, we thus have

$$\frac{(R_1 - t_{s_1}p_1(d_1,d_2)-\epsilon_1)^+}{p_r(d_1,d_2)} + \frac{(R_2 - t_{s_2}p_2(d_2)-\epsilon_2)^+}{p_r(d_2)} \le t_r\cdot \frac{\sum_{q_t=1}^n \frac{1}{n}\sum_{\forall [z]_1^{q_t-1}} p([z]_1^{q_t-1},1)\cdot H(X_r(q_t)|[z]_1^{q_t-1},1)}{t_r} \tag{120}$$
$$\le t_r, \tag{121}$$

where (121) is based on the following observations. We first note that by definition

$$t_r = \sum_{q_t=1}^n \frac{1}{n}\mathsf{Prob}(\sigma(q_t)=r) = \sum_{q_t=1}^n \frac{1}{n}\sum_{\forall [z]_1^{q_t-1}} p([z]_1^{q_t-1},1).$$

Therefore, the fraction term in (120) can be viewed as the normalization of the conditional entropy $H(X_r(q_t)|[z]_1^{q_t-1},1)$. Since each conditional entropy is no larger than 1 (with the base of the logarithm being q), we thus have (121).

(121) holds for arbitrary $\epsilon > 0$. Letting $\epsilon \to 0$[13], we thus have the following final inequality.

$$\frac{(R_1 - t_{s_1} p_1(d_1,d_2))^+}{p_r(d_1,d_2)} + \frac{(R_2 - t_{s_2} p_2(d_2))^+}{p_r(d_2)} \le t_r,$$

which gives us (13). (12) can be proven by symmetry. The proof of the outer bound is thus complete.

B. Detailed Achievability Analysis

In this appendix, we finish the discussion about the policy feasibility in Section V-A. The feasibility for Policy $\Gamma_{s_1,0}$ and Policy $\Gamma_{s_1,1}$ has been proven in Section V-A. In the following discussion about the space dimensions, we again rely on the first order, expectation-based analysis and assume the application of the law of large numbers implicitly.

Policy $\Gamma_{s_1,2}$**:** Similar to the analysis for Policy $\Gamma_{s_1,1}$, assuming $q \ge 2$, the condition that (28) being non-empty is equivalent to whether the following dimension-based inequality is satisfied.

$$\begin{aligned}&\mathsf{dim}(S_2) - \mathsf{dim}(S_2 \cap (S_1 \oplus S_r))\\ =&\mathsf{dim}(S_1 \oplus S_2 \oplus S_r) - \mathsf{dim}(S_1 \oplus S_r) > 0, \end{aligned} \tag{122}$$

where (122) follows from Lemma 2.

Similar to the discussion in $\Gamma_{s_1,0}$ and $\Gamma_{s_1,1}$, we will quantify individual dimension at the end of $\Gamma_{s_1,2}$, the policy of interest, and prove that even in the end of $\Gamma_{s_1,2}$, the dimension difference in (122) is strictly larger than 0. Therefore, throughout the entire duration of $\Gamma_{s_1,2}$, (122) is larger than 0 and $\Gamma_{s_1,2}$ is always feasible.

We first focus on $\mathsf{dim}(S_1 \oplus S_2 \oplus S_r)$. Since $S_1 \oplus S_2 \oplus S_r$ is a subset of the exclusion set in $\Gamma_{s_1,0}$, every time a $\Gamma_{s_1,0}$ packet is received by one of d_1, d_2, and r, $\mathsf{dim}(S_1 \oplus S_2 \oplus S_r)$ will increase by one. On the other hand, notice that $S_1 \oplus S_2 \oplus S_r$ is a superset of the inclusion set in $\Gamma_{s_1,1}$ and $\Gamma_{s_1,2}$. Hence $\mathsf{dim}(S_1 \oplus S_2 \oplus S_r)$ remains the same throughout $\Gamma_{s_1,1}$ and $\Gamma_{s_1,2}$. As a result, in the end of policy $\Gamma_{s_1,2}$, we have

$$\mathsf{E}\{\mathsf{dim}(S_1 \oplus S_2 \oplus S_r)\} = n\omega^0_{s_1} p_1(d_1,d_2,r). \tag{123}$$

We now focus on $\mathsf{dim}(S_1 \oplus S_r)$. Since $S_1 \oplus S_r$ is a subset of the exclusion sets of $\Gamma_{s_1,0}$, $\Gamma_{s_1,1}$ and $\Gamma_{s_1,2}$, every time a packet of $\Gamma_{s_1,0}$, $\Gamma_{s_1,1}$, or $\Gamma_{s_1,2}$ is received by one of d_1 and r, $\mathsf{dim}(S_1 \oplus S_r)$ will increase by one. As a result, in the end of policy $\Gamma_{s_1,2}$, we have

$$\mathsf{E}\{\mathsf{dim}(S_1 \oplus S_r)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1}) p_1(d_1,r). \tag{124}$$

Jointly, (123), (124), and (20) imply (122) in the end of $\Gamma_{s_1,2}$.

