On Throughput Optimality With Delayed Network-State Information

Lei Ying, *Member, IEEE*, and Sanjay Shakkottai, *Senior Member, IEEE*

***Abstract*—The problem of routing/scheduling in a wireless network with partial/delayed network (channel and queue) state information (NSI) is studied in this paper. Two cases are considered: (i) centralized routing/scheduling, where a central controller obtains heterogeneously delayed information from each of the nodes (thus, the controller has NSI with different delays from different nodes), and makes routing/scheduling decisions; (ii) decentralized routing/scheduling, where each node makes a decision based on its current channel and queue states along with homogeneous delayed NSI from other nodes. For each of the cases (with additional flow restrictions for the decentralized routing/scheduling case), the optimal network throughput regions are characterized under the above described NSI models and it is shown that the throughput regions shrinks with the increase of delay. Further, channel and queue length based routing/scheduling algorithms that achieve the above throughput regions are proposed in this paper.**

***Index Terms*—Delayed network state information (NSI), routing, scheduling, throughput optimality, wireless networks.**

I. Introduction

The wireless network provides a versatile platform to support a diverse set of applications such as voice, multimedia, data and messaging. These applications can exist over many configurations—citywide mesh networks, cellular deployments for mobile users, and *ad-hoc* battlefield networks. A key component in engineering such networks is the routing/scheduling algorithm. So far, throughput-optimal routing/scheduling algorithms have been developed assuming that *complete* network state information (NSI—the channel and queue state information of the entire network) is instantaneously available [2]–[12]. We refer to [13], [14] for comprehensive surveys. While these algorithms optimize network performance with complete and instantaneous NSI; they could be inefficient in practical deployments because obtaining complete and instantaneous NSI not only incurs significant communication/computation overhead, but also is hard to come by when channels vary rapidly in a dynamic environment. Delayed NSI could ravage the performance of those algorithms requiring complete NSI. In this paper, we consider practical scenarios where only delayed NSI can be explored. We will study the fundamental impact of delayed NSI on network throughput, and develop throughput-optimal routing/scheduling algorithms with delayed NSI.

Notably, there have been recent studies on wireless scheduling with partial/delayed NSI. To the best of our knowledge, the earliest work to consider delayed queue-length information and its impact on stability of back-pressure algorithms is [15]. The decentralized channel-aware ALOHA has been developed in [16]–[18] for uplink networks, where each node transmits based on their own NSI. For downlink networks, the authors in [19] have considered a scenario where the base-station can only access the NSI of a subset of mobiles, and have developed a variant of the MaxWeight rule [3], which is throughput-optimal. Furthermore, the authors in [20]–[24] have taken account of the cost of channel probing, and have studied the trade-off between probing cost and scheduling gain. Joint channel-probing and transmission-scheduling algorithms have been developed to maximize network throughput. In [25], the authors have studied joint routing and scheduling with noisy channel estimates in the context of channels that vary i.i.d. across time. In [26], the authors consider a base-station connected to a collection of mobiles, and study the scheduling problem (uplink and downlink) when the channel and queue states are known periodically. They develop throughput optimal policies where a (matching) decision is made at each slot based on this globally known delayed information.

While partial/delayed NSI has been studied in the prior work, there are two distinguishing features in our study. First, we study the case of a general network topology with *heterogeneous* delays from each of the nodes in the network to the central controller (the decision maker). Second, and perhaps more important, we study the case of decentralized decision making where each node makes a transmission decision based on locally known information, and the information known at each node is different (i.e., different nodes have different "views" of the network). The resulting information inconsistency can cause nodes to make potentially conflicting decisions (e.g., two nodes on the same collision domain may decide to transmit simultaneously because they have different delayed versions of the channel-states and cannot coordinate their decisions with each other instantaneously). We also note that optimal control with information delays has been studied

Manuscript received September 04, 2008; revised August 27, 2010; accepted January 31, 2011. Date of current version July 29, 2011. This work was partially supported by the NSF by Grants CNS-0347400, CNS-0721380, CNS-0831756, CNS-0953165, and CNS-1017549, the DARPA ITMANET program, and the DTRA Grants HDTRA1-08-1-0016 and HDTRA1-09-1-0055. A shorter version of this paper appears in the Proceedings of the Information Theory and Applications Workshop, San Diego, CA, February 2008.

L. Ying is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: leiying@iastate.edu).

S. Shakkottai is with the Department of Electrical and Computer Engineering and the Wireless Communications and Networking Group (WNCG), The University of Texas at Austin, Austin, TX 78712-0240 USA (e-mail: shakkott@ece.utexas.edu).

Communicated by S. Ulukus, Associate Editor for Communication Networks.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2158879

since the classic work of Witsenhausen [34]–[37]; it is known that even simple control problems without information delays become much harder with inconsistent delayed state.

The two cases to be considered in this paper are summarized here:

(i) There is a central controller that selects a subset of links to transmit based on heterogeneous delayed NSI at each time slot.

(ii) There is no central controller, and each node has to make transmission decision based on their own instantaneous NSI and delayed NSI from other nodes. Uplink networks and ad-hoc networks are such examples. In this scenario, on one hand, the nodes have inconsistent views of the network because they have different local NSI; but on the other hand, we need to exploit the local instantaneous NSI since it provides more accurate network information. Developing a consistent policy based on the inconsistent information is the key challenge in this scenario.

In both of these contexts, we study routing/scheduling algorithms with delayed NSI. For the centralized NSI case and decentralized NSI case with unit delay, we both consider multihop traffic flows and study joint routing and scheduling. For the decentralized NSI case with delay larger than one, we only consider single-hop flows so our focus is primarily on scheduling and routing is not considered in this case. The main contributions of this paper include:

(i) We characterize the network throughput regions with delayed NSI for both the centralized and decentralized cases (under additional assumptions on delays and flows, see Section V). We present examples in Sections IV and V where we explicitly compute the throughput-optimal policy. We will see from these examples that the network throughput region is determined by the available NSI, and shrinks with the increase of the delay in NSI.

(ii) We also develop throughput-optimal routing/scheduling algorithms based on the delayed NSI. For the centralized case, the algorithm is a variant of the back-pressure algorithm proposed in [2], which uses the expected channel-states, conditioned on the delayed channel-states, in the routing/scheduling. For the decentralized case, the proposed algorithm contains two parts: each node first calculates a threshold vector based on the global delayed NSI; and then makes a transmission decision based on the local instantaneous NSI and the threshold vector.

We remark that distributed implementations of throughput-optimal routing/scheduling algorithms (e.g., back-pressure algorithm) have been extensively studied recently [27]–[33]. These results, however, assume either channels are time-invariant so that channel states are known, or channels are orthogonalized so that routing/scheduling decisions for different links are decoupled. Therefore, for both cases, the channel state information required for routing/scheduling is known exactly. In this paper, we consider a wireless network with time-varying channels, and assume only delayed NSI is available for routing/scheduling. This paper focuses on understanding the impact of incomplete and inconsistent NSI on throughput; while the distributed implementations of throughput-optimal routing/scheduling algorithms mainly focus on reducing the computational complexity of these algorithms.

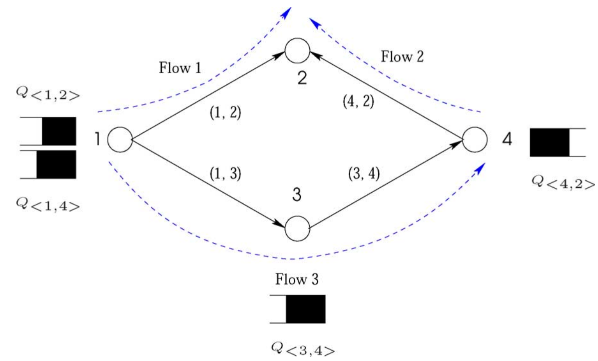

Fig. 1. A network example.

II. Basic Model

We first introduce the basic models in this section.

Network model: Consider a network represented by a graph $\mathcal{G} = (N, L)$, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of directed links (channels). Denote by (m,n) a link from node m to node n, where m is the transmitter and n is the receiver.

Traffic model: We let $[s,d]$ denote a flow in the network, where s denotes the source of the flow, and d denotes the destination of the flow. Further, let $\mathcal{F}$ denote the set of all flows in the network. Assume that time is discretized. Denote by $A_{[s,d]}[t]$ the number of packets injected into node s and destined to node d at time slot t, which is a nonnegative integer. We assume that $\{A_{[s,d]}[t]\}_{[s,d]\in\mathcal{F}}$ are stationary random variables independent across time and flows, $\mathbf{E}[A_{[s,d]}[t]] = a_{[s,d]}$, and $A_{[s,d]}[t] \le A_{\max}$ for all $[s,d] \in \mathcal{F}$ and t. We further assume that $\Pr(A_{[s,d]}[t] = 1) > 0$ and $\Pr(A_{[s,d]}[t] = 0) > 0$.[1] A network with four nodes, four links, and three flows is illustrated in Fig. 1.

Channel-fading model: We denote by $C_{(m,n)}[t]$ the state of channel (m,n) at time slot t, i.e., the number of packets that can be transmitted over link (m,n) during time slot t when there is no other simultaneous transmissions in the network. We assume that $C_{(m,n)}[t]$ is random and can be described as a finite-state Markov chain,[2] i.e.,

$$\begin{aligned}&\Pr\left(C_{(m,n)}[t]\middle|\, C_{(m,n)}[t-1],\ldots,C_{(m,n)}[0]\right)\\ &=\Pr\left(C_{(m,n)}[t]\middle|\, C_{(m,n)}[t-1]\right)\end{aligned}$$

and $C_{(m,n)}[t] \in \mathcal{C}$ for $|\mathcal{C}| < \infty$.[3] We assume that $C_{(m,n)}[t] \le C_{\max}$ for all $(m,n) \in \mathcal{L}$ and t. The one-step transition probabilities of the Markov chains are assumed to be known at all nodes,

[1]These assumptions guarantee that the Markov chains defined by the queue-lengths and channel states are irreducible and aperiodic. The definitions of the Markov chains used in stability analysis will be presented in Sections VI and VII.

[2]Finite-state Markov chain models have been shown to be very useful for modeling wireless links (e.g., [38]–[40]).

[3]In this paper, whenever there is no ambiguity, we abuse notation to let $\Pr(C_{(m,n)}[t]\,|\,C_{(m,n)}[t-1]) := \Pr(C_{(m,n)}[t] = c_1\,|\,C_{(m,n)}[t-1] = c_2)$ (i.e., we do not explicitly indicate the values the random variables take in conditional probability expressions).

and $\{C_{(m,n)}[t]\}_{(m,n)\in\mathcal{L}}$ are independent across links. We also assume that $\Pr(C_{(m,n)}[t] = C_{(m,n)}[t-1]) > 0$.[1]

Interference model: We assume a general collision model in this paper. If two links interfere with each other, simultaneous transmissions on the two links will lead to a collision and no information (packet) can get through. For link (m,n), we denote by $\mathcal{I}_{(m,n)}$ the set of links that interfere with link (m,n). For example, in downlink/uplink networks, where only one link can successfully transmit at a time, $\mathcal{I}_{(m,n)} = \mathcal{L}\backslash\{(m,n)\}$. Note that the half-duplex constraint can be easily included in this general collision model by assuming that $(m,i) \in \mathcal{I}_{(n,m)}$, i.e., when node m is receiving from node n, node m cannot transmit at the same time.

Queue dynamics: We assume that each node maintains a separate queue for each destination. Denote by $Q_{\langle m,d\rangle}[t]$ the length of the queue maintained at node m for destination d. It has been shown in [3] that per-destination queues are sufficient for stability. In most of this paper, we assume that hop-by-hop feedback is available so that node m knows those packets successfully transmitted over link (m,n), and immediately removes them from the queues. Thus, the dynamics of queue $\langle m,d\rangle$ can be described as follows:

$$Q^{\mathcal{P}}_{\langle m,d\rangle}[t+1] = Q^{\mathcal{P}}_{\langle m,d\rangle}[t] + A_{[m,d]}[t] + \sum_{(k,m)\in\mathcal{L}} \nu^{\mathcal{P}}_{(k,m),d}[t] - \sum_{(m,n)\in\mathcal{L}} \nu^{\mathcal{P}}_{(m,n),d}[t]$$

where $\nu^{\mathcal{P}}_{(m,n),d}[t]$ is the number of packets successfully transmitted over link (m,n) at time t with destination d. The superscript $\mathcal{P}$ indicates the policy used. (However, for one of the cases to be considered, somewhat different queue dynamics will be assumed, which will be described explicitly when the case is presented.)

Stochastically stable: Given traffic $\{A_{[s,d]}[t]\}_{[s,d]\in\mathcal{F}}$ and a routing/schedulig policy, we say the network is *stochastically stable* if a Markov chain defined by queue-lengths and channel states is positive recurrent (the Markov chains will be defined in Sections VI and VII). Traffic $\{A_{[s,d]}[t]\}_{[s,d]\in\mathcal{F}}$ is said to be *supportable* if there exists any routing/scheduling policy (without packet dropping), under which the network is stochastically stable.

III. More Notations

To simplify our notations, we let $\mathbf{C}[t]$ denote the states of all channels at time slot t, i.e.

$$\mathbf{C}[t] \triangleq \left\{C_{(m,n)}[t]\right\}_{(m,n)\in\mathcal{L}}.$$

We also let $\mathbf{C}[t](\tau_1:\tau_2)$ denote the states of all channels from time slot $t-\tau_2$ to $t-\tau_1$ ($\tau_2 \geq \tau_1$), i.e.

$$\mathbf{C}[t](\tau_1:\tau_2) \triangleq \left\{C_{(m,n)}[t-s]\right\}_{(m,n)\in\mathcal{L}, s\in[\tau_1,\tau_2]}.$$

Similarly, we define $\mathbf{Q}[t]$ and $\mathbf{Q}[t](\tau_1:\tau_2)$.

