Throughput Optimal Decentralized Scheduling of Multihop Networks With End-to-End Deadline Constraints: Unreliable Links

Rahul Singh, *Member, IEEE*, and P. R. Kumar, *Fellow, IEEE*

***Abstract*—We consider multihop networks serving multiple flows in which packets have hard deadlines. Packets not delivered to their destinations by their deadlines are of no value. The throughput of packets delivered within their deadlines is called the timely throughput. We address the design of packet scheduling, transmit power control, and routing policies that maximize any specified weighted average of the timely throughputs of the multiple flows. We determine a tractable linear program (LP) whose solution yields an optimal routing, scheduling, and power control policy, when nodes have average-power constraints. The optimal policy is fully decentralized, with decisions regarding any packet's transmission scheduling, transmit power level, and routing, based solely on the age and location of that packet. No knowledge of states of any other packets in the network is needed. This resolves a fundamental obstacle that arises whenever one attempts to optimally schedule networks. The number of variables in the LP is bounded by the product of the square of the number of nodes, the number of flows, the maximum relative deadline, and the number of transmit power levels. This solution is obtained from decomposition of the Lagrangian of the constrained Markov decision process describing the complete network state. Global coordination is achieved through a price for energy usage paid by a packet each time that its transmission is attempted at a node. It is fundamentally different from the decomposition of the fluid model used to derive the backpressure policy, which is throughput optimal when packets have no deadlines, where prices are related to queue lengths. If nodes instead have peak-power constraints, then a decentralized policy obtained by simple truncation is near optimal as link capacities increase in a proportional way.**

***Index Terms*—Communication networks, delay guarantees in networks, quality of service, scheduling networks, wireless networks.**

Manuscript received July 18, 2017; revised January 2, 2018 and January 5, 2018; accepted February 5, 2018. Date of publication October 8, 2018; date of current version December 24, 2018. This work was supported in part by the National Science Foundation under Contract CNS-1719384, Contract CCF-1619085, and Contract CNS-1302182; in part by the Science and Technology Center under Grant CCF-0939370; in part by the Office of Naval Research under Contract N0014-18-1-2048; and in part by USARO under Contract Nos. W911NF-18-10331 and W911NF-15-1-0279. Recommended by Associate Editor L. Wu. *(Corresponding author: P. R. Kumar.)*

R. Singh is with Intel SC12, Santa Clara, CA 95054 USA (e-mail: rahulsiitk@gmail.com).

P. R. Kumar is with Texas A&M University, College Station, TX 77843-3259 USA (e-mail: prk@tamu.edu).

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/TAC.2018.2874671

I. Introduction

Recently, the U.S. Federal Communications Commission released 3.85 GHz of licensed spectrum and 7 GHz of unlicensed spectrum in the band above 24 GHz [1]. With this, they aim to enable applications with high data rates including those requiring low latency. Transmissions in this millimeter wave band are directional and not subject to typical omnidirectional interference [2], but are subject to absorption. Multiple hops may be needed to traverse longer distances. This motivates the problem studied in this paper: How to maximize the throughput of packets meeting hard end-to-end delay bounds over multihop wireless networks of unreliable links?

The past quarter century has seen the pioneering work of Tassiulas and Ephremides [3], Lin and Shroff [4], Lin *et al.* [5], Eryilmaz and Srikant [6], and Neely *et al.* [7] on max-weight and backpressure-based scheduling policies for communication networks that are provably throughput optimal, attaining any desired maximal throughput vector on the Pareto frontier of the feasible throughput region, when packets have no delay constraints.

However, fluid-based policies such as the backpressure policy should not be expected to provide delay optimality, and are indeed not optimal when packets have hard delay constraints. They can in fact perform poorly with regard to delay performance [8]. This is especially noticeable at light traffic when there is no "pressure" and packets just move randomly in the absence of any pressure driving them. Delay depends on fluctuations, as illustrated, for example, by the Pollaczek–Khinchine formula. The difference between throughput and delay is akin to the difference between the law of large numbers and the central limit theorem.

For optimal delay performance, one needs to start with a fundamentally stochastic framework that takes all randomness into account, such as the complete Markov decision process (MDP) describing the evolution of the complete network state. How to analyze such a complete stochastic model, and obtain a tractably computable, fully decentralized solution that is optimal, is addressed in this paper.

II. Problem Studied

We consider multihop, multiflow networks in which packets of flows have hard end-to-end relative deadlines. The deadline is the time by which a packet is to be delivered to its destination

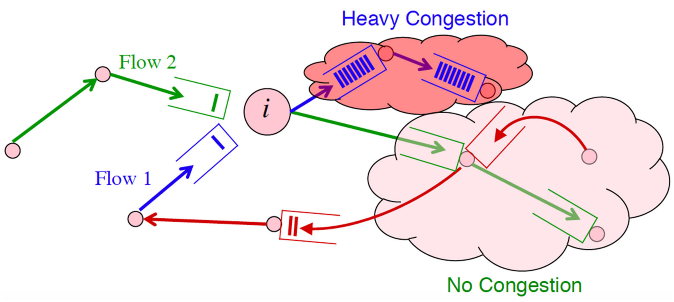


Fig. 1. Possible usefulness of network state. Suppose that node i wishes to transmit one packet, either a packet of flow 1 or a packet of flow 2. Suppose packets of flow 1 are experiencing heavy downstream congestion, but packets of flow 2 face no downstream congestion. Then, one expects that serving a packet of flow 1 is not useful since it will just get stuck in downstream congestion and not make it to its destination in time. Hence, node i should serve a packet of flow 2. Therefore, one expects that knowing downstream network state, i.e., which packets are at which nodes, is useful information.