Policy $\Gamma_{s_1,3}$**:** Similar to the analysis of the previous policies, assuming $q \ge 2$, the condition that (29) being non-empty is equivalent to whether the following dimension-based inequality is satisfied.

$$\begin{aligned}&\mathsf{dim}(S_r) - \mathsf{dim}(((S_2 \cap S_r) \oplus S_1) \cap S_r)\\ =&\mathsf{dim}((S_2 \cap S_r) \oplus S_1 \oplus S_r) - \mathsf{dim}((S_2 \cap S_r) \oplus S_1) && (125)\\ =&\mathsf{dim}(S_1 \oplus S_r) - (\mathsf{dim}(S_1) + \mathsf{dim}(S_2 \cap S_r)\\ &- \mathsf{dim}(S_1 \cap S_2 \cap S_r)) && (126)\\ =&\mathsf{dim}(S_1 \oplus S_r) - \mathsf{dim}(S_1) - (\mathsf{dim}(S_2) + \mathsf{dim}(S_r)\\ &- \mathsf{dim}(S_2 \oplus S_r)) + \mathsf{dim}(S_1 \cap S_2 \cap S_r) > 0, && (127)\end{aligned}$$

where (125) follows from Lemma 2; (126) follows from simple set operations and from Lemma 2; and (127) follows from Lemma 2.

Similar to the previous discussion, we will quantify individual dimension at the end of $\Gamma_{s_1,3}$, the policy of interest and prove that even in the end of $\Gamma_{s_1,3}$, the dimension difference in (127) is strictly larger than 0. Therefore, throughout the entire duration of $\Gamma_{s_1,3}$, (127) is larger than 0 and $\Gamma_{s_1,3}$ is always feasible.

By similar analysis,[14] in the end of $\Gamma_{s_1,3}$ we have

$$\begin{aligned}&\mathsf{E}\{\mathsf{dim}(S_1)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1} + \omega^3_{s_1}) p_1(d_1), && (128)\\ &\mathsf{E}\{\mathsf{dim}(S_2)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^3_{s_1}) p_1(d_2), && (129)\\ &\mathsf{E}\{\mathsf{dim}(S_r)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1}) p_1(r), && (130)\\ &\mathsf{E}\{\mathsf{dim}(S_1 \oplus S_r)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1}) p_1(d_1,r), && (131)\\ &\mathsf{E}\{\mathsf{dim}(S_2 \oplus S_r)\} = n\omega^0_{s_1} p_1(d_2,r). && (132)\end{aligned}$$

What remains to be decided is the value of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ at the end of Policy $\Gamma_{s_1,3}$. To proceed, we introduce an auxiliary node a in the following way. Whenever a vector $\mathbf{v}$ sent by s_1 is received by both d_1 and r, we let the auxiliary node a observe such $\mathbf{v}$ as well. The knowledge space of a, denoted by S_a is thus the linear span of all vectors received by both d_1 and r.

We first argue that $S_a = S_1 \cap S_r$ in the end of policy $\Gamma_{s_1,2}$. Since a only observes those vectors commonly available at both d_1 and r, the knowledge space of S_a is a subset of $S_1 \cap S_r$. Knowing $S_a \subseteq S_1 \cap S_r$, we can quickly check that S_a is a subset of the exclusion sets in Policies $\Gamma_{s_1,0}$, $\Gamma_{s_1,1}$, and $\Gamma_{s_1,2}$. Therefore, every time node a receives a packet during policies $\Gamma_{s_1,0}$, $\Gamma_{s_1,1}$, and $\Gamma_{s_1,2}$, the dimension of S_a will increase by one. Therefore, we have

$$\mathsf{E}\{\mathsf{dim}(S_a)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1}) p_1(d_1 r) \tag{133}$$

in the end of $\Gamma_{s_1,2}$. On the other hand, by similar analysis as before, we have

$$\begin{aligned}&\mathsf{E}\{\mathsf{dim}(S_1)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1}) p_1(d_1),\\ &\mathsf{E}\{\mathsf{dim}(S_r)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1}) p_1(r),\\ &\mathsf{E}\{\mathsf{dim}(S_1 \oplus S_r)\} = n(\omega^0_{s_1} + \omega^1_{s_1} + \omega^2_{s_1}) p_1(d_1,r),\end{aligned}$$

[13] As a result, $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$.

[14] The derivation of (129) for the case of Policy $\Gamma_{s_1,3}$ uses the following inequality as well.

$$(29) \subseteq (S_r \backslash (S_2 \cap S_r)) = (S_r \backslash S_2).$$

in the end of policy $\Gamma_{s_1,2}$. By Lemma 2, we thus have $\mathsf{dim}(S_a) = \mathsf{dim}(S_1 \cap S_r)$. As a result, we have proven $S_a = (S_1 \cap S_r)$ in the end of $\Gamma_{s_1,2}$.

By the above analysis, we thus have $(S_1 \cap S_2 \cap S_r) = S_a \cap S_2$. By similarly dimension-based analysis, in the end of $\Gamma_{s_1,2}$ we have

$$\mathsf{E}\{\mathsf{dim}(S_2)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1)p_1(d_2), \tag{134}$$

$$\mathsf{E}\{\mathsf{dim}(S_2 \oplus S_a)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1)p_1(d_2, d_1 r), \tag{135}$$

where $p_1(d_2, d_1 r)$ in (135) is the probability that at least one of node d_2 and node a receives the packet and (135) follows from the observation that $S_2 \oplus S_a$ is a subset of the exclusion sets of $\Gamma_{s_1,0}$, $\Gamma_{s_1,1}$ and is a superset of the inclusion set of $\Gamma_{s_1,2}$. By (133), (134), and (135), we have thus proven that

$$\begin{aligned}&\mathsf{E}\{\mathsf{dim}(S_1 \cap S_2 \cap S_r)\} = \mathsf{E}\{\mathsf{dim}(S_a \cap S_2)\} \\ =&\mathsf{E}\{\mathsf{dim}(S_2)\} + \mathsf{E}\{\mathsf{dim}(S_a)\} - \mathsf{E}\{\mathsf{dim}(S_2 \oplus S_a)\} \\ =&n\left(\omega_{s_1}^0 + \omega_{s_1}^1\right)p_1(d_1 d_2 r) + n\omega_{s_1}^2 p_1(d_1 r)\end{aligned} \tag{136}$$

in the end of $\Gamma_{s_1,2}$.