If the network is stochastically stable under routing/scheduling policy $\mathcal{P}$, we denote by $\mathbf{C}[\infty]$ and $\mathbf{Q}^{\mathcal{P}}[\infty]$ the steady-state channel and queue states under policy $\mathcal{P}$. We further define

$$\mathbf{C}(\tau_1,\tau_2) \triangleq \lim_{t\to\infty} \left\{C_{(m,n)}[t-s]\right\}_{(m,n)\in\mathcal{L}, s\in[\tau_1,\tau_2]}$$

and $\mathbf{Q}^{\mathcal{P}}(\tau_1,\tau_2)$ is defined similarly.

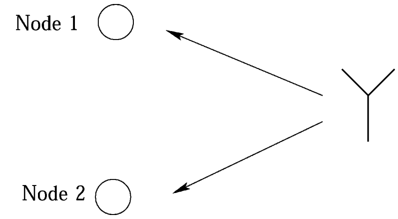

Fig. 2. A downlink network.

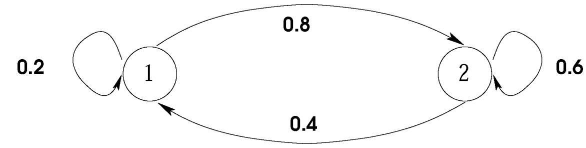

Fig. 3. A two-state Markov chain: State 1 supports 50 packets/slot, and State 2 supports 0 packets/slot.

Furthermore, let $D^{\mathcal{P}}_{(m,n)}[t]$ denote the transmission decision on link (m,n) under policy $\mathcal{P}$, i.e., $D^{\mathcal{P}}_{(m,n)}[t] = 1$ if node m attempts to transmit over link (m,n) at time slot t; and $D^{\mathcal{P}}_{(m,n)}[t] = 0$ otherwise. We also let $S^{\mathcal{P}}_{(m,n)}[t]$ denote the achievable rate over link (m,n) at time t. Assuming that a node can only transmit over one link at a time, we have

$$S^{\mathcal{P}}_{(m,n)}[t] = C_{(m,n)}[t] D^{\mathcal{P}}_{(m,n)}[t] \prod_{(k,l)\in\mathcal{I}_{(m,n)}} \left(1 - D^{\mathcal{P}}_{(k,l)}[t]\right).$$

Note that $S^{\mathcal{P}}_{(m,n)}[t] \geq \nu^{\mathcal{P}}_{(m,n)}[t]$, where the strict inequality holds when there is no enough packets in the queue.

Finally, we say a set of links $\mathcal{M}$ is an independent-set if for any link $(m,n)\in\mathcal{M}$, $\mathcal{I}_{(m,n)} \cap \mathcal{M} = \emptyset$.

IV. Centralized Routing/Scheduling With Delayed NSI: An Example and the Main Result

We first consider the centralized case where the central controller knows $\mathbf{C}[t-\boldsymbol{\tau}]$ and $\mathbf{Q}[t-\boldsymbol{\tau}]$, and makes routing/scheduling decisions based on this delayed NSI.[4] Here, we abuse notation and let $\boldsymbol{\tau}$ denote the vector of delays. In particular, $\boldsymbol{\tau} = \{\tau_{(m,n)}\}_{(m,n)\in\mathcal{L}}$ when associated with the channel-states and $\boldsymbol{\tau} = \{\tau_{\langle m,d\rangle}\}_{m,d\in\mathcal{N}}$ when associated with the queue-states. We also assume that the routing/scheduling decisions are executed in the network immediately without delay.

A. An Illustrative Example

Consider a wireless downlink network with two nodes and a single base-station as shown in Fig. 2. Denote by $(0,1)$ the channel from the base-station to node 1 and by $(0,2)$ the channel from the base-station to node 2. Assume that the channels are represented by the two-state Markov chain as shown in Fig. 3.

[4]In this paper, we assume that the central controller knows the channel and queue trajectories with some (node dependent) delay. This model can easily be relaxed to the case where as before, the controller knows the channels with delay; however, it knows the queue lengths only *roughly periodically* with (node-dependent) delay. In other words, at each time t and for each queue $Q_{\langle m,d\rangle}$, it suffices for the controller to know the queue length only at some time s in the past (as opposed to the entire trajectory) in the finite interval $s \in (t-\bar{d}_{\langle m,d\rangle}, t-\underline{d}_{\langle m,d\rangle})$, where $0 \leq \tau_{m,n} \leq \underline{d}_{\langle m,d\rangle} \leq \bar{d}_{\langle m,d\rangle} < \infty$. However, for notational ease, in this paper, we only discuss the case where $\underline{d}_{\langle m,d\rangle} = \bar{d}_{\langle m,d\rangle} = \tau_{\langle m,d\rangle}$.

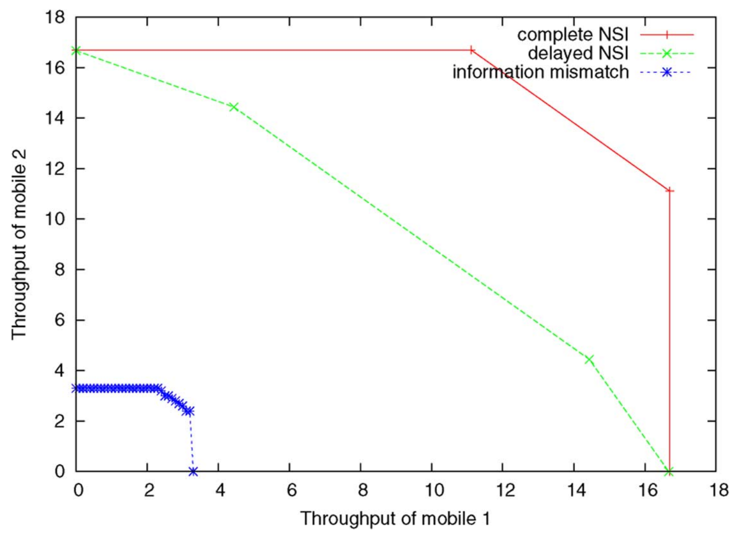

Fig. 4. The long-term throughput regions of the downlink network.

We consider three different cases:

(i) **Complete NSI:** The base-station has the instantaneous NSI. In this case, we analytically compute the network throughput region, which is depicted using the "+"-line in Fig. 4. The following queue-length-based throughput-optimal scheduling algorithm has been proposed in [2]: At time slot t, the base-station transmits to node $i^*[t]$ such that

$$i^*[t] \in \arg\max_{i=1,2} Q_{\langle 0,i\rangle}[t]C_{(0,i)}[t]. \quad (1)$$

(ii) **Delayed NSI:** The base-station has one time slot delayed NSI, i.e., the base-station only has $C_{(0,i)}[t-1]$ and $Q_{\langle 0,i\rangle}[t-1]$. The throughput region in this case is analytically computed based on the theoretical characterization Λ_τ, which will be presented in Section VI. The throughput region is depicted using the "$\times$"-line Fig. 4. A throughput optimal policy is as follows: At time slot t, the base-station transmits to node $i^*[t]$ such that

$$i^*[t] \in \arg\max_{i=1,2} Q_{\langle 0,i\rangle}[t-1]\mathbf{E}\left[C_{(0,i)}[t]\,\middle|\,C_{(0,i)}[t-1]\right]. \quad (2)$$

The throughput optimality of the algorithm above will be proved in Section VI.

(iii) **Information Mismatch:** In this case, the base-station has one time slot delayed NSI, which is the same as the delayed NSI case. However, the base-station simply treats the delayed NSI as the instantaneous NSI, and use the following naive scheduling algorithm: At time slot t, the base-station transmits to node $i^*[t]$ such that

$$i^*[t] \in \arg\max_{i=1,2} Q_{\langle 0,i\rangle}[t-1]C_{(0,i)}[t-1].$$

We call this case *information mismatch* because the scheduling algorithm uses the delayed information as instantaneous information. In this case, we use simulations to plot the rate region of the naive scheduling under deterministic arrivals, i.e., at the beginning of each time slot, λ_1 packets arrive at mobile one and λ_2 packets arrive at mobile two. The rate region is the region depicted using the "$*$"-line in Fig. 4.

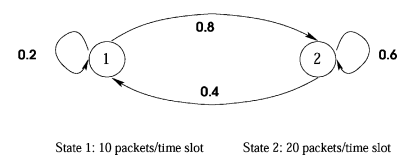

Fig. 5. The twisted two-state Markov chain.

Remark 1: Comparing case (i) and (ii), we can see that the delay in NSI leads to a throughput degradation. The intuition is easy to understand. While, without delay, the base-station can transmit 50 packets/slot over link $(0,i)$ when $C_{(0,i)}[t] = 50$; only 10 packets/slot in average can be successfully transmitted if the base-station does not have $C_{(0,i)}[t]$, but only knows $C_{(0,i)}[t-1] = 50$. With the one-time-slot delay, the channels revealed to the base-station can be modeled as the Markov chain illustrated in Fig. 5. Such a channel twist due to information delay reduces the achievable throughput region. We remark that qualitatively similar observations have been made in the context of multichannel access point networks for uplink and downlink scheduling[5] in [26].

Remark 2: Furthermore, we would like to comment that incorporating the delay in scheduling is critical to achieve the maximum network throughput. For example, we can see that the information mismatch case—case (iii), which uses the delayed NSI as the current NSI, leads to a significant performance loss compared to the throughput optimal algorithm (case (ii)).

[5]The model studied in [26] is in the context where the base-station (a central controller) periodically gets all channel and queue states together (homogeneous delay).

B. Main Result

For the centralized scenario, we propose the following on-off routing/scheduling based on conditional expectation.

On-Off Routing/Scheduling: At time slot t.

(1) The controller first computes the optimal independent-link-set $\mathcal{M}^*[t]$ which maximizes

$$\sum_{(m,n)\in\mathcal{M}} \mathbf{E}\left[C_{(m,n)}[t] \middle| C_{(m,n)}\left[t-\tau_{(m,n)}\right]\right] P_{(m,n)}[t]$$

where

$$P_{(m,n)}[t] = \max_{d\in\mathcal{N}} \left(Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right] - Q_{\langle n,d\rangle}[t-\tau_{\langle n,d\rangle}]\right)$$

(2) Node m transmits the packets from queue $d^*_{(m,n)}[t]$ over link (m,n) with a rate $C_{(m,n)}[t]$ if $(m,n)\in\mathcal{M}^*[t]$, where

$$d^*_{(m,n)}[t] \in \arg\max_{d} \left(Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right] - Q_{\langle n,d\rangle}[t-\tau_{\langle n,d\rangle}]\right).$$

Remark: We can see the on-off routing/scheduling algorithm is similar to the back-pressure algorithm. The main difference is that in the on-off routing/scheduling algorithm, we use the conditional expectation $\mathbf{E}[C_{(m,n)}[t]\,|\,C_{(m,n)}[t-\tau_{(m,n)}]]$ instead of the instantaneous channel rate $C_{(m,n)}[t]$. Note that $\mathbf{E}[C_{(m,n)}[t]\,|\,C_{(m,n)}[t-\tau_{(m,n)}]]$ is the mean of the achievable channel rate given that the delayed channel state information $C_{(m,n)}[t-\tau_{(m,n)}]$. The quantity $E[C_{(m,n)}[t]\,|\,C_{(m,n)}[t-\tau_{(m,n)}]]$ is computed based on the transition probability of the Markov chain and the delayed channel state information $C_{(m,n)}[t-\tau_{(m,n)}]$. This is a time-varying quantity because the delayed channel state information is time-varying.

We will prove in Section VI that this on-off routing/scheduling is throughput-optimal if the delays in channel feedbacks are smaller than the delays in queue feedbacks.

Theorem 1: Assume that the delays satisfy $\max_{(m,n)\in\mathcal{L}} \tau_{(m,n)} < \min_{m,d\in\mathcal{N}} \tau_{\langle m,d\rangle}$ and $(1+\epsilon)\mathbf{A}[t]$ is supportable, the network is stochastically stable under the on-off routing/scheduling algorithm. □

V. Decentralized Routing/Scheduling With Delayed NSI: Examples and Main Results

In this section, we consider the decentralized scenario where a node has its own local instantaneous NSI and a network-wide delayed NSI. Thus, the nodes have different views of the network. The challenge is to develop a consistent policy based on the inconsistent NSI.

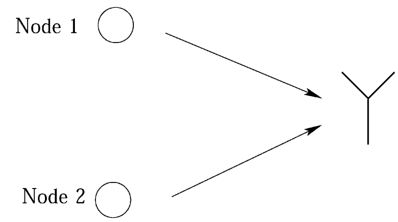

Fig. 6. Uplink network.

A. Illustrative Example: A Single-Hop Network

Consider an uplink network with two nodes and a single base-station as illustrated in Fig. 6. The channels between the base-station and the nodes are each described by a two-state Markov chain as illustrated in Fig. 3.

To illustrate the impact of delayed NSI on the network throughput, we also show the throughput region of the following four scenarios in Fig. 7.

(i) **Complete NSI:** Both mobiles have the complete instantaneous NSI, i.e., both of them know $\{C_{(i,0)}[t], Q_{\langle i,0\rangle}[t]\}_{i=1,2}$. The throughput region is analytically computed and depicted using the "+"-line in Fig. 7.

(ii) **Decentralized NSI:** In this example, we assume that each of the two nodes has exact knowledge of its channel and queue state (over the current and previous time-slots), but has one-time-slot delay in the information from the other node. Equivalently, both of the nodes have delayed NSI $\{C_{(i,0)}[t-1], Q_{\langle i,0\rangle}[t-1]\}_{i=1,2}$, and additionally, each node i only knows $C_{(i,0)}[t]$ and $Q_{\langle i,0\rangle}[t]$.

In this case, the throughput region is analytically computed based on the theoretical characterization $\tilde{\Lambda}_\tau$, which will be presented in Section VII. The region is depicted using the "×"-line in Fig. 7. A throughput optimal policy is as follows: At time slot t,

(1) Node 1 and 2 first calculate a threshold vector $\mathbf{T}^*[t]$ by solving the optimization problem (3). We use "∈" in (3) to indicate that there may be multiple solutions to the optimization problem. Throughout this paper, we assume that all nodes break the tie in the same way so that all nodes obtain the same $\mathbf{T}^*[t]$ [see (3) at the bottom of the page].

(2) Node i transmits to the base-station if $C_{(i,0)}[t] \geq T^*_{(i,0)}[t]$.

$$\mathbf{T}^*[t] \in \arg\max_{\mathbf{T}\geq 0} Q_{\langle 1,0\rangle}[t-1]\mathbf{E}\left[C_{(1,0)}[t]\mathbf{1}_{C_{(1,0)}[t]\geq T_{(1,0)},C_{(2,0)}[t]<T_{(2,0)}} \middle| \mathbf{C}[t-1]\right] + Q_{\langle 2,0\rangle}[t-1]\mathbf{E}\left[C_{(2,0)}[t]\mathbf{1}_{C_{(2,0)}[t]\geq T_{(2,0)},C_{(1,0)}[t]<T_{(1,0)}} \middle| \mathbf{C}[t-1]\right] \quad (3)$$

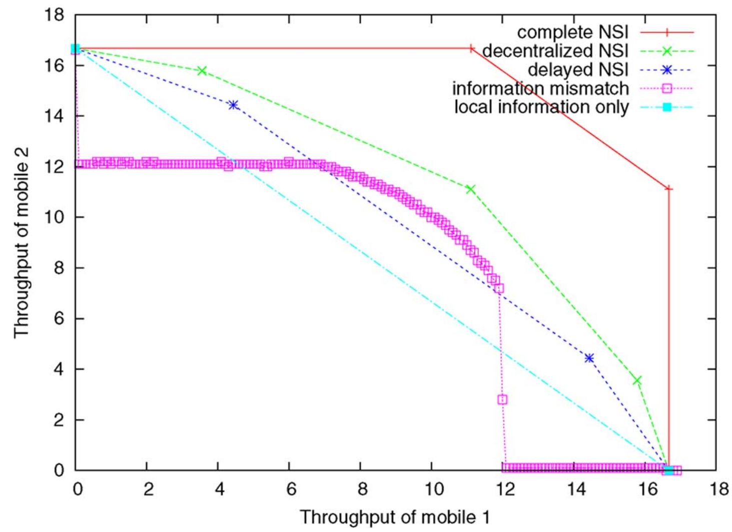

Fig. 7. The long-term throughput regions of the single-hop network.

(iii) **Delayed NSI:** Both nodes have $\{C_{(i,0)}[t-1], Q_{\langle i,0\rangle}[t-1]\}$, but no instantaneous NSI is available. The throughput region is analytically computed and depicted using the "$*$"-line.

(iv) **Information Mismatch:** In this case, we consider a strategy where we ignore the delays in NSI, and use the algorithm that is optimal with the complete and instantaneous NSI: Node 1 attempts to transmit if $C_{(1,0)}[t]Q_{(1,0)}[t] > C_{(2,0)}[t-1]Q_{(2,0)}[t-1]$, and node 2 attempts to transmit if $C_{(2,0)}[t]Q_{(2,0)}[t] \geq C_{(1,0)}[t-1]Q_{(1,0)}[t-1]$. Assuming deterministic arrivals, the rate region is obtained using the simulation, and depicted using "□"-line in Fig. 7. We can see that the information mismatch leads to a significant performance loss.

(v) **Local Information Only:** In this case, each mobile only uses its own channel and queue information, and views the channel as a static channel with channel rate 16.67 packets/time slot. The network is equivalent to a network with two static channels, each with channel rate 16.67 packets/time slot. The throughput region is depicted using "■" in Fig. 7. We can see this region is significantly smaller than that under either delayed NSI or decentralized NSI, which demonstrates the importance of exploiting delayed NSI in scheduling.

Remark 1: The throughput optimal algorithm for the decentralized NSI consists of two steps:

(1) Determine the threshold vector based on *global* delayed NSI.

(2) Make the transmission decisions based on *local* instantaneous (current) NSI.

This algorithm is a combination of greedy-contention scheduling and collision-avoidance scheduling. For example, given $C_{(1,0)}[t-1] = C_{(2,0)}[t-1] = 50$,

- If $Q_{\langle 1,0\rangle}[t-1] > 4Q_{\langle 2,0\rangle}[t-1]$, then $\mathbf{T}^*[t] = [0, 51]$. Thus, node 2 keeps silent, and node 1 attempts to transmit. This corresponds to a "conservative" collision-avoidance scheduling, where node 2 does not transmit irrespective of its current local channel state (i.e., $C_{(2,0)}[t]$). On the other hand, node 1 transmits (successfully) only if $C_{(1,0)}[t] = 50$.
- If $Q_{\langle 1,0\rangle}[t-1]/Q_{\langle 2,0\rangle}[t-1] \in [0.25, 4]$, then $\mathbf{T}^*[t] = [50, 50]$. In this case, both nodes attempt to transmit when $C_{(i,0)}[t] = 50, i = 1, 2$. Further, if any one of the users' current channel is in state 2 (i.e., no packet can be transmitted), that node does not attempt to transmit. This corresponds to a greedy contention scheduling mechanism, where the randomization is provided by the channel (i.e., both channels are in state 50 in the previous time-slot, and there is a positive probability for each of the four cases $(50,50), (50,0)(0,50, (0,0)$ to occur in the current time-slot).

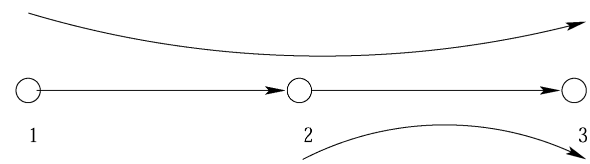

Fig. 8. Two-hop network.

Remark 2: We can see that the network throughput region shrinks with the increase of the delay. Further, the information mismatch leads to a significant performance loss compared to the throughput-optimal algorithm. Thus, to maximize the network performance under delayed NSI, it is important that we understand the impact of the delayed NSI, and intelligently use the delayed NSI in routing/scheduling.

B. Illustrative Example: A Multihop Network

To further demonstrate the impact of delayed NSI. We consider a two-hop network with two flows as in Fig. 8. We consider three different cases:

(i) **Complete NSI:** The complete NSI is available at node 1 and 2. The network throughput region is analytically computed and depicted using the "$+$"-line in Fig. 9.

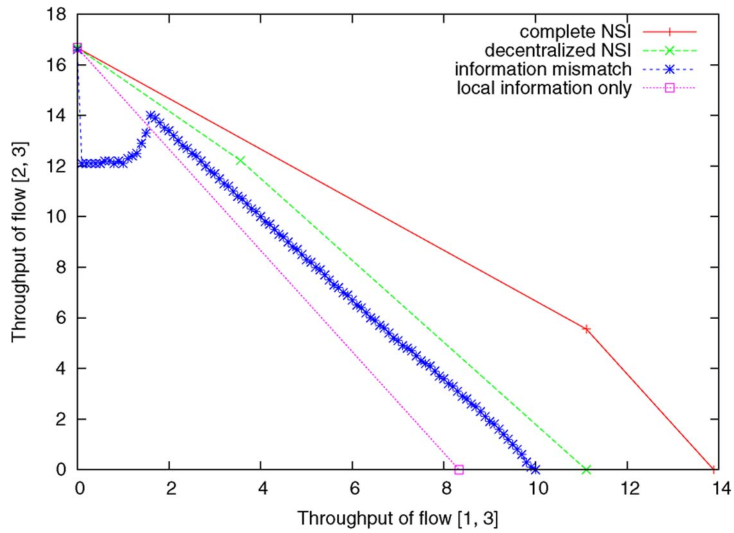

Fig. 9. The long-term throughput regions of the two-hop network.

(ii) **Decentralized NSI:** Node 1 has $C_{(1,2)}[t]$, $Q_{\langle 1,2\rangle}[t]$, $Q_{\langle 1,3\rangle}[t]$, $C_{(2,3)}[t-1]$, and $Q_{\langle 2,3\rangle}[t-1]$. Node 2 has $C_{(2,3)}[t]$, $Q_{\langle 2,3\rangle}[t]$, $C_{(1,2)}[t-1]$, $Q_{\langle 1,2\rangle}[t-1]$, and $Q_{\langle 1,3\rangle}[t-1]$. The network throughput region is analytically computed and depicted using the "$\times$"-line in Fig. 9. In this scenario, the throughput-optimal algorithm can be viewed as a "back-pressure" like algorithm [2], [3] which determines the policy based on *consistent* delayed NSI and makes transmission decisions based on the best available NSI (we refer to Section V-C and (4) for additional details).

(iii) **Information Mismatch:** We also consider an information mismatch scenario, where we use the algorithm that is optimal with complete NSI (case (i)) for the decentralized NSI (case (ii)). Assuming deterministic arrivals, the rate region is obtained using the simulation, and depicted using "$*$"-line in Fig. 9. We can see that such a information mismatch also leads to a performance loss.

(v) **Local Information Only:** In this case, each link only uses its own channel and queue information, and views the channels as static channels with channel rate 16.67 packets/time slot. The network is equivalent to a network with two static links with rate 16.67 packets/time slot. The throughput region is depicted using "□"-line in Fig. 9. We can see this region is significantly smaller than that under decentralized NSI.

C. Main Result

We now study multihop networks, where each node has its local instantaneous NSI, and homogeneous delayed NSI from other nodes. For networks with infrastructures such as base-stations, the base-stations can collect the information and broadcast to all the nodes. In ad hoc networks, nodes can form clusters and elect cluster-head to collect and broadcast the NSI. Since it takes a certain amount time for the information to be propagated to all the nodes in the network, we assume that all nodes have only delayed NSI of other nodes. We call this case decentralized case because each node needs to make their own decision based on the available information including both global delayed NSI and local instantaneous NSI. Note that nodes have inconsistent NSI in this case. We propose the following threshold-based routing/scheduling algorithm.

Threshold-Based Routing/Scheduling: At time slot t,

(1) All nodes first compute a common threshold vector $\mathbf{T}^*[t]$ by solving the optimization problem (4), shown at the bottom of the page.

(2) Node m attempts to transmit packets from queue $d^*_m[t]$ over link (m,n) if $C_{(m,n)} \geq T^*_{(m,n)}[t]$, where

$$d^*_{(m,n)}[t] \in \arg\max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right).$$

We consider two different NSI structures:

$S1$ Node m has $(C_{(m,n)}[t], Q_{\langle m,d\rangle}[t])$ for all $(m,n)\in\mathcal{L}$ and $d\in\mathcal{N}$, and $(C_{(l,k)}[t-1], Q_{\langle l,d\rangle}[t-1])$ for all $(l,k)\in\mathcal{L}$ for all $l,d\in\mathcal{N}$. In other words, each node has their instantaneous local NSI, and one time slot delayed NSI of the entire network.

$S2$ All nodes have $\mathbf{C}[t-\tau]$ and $\mathbf{Q}[t-\tau]$. Node m also has $C_{(m,n)}[t](0:\tau-1)$ and $Q_{\langle m,d\rangle}[t](0:\tau-1)$ for $(m,n)\in\mathcal{L}$ and $d\in\mathcal{N}$. We assume that all flows are single-hop flows, and a transmitter is disjoint from other

$$\mathbf{T}^*[t] \in \arg\max_{\mathbf{T}\geq 0} \sum_{(m,n)\in\mathcal{L}} \max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right) \times \mathbf{E}\left[C_{(m,n)}[t]\mathbf{1}_{C_{(m,n)}[t]\geq T_{(m,n)}}\Pi_{(l,k)\in\mathcal{I}_{(m,n)}}\mathbf{1}_{C_{(l,k)}[t]<T_{(l,k)}}\,\middle|\,\mathbf{C}[t-\tau]\right]. \quad (4)$$

transmitters and receivers. It is easy to see that only a single queue is needed for each flow. Furthermore, we assume that transmitters cannot detect collisions (it needs to explicitly wait for an acknowledgment (ACK) from the receiver to determine success/failure of the transmission), and a source does not delete transmitted packets until it receives an acknowledgment from the destination (after which it can flush packets in its queue corresponding to the successful ACKs). Thus, the queue dynamics can be described as follows:

$$Q_{\langle m,d\rangle}[t+1] = \left(Q_{\langle m,d\rangle}[t] + A_{[m,d]}[t] - \nu_{\langle m,d\rangle}[t-\tau_f]\right)^+ \tag{5}$$

where τ_f is the feedback delay, and assumed to be larger than τ. We need this dynamic and the condition $\tau_f > \tau$ to make sure that the hidden information in ACKs is older than the delayed NSI. Otherwise, we need to learn NSI from ACKs and quantify the amount of knowledge of $\mathbf{C}[t-r]$ $(r \leq \tau)$ we can learn. In the rest of this paper, we focus on $S2$ For the decentralized scenario since the analysis of NSI structure $S1$ is similar.

The main result, given the two NSI structures above, is as follows.