In the following, we will quantify the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{s_1,3}$. To that end, we introduce two more auxiliary nodes b and c. In the beginning of $\Gamma_{s_1,3}$, we let node b (resp. c) be aware of the knowledge space $S_1 \cap S_r$ (resp. $S_2 \cap S_r$). During $\Gamma_{s_1,3}$, whenever a packet is received by d_1 (resp. d_2), we let the auxiliary node b (resp. c) observe such a packet as well. From the construction, it is clear that the following equalities hold in the beginning of $\Gamma_{s_1,3}$.

$$S_b = S_1 \cap S_r, \tag{137}$$

$$S_c = S_2 \cap S_r. \tag{138}$$

We will prove that (137) and (138) hold even in the end of $\Gamma_{s_1,3}$ as well.

In the following, we will prove that (137) holds in the end of $\Gamma_{s_1,3}$. We first note that by our construction, we always have $S_1 \supset S_b \supset (S_1 \cap S_r)$. Knowing that S_b is always a subset of S_1, and S_1 is a subset of the exclusion sets in $\Gamma_{s_1,3}$, we can see that every time d_1 receives a packet during policy $\Gamma_{s_1,3}$, $\mathsf{dim}(S_b)$ will increase by one. Moreover, only when d_1 receives a packet during policy $\Gamma_{s_1,3}$ will $\mathsf{dim}(S_b)$ increase. As a result, the increment of $\mathsf{dim}(S_b)$ during $\Gamma_{s_1,3}$ equals the number of times d_1 receives a packet during $\Gamma_{s_1,3}$. On the other hand, $\mathsf{dim}(S_1 \cap S_r) = \mathsf{dim}(S_1) + \mathsf{dim}(S_r) - \mathsf{dim}(S_1 \oplus S_r)$. Since both S_r and $S_1 \oplus S_r$ are supersets of the inclusion set of $\Gamma_{s_1,3}$, both $\mathsf{dim}(S_r)$ and $\mathsf{dim}(S_1 \oplus S_r)$ remain identical during $\Gamma_{s_1,3}$. Therefore, the increment of $\mathsf{dim}(S_1 \cap S_r)$ is identical to the increment of $\mathsf{dim}(S_1)$ during $\Gamma_{s_1,3}$. As a result, the increment of $\mathsf{dim}(S_1 \cap S_r)$ during $\Gamma_{s_1,3}$ equals the number of times d_1 receives a packet during $\Gamma_{s_1,3}$. We have thus proven $\mathsf{dim}(S_b) = \mathsf{dim}(S_1 \cap S_r)$ in the end of $\Gamma_{s_1,3}$, which implies (137). (138) can be proven by symmetry.

To quantify the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{s_1,3}$, we notice that $\mathsf{dim}(S_1 \cap S_2 \cap S_r) = \mathsf{dim}(S_b \cap S_c) = \mathsf{dim}(S_b) + \mathsf{dim}(S_c) - \mathsf{dim}(S_b \oplus S_c)$. As a result, the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during policy $\Gamma_{s_1,3}$ is the summation of the increments of $\mathsf{dim}(S_b)$ and $\mathsf{dim}(S_c)$ minus the increment of $\mathsf{dim}(S_b \oplus S_c)$ during $\Gamma_{s_1,3}$. By our construction, the increments of S_b, S_c, and $S_b \oplus S_c$ during $\Gamma_{s_1,3}$ are simply $n\omega_{s_1}^3 p_1(d_1)$, $n\omega_{s_1}^3 p_1(d_2)$, and $n\omega_{s_1}^3 p_1(d_1, d_2)$, respectively. As a result, the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{s_1,3}$ is simply $n\omega_{s_1}^3 p_1(d_1 d_2)$.

Combining (136), we have thus proven that

$$\begin{aligned}&\mathsf{E}\{\mathsf{dim}(S_1 \cap S_2 \cap S_r)\} \\ =&n\left(\omega_{s_1}^0 + \omega_{s_1}^1\right)p_1(d_1 d_2 r) + n\omega_{s_1}^2 p_1(d_1 r) + n\omega_{s_1}^3 p_1(d_1 d_2)\end{aligned} \tag{139}$$

in the end of Policy $\Gamma_{s_1,3}$.

Jointly, (128) to (132), (139), and (21) imply (127) in the end of $\Gamma_{s_1,3}$.

Policy $\Gamma_{s_1,4}$: Similar to the analysis of the previous policies, the condition that (30) being non-empty is equivalent to whether the following dimension-based inequality is satisfied in the end of $\Gamma_{s_1,4}$.

$$\begin{aligned}&\mathsf{dim}(S_2 \cap S_r) - \mathsf{dim}(S_1 \cap S_2 \cap S_2) \\ =&(\mathsf{dim}(S_2) + \mathsf{dim}(S_r) - \mathsf{dim}(S_2 \oplus S_r)) \\ &- \mathsf{dim}(S_1 \cap S_2 \cap S_r) > 0.\end{aligned} \tag{140}$$

Similar to the previous discussion, we will quantify individual dimension at the end of $\Gamma_{s_1,4}$ and prove that (140) holds in the end of $\Gamma_{s_1,4}$.

By similar analysis, we have

$$\mathsf{E}\{\mathsf{dim}(S_2)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^3)p_1(d_2), \tag{141}$$

$$\mathsf{E}\{\mathsf{dim}(S_r)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^2)p_1(r), \tag{142}$$

$$\mathsf{E}\{\mathsf{dim}(S_2 \oplus S_r)\} = n\omega_{s_1}^0 p_1(d_2, r), \tag{143}$$

in the end of $\Gamma_{s_1,4}$. What remains to be decided is the value of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ at the end of Policy $\Gamma_{s_1,4}$. In (139), we have already quantified $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ in the end of $\Gamma_{s_1,3}$. In the following, we will quantify the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{s_1,4}$. By (30), we can see that every time d_1 receives a packet during $\Gamma_{s_1,4}$, $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ will increase by one. As a result, the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{s_1,4}$ is $n\omega_{s_1}^4 p_1(d_1)$. Together with (139), we have proven that

$$\begin{aligned}&\mathsf{E}\{\mathsf{dim}(S_1 \cap S_2 \cap S_r)\} \\ =&n\left(\omega_{s_1}^0 + \omega_{s_1}^1\right)p_1(d_1 d_2 r) + n\omega_{s_1}^2 p_1(d_1 r) \\ &+ n\omega_{s_1}^3 p_1(d_1 d_2) + n\omega_{s_1}^4 p_1(d_1)\end{aligned} \tag{144}$$

in the end of $\Gamma_{s_1,4}$. Jointly, (141) to (144) and (22) imply that (140) holds in the end of $\Gamma_{s_1,4}$.

The feasibility of policy $\Gamma_{s_2,k}$, $k = 0, 1, 2, 3, 4$, can be proven by symmetry.

Policy $\Gamma_{r,1}$: We first notice that the inclusion space and exclusion space of Policy $\Gamma_{r,1}$ are the same as of Policy $\Gamma_{s_1,3}$. Hence to prove the feasibility of Policy $\Gamma_{r,1}$, we need to prove that (127) holds in the end of $\Gamma_{r,1}$. By similar analysis, we have

$$\mathsf{E}\{\mathsf{dim}(S_1)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^2 + \omega_{s_1}^3 + \omega_{s_1}^4)p_1(d_1) + n\omega_{r,\mathrm{N}}^1 p_r(d_1), \tag{145}$$

$$\mathsf{E}\{\mathsf{dim}(S_2)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^3)p_1(d_2) + n\omega_{r,\mathrm{N}}^1 p_r(d_2), \tag{146}$$

$$\mathsf{E}\{\mathsf{dim}(S_r)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^2)p_1(r), \tag{147}$$

$$\mathsf{E}\{\mathsf{dim}(S_1 \oplus S_r)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^2)p_1(d_1, r), \tag{148}$$

$$\mathsf{E}\{\mathsf{dim}(S_2 \oplus S_r)\} = n\omega_{s_1}^0 p_1(d_2, r), \tag{149}$$

in the end of $\Gamma_{r,1}$.

What remains to be decided is the value of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ at the end of Policy $\Gamma_{r,1}$. In (144) we have computed the value of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ in the end of $\Gamma_{s_1,4}$. As a result, we only need to quantify the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{r,1}$. By the same analysis as used when we quantify the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{s_1,3}$, the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{r,1}$ is $n\omega_{r,\mathrm{N}}^1 p_r(d_1 d_2)$. By (144), we have shown that

$$\begin{aligned}
&\mathsf{E}\{\mathsf{dim}(S_1 \cap S_2 \cap S_r)\} \\
=&n\left(\omega_{s_1}^0 + \omega_{s_1}^1\right)p_1(d_1 d_2 r) + n\omega_{s_1}^2 p_1(d_1 r) \\
&+ n\omega_{s_1}^3 p_1(d_1 d_2) + n\omega_{s_1}^4 p_1(d_1) + n\omega_{r,\mathrm{N}}^1 p_r(d_1 d_2)
\end{aligned} \tag{150}$$

in the end of $\Gamma_{r,1}$. Jointly, (145) to (150) and (23) imply that (127) holds in the end of $\Gamma_{r,1}$.

The discussion of Policy $\Gamma_{r,2}$ follows symmetrically.

Policy $\Gamma_{r,3}$ for $\mathbf{v}^{(1)}$: We will prove that for the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$, we can always choose $\mathbf{v}^{(1)}$ according to (38). To that end, we first notice that the inclusion space and exclusion space in (38) are the same as those of Policy $\Gamma_{s_1,4}$. Hence to prove that (38) remains non-empty during the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$, we need to prove that (140) holds in the end of the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$. By similar analysis as used in the previous policies, we have

$$\mathsf{E}\{\mathsf{dim}(S_2)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^3)p_1(d_2) + n\omega_{r,\mathrm{N}}^1 p_r(d_2), \tag{151}$$

$$\mathsf{E}\{\mathsf{dim}(S_r)\} = n(\omega_{s_1}^0 + \omega_{s_1}^1 + \omega_{s_1}^2)p_1(r), \tag{152}$$

$$\mathsf{E}\{\mathsf{dim}(S_2 \oplus S_r)\} = n\omega_{s_1}^0 p_1(d_2, r), \tag{153}$$

in the end of the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$. What remains to be decided is the value of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ at the end of the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$. In (150) we have computed the value of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ in the end of $\Gamma_{r,1}$. As a result, we only need to quantify the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$. By the same analysis as used when we quantify the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during $\Gamma_{s_1,4}$, the increment of $\mathsf{dim}(S_1 \cap S_2 \cap S_r)$ during the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$ is $n\omega_{r,\mathrm{C}}^1 p_r(d_1)$. By (150), we have shown that

$$\begin{aligned}
&\mathsf{E}\{\mathsf{dim}(S_1 \cap S_2 \cap S_r)\} \\
=&n\left(\omega_{s_1}^0 + \omega_{s_1}^1\right)p_1(d_1 d_2 r) + n\omega_{s_1}^2 p_1(d_1 r) \\
&+ n\omega_{s_1}^3 p_1(d_1 d_2) + n\omega_{s_1}^4 p_1(d_1) + n\omega_{r,\mathrm{N}}^1 p_r(d_1 d_2) \\
&+ n\omega_{r,\mathrm{C}}^1 p_r(d_1)
\end{aligned} \tag{154}$$

in the end of the first $n\omega_{r,\mathrm{C}}^1$ time slots of Policy $\Gamma_{r,3}$. Jointly, (151) to (154) and (24) imply that (140) holds in the end of the first $n\omega_{r,\mathrm{C}}^1$ time slots of $\Gamma_{r,3}$.