Theorem 2: Given NSI structure $S1$ or $S2$, and traffic $\mathbf{A}[t]$ such that $(1+\epsilon)\mathbf{A}[t]$ is supportable, the network is stochastically stable under the threshold-based routing/scheduling algorithm. □

VI. Centralized Algorithm With Delayed NSI: Detailed Analysis

A. Throughput Region

In this subsection, we will characterize the optimal throughput region with a central controller and delayed NSI. First, given

$$\mathbf{C}[0] = \left\{C_{(m,n)}[0]\right\}_{(m,n)\in\mathcal{L}}$$

and an independent-link-set $\mathcal{M}$, we define

$$\begin{aligned}\mathbf{S}(\mathbf{C}[0],\mathcal{M}) &= \left\{S_{(m,n)}\left(C_{(m,n)}[0],\mathcal{M}\right)\right\}_{(m,n)\in\mathcal{L}} \\ &= \left\{\mathbf{E}\left[C_{(m,n)}[\tau_{(m,n)}]\middle| C_{(m,n)}[0]\right]\mathbf{1}_{(m,n)\in\mathcal{M}}\right\}_{(m,n)\in\mathcal{L}}.\end{aligned}$$

Further, we define $\eta(\mathbf{C}[0])$ to be the convex hull of $\mathbf{S}(\mathbf{C}[0],\mathcal{M})$ over all independent-link set $\mathcal{M}$

$$\eta(\mathbf{C}[0]) = \mathcal{CH}_{\mathcal{M}}(\mathbf{S}(\mathbf{C}[0],\mathcal{M})).$$

Then, we define

$$\Lambda_\tau = \left\{\boldsymbol{\eta} : \boldsymbol{\eta} = \sum_{\mathbf{c}\in\mathcal{C}^{|\mathcal{L}|}} \Pr(\mathbf{C}[0]=\mathbf{c})\boldsymbol{\eta}(\mathbf{c}), \boldsymbol{\eta}(\mathbf{c}) \in \eta(\mathbf{C}[0]=\mathbf{c})\right\}. \tag{6}$$

In the following analysis, we will prove that Λ_τ is the optimal throughput region. To simplify our notation, we define

$$\mathbf{C}[t-\boldsymbol{\tau}] = \left\{C_{(m,n)}\left[t-\tau_{(m,n)}\right]\right\}_{(m,n)\in\mathcal{L}}$$

and

$$\mathbf{Q}[t-\boldsymbol{\tau}] = \left\{Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\right\}_{m,d\in\mathcal{N}}.$$

Furthermore, we define

$$\mathbf{Y}[t] = \left\{Q_{\langle m,d\rangle}[t](0:\tau_{\langle m,d\rangle}), C_{(m,n)}[t]\left(0:\tau_{(m,n)}\right)\right\}$$

for all $m,d \in \mathcal{N}$ and $(m,n) \in \mathcal{L}$. It is easy to see that $\mathbf{Y}[t]$ is Markovian. Note that under the assumptions $\Pr(A_{[s,d]}[t]=1) > 0$, $\Pr(A_{[s,d]}[t]=0) > 0$ and $\Pr(C_{(m,n)}[t] = C_{(m,n)}[t-1]\,|\,C_{(m,n)}[t-1]) > 0$, we can easily verify the Markov chain is irreducible and aperiodic.

Now, given arrivals $\mathbf{A}[t] = \{A_{[s,d]}[t]\}_{[s,d]\in\mathcal{F}}$, the network is stochastically stable under a scheduling policy $\mathcal{P}$ if $\mathbf{Y}^{\mathcal{P}}[t]$ is positive recurrent, where the superscript indicates the scheduling policy. Thus, given a stabilizing policy $\mathcal{P}$, we define the steady-state distribution $\mathbf{Y}^{\mathcal{P}}[\infty]$ as follows:

$$\begin{aligned}&\left\{\left(Q^{\mathcal{P}}_{\langle m,d\rangle}\left(0:\tau_{\langle m,d\rangle}\right), C_{(m,n)}\left(0:\tau_{(m,n)}\right)\right)\right\} \\ &\triangleq \lim_{t\to\infty}\left\{Q^{\mathcal{P}}_{\langle m,d\rangle}[t]\left(0:\tau_{\langle m,d\rangle}\right), C_{(m,n)}[t]\left(0:\tau_{(m,n)}\right)\right\} \\ &= \lim_{t\to\infty}\mathbf{Y}^{\mathcal{P}}[t] \\ &= \mathbf{Y}^{\mathcal{P}}[\infty].\end{aligned}$$

Now denote by $\mathbf{D}^{\mathcal{P}}(\mathbf{Q}[t-\boldsymbol{\tau}]=\mathbf{q},\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c})$ the decision under policy $\mathcal{P}$ given delayed NSI $(\mathbf{Q}[t-\boldsymbol{\tau}]=\mathbf{q},\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c})$. Based on a stabilizing policy $\mathcal{P}$, we now define a corresponding (probabilistic) time-sharing policy $\mathcal{P}_s$ which uses the knowledge of the *steady-state* distribution of the queue lengths and channel states under policy $\mathcal{P}$ (denoted as before by $\mathbf{Y}^{\mathcal{P}}[\infty]$) along with the delayed channel states $\mathbf{c}$ as follows:

Time-sharing Policy $\mathcal{P}_s$: Given the delayed channel-state information $\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c}$, we let

$$r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}} = \Pr(\mathbf{Q}(\boldsymbol{\tau}:\boldsymbol{\tau})=\mathbf{q}\,|\,\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})=\mathbf{c}).$$

Then, at each time when the delayed channel state is $\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c}$, the policy $\mathcal{P}^s$ probabilistically makes decisions (denoted by $\mathbf{D}^{\mathcal{P}_s}$) as follows:

Let $\mathcal{Q} = \{\mathbf{q} : r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}} > 0\}$. The decision is to choose

$$\mathbf{D}^{\mathcal{P}_s}(\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c}) = \mathbf{D}^{\mathcal{P}}(\mathbf{Q}[t-\boldsymbol{\tau}]=\mathbf{q},\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c})$$

with probability $r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}}$.

In other words, for each delayed channel state $\mathbf{c}$, the policy $\mathcal{P}_s$ first randomly (with probability $r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}}$) picks a possible steady-state queue lengths *under policy* $\mathcal{P}$ and uses the resulting pair $(\mathbf{q},\mathbf{c})$ as the "input" to policy $\mathcal{P}$ (which is known to be stabilizing) in order to get a decision. We note that this rule is in the same spirit as the Static Service Split (SSS) Rule in [41].

Next we will show that the service rates obtained under $\mathcal{P}$ or $\mathcal{P}_s$ are the same.

Lemma 3: Assume that the queue-length information is collected with larger delays than channel state information, i.e.

$$\max_{(m,n)\in\mathcal{L}} \tau_{(m,n)} < \min_{k,d\in\mathcal{N}} \tau_{\langle k,d\rangle}.$$

Then conditioned on $\mathbf{C}[t-\boldsymbol{\tau}]$, queue-length $Q_{\langle k,d\rangle}[t-\tau_{\langle k,d\rangle}]$ and channel state $C_{(m,n)}[t-s]$ are independent for any $s < \tau_{(m,n)}$, $(m,n)\in\mathcal{L}$, and $k,d\in\mathcal{N}$.

Proof: Recall that $Q_{\langle k,d\rangle}[t-\tau_{\langle k,d\rangle}]$ is queue-length maintained for destination d at node k at time $t-\tau_{\langle k,d\rangle}$. The value of $Q_{\langle k,d\rangle}[t-\tau_{\langle k,d\rangle}]$ is determined by the arrivals injected into the queue before $t-\tau_{\langle k,d\rangle}$, i.e., $\mathbf{A}[t](\tau_{\langle k,d\rangle},t))$, channel states before $t-\tau_{\langle k,d\rangle}$, i.e., $\mathbf{C}[t](\tau_{\langle k,d\rangle},t)$, and scheduling decisions before $t-\tau_{\langle k,d\rangle}$, i.e., $\mathbf{D}^{\mathcal{P}}[t](\tau_{\langle k,d\rangle},t)$. Since the value of $Q_{\langle k,d\rangle}[t-\tau_{\langle k,d\rangle}]$ is determined by arrivals, channel states, and scheduling decisions before $[t-\tau_{\langle k,d\rangle}]$ and $\tau_{\langle k,d\rangle}\geq\tau_{(m,n)}$, we can conclude that $C_{(m,n)}[t-s]$ is independent of $Q_{\langle k,d\rangle}[t-\tau_{\langle k,d\rangle}]$ for $s\leq\tau_{(m,n)}$. ■

Lemma 4: Given $\max_{(m,n)\in\mathcal{L}}\tau_{(m,n)} < \min_{k,d\in\mathcal{N}}\tau_{\langle k,d\rangle}$ and policy $\mathcal{P}$ that stabilizes traffic $\mathbf{A}[t]$, we have that

$$\lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}}_{(m,n)}[t]\right] = \lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}_s}_{(m,n)}[t]\right]$$

where $\mathcal{P}_s$ is the corresponding time-sharing policy.

Proof: The lemma is proved in (7), at the bottom of the page, where (a) holds due to Lemma 3, and (b) yields from the definition of the time-sharing policy $\mathcal{P}_s$. ■

Lemma 5: Consider the centralized case where the central controller makes routing/scheduling decisions based on $\mathbf{C}[t-\boldsymbol{\tau}]$ and $\mathbf{Q}[t-\boldsymbol{\tau}]$. Given the delays satisfying $\max_{(m,n)\in\mathcal{L}}\tau_{(k,n)} < \min_{m,d\in\mathcal{N}}\tau_{\langle k,d\rangle}$, traffic $\mathbf{A}[t]$ is supportable if and only if there exists $\boldsymbol{\eta}\in\Lambda_\tau$ such that

$$\mathbf{E}\left[A_{[s,d]}[t]\right]1_{s=n} + \sum_{m:(m,n)\in\mathcal{L}}\boldsymbol{\eta}_{(m,n)} \leq \sum_{k:(n,k)\in\mathcal{L}}\boldsymbol{\eta}_{(n,k)} \quad (8)$$

holds for all n.

Proof: First if $\mathbf{A}[t]$ is supportable, then there exists a policy $\mathcal{P}$, under which the queueing system is positive recurrent. From Lemma 4, we can find a corresponding time-sharing policy $\mathcal{P}^s$ which allocates the same amount service rate to each link. Note that

$$\lim_{t\to\infty}\mathbf{E}\left[\mathbf{S}^{\mathcal{P}_s}[t]\middle|\,\mathbf{C}[t-\tau]=\mathbf{c}\right]\in\eta(\mathbf{C}[0]=\mathbf{c}).$$

We define

$$\bar{\eta}(\mathbf{c}) = \lim_{t\to\infty}\mathbf{E}\left[\mathbf{S}^{\mathcal{P}_s}[t]\middle|\,\mathbf{C}[t-\tau]\right],$$

i.e., the stationary service rates conditioned on $\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c}$. Define

$$\boldsymbol{\eta} = \sum_{\mathbf{c}\in\mathcal{C}^{|\mathcal{L}|}}\Pr(\mathbf{C}[0]=\mathbf{c})\bar{\boldsymbol{\eta}}(\mathbf{c}).$$

According to the definition of Λ_τ, we can easily verify that $\boldsymbol{\eta}\in\Lambda_\tau$. Further, since the network is stochastically stable, (8) holds for all n because otherwise, some queue goes to infinity and the Markov is not positive recurrent.

The other direction is immediate. In particular, given the arrival rates, we can define a channel state dependent time-sharing rule over appropriate vertices of Λ_τ that will stabilize (support) the process $\mathbf{A}[t]$ (see [41, Theorem 1] for an analogous proof).■

B. Throughput Optimality

In this subsection, we prove that the on-off routing/scheduling is throughput optimal.

Theorem 1: Given the delays satisfying $\max_{(m,n)\in\mathcal{L}}\tau_{(m,n)} < \min_{m,d\in\mathcal{N}}\tau_{\langle m,d\rangle}$ and traffic $\mathbf{A}[t]$ such that $(1+\epsilon)\mathbf{A}[t]$ is supportable, the network is stochastically stable under the on-off routing/scheduling algorithm.

Proof: As we have shown in Section VI-A, $\mathbf{Y}[t]$ is a Markov chain. Therefore $\tilde{\mathbf{Y}}[n]=\mathbf{Y}[nT]$ is also a Markov chain, where T is a constant. We define T_m to be a constant such that

$$|\Pr(\mathbf{C}[nT+T_m]|\tilde{\mathbf{Y}}[n]) - \Pr(\mathbf{C}[nT+T_m])| \leq \epsilon/2. \quad (9)$$

Note that T_m is at the same time-scale of the mixing time of the Markov chain $\mathbf{C}[t]$. We further assume that

$$T \geq \frac{(1+2\epsilon)(1-\epsilon/2)(T_m+\max_{m,d}\tau_{\langle m,d\rangle})}{3\epsilon-\epsilon^2}.$$

Define a Lyapunov function $V[t]$ such that

$$V[n] = \sum_{m,d\in\mathcal{N}} Q^2_{\langle m,d\rangle}[nT]$$

$$\begin{aligned}
\mathbf{E}\left[S^{\mathcal{P}}_{(m,n)}[\infty]\right] &= \mathbf{E}\left[\mathbf{E}\left[C_{(m,n)}(0:0)D^{\mathcal{P}}_{(m,n)}\left(\mathbf{Q}(\boldsymbol{\tau}:\boldsymbol{\tau}),\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right)\middle|\,\mathbf{Q}(\boldsymbol{\tau}:\boldsymbol{\tau}),\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right]\right]\\
&= \mathbf{E}\left[D^{\mathcal{P}}_{(m,n)}\left(\mathbf{Q}(\boldsymbol{\tau}:\boldsymbol{\tau}),\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right)\mathbf{E}\left[C_{(m,n)}(0:0)\left|\mathbf{Q}(\boldsymbol{\tau}:\boldsymbol{\tau}),\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right.\right]\right]\\
&=_{(a)} \mathbf{E}\left[D^{\mathcal{P}}_{(m,n)}\left(\mathbf{Q}(\boldsymbol{\tau}:\boldsymbol{\tau}),\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right)\mathbf{E}\left[C_{(m,n)}(0:0)\left|C_{(m,n)}\left(\tau_{(m,n)}:\tau_{(m,n)}\right)\right.\right]\right]\\
&= \mathbf{E}\left[\mathbf{E}\left[C_{(m,n)}(0:0)\left|C_{(m,n)}\left(\tau_{(m,n)}:\tau_{(m,n)}\right)\right.\right]\mathbf{E}\left[D^{\mathcal{P}}_{(m,n)}\left(\mathbf{Q}(\boldsymbol{\tau}:\boldsymbol{\tau}),\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right)\middle|\,\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right]\right]\\
&=_{(b)} \mathbf{E}\left[\mathbf{E}\left[C_{(m,n)}(0:0)\left|C_{(m,n)}(\tau_{(m,n)}:\tau_{(m,n)})\right.\right]\mathbf{E}\left[D^{\mathcal{P}_s}_{(m,n)}\left(\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right)\middle|\,\mathbf{C}(\boldsymbol{\tau}:\boldsymbol{\tau})\right]\right]\\
&= \mathbf{E}\left[S^{\mathcal{P}_s}_{(m,n)}[\infty]\right] \quad (7)
\end{aligned}$$

so we have

$$\begin{aligned}&\mathbf{E}[V[n+1]-V[n]\,|\,\tilde{\mathbf{Y}}[n]]\\&=\mathbf{E}\left[\sum_{m,d}(Q_{\langle m,d\rangle}[(n+1)T]-Q_{\langle m,d\rangle}[nT])\times\right.\\&\qquad\left.(Q_{\langle m,d\rangle}[(n+1)T]+Q_{\langle m,d\rangle}[nT])\middle|\tilde{\mathbf{Y}}[n]\right].\end{aligned}$$

Note that $Q_{\langle m,d\rangle}[t]-Q_{\langle m,d\rangle}[s]\leq|t-s|(A_{\max}+C_{\max})$. Thus, there exists a constant K, which depends on T, but independent of $\mathbf{Q}[nT]$, such that (10) (see the bottom of the page) holds.

Now we consider

$$\mathbf{E}\left[\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(Q_{\langle m,d\rangle}[t+1]-Q_{\langle m,d\rangle}[t]\right)\middle|\mathbf{Q}[t-\boldsymbol{\tau}],\tilde{\mathbf{Y}}[n]\right]$$

and obtain (11) (shown at the bottom of the page), where $\mathcal{M}^*[t]$ and $d^*_{(m,n)}[t]$ are as defined in the on-off scheduling algorithm.

Since $(1+\epsilon)\mathbf{A}[t]$ is supportable, we have $(1+\epsilon)\mathbf{E}[\mathbf{A}[t]]\in\Lambda_\tau$. According to Lemma 5 and the definition of Λ_τ [(6)]), there exists a set of $\bar{\boldsymbol{\eta}}(\mathbf{c})$ such that $\bar{\boldsymbol{\eta}}(\mathbf{c})\in\eta(\mathbf{C}[0]=\mathbf{c})$ for all $\mathbf{c}$ and

$$\sum_{\mathbf{c}\in\mathcal{C}^{|\mathcal{L}|}}\Pr(\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c})\left(\sum_{d\in\mathcal{N}}(1+\epsilon)a_{[m,d]}1_{[m,d]\in\mathcal{F}}+\sum_{(k,m)\in\mathcal{L}}\bar{\eta}_{(k,m)}(\mathbf{c})-\sum_{(m,n)\in\mathcal{L}}\bar{\eta}_{(m,n)}(\mathbf{c})\right)\leq 0 \quad (12)$$

for all $m,d\in\mathcal{N}$.

Now consider t such that $t\geq T_m+\max_{\langle m,d\rangle}\tau_{\langle m,d\rangle}$. According to the definition of T_m, and (12), we can obtain (13), shown at the bottom of the next page.

To that end, we can finally obtain (14) (at the bottom of the next page) where the last inequality holds due to the definition of T. Finally, we can obtain the theorem by invoking the Foster's Criterion [42]. ■

VII. Decentralized Routing/Scheduling: Detailed Analysis

In this section, we study the decentralized routing/scheduling. We only present the detailed analysis for NSI structure $S2$ since the analysis of NSI structure $S1$ is similar.

A. Throughput Region

First, given a threshold vector $\mathbf{T}=\{T_{(m,n)}\}_{(m,n)\in\mathcal{L}}$ and a probability vector $\mathbf{p}=\{p_{(m,n)}\}_{(m,n)\in\mathcal{L}}$, we define a threshold

$$\begin{aligned}&\mathbf{E}[V[(n+1)T]-V[nT]\,|\,\tilde{\mathbf{Y}}[n]]\\&\leq K+2\mathbf{E}\left[\sum_{t=nT}^{(n+1)T-1}\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(Q_{\langle m,d\rangle}[t+1]-Q_{\langle m,d\rangle}[t]\right)\middle|\tilde{\mathbf{Y}}[n]\right]\\&=K+2\sum_{t=nT}^{(n+1)T-1}\mathbf{E}\left[\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(Q_{\langle m,d\rangle}[t+1]-Q_{\langle m,d\rangle}[t]\right)\middle|\tilde{\mathbf{Y}}[n]\right]\\&=K+2\sum_{t=nT}^{(n+1)T-1}\mathbf{E}\left[\mathbf{E}\left[\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(Q_{\langle m,d\rangle}[t+1]-Q_{\langle m,d\rangle}[t]\right)\middle|\mathbf{Q}[t-\boldsymbol{\tau}]\right]\middle|\tilde{\mathbf{Y}}[n]\right]\end{aligned} \quad (10)$$

$$\begin{aligned}&\mathbf{E}\left[\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(Q_{\langle m,d\rangle}[t+1]-Q_{\langle m,d\rangle}[t]\right)\middle|\mathbf{Q}[t-\boldsymbol{\tau}],\tilde{\mathbf{Y}}[n]\right]\\&=\mathbf{E}\left[\mathbf{E}\left[\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(Q_{\langle m,d\rangle}[t+1]-Q_{\langle m,d\rangle}[t]\right)\middle|\mathbf{C}[t-\boldsymbol{\tau}]\right]\middle|\mathbf{Q}[t-\boldsymbol{\tau}],\tilde{\mathbf{Y}}[n]\right]\\&=\sum_{\mathbf{c}\in\mathcal{C}^{|\mathcal{L}|}}\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(a_{[m,d]}1_{[m,d]\in\mathcal{F}}+\sum_{(k,m)\in\mathcal{L}}\mathbf{E}[C_{(k,m)}[t]\,|\,C_{(k,m)}\left[t-\tau_{(k,m)}\right]]1_{(k,m)\in\mathcal{M}^*[t],d^*_{(k,m)}[t]=d}\right.\\&\qquad\left.-\sum_{(m,n)\in\mathcal{L}}\mathbf{E}\left[C_{(m,n)}[t]\middle|C_{(m,n)}\left[t-\tau_{(m,n)}\right]\right]1_{(m,n)\in\mathcal{M}^*[t],d^*_{(m,n)}[t]=d}\right)\Pr\left(\mathbf{C}[t-\boldsymbol{\tau}]=\mathbf{c}\,|\,\mathbf{Q}[t-\boldsymbol{\tau}],\tilde{\mathbf{Y}}[n]\right).\end{aligned} \quad (11)$$

policy $\mathcal{T}(\mathbf{T},\mathbf{p})$ as follows:

Threshold Policy $\mathcal{T}(\mathbf{T},\mathbf{p})$:

(i) If $C_{(m,n)}[t] > T_{(m,n)}$, node m transmits over link (m,n) at time slot t.

(ii) If $C_{(m,n)}[t] = T_{(m,n)}$, node m transmits over link (m,n) with probability $p_{(m,n)}$ at time slot t.

(iii) If $C_{(m,n)}[t] < T_{(m,n)}$, node m keeps silent at time slot t.

Next, we denote by $S_{(m,n)}(\mathcal{T}(\mathbf{T},\mathbf{p}))[t]$ the achievable rate over link (m,n) at time slot t under the threshold policy $\mathcal{T}(\mathbf{T},\mathbf{p})$. Note that $S_{(m,n)}(\mathcal{T}(\mathbf{T},\mathbf{p}))[t] = C_{(m,n)}[t]$ if no link in $\mathcal{I}_{(m,n)}$ is active simultaneously; and $S_{(m,n)}(\mathcal{T}(\mathbf{T},\mathbf{p}))[t] = 0$ otherwise. Given $\mathbf{C}[t-\tau] = \mathbf{c}$, the expected achievable rate over link (m,n) is

$$\mathbf{E}\left[S_{(m,n)}(\mathcal{T}(\mathbf{T},\mathbf{p}))[t]\middle|\,\mathbf{C}[t-\tau]=\mathbf{c}\right].$$

Then, we define $\boldsymbol{\eta}(\mathbf{c})$ to be the convex hull of $\mathbf{E}[\mathbf{S}(\mathcal{T}(\mathbf{T},\mathbf{p}))[t]\,|\,\mathbf{C}[t-\tau]=\mathbf{c}]$ over all $\mathbf{T}$ and $\mathbf{p}$, i.e.

$$\boldsymbol{\eta}(\mathbf{c}) = \mathcal{CH}_{\mathbf{T},\mathbf{p}}(\mathbf{E}[\mathbf{S}(\mathcal{T}(\mathbf{T},\mathbf{p}))[t]\,|\,\mathbf{C}[t-\tau]=\mathbf{c}]).$$

Furthermore, we define $\tilde{\Lambda}_\tau$ such that

$$\tilde{\Lambda}_\tau = \left\{\boldsymbol{\eta} : \boldsymbol{\eta} = \sum_{\mathbf{c}\in\mathcal{C}^{|\mathcal{L}|}} \Pr(\mathbf{C}[t-\tau]=\mathbf{c})\boldsymbol{\eta}_{\mathbf{c}}, \boldsymbol{\eta}_{\mathbf{c}} \in \boldsymbol{\eta}(\mathbf{c})\right\}.$$

Next define

$$\mathbf{Y}[t] = (\mathbf{C}[t](0:\tau+\tau_f), \mathbf{Q}[t](0:\tau+\tau_f)).$$

In the following lemma, we show that $\mathbf{Y}[t]$ is Markovian.

Lemma 6: Assume queues evolve as described in (5), then $\mathbf{Y}[t]$ is Markovian.

Proof: First, it is easy to see that

$$\begin{aligned}\Pr(\mathbf{C}[t+1]\,|\,\mathbf{Y}[t](0:t)) &= \Pr(\mathbf{C}[t+1]\,|\,\mathbf{C}[t]) \\ &= \Pr(\mathbf{C}[t+1]\,|\,\mathbf{Y}[t]). \quad (15)\end{aligned}$$

Next, from queue dynamics (5), we can see that $\mathbf{Q}[t+1]$ is a function of $\mathbf{A}[t]$, $\mathbf{Q}[t]$, and $\mathbf{S}[t-\tau_f]$. Note that $\mathbf{A}[t]$ is independent across time, $\mathbf{Q}[t]$ is included in $\mathbf{Y}[t]$, and $\mathbf{S}[t-\tau_f]$ is determined by $\mathbf{Q}[t](\tau_f : \tau_f+\tau)$ and $\mathbf{C}[t](\tau_f : \tau_f+\tau)$, which are also included in $\mathbf{Y}[t]$. Thus, we can conclude that

$$\Pr(\mathbf{Q}[t+1]\,|\,\mathbf{Y}[t](0:t)) = \Pr(\mathbf{Q}[t+1]\,|\,\mathbf{Y}[t]). \quad (16)$$

From (15) and (16), we can conclude that

$$\Pr(\mathbf{Y}[t+1]\,|\,\mathbf{Y}[t](0:t)) = \Pr(\mathbf{Y}[t+1]\,|\,\mathbf{Y}[t]),$$

and the lemma holds. ∎

In the next lemma, we will show that given traffic $\mathbf{A}[t]$, if the network can be stabilized by a scheduling algorithm $\mathcal{P}$, then there exists a time-sharing policy $\mathcal{P}_s$ that stabilizes the network as well. This is analogous to the development of the

$$\begin{aligned}
&\mathbf{E}\left[\sum_{m,d} Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(Q_{\langle m,d\rangle}[t+1]-Q_{\langle m,d\rangle}[t]\right)\middle|\,\mathbf{Q}[t-\boldsymbol{\tau}],\tilde{\mathbf{Y}}[n]\right]\\
&\le 2\sum_{m,d}(1-\epsilon/2)Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]\left(a_{[m,d]}1_{[m,d]\in\mathcal{F}} + \sum_{(k,m)\in\mathcal{L}}\bar{\eta}_{(k,m)}\left(\mathbf{C}[t-\boldsymbol{\tau}]\right) - \sum_{(m,n)\in\mathcal{L}}\bar{\eta}_{(m,n)}\left(\mathbf{C}[t-\boldsymbol{\tau}]\right)\right)\\
&\quad -2(1-\epsilon/2)\sum_{(m,n),d}\left(\mathbf{E}\left[C_{(m,n)}[t]\middle|\,C_{(m,n)}\left[t-\tau_{(m,n)}\right]\right]1_{(m,n)\in\mathcal{M}^*[t],d^*_{(m,n)}[t]=d} - \bar{\eta}_{(m,n)}\left(\mathbf{C}[t-\boldsymbol{\tau}]\right)\right)\times\\
&\qquad\left(Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right] - Q_{\langle n,d\rangle}[t-\tau_{\langle n,d\rangle}]\right)\\
&\le -2\epsilon(1-\epsilon/2)\sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]a_{[m,d]\in\mathcal{F}}. \qquad (13)
\end{aligned}$$

$$\begin{aligned}
&\mathbf{E}[V[(n+1)T]-V[nT]|\tilde{\mathbf{Y}}[n]]\\
&\le K - 2\epsilon(1-\epsilon/2)\sum_{t=nT+T_m+\max_{\langle m,d\rangle}\tau_{\langle m,d\rangle}}^{(n+1)T-1}\ \sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right]a_{[m,d]\in\mathcal{F}} +\\
&\quad\sum_{t=nT}^{nT+T_m+\max_{\langle m,d\rangle}\tau_{\langle m,d\rangle}-1}\ \sum_{m,d}Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right](A_{\max+C_{\max}})\\
&\le K - 2\epsilon(1-\epsilon/2)(T-T_m-\max_{\langle m,d\rangle}\tau_{\langle m,d\rangle})\sum_{m,d}Q_{\langle m,d\rangle}[nT] + (T_m+\max_{\langle m,d\rangle}\tau_{\langle m,d\rangle})(A_{\max+C_{\max}})\sum_{m,d}Q_{\langle m,d\rangle}[nT]\\
&\le K - \epsilon T\sum_{m,d}Q_{\langle m,d\rangle}[nT] \qquad (14)
\end{aligned}$$

time-sharing policy in Section VI. To demonstrate this, we first define a time-sharing policy for a given stabilizing $\mathcal{P}$.