The discussion of the first $n\omega_{r,\mathrm{C}}^2$ time slots of $\Gamma_{r,3}$ follows symmetrically.

The above analysis completes the achievability proof stated in Section V-A.

C. The Proof of Lemma 3

In this appendix, we will show that the assignment of $\omega_{s_i}^0$ to $\omega_{r,\mathrm{C}}^i$ in (62) to (68) is non-negative and satisfies (14), (15), (54) to (61).

We first show that the $\omega_{s_i}^0$ to $\omega_{r,\mathrm{C}}^i$ are non-negative. By the definitions, $\omega_{s_i}^0$, $\omega_{s_i}^2$, $\omega_{s_i}^4$, $\omega_{r,\mathrm{N}}^i$ are non-negative. Since $\max\{p_i(r), p_i(d_j)\} \leq p_i(r, d_j)$, $\omega_{s_i}^1$ is non-negative. By (7), we have $R_i \leq t_{s_i} p_i(r)$. Therefore, $\omega_{s_i}^3$ is non-negative. There are two terms in the definition of $\omega_{r,\mathrm{C}}^i$ in (68). The numerator of the first term is no less than the numerator of the second term. The denominator of the first term is no larger than the denominator of the second term. As a result, $\omega_{r,\mathrm{C}}^i$ is non-negative.

By (62) to (65), we have

$$\begin{aligned}
\sum_{k=0}^{3} \omega_{s_i}^k =& \frac{R_i}{p_i(d_j, r)} \\
&+ R_i\left(\min\left\{\frac{1}{p_i(r)}, \frac{1}{p_i(d_j)}\right\} - \frac{1}{p_i(d_j, r)}\right) \\
&+ R_i\left(\frac{1}{p_i(r)} - \frac{1}{p_i(d_j)}\right)^+ \\
&+ \min\left\{R_i\left(\frac{1}{p_i(d_j)} - \frac{1}{p_i(r)}\right)^+, t_{s_i} - \frac{R_i}{p_i(r)}\right\}
\end{aligned} \tag{155}$$

$$\begin{aligned}
=&\frac{R_i}{p_i(r)} \\
&+ \min\left\{R_i\left(\frac{1}{p_i(d_j)} - \frac{1}{p_i(r)}\right)^+, t_{s_i} - \frac{R_i}{p_i(r)}\right\}
\end{aligned} \tag{156}$$

$$\leq t_{s_i}, \tag{157}$$

where (156) is based on the fact that $\min(x, y) + (x - y)^+ = x$ for arbitrary real valued x and y; and (157) follows from the fact that the minimum of two values is no more than any of two. (157) thus shows that our assignment satisfies (14).

To prove that (15) holds, we observe that our assignment leads to

$$\begin{aligned}
&\omega_{r,\mathrm{N}}^i + \omega_{r,\mathrm{N}}^j + \omega_{r,\mathrm{C}}^i \\
=&\frac{(R_i - t_{s_i} p_i(d_j))^+}{p_r(d_i, d_j)} + \frac{(R_j - t_{s_j} p_{s_j}(d_i))^+}{p_r(d_i, d_j)} \\
&+ \frac{R_i}{p_r(d_i)} - \frac{(R_i - t_{s_i} p_i(d_j))^+}{p_r(d_i, d_j)} \\
=&\frac{R_i}{p_r(d_i)} + \frac{(R_j - t_{s_j} p_{s_j}(d_i))^+}{p_r(d_i, d_j)} \leq t_r.
\end{aligned}$$

Hence the assignment satisfies (15). By noticing that

$$\omega_{s_i}^0 p_i(d_j,r) = \frac{R_i}{p_i(d_j,r)} p_i(d_j,r) = R_i,$$

we have shown that the assignment satisfies (54). We now consider (55) and notice that

$$\begin{aligned}
\text{LHS of (55)} =& R_i \left(\min\left\{ \frac{1}{p_i(r)}, \frac{1}{p_i(d_j)} \right\} - \frac{1}{p_i(d_j,r)} \right) p_i(r) \\
=& R_i \left(\min\left\{ 1, \frac{p_i(r)}{p_i(d_j)} \right\} - \frac{p_i(r)}{p_i(d_j,r)} \right) \\
\leq& R_i - R_i \frac{p_i(r)}{p_i(d_j,r)}, \qquad (158)\\
\text{RHS of (55)} =& R_i \frac{p_i(d_j\overline{r})}{p_i(d_j,r)} = R_i - R_i \frac{p_i(r)}{p_i(d_j,r)}, \qquad (159)
\end{aligned}$$

where (158) follows from the definition of the minimum, and (159) follows from $p_i(d_j,r) - p_i(r) = p_i(d_j\overline{r})$. Since with the assignment, the left hand side (LHS) of (55) is no larger than the right hand side (RHS) of (55), the assignment satisfies (55). We now consider (56) and notice that