Note that if $\mathbf{A}[t]$ is supportable under $\mathcal{P}$, then $\mathbf{Y}^{\mathcal{P}}[t]$ is positive recurrent. Thus, we can define

$$\begin{aligned} & (\mathbf{C}(0:\tau+\tau_f), \mathbf{Q}^{\mathcal{P}}(0:\tau+\tau_f)) \\ \triangleq & \lim_{t\to\infty} \left(\mathbf{C}[t](0:\tau+\tau_f), \mathbf{Q}^{\mathcal{P}}[t](0:\tau+\tau_f)\right) \\ = & \lim_{t\to\infty} \mathbf{Y}^{\mathcal{P}}[t]. \end{aligned} \tag{17}$$

Next we define a time-sharing policy related to $\mathcal{P}$.

Time-sharing Policy $\mathcal{P}_s$: Given delayed information $\mathbf{C}[t-\tau]=\mathbf{c}$, we let

$$r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}} = \Pr(\mathbf{Q}^{\mathcal{P}}(0:\tau)=\mathbf{q}(0:\tau)\,|\,\mathbf{C}(\tau:\tau)=\mathbf{c}).$$

Then, at each time when the delayed channel state is $\mathbf{C}[t-\tau]=\mathbf{c}$, the policy $\mathcal{P}^s$ probabilistically make decisions (denoted by $\mathcal{D}^{\mathcal{P}_s}$) as follows: Let $\mathcal{Q}=\{\mathbf{q}: r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}}>0\}$. The policy $\mathcal{P}^s$ selects a $\mathbf{q}\in\mathcal{Q}$ with probability $r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}}$, and makes the corresponding decision be given by

$$\begin{aligned} &\mathbf{D}^{\mathcal{P}_s}(\mathbf{C}[t](0:\tau-1), \mathbf{C}[t-\tau]=\mathbf{c}) = \\ &\mathbf{D}^{\mathcal{P}}(\mathbf{C}[t](0:\tau-1), \mathbf{C}[t-\tau]=\mathbf{c}, \mathbf{Q}[t](0:\tau)=\mathbf{q}(0:\tau)). \end{aligned}$$

Note that this time-sharing policy is analogous to that described in Section VI, in the sense that it generates the decisions using a known stabilizing policy $\mathcal{P}$ by "feeding" it inputs which are randomly chosen (according to the stationary conditional distribution $r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}}$).

Example: Consider the two-user example studied in Section V-A, where, at each of the nodes, there is one-time-slot delay in the channel and queue state information from the other node. In this case, suppose that for given arrivals and channel model (i.e., Markov chain transition probabilities), a stabilizing queue and channel state based scheduling policy exists (denoted by $\mathcal{P}$) that makes transmission attempt decisions as described below. Note that $\mathcal{P}$ makes decision based on the common delayed NSI $(C_{(1,0)}[t-1], C_{(2,0)}[t-1])$ and $(Q_{(1,0)}[t-1], Q_{(2,0)}[t-1])$, along with the *local* current channel state information at each of the nodes[6])

For the delayed channel state $C_{(1,0)}[t-1]=C_{(2,0)}[t-1]=50$, the stabilizing policy $\mathcal{P}$ makes decisions as follows:

(a) If $Q_{(1,0)}[t-1]>4Q_{(2,0)}[t-1]$, node 1 attempts to transmit if its channel is nonzero (in this case, if $C_{(1,0)}[t]=50$), and remains silent if it's current channel is zero (i.e., $C_{(1,0)}[t]=0$). Node 2 remains silent irrespective of its current local channel state $(C_{(2,0)}[t])$. Note that in this case, it is possible that neither of the nodes transmit even though a *centralized scheduler* that knows all the channels at each time might make a transmission attempt (this happens when $(C_{(1,0)}[t]=0, C_{(2,0)}[t]=50)$).

(b) If $Q_{(2,0)}[t-1]>4Q_{(1,0)}[t-1]$, node 2 attempts to transmit and node 1 remains silent. The above decision happens irrespective of the current local channel states.

(c) Finally, if $Q_{(1,0)}[t-1]/Q_{(2,0)}[t-1]\in[0.25,4]$, node i $(i=1,2)$ attempts to transmit if $C_{(i,0)}[t]=50$. In this case, a collision could occur if both the (local) channels are at state 50.

Similarly, suppose that for each of the other cases of delayed channel state (i.e., for each possible tuple of values of the random variables $(C_{(1,0)}[t-1], C_{(2,0)}[t-1])$), the policy $\mathcal{P}$ can be described.

Further, suppose that under this policy $\mathcal{P}$ the following holds in steady-state (stationarity):

$$\Pr\left(\frac{Q^{\mathcal{P}}_{(1,0)}(1:1)}{Q^{\mathcal{P}}_{(2,0)}(1:1)}>4\,\middle|\,\mathbf{C}(1:1)=(50,50)\right)=0.3$$

$$\Pr\left(\frac{Q^{\mathcal{P}}_{(1,0)}(1:1)}{Q^{\mathcal{P}}_{(2,0)}(1:1)}<\frac{1}{4}\,\middle|\,\mathbf{C}(1:1)=(50,50)\right)=0.3$$

$$\Pr\left(\frac{Q^{\mathcal{P}}_{(2,0)}(1:1)}{Q^{\mathcal{P}}_{(1,0)}(1:1)}\in[0.25,4]\,\middle|\,\mathbf{C}(1:1)=(50,50)\right).=0.4.$$

Similarly, we can describe the conditional probabilities for other values of the delayed channel states. We recall that in the context of this example, $\mathbf{C}(1:1)$ corresponds to the channel state of the pair of users with one-time-slot delay under system stationarity. Then, the corresponding time-sharing policy is as follows:

Time-sharing Policy: For each channel state $(C_{(1,0)}[t-1], C_{(2,0)}[t-1])$, the time-sharing policy $\mathcal{P}_s$ will make a randomized decision depending on the decision taken by $\mathcal{P}$. Suppose that $C_{(1,0)}[t-1]=C_{(2,0)}[t-1]=50$. Then, under the time-sharing policy $\mathcal{P}_s$,

(a) With probability 0.3, node 1 attempts to transmit and node 2 keeps silent.

(b) With probability 0.3, node 2 attempts to transmit and node 1 keeps silent.

(c) With probability 0.4, both nodes $(i=1,2)$ attempt to transmit if $C_{(i,0)}[t]=50$ (note that in this case, collision could occur).

Similarly, the time-sharing policy can be described for various other values of the delayed channel states.

In the following lemma, we will show that the link rates under $\mathcal{P}$ is the same as the service rate obtained under the corresponding $\mathcal{P}_s$.

Lemma 7: Given policy $\mathcal{P}$ which stabilizes the network with traffic $\mathbf{A}[t]$, we have that

$$\lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}}_{(m,n)}[t]\right]=\lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}_s}_{(m,n)}[t]\right]$$

where $\mathcal{P}_s$ is the corresponding time-sharing policy.

Proof: First, we can obtain (18), shown in the first equation at the bottom of the next page, where the second equality yields from the definition of the time-sharing policy.

Note that $Q_{(m,n)}[t]$ is a function of $A_{(m,n)}[t](0:t)$ and $S_{(m,n)}[t](0:\tau_f)$, and $S_{(m,n)}[t-\tilde{\tau}]$ is determined by $\mathbf{Q}[t-\tilde{\tau}](0:\tau)$ and $\mathbf{C}[t-\tilde{\tau}](0:\tau)$. Since $\tau_f>\tau$, we can conclude that, conditioned on $C_{(m,n)}[t-\tau]$, $C_{(m,n)}[t-r]$ $(r\le\tau)$ is independent of $Q_{(l,k)}[s]$ for any $(l,k)\in\mathcal{L}$ and $s\le t$. In other words, under policy $\mathcal{P}$

$$\begin{aligned} &\Pr(\mathbf{C}(0:\tau-1)\,|\,\mathbf{C}(\tau:\tau), \mathbf{Q}^{\mathcal{P}}(0:\tau)) \\ &=\Pr(\mathbf{C}(0:\tau-1)\,|\,\mathbf{C}(\tau:\tau)). \end{aligned} \tag{19}$$

[6]While policy $\mathcal{P}$ can potentially use the *current local* queue length information at each of the nodes, in this example, it chooses not to do so.

Thus, we can further obtain that

$$\lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}}_{(m,n)}[t]\middle|\mathbf{C}[t-\tau]\right]=\lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}_s}_{(m,n)}[t]\middle|\mathbf{C}[t-\tau]\right]$$

as demonstrated in (20), at the bottom of the page, where the last equality holds due to (19). ∎

Now, given $(\mathbf{Q}^{\mathcal{P}}(0:\tau)=\mathbf{q}(0:\tau),\mathbf{C}(\tau:\tau)=\mathbf{c})$, there exist $\mathbf{T}(\mathbf{c},\mathbf{q}(0:\tau))$ and $\mathbf{p}(\mathbf{c},\mathbf{q}(0:\tau))$ such that (21) (the equation at the bottom of the page) holds for all $(m,n)\in\mathcal{L}$. In other words, there exist $\mathbf{T}(\mathbf{c},\mathbf{q}(0:\tau))$ and $\mathbf{p}(\mathbf{c},\mathbf{q}(0:\tau))$ such that, given the queue lengths and delayed channel states, the probability the channel state of (m,n) is greater than or equal to the threshold is the same as the probability that link (m,n) is active under the time-sharing policy. Then, we can further define a threshold-based time-sharing policy.

Threshold-Based Time-sharing Policy $\mathcal{P}_{ts}$: Given delayed channel-states $\mathbf{C}[t-\tau]=\mathbf{c}$, we let

$$r^{\mathcal{P}}_{\mathbf{q}\,|\,\mathbf{c}}=\Pr(\mathbf{Q}^{\mathcal{P}}(0:\tau)=\mathbf{q}(0:\tau)\,|\,\mathbf{C}(\tau:\tau)=\mathbf{c}).$$

Then, at each time when the delayed channel state is $\mathbf{C}[t-\tau]=\mathbf{c}$, the policy $\mathcal{P}_{ts}$ probabilistically make decisions (denoted by

$$\begin{aligned}
&\mathbf{E}\left[S^{\mathcal{P}_s}_{(m,n)}[t]\,|\,\mathbf{C}[t-\tau]=\mathbf{c}\right]\\
&=\mathbf{E}\Bigg[C_{(m,n)}[t]D^{\mathcal{P}_s}_{(m,n)}\left(C_{(m,n)}[t](0:\tau-1),\mathbf{C}[t-\tau]\right)\times\\
&\qquad\prod_{(l,k)\in\mathcal{I}_{(m,n)}}\left(1-D^{\mathcal{P}_s}_{(l,k)}\left(C_{(l,k)}[t](0:\tau-1),\mathbf{C}[t-\tau]\right)\right)\Bigg|\mathbf{C}[t-\tau]=\mathbf{c}\Bigg]\\
&=\sum_{\mathbf{q}(0:\tau)}\Pr(\mathbf{Q}^{\mathcal{P}}(0:\tau)=\mathbf{q}(0:\tau)\,|\,\mathbf{C}(\tau:\tau)=\mathbf{c})\times\\
&\quad\mathbf{E}\Bigg[C_{(m,n)}[t]D^{\mathcal{P}}_{(m,n)}\left(C_{(m,n)}[t](0:\tau-1),\mathbf{C}[t-\tau],\mathbf{Q}[t](0:\tau)=\mathbf{q}(0:\tau)\right)\times\\
&\qquad\prod_{(l,k)\in\mathcal{I}_{(m,n)}}\left(1-D^{\mathcal{P}}_{(l,k)}\left(C_{(l,k)}[t](0:\tau-1),\mathbf{C}[t-\tau],\mathbf{Q}[t](0:\tau)=\mathbf{q}(0:\tau)\right)\right)\Bigg|\mathbf{C}[t-\tau]=\mathbf{c}\Bigg].
\end{aligned}\tag{18}$$

$$\begin{aligned}
&\lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}_s}_{(m,n)}[t]\middle|\mathbf{C}[t-\tau]=\mathbf{c}\right]\\
&=\sum_{\mathbf{q}(0:\tau)}\Pr(\mathbf{Q}^{\mathcal{P}}(0:\tau)=\mathbf{q}(0:\tau)\,|\,\mathbf{C}(\tau:\tau)=\mathbf{c})\mathbf{E}\Bigg[C_{(m,n)}(0:0)D^{\mathcal{P}}_{(m,n)}\left(C_{(m,n)}(0:\tau-1),\mathbf{C}(\tau:\tau),\mathbf{Q}(0:\tau)\right)\times\\
&\qquad\prod_{(l,k)\in\mathcal{I}_{(m,n)}}\left(1-D^{\mathcal{P}}_{(l,k)}\left(C_{(l,k)}(0:\tau-1),\mathbf{C}(\tau:\tau),\mathbf{Q}(0:\tau)\right)\right)\Bigg|\mathbf{C}(\tau:\tau)=\mathbf{c},\mathbf{Q}(0:\tau)=\mathbf{q}(0:\tau)\Bigg]\\
&=\lim_{t\to\infty}\mathbf{E}\left[\mathbf{E}\left[S^{\mathcal{P}}_{(m,n)}[t]\middle|\mathbf{Q}^{\mathcal{P}}(0:\tau)\right]\middle|\mathbf{C}[t-\tau]=\mathbf{c}\right]\\
&=\lim_{t\to\infty}\mathbf{E}\left[S^{\mathcal{P}}_{(m,n)}[t]\middle|\mathbf{C}[t-\tau]=\mathbf{c}\right].
\end{aligned}\tag{20}$$

$$\begin{aligned}
&\Pr\left(C_{(m,n)}(0:0)>T_{(m,n)}(\mathbf{c},\mathbf{q}(0:\tau))\middle|\mathbf{C}(\tau:\tau)=\mathbf{c}\right)\\
&+p_{(m,n)}(\mathbf{c},\mathbf{q}(0:\tau))\Pr\left(C_{(m,n)}(0:0)=T_{(m,n)}(\mathbf{c},\mathbf{q}(0:\tau))\middle|\mathbf{C}(\tau:\tau)=\mathbf{c}\right)=\\
&\Pr\left(D^{\mathcal{P}}_{(m,n)}\left(C_{(m,n)}(0:\tau-1),\mathbf{C}(\tau:\tau),Q_{\langle m,n\rangle}(0:\tau-1)=q_{\langle m,n\rangle}(0:\tau-1),\mathbf{Q}(\tau:\tau)=\mathbf{q}(\tau:\tau)\right)=1\middle|\mathbf{C}(\tau:\tau)=\mathbf{c}\right).
\end{aligned}\tag{21}$$

$\mathcal{D}^{\mathcal{P}_{ts}}$) as follows: Let $\mathcal{Q} = \{\mathbf{q} : r^{\mathcal{P}}_{\mathbf{q}|\mathbf{c}} > 0\}$. Policy $\mathcal{P}_{ts}$ chooses $\mathbf{q} \in \mathcal{Q}$, with probability $r^{\mathcal{P}}_{\mathbf{q}|\mathbf{c}}$, and lets the corresponding decision be given by

$$D^{\mathcal{P}_{ts}}_{(m,n)}\left(C_{(m,n)}[t](0:\tau-1), \mathbf{C}[t-\tau]=\mathbf{c}\right) = 1_{C_{(m,n)}[t] > T_{(m,n)}(\mathbf{c},\mathbf{q}(0:\tau))} + z 1_{C_{(m,n)}[t] = T_{(m,n)}(\mathbf{c},\mathbf{q}(0:\tau))}$$

where

$$\Pr(z=1) = p_{(m,n)}(\mathbf{c}, \mathbf{q}(0:\tau)).$$

In other words, for each delayed channel state $\mathbf{c}$, the policy $\mathcal{P}_{ts}$ first randomly (with probability $r^{\mathcal{P}}_{\mathbf{q}|\mathbf{c}}$) picks a possible stationary queue length *under policy* $\mathcal{P}$. Then the policy uses the resulting pair $(\mathbf{q}, \mathbf{c})$ to determine the threshold $T_{(m,n)}(\mathbf{c}, \mathbf{q})$ and the probability $p_{(m,n)}(\mathbf{c}, \mathbf{q})$, and uses the threshold and the probability to make the transmission decision.

In the following lemma, we will show that considering $\mathcal{P}_s$ and $\mathcal{P}_{ts}$, which are related to the same stabilizing policy $\mathcal{P}$, the service rate obtained by link (m,n) under the threshold time-sharing policy is no less than the one under the time sharing policy.

Lemma 8: Assuming that $\mathcal{P}_s$ and $\mathcal{P}_{ts}$ are the time-sharing policy and the threshold-based time-sharing policy defined based the stabilizing policy $\mathcal{P}$, then

$$\lim_{t\to\infty} \mathbf{E}\left[S^{\mathcal{P}_s}_{(m,n)}[t]\right] \leq \lim_{t\to\infty} \mathbf{E}\left[S^{\mathcal{P}_{ts}}_{(m,n)}[t]\right].$$

Proof: From (21), we can obtain (22) (see the lower equation at the bottom of the page), holds for any link (m,n). We note that (22) yields from the assumptions: (i) independence of channel fading and (ii) collision-type interference, i.e., two simultaneous transmissions either succeed or collide. Therefore the interference of a node to other nodes only depends on the probability that the node is on, and is independent of the specific channel states in which the node is on. Combining (21) and (22), we can conclude that the lemma holds. ∎

Lemma 9: Considering NSI structure $S2$, traffic $\mathbf{A}[t]$ is supportable if and only if $\mathbf{A}[t] \in \tilde{\Lambda}_\tau$.

Proof: From Lemma 7 and 8, we can conclude that

$$\mathbf{E}\left[A_{(m,n)}[\infty]\right] \leq \lim_{t\to\infty} \mathbf{E}\left[S^{\mathcal{P}_{ts}}_{(m,n)}[\infty]\right]$$

if $\mathbf{A}[t]$ is supportable. From the definition of the threshold-based time-sharing policy, we also have that

$$\mathbf{E}[\mathbf{S}^{\mathcal{P}_{ts}}[\infty]] \in \tilde{\Lambda}_\tau,$$

which implies that $\mathbf{E}[\mathbf{A}[t]] \in \tilde{\Lambda}_\tau$ if $\mathbf{A}[t]$ is supportable.

Now assume $\mathbf{E}[\mathbf{A}[t]] \in \tilde{\Lambda}_\tau$, then from the definition of $\tilde{\Lambda}_\tau$, there exists a threshold time-sharing policy that stabilizes the network. ∎

B. Throughput Optimality

Next we show the threshold routing/scheduling algorithm is throughput-optimal for case (ii).

Theorem 2: Given NSI structure $S2$, and traffic $\mathbf{A}[t]$ such that $(1+\epsilon)\mathbf{E}[\mathbf{A}[t]] \in \tilde{\Lambda}_\tau$, the network is stochastically stable under the threshold-based routing/scheduling algorithm.

Proof: Define $\tilde{Q}_{\langle m,d\rangle}[t]$ to be the number of packets that are queued at node m, but have not been delivered to node d. Define a Lyapunov function

$$V[nT] = \sum_{(m,n)} \tilde{Q}^2_{\langle m,n\rangle}[nT].$$

Similar to the analysis in Theorem 1, we can obtain that there exists $K > 0$ and $T > 0$ such that

$$\mathbf{E}[V[(n+1)T] - V[nT] \mid \mathbf{Y}[nT]] \leq -\epsilon \sum_{(m,n)} a_{(m,n)} \tilde{Q}_{\langle m,n\rangle}[nT] 1_{(m,n)\in\mathcal{F}} + K,$$

and the theorem holds. ∎

VIII. Discussions and Conclusion

In this paper, we studied the impact of delayed NSI on network throughput. We provided the relations between the delays in NSI and the network throughput region. We also developed throughput-optimal scheduling algorithms that incorporate the delayed NSI, both in the context of centralized and decentralized scheduling. In the decentralized scenario, we characterized the impact for two special cases (homogeneous delayed NSI from other nodes, with flow/delay restrictions). Our results were obtained based on several assumptions. Next, we will discuss the limitations of our model and several possible extensions.

A. Correlated Channels

We assume channels are independent across links in this paper. For the centralized case, our result can be easily extended to the networks with correlated channels. It can be shown that the following algorithm is throughput optimal.

On-Off Routing/Scheduling: At time slot t,

(1) The controller first computes the optimal independent-link-set $\mathcal{M}^*[t]$ which maximizes

$$\sum_{(m,n)\in\mathcal{M}} \mathbf{E}\left[C_{(m,n)}[t] \middle| \mathbf{C}[t-\boldsymbol{\tau}]\right] P_{(m,n)}[t]$$

$$\mathbf{E}\left[C_{(m,n)}[t] 1_{C_{(m,n)}[t] > T_{(m,n)}(\mathbf{c},\mathbf{q}(0:\tau))} + C_{(m,n)}[t] z 1_{C_{(m,n)}[t] > T_{(m,n)}(\mathbf{c},\mathbf{q}(0:\tau))} \middle| \mathbf{C}(\tau:\tau) = \mathbf{c}\right] \geq$$
$$\mathbf{E}\left[C_{(m,n)}[t] D^{\mathcal{P}}_{(m,n)}\left(C_{(m,n)}(0:\tau-1), \mathbf{C}(\tau:\tau), Q_{\langle m,n\rangle}(0:\tau-1) = q_{\langle m,n\rangle}(0:\tau-1), \mathbf{Q}(\tau:\tau) = \mathbf{q}(\tau:\tau)\right) \middle| \mathbf{C}(\tau:\tau) = \mathbf{c}\right]. \quad (22)$$

where

$$P_{(m,n)}[t] = \max_{d\in\mathcal{N}} \left(Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right] - Q_{\langle n,d\rangle}\left[t-\tau_{\langle n,d\rangle}\right]\right)$$

(2) Node m transmits the packets from queue $d^*_{(m,n)}[t]$ over link (m,n) with a rate $C_{(m,n)}[t]$ if $(m,n)\in\mathcal{M}^*[t]$, where

$$d^*_{(m,n)}[t] \in \arg\max_d \left(Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right] - Q_{\langle n,d\rangle}\left[t-\tau_{\langle n,d\rangle}\right]\right).$$

Note that the key modification is to compute the expected channel rate based on the complete delayed information $\mathbf{C}[t-\boldsymbol{\tau}]$ instead of $C_{(m,n)}[t-\tau_{(m,n)}]$. The throughput-optimality proof follows the analysis in Section VI with minor modifications. For example, in (a) of (7), $\mathbf{E}[C_{(m,n)}(0:0)\,|\,C_{(m,n)}(\tau_{(m,n)}:\tau_{(m,n)})]$ should be replaced by $\mathbf{E}[C_{(m,n)}(0:0)\,|\,\mathbf{C}(\tau:\tau)]$.

For the decentralized case, the following algorithm can be shown to be throughput optimal:

Routing/Scheduling for Correlated Channels: At time slot t,

(1) All nodes first compute scheduling decision functions $\mathbf{D}^*(\mathbf{C}[t],\mathbf{C}[t-\tau]) = \{D^*_{(m,n)}(C_{(m,n)}[t],\mathbf{C}[t-\tau])\}_i$ by solving the optimization problem (23), shown at the bottom of the page, where $D^*_{(m,n)}(C_{(m,n)}[t],\mathbf{C}[t-\tau])\in\{0,1\}$.

(2) Node m attempts to transmit packets from queue $d^*_m[t]$ over link (m,n) if $D^*_{(m,n)}(C_{(m,n)}[t],\mathbf{C}[t-\tau])=1$, where

$$d^*_{(m,n)}[t] \in \arg\max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right).$$

The throughput optimality of this algorithm can be proved by showing: (i) any traffic load that is strictly within the throughput region can be supported by a time-sharing policy (*not a threshold-based time-sharing policy*), and then (ii) the proposed algorithm guarantees $\mathbf{E}[V[(n+1)T]-V[nT]\,|\,\mathbf{Y}(nT)]<0$ for properly chosen $\mathbf{V}$ and $\mathbf{Y}$ if it is negative under some time-sharing policy. This algorithm however does not have the threshold structure as the threshold-based routing/scheduling algorithm, so the optimization problem (23) is much harder to solve. The independent channel assumption is critical for proving that the threshold-based routing/scheduling algorithm is throughput optimal as we explained in the proof of Lemma 8.

B. Packet Capture Model

We assumed a specific collision model in this paper. Our results, however, can be extended to more general interference models such as partial reception channels: when a collision happens, node m can transmit to node n with a rate $\alpha_{(m,n)}C_{(m,n)}[t]$, where $\alpha_{(m,n)}$ can be different for different links. Given this partial reception channel model, the following algorithm is throughput-optimal for the centralized scheduling case:

On-Off Routing/Scheduling for Partial Reception Channel: At time slot t,

(1) The controller first computes the optimal link-set $\mathcal{M}^*[t]$ which maximizes (24).

(2) Node m transmits the packets from queue $d^*_{(m,n)}[t]$ over link (m,n) with a rate $C_{(m,n)}[t]$ if $(m,n)\in\mathcal{M}^*[t]$, where

$$d^*_{(m,n)}[t] \in \arg\max_d \left(Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right] - Q_{\langle n,d\rangle}\left[t-\tau_{\langle n,d\rangle}\right]\right).$$

We can also show that given the decentralized NSI with structure $S1$ and $S2$, the following routing/scheduling algorithm is throughput optimal.

Threshold-Based Routing/Scheduling for Partial Reception Channel: At time slot t,

(1) All nodes first compute a common threshold vector $\mathbf{T}^*[t]$ by solving the optimization problem (25) (shown at the bottom of the page).

$$\mathbf{D}^*(\mathbf{C}[t],\mathbf{C}[t-\tau]) \in \arg\max_{\mathbf{D}(\mathbf{C}[t],\mathbf{C}[t-\tau])} \sum_{(m,n)\in\mathcal{L}} \max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right)\times \mathbf{E}\left[C_{(m,n)}[t]D_{(m,n)}(C_{(m,n)}[t],\mathbf{C}[t-\tau])\Pi_{(l,k)\in\mathcal{I}_{(m,n)}}\left(1-D_{(m,n)}(C_{(m,n)}[t],\mathbf{C}[t-\tau])\right)\Big|\,\mathbf{C}[t-\tau]\right]. \quad (23)$$

$$\sum_{(m,n)\in\mathcal{M}} \mathbf{E}\left[\left(1-\alpha_{(m,n)}\right)C_{(m,n)}[t] + \left(1-\alpha_{(m,n)}\right)C_{(m,n)}[t]\mathbf{1}_{\mathcal{M}\cap\mathcal{I}_{(m,n)}=\emptyset}\Big|\,C_{(m,n)}\left[t-\tau_{(m,n)}\right]\right]P_{(m,n)}[t]$$

where

$$P_{(m,n)}[t] = \max_{d\in\mathcal{N}}\left(Q_{\langle m,d\rangle}\left[t-\tau_{\langle m,d\rangle}\right] - Q_{\langle n,d\rangle}[t-\tau_{\langle n,d\rangle}]\right) \quad (24)$$

$$\mathbf{T}^*[t] \in \arg\max_{\mathbf{T}} \sum_{(m,n)\in\mathcal{L}} \max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right)\times \mathbf{E}\left[\alpha_{(m,n)}C_{(m,n)}[t]\mathbf{1}_{C_{(m,n)}[t]\geq T_{(m,n)}} + \left(1-\alpha_{(m,n)}\right)C_{(m,n)}[t]\mathbf{1}_{C_{(m,n)}[t]\geq T_{(m,n)}}\prod_{(l,k)\in\mathcal{I}_{(m,n)}}\mathbf{1}_{C_{(l,k)}[t]<T_{(l,k)}}\Big|\,\mathbf{C}[t-\tau]\right]. \quad (25)$$

$$\mathbf{P}^*(\mathbf{C}[t], \mathbf{C}[t-\tau]) \in \arg\max_{\mathbf{P}(\mathbf{C}[t],\mathbf{C}[t-\tau])} \sum_{(m,n)\in\mathcal{L}} \max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right) \times \mathbf{E}\left[\sigma_{(m,n)}\left(\mathbf{P}\left(\mathbf{C}[t],\mathbf{C}[t-\tau]\right)\right)\middle| \mathbf{C}[t-\tau]\right]. \quad (26)$$

(2) Node m attempts to transmit packets from queue $d_m^*[t]$ over link (m,n) if $C_{(m,n)} \geq T^*_{(m,n)}[t]$, where

$$d^*_{(m,n)}[t] \in \arg\max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right).$$

The key modification of the algorithms is including the achievable rate under collision, i.e., $\alpha_{(m,n)}C_{(m,n)}[t]$ in independent-set selection and threshold calculation. The throughput-optimality proofs of these two algorithms are almost the same as those under the general collision model, which are omitted to avoid more complicated notations.