$$\begin{aligned}
\text{LHS of (56)} =& R_i \left(\min\left\{ \frac{1}{p_i(r)}, \frac{1}{p_i(d_j)} \right\} - \frac{1}{p_i(d_j,r)} \right) p_i(d_j) \\
=& R_i \left(\min\left\{ \frac{p_i(d_j)}{p_i(r)}, 1 \right\} - \frac{p_i(d_j)}{p_i(d_j,r)} \right) \\
\leq& R_i - R_i \frac{p_i(d_j)}{p_i(d_j,r)}, \qquad (160)\\
\text{RHS of (56)} =& R_i \frac{p_i(r\overline{d_j})}{p_i(d_j,r)} = R_i - R_i \frac{p_i(d_j)}{p_i(d_j,r)}, \qquad (161)
\end{aligned}$$

where (160) follows from the definition of the minimum, and (161) follows from $p_i(d_j,r) - p_i(d_j) = p_i(r\overline{d_j})$. Since with the assignment, the LHS of (56) is no larger than the RHS of (56), the assignment satisfies (56). We now consider (57) and notice that

$$\begin{aligned}
\text{LHS of (57)} =& R_i \left(\frac{1}{p_i(r)} - \frac{1}{p_i(d_j)} \right)^+ p_i(r) \\
=& R_i \left(1 - \frac{p_i(r)}{p_i(d_j)} \right)^+, \\
\text{RHS of (57)} =& R_i \frac{p_i(d_j\overline{r})}{p_i(d_j,r)} \\
&- R_i \left(\min\left\{ \frac{1}{p_i(r)}, \frac{1}{p_i(d_j)} \right\} - \frac{1}{p_i(d_j,r)} \right) p_i(r) \\
=& R_i \left(1 - \frac{p_i(r)}{p_i(d_j,r)} \right) \\
&- R_i \left(\min\left\{ 1, \frac{p_i(r)}{p_i(d_j)} \right\} - \frac{p_i(r)}{p_i(d_j,r)} \right) \qquad (162)\\
=& R_i - R_i \min\left\{ 1, \frac{p_i(r)}{p_i(d_j)} \right\} \\
=& R_i \left(1 - \frac{p_i(r)}{p_i(d_j)} \right)^+, \qquad (163)
\end{aligned}$$

where (162) follows from $p_i(d_j,r) - p_i(r) = p_i(d_j\overline{r})$, and (163) follows from $1 - \min\left\{\frac{p_i(r)}{p_i(d_j)}, 1\right\} = \left(1 - \frac{p_i(r)}{p_i(d_j)}\right)^+$. Since with the assignment, the LHS of (57) is no larger than the RHS of (57), the assignment satisfies (57). We now consider (58) and notice that

$$\begin{aligned}
\text{LHS of (58)} =& \min\left\{ R_i \left(1 - \frac{p_i(d_j)}{p_i(r)} \right)^+, \right. \\
&\left. t_{s_i} p_i(d_j) - R_i \frac{p_i(d_j)}{p_i(r)} \right\}, \\
\text{RHS of (58)} =& R_i \frac{p_i(r\overline{d_j})}{p_i(d_j,r)} \\
&- R_i \left(\min\left\{ \frac{1}{p_i(r)}, \frac{1}{p_i(d_j)} \right\} - \frac{1}{p_i(d_j,r)} \right) p_i(d_j) \\
=& R_i \left(1 - \frac{p_i(d_j)}{p_i(d_j,r)} \right) \\
&- R_i \left(\min\left\{ \frac{p_i(d_j)}{p_i(r)}, 1 \right\} - \frac{p_i(d_j)}{p_i(d_j,r)} \right) \qquad (164)\\
=& R_i - R_i \min\left\{ \frac{p_i(d_j)}{p_i(r)}, 1 \right\} \\
=& R_i \left(1 - \frac{p_i(d_j)}{p_i(r)} \right)^+. \qquad (165)
\end{aligned}$$

where (164) and (165) follow similar reasons as in (162) and (163). Hence the assignment satisfies (58). We now consider (59) and we notice that

$$\begin{aligned}
\text{LHS} =& \left(R_i - t_{s_i} p_i(d_j) \right)^+, \\
\text{RHS} =& R_i \frac{p_i(r\overline{d_j})}{p_i(d_j,r)} \\
&- R_i \left(\min\left\{ \frac{1}{p_i(r)}, \frac{1}{p_i(d_j)} \right\} - \frac{1}{p_i(d_j,r)} \right) p_i(d_j) \\
&- \min\left\{ R_i \left(1 - \frac{p_i(d_j)}{p_i(r)} \right)^+, t_{s_i} p_i(d_j) - R_i \frac{p_i(d_j)}{p_i(r)} \right\} \\
=& R_i \left(1 - \frac{p_i(d_j)}{p_i(d_j,r)} \right) \\
&- R_i \left(\min\left\{ \frac{p_i(d_j)}{p_i(r)}, 1 \right\} - \frac{p_i(d_j)}{p_i(d_j,r)} \right) \\
&- \min\left\{ R_i \left(1 - \frac{p_i(d_j)}{p_i(r)} \right)^+, t_{s_i} p_i(d_j) - R_i \frac{p_i(d_j)}{p_i(r)} \right\} \qquad (166)\\
=& R_i - R_i \min\left\{ \frac{p_i(d_j)}{p_i(r)}, 1 \right\} \\
&- \min\left\{ R_i \left(1 - \frac{p_i(d_j)}{p_i(r)} \right)^+, t_{s_i} p_i(d_j) - R_i \frac{p_i(d_j)}{p_i(r)} \right\} \\
=& R_i \left(1 - \frac{p_i(d_j)}{p_i(r)} \right)^+ \\
&- \min\left\{ R_i \left(1 - \frac{p_i(d_j)}{p_i(r)} \right)^+, t_{s_i} p_i(d_j) - R_i \frac{p_i(d_j)}{p_i(r)} \right\} \qquad (167)\\
\geq& \left(R_i - t_{s_i} p_i(d_j) \right)^+, \qquad (168)
\end{aligned}$$

where (166) follows from $p_i(d_j,r) - p_i(d_j) = p_i(r\overline{d_j})$; and

(167) follows from the definition of $(\cdot)^+$. (168) follows from the following arguments.

$$\begin{aligned}(167) =& \max\left\{0, R_i\left(1-\frac{p_i(d_j)}{p_i(r)}\right)^+ - t_{s_i}p_i(d_j) + R_i\frac{p_i(d_j)}{p_i(r)}\right\}\\ \geq& \max\left\{0, R_i\left(1-\frac{p_i(d_j)}{p_i(r)}\right) - t_{s_i}p_i(d_j) + R_i\frac{p_i(d_j)}{p_i(r)}\right\}\\ =&(168).\end{aligned}$$

Since with the assignment, the LHS of (59) is no greater than the RHS of (59), the assignment satisfies (59). We now consider (60) and notice that

$$\text{LHS} = R_i - (R_i - t_{s_i}p_i(d_j))^+\frac{p_r(d_i)}{p_r(d_i,d_j)},$$

$$\begin{aligned}\text{RHS} =& R_i\frac{p_i(d_jr)}{p_i(d_j,r)}\\ &+ (p_i(d_j)+p_i(r))\\ &\cdot R_i\left(\min\left\{\frac{1}{p_i(r)},\frac{1}{p_i(d_j)}\right\} - \frac{1}{p_i(d_j,r)}\right)\\ &+ R_i\left(1-\frac{p_i(r)}{p_i(d_j)}\right)^+\\ &+ \min\left\{R_i\left(1-\frac{p_i(d_j)}{p_i(r)}\right)^+, t_{s_i}p_i(d_j) - R_i\frac{p_i(d_j)}{p_i(r)}\right\}\\ &+ \frac{(R_i - t_{s_i}p_i(d_j))^+}{p_r(d_i,d_j)}(p_r(d_i,d_j)-p_r(d_i))\\ =& R_i\left(\frac{p_i(d_jr)-p_i(d_j)-p_i(r)}{p_i(d_j,r)}\right)\\ &+ R_i\min\left\{1+\frac{p_i(d_j)}{p_i(r)}, 1+\frac{p_i(r)}{p_i(d_j)}\right\}\\ &+ R_i\left(1-\frac{p_i(r)}{p_i(d_j)}\right)^+\\ &+ \min\left\{R_i\left(1-\frac{p_i(d_j)}{p_i(r)}\right)^+, t_{s_i}p_i(d_j) - R_i\frac{p_i(d_j)}{p_i(r)}\right\}\\ &+ \frac{(R_i - t_{s_i}p_i(d_j))^+}{p_r(d_i,d_j)}(p_r(d_i,d_j)-p_r(d_i))\\ =& R_i\min\left\{\frac{p_i(d_j)}{p_i(r)},\frac{p_i(r)}{p_i(d_j)}\right\} + R_i\left(1-\frac{p_i(r)}{p_i(d_j)}\right)^+\\ &+ \min\left\{R_i\left(1-\frac{p_i(d_j)}{p_i(r)}\right)^+, t_{s_i}p_i(d_j) - R_i\frac{p_i(d_j)}{p_i(r)}\right\}\\ &+ (R_i - t_{s_i}p_i(d_j))^+\left(1-\frac{p_r(d_i)}{p_r(d_i,d_j)}\right), \qquad (169)\end{aligned}$$

where (169) follows from $p_i(d_jr) - p_i(d_j) - p_i(r) = -p_i(d_j,r)$. We now consider two cases. Case 1: $p_i(d_j) \geq p_i(r)$. In this case, we can rewrite (169) as follows.

$$\begin{aligned}(169) =& R_i\frac{p_i(r)}{p_i(d_j)} + R_i\left(1-\frac{p_i(r)}{p_i(d_j)}\right) + 0\\ &+ (R_i - t_{s_i}p_i(d_j))^+\left(1-\frac{p_r(d_i)}{p_r(d_i,d_j)}\right)\\ \geq& R_i - (R_i - t_{s_i}p_i(d_j))^+\frac{p_r(d_i)}{p_r(d_i,d_j)}. \qquad (170)\end{aligned}$$

Case 2: $p_i(d_j) < p_i(r)$. In this case, we can rewrite (169) as follows.

$$\begin{aligned}(169) =& R_i\frac{p_i(d_j)}{p_i(r)} + 0\\ &+ \min\left\{R_i\left(1-\frac{p_i(d_j)}{p_i(r)}\right), t_{s_i}p_i(d_j) - R_i\frac{p_i(d_j)}{p_i(r)}\right\}\\ &+ (R_i - t_{s_i}p_i(d_j))^+(1-\frac{p_r(d_i)}{p_r(d_i,d_j)})\\ =& \min\{R_i, t_{s_i}p_i(d_j)\}\\ &+ (R_i - t_{s_i}p_i(d_j))^+\left(1-\frac{p_r(d_i)}{p_r(d_i,d_j)}\right)\\ =& R_i - (R_i - t_{s_i}p_i(d_j))^+\frac{p_r(d_i)}{p_r(d_i,d_j)}, \qquad (171)\end{aligned}$$

where (171) follows from the equality $\min\{x,y\}+(x-y)^+ = x$ for any real valued x, y. Plugging (171) and (170) into (169), we thus see that the LHS of (60) is no larger than the RHS of (60) with the assignment. Hence the assignment satisfies (60). We now consider (61) and notice that

$$p_r(d_i)(\omega^i_{r,\mathrm{N}} + \omega^i_{r,\mathrm{C}}) = R_i$$

which satisfies (61).