It is also possible to extend our results to an SINR-based interference model, where each link decides a transmission power $P_{(m,n)}$, and the rate achieved at link (m,n) is a function of the transmission powers of all links, i.e., $R_{(m,n)}(t) = \sigma_{(m,n)}(\mathbf{P}(t))$. For example, for the decentralized case, a throughput optimal algorithm is as follows:

Routing/Scheduling for SINR-based Interference Model: At time slot t,

(1) All nodes first compute scheduling power functions $\mathbf{P}^*(\mathbf{C}[t], \mathbf{C}[t-\tau]) = \{P^*_{(m,n)}(C_{(m,n)}[t], \mathbf{C}[t-\tau])\}_i$ by solving the optimization problem (26), (shown at the top of the page), where $P^*_{(m,n)}(C_{(m,n)}[t], \mathbf{C}[t-\tau]) \in [0, P_{\max}]$ and $P_{\max}$ is the maximum power a link allows to use.

(2) Node m transmits packets from queue $d_m^*[t]$ over link (m,n) with power $P^*_{(m,n)}(C_{(m,n)}[t], \mathbf{C}[t-\tau])$, where

$$d^*_{(m,n)}[t] \in \arg\max_d \left(Q_{\langle m,d\rangle}[t-\tau] - Q_{\langle n,d\rangle}[t-\tau]\right).$$

But again, we cannot identify any structure property of the solution to optimization problem (26). An interesting future research problem is to extend this algorithm to a general packet capture model where the partial reception may depend on the number of simultaneous transmissions.

C. Routing

For the centralized NSI case and decentralized NSI case with $\tau = 1$, we both consider multihop traffic flows and our algorithms exploit back-pressure routing, which is proved to be throughput optimal. For the decentralized NSI case with $\tau > 1$, the throughput-optimality of the proposed algorithm is valid only for the networks with single-hop flows and a specific queue-dynamic [defined by (5)]. In our future research, we will study decentralized routing/scheduling with $\tau > 1$ and multihop traffic flows.

D. Signalling Overhead

While scheduling based on delayed NSI achieves higher throughput than scheduling algorithms that only exploit local NSI as shown in Figs. 7 and 9, collecting delayed NSI consumes network resources. Thus there is a clear tradeoff between network throughput and the amount of resource required to collect delayed NSI (signalling overhead). It is important to quantitatively study this tradeoff to understand the actual performance gain of exploiting delayed NSI. However, the signalling overhead is determined by the signalling technique used to collect NSI (e.g., special control channels could be used, or packet headers could be modified to include NSI so that nodes can "opportunistically" collect NSI from the headers of all packets that they can decode). Hence quantifying the tradeoff between throughput and signalling overhead is out the scope of this paper, and is one of our future research topics.

E. Delay

In this paper, we have focused on the impact of delayed NSI on network throughput. Another interesting future research topic is to understand the impact of delayed NSI on end-to-end transmission delays. For example, algorithms that significantly improve the delay performance of the back-pressure algorithm have been developed in [43]–[49]. It would be interesting to study the impact of delayed NSI on the delay performance of these algorithms.

REFERENCES

[1] L. Ying and S. Shakkottai, "On Throughput Optimality With Delayed Network-State Information," presented at the Information Theory and Applications Workshop, San Diego,, CA, Feb. 2008.

[2] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 4, pp. 1936–1948, Dec. 1992.

[3] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. Inf. Theory*, vol. 39, pp. 466–478, Mar. 1993.

[4] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queueing system with asynchronously varying service rates," *Probabil. Eng. Informat. Sci.*, vol. 18, pp. 191–217, 2004.

[5] A. Stolyar and K. Ramanan, "Largest weighted delay first scheduling: Large deviations and optimality," *Adv. Appl. Prob.*, vol. 11, pp. 1–48, 2001.

[6] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," *Ann. Math. Statist.*, vol. 207, pp. 185–202, 2000.

[7] X. Lin and N. Shroff, "Joint rate control and scheduling in multihop wireless networks," in *Proc. Conf. on Decision Contr.*, Paradise Island, Bahamas, Dec. 2004.

[8] S. Shakkottai, R. Srikant, and A. Stolyar, "Pathwise optimality of the exponential scheduling rule for wireless channels," *Adv. Appl. Prob.*, vol. 36, no. 4, pp. 1021–1045, Dec. 2004.

[9] A. Eryilmaz, R. Srikant, and J. Perkins, "Stable scheduling policies for fading wireless channels," *IEEE/ACM Trans. Netw.*, vol. 13, pp. 411–424, 2005.

[10] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," in *Proc. IEEE Infocom*, 2005.

[11] M. Neely, E. Modiano, and C. Li, "Fairness and optimal stochastic control for heterogeneous networks," in *Proc. IEEE Infocom*, Miami, FL, Mar. 2005, vol. 3, pp. 1723–1734.

[12] A. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Syst.*, vol. 50, no. 4, pp. 401–457, Aug. 2005.

[13] L. Georgiadis, M. J. Neely, and L. Tassiulas, *Resource Allocation and Cross-Layer Control in Wireless Networks*. New York: NOW, 2006, Found. Trends in Netw..

[14] 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, 2006.
[15] L. Tassiulas and A. Ephremides, "Throughput properties of a queueing network with distributed dynamic routing and flow control," *Adv. Appl. Prob.*, vol. 28, pp. 285–307, 1996.
[16] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks," in *Proc. IEEE Infocom*, Mar. 2004.
[17] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks with heterogenous users," in *Proc. Conf. on Inf. Sci. Syst. (CISS)*, Mar. 2004.
[18] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *IEEE Trans. Inf. Theory*, no. 2, Feb. 2005.
[19] A. Gopalan, C. Caramanis, and S. Shakkottai, "On wireless scheduling with partial channel-state information," in *Proc. Ann. Allerton Conf. Commun., Contr. Comput.*, Urbana, IL, Sep. 2007.
[20] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad-hoc communications: An optimal stopping approach," in *Proc. ACM MobiHoc*, 2007.
[21] N. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," in *Proc. ACM Int. Conf. on Mobile Comput. Netw. (MobiCom)*, Sep. 2007.
[22] Z. Ji, Y. Yang, J. Zhou, M. Takai, and R. Bagrodia, "Exploiting medium access diversity in rate adaptive wireless LANs," in *Proc. ACM Int. Conf. on Mobile Comput. Netw. (MobiCom)*, 2004.
[23] A. Sabharwal, A. Khoshnevis, and E. Knightly, "Opportunistic spectral usage: Bounds and a multiband CSMA/CA protocol," *IEEE/ACM Trans. Netw.*, 2006.
[24] S. Guha, K. Munagala, and S. Sarkar, Performance Guarantees Through Partial Information Based Control in Multichannel Wireless Networks Univ. Penn., Tech. Rep., 2006 [Online]. Available: http://www.seas.upenn.edu/~swati/report.pdf
[25] A. Pantelidou, A. Ephremides, and A. L. Tits, "Joint scheduling and routing for ad-hoc networks under channel state uncertainty," in *Proc. 5th Int. Symp. Model. Optimiz. Mobile, Ad-Hoc and Wireless Netw. (WiOpt)*, Apr. 2007.
[26] K. Kar, X. Luo, and S. Sarkar, "Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements," in *Proc. IEEE Infocom*, Anchorage, AK, May 2007.
[27] X. Lin and S. Rasool, "Constant-time distributed scheduling policies for ad hoc wireless networks," in *Proc. Conf. Decision Contr.*, 2006.
[28] X. Wu and R. Srikant, "Scheduling efficiency of distributed greedy scheduling algorithms in wireless networks," in *Proc. IEEE Infocom*, 2006.
[29] A. Eryilmaz, A. Ozdaglar, and E. Modiano, "Polynomial complexity algorithms for full utilization of multihop wireless networks," in *Proc. IEEE Infocom*, 2007.
[30] A. Gupta, X. Lin, and R. Srikant, "Low-complexity distributed scheduling algorithms for wireless networks," in *Proc. IEEE Infocom*, 2007.
[31] S. Sanghavi, L. Bui, and R. Srikant, "Distributed link scheduling with constant overhead," in *Proc. Ann. ACM SIGMETRICS Conf.*, San Diego, CA, Jun. 2007.
[32] C. Joo, X. Lin, and N. B. Shroff, "Understanding the capacity region of the greedy maximal scheduling algorithm in multihop wireless networks," in *Proc. IEEE Infocom*, Phoenix, AZ, Apr. 2008.
[33] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89–103, 2005.
[34] H. S. Witsenhausen, "A counter-example in stochastic optimal control," *SIAM J. Contr.* , vol. 6, 1968.
[35] H. S. Witsenhausen, "Information structures, feedback and causality," *SIAM J. Contr.*, vol. 8, 1972.
[36] N. Sandell and M. Athans, "Solution of some nonclassical lqg stochastic decision problems," *IEEE Trans. Autom. Control*, vol. 19, no. 2, pp. 108–116, Apr. 1974.
[37] P. Varaiya and J. Walrand, "On delayed sharing patterns," *IEEE Trans. Autom. Control*, vol. 23, no. 3, pp. 443–445, Jun. 1978.
[38] H. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, 1995.
[39] T. Su, H. Ling, and W. Vogel, "Markov modeling of slow fading in wireless mobile channels at 1.9 GHz," *IEEE Trans. Antennas Propag.*, vol. 46, no. 6, pp. 947–948, 1998.
[40] Y. Kim and S. Li, "Capturing important statistics of a fading/shadowing channel for network performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 5, pp. 888–901, 1999.
[41] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," *Bell Labs Tech. Memo*, Apr. 2000.
[42] S. Asmussen*, Applied Probability and Queues*. New York: Springer-Verlag, 2003.
[43] M. J. Neely, "Dynamic power allocation and routing for satellite and wireless networks with time varying channels," Ph.D. dissertation, Mass. Inst. Technol., , Nov. 2003.
[44] P. Gupta and T. Javidi, "Towards throughput and delay-optimal routing for wireless ad-hoc networks," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007.
[45] L. Bui, R. Srikant, and A. L. Stolyar, "Novel architectures and algorithms for delay reduction in back-pressure scheduling and routing," in *Proc. IEEE Infocom Mini Conf.*, 2009.
[46] M. Naghshvar, H. Zhuang, and T. Javidi, "A general class of throughput optimal routing policies in multihop wireless networks," in *Proc. Ann. Allerton Conf. Commun., Contr. Comput.*, 2009.
[47] L. Ying, S. Shakkottai, and A. Reddy, "On combining shortest-path and back-pressure routing over multihop wireless networks," in *Proc. IEEE Infocom*, Rio de Janeiro, Brazil, 2009.
[48] L. Huang and M. Neely, "Delay reduction via Lagrange multipliers in stochastic network optimization," in *Proc. Int. Symp. Model. Optimiz. Mobile, Ad Hoc, and Wireless Netw. (WiOpt)*, 2009.
[49] L. Huang and M. Neely, "Delay efficient scheduling via redundant constraints in multihop networks," in *Proc. Int. Symp. Model. Optimiz. Mobile, Ad Hoc, and Wireless Netw. (WiOpt)*, 2010, pp. 142–151.

Lei Ying (M'08) received the B.E. degree from Tsinghua University, Beijing, in 2001, the M.S. and Ph.D. in electrical engineering from the University of Illinois at Urbana-Champaign in 2003 and 2007, respectively.

During fall 2007, he worked as a Postdoctoral Fellow with the University of Texas at Austin. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Iowa State University, Ames. His research interest is broadly in the area of information networks, including wireless networks, mobile *ad hoc* networks, P2P networks, and social networks.

Dr. Ying received a Young Investigator Award from the Defense Threat Reduction Agency (DTRA) in 2009, the NSF CAREER Award in 2010, and is named Litton Assistant Professor at the Department of Electrical and Computer Engineering, Iowa State University, for 2010–2012.

Sanjay Shakkottai (M'02–SM'11) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, in 2002.

He is with The University of Texas at Austin, where he is currently an Associate Professor and the Engineering Foundation Centennial Teaching Fellow with the Department of Electrical and Computer Engineering. His current research interests include network architectures, algorithms and performance analysis for wireless and sensor networks.

Dr. Shakkottai received the NSF CAREER award in 2004.