The above proof shows that the proposed assignment satisfies (14), (15), (54) to (61).

References

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

[2] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: Practical wireless network," in *Proc. ACM Special Interest Group on Data Commun. (SIGCOMM)*, 2006.

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

[4] A. Eryilmaz, D. Lun, and B. Swapna, "Control of multi-hop communication networks for inter-session network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 1092–1110, Feb. 2011.

[5] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, July 1984.

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

[7] C.-C. Wang, "On the capacity of 1-to-K broadcast packet erasure channels with channel output feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 931–956, Feb 2012.

[8] P. Chaporkar and A. Proutiere, "Adaptive network coding and scheduling for maximizing throughput in wireless networks," in *Proc. Ann. ACM Int. Conf. on Mobile Computing & Networking (MobiCom)*, Montreal, QC, Canada, Sept 2007, pp. 135–146.

[9] T. Cui, L. Chen, and T. Ho, "Energy efficient opportunistic network coding for wireless networks," in *Proc. 27th IEEE Conf. Computer Communications (INFOCOM)*, Phoenix, AZ, Apr 2008.

[10] 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. Symp. Inform. Theory*, Saint Petersburg, Russia, Aug 2011.

[11] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, WA, 2006.

[12] K. Salamatian and R. Khalili, "An information theory for erasure channels," in *Proc. 43rd Ann. Allerton Conf. on Comm.*, Monticello, IL, Sept 2005.

[13] C. Suh and K. Ramachandran, "Exact-repair MDS code construction using interference alignment," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1425–1442, Mar 2011.

[14] J. Yoo, T. Liu, and F. Xue, "Gaussian broadcast channels with receiver message side information," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, June 2009.

[15] S. Rayanchu, S. Sen, J. Wu, S. Banerjee, and S. Sengupta, "Loss-aware network coding for unicast wireless sessions: design, implementation, and performance evaluation," in *Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, Annapolis, MD, USA, 2008, pp. 85–96.

[16] S. Sengupta, S. Rayanchu, and S. Banerjee, "An analysis of wireless network coding for unicast sessions: The case for coding-aware routing," in *Proc. 26th IEEE Conf. Computer Communications (INFOCOM)*, Anchorage, Alaska, USA, May 2007, pp. 1028–1036.

[17] H. Seferoglu, A. Markopoulou, and K. K. Ramakrishnan, "I^{2}NC: Intra- and inter-session network coding for unicast flows in wireless networks," in *Proc. 30th IEEE Conf. Computer Communications (INFOCOM)*, Shanghai, China, April 2011, pp. 1035–1043.

[18] J. Korner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, Jan 1997.

[19] G. Kramer and S. Shamai, "Capacity for classes of broadcast channels with receiver side information," in *Proc. IEEE Inform. Theory Workshop*, Lake Tahoe, CA, Sept 2007, pp. 313–318.

[20] B. Smith and B. Hassibi, "Wireless erasure networks with feedback," in *Proc. IEEE Int. Symp. Inform. Theory*, July 2008, pp. 339–343.

[21] C.-C. Wang and N. Shroff, "Beyond the butterfly—a graph-theoretic characterization of the feasibility of network coding with two simple unicast sessions," in *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, June 2007, pp. 121–125.

[22] F. Xue and X. Yang, "Network coding and packet-erasure broadcast channel," in *Proc. 5th IEEE Ann. Communications Society Conf. on Sensor, Mesh and Ad Hoc Communications and Networks Workshops (SECON)*, San Francisco, CA, July 2008.

[23] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," in *Proc. ACM Special Interest Group on Data Commun. (SIGCOMM)*, Kyoto, Japan, Aug 2007.

[24] D. Koutsonikolas, C.-C. Wang, and Y. Hu, "CCACK: Efficient network coding based opportunistic routing through cumulative coded acknowledgments," in *Proc. 29th IEEE Conf. Computer Communications (INFOCOM)*, San Diego, CA, Mar 2010, pp. 1–9.

[25] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, Aug 2006.

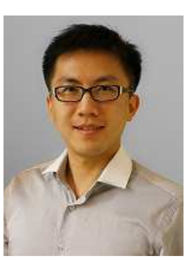
[26] S. Biswas and R. Morris, "ExOR: opportunistic multi-hop routing for wireless networks," vol. 35, no. 4, Oct. 2005, pp. 133–143.

[27] V. Anantharam and S. Verdtu, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, pp. 4–18, 1996.

[28] T. Lutz, G. Kramer, and C. Hausl, "Capacity for half-duplex line networks with two sources," in *Proc. IEEE Int'l Symp. Inform. Theory*, Austin, Texas, USA, June 2010.

[29] A. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, Mar. 2006.

[30] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 379–381, March 1978.



Wei-Cheng Kuo received the B.S. degree in electrical engineering from Nation Chiao Tung University, Hsinchu, Taiwan, in 2008. He is currently working toward the Ph.D. degree in electrical and computer engineering at Purdue University, Indiana, USA. His research interests are in Shannon capacity characterization of wireless network coding and scheduling algorithms for network coded traffic.

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. His current research interests are in the graph-theoretic and algorithmic analysis of iterative decoding and of 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.