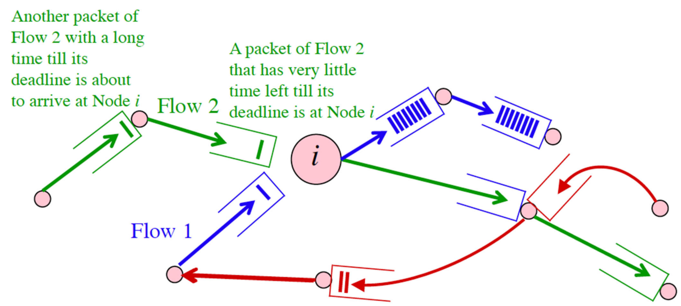


Fig. 2. Challenge of decentralized scheduling. Similarly, it can conceivably also be useful to know upstream network state information. Suppose there is a packet of flow 2 just about to arrive at node i that has a much larger time till its deadline. Then, node i may prefer to wait for that yet-to-arrive packet and use its valuable energy on its transmission, instead of transmitting the nearly late packet that is already at node i. Therefore, one expects that complete network state is useful information. This however creates a chicken and egg situation since instantaneously obtaining network state information to make optimal decisions requires zero-delay communication of information over the network, while the very purpose of obtaining network state information is to provide low delay communication. A fundamental result that we prove is that this conundrum does not arise, and that network state is irrelevant and not necessary to know in order to do optimal scheduling, when nodes have average-power constraints.

if it is to be useful, whereas the relative deadline is the remaining time-till-deadline when a packet arrives. The throughput of packets of a flow that meet their end-to-end relative deadline constraint is called its timely throughput. The vector comprised of the timely throughputs of all the flows is called the timely throughput vector.

Nodes can transmit and receive packets simultaneously. Nodes can transmit packets at varying power levels. We consider the following two types of nodal power constraints: 1) an average-power constraint at each node, or 2) a link-capacity constraint on each network link, which bounds the number of concurrent packets that can be transmitted on it at any given time t, or, equivalently, a peak-power constraint at each node.

Since the wireless channel is unreliable, the outcome of packet transmissions is modeled as a random process.[1]

Our goal is to design a decentralized, joint transmission scheduling, power control, and routing policy, which maximizes the weighted sum of the timely throughputs of the flows, for any given nonnegative choice of weights (i.e., attain any point on the Pareto frontier of the set of achievable timely throughput vectors).

In general, optimally controlling a distributed system such as a multihop network can be a challenge, since, as shown in Figs. 1 and 2, one generally *expects* that the policy should be dynamic enough and take into account in an online fashion the following factors.

A. *Routing*: The policy will need to dynamically route packets so as to:
   - A.i. avoid nodes with lower power budgets; and
   - A.ii. avoid nodes that are currently congested and experiencing a higher delay.

B. *Scheduling and Power Control*: The policy will need to schedule a packet's transmission time and transmission power at a node:
   - B.i. based on its own time-till-deadline and the reliability of the channel; and
   - B.ii. based on the times-till-deadline of other packets in the pipeline.

Consideration (A.i) requires only static nonchanging information, and (B.i) requires only local information that a packet has about itself and the node it is at. In contrast, considerations (A.ii) and (B.ii) require information about the states of all other packets at all nodes. Obtaining such instantaneous information about the state of the entire network—where all other packets are, and their times-till-deadline—is a major obstacle in optimal scheduling of networks since it itself requires communication of information across the network, while the very purpose of obtaining network state information is to transfer information-bearing packets from sources to destinations. That is, it gives rise to a chicken and egg situation. This is a major difficulty that has stymied the field of optimal scheduling of networks.

We solve this challenge by proving the surprising result that (A.ii) and (B.ii) are unnecessary, when nodes have average-power constraints. This eliminates the need for knowing network state in order to optimally schedule all packet transmissions, their transmit power levels, and dynamically route them. This allows us to obtain an optimal policy that is decentralized. Moreover, we show that this decentralized optimal policy can be computed offline with tractable complexity, simply by solving a low dimensional linear program (LP).

Central to our solution is the result that packets can make decisions regarding their scheduling, power control, and routing, independently of other packets, in a totally decoupled way. We show that each packet can follow its own MDP, which governs its actions, oblivious to all other traffic or network state, in the case of average-power constraints at each node. In this packet-level standalone problem, called an "optimal single-packet transportation problem," a packet optimizes its progress

[1]We do not consider mobility, which can lead to changing network topology. However, if the time scale of mobility is minutes, whereas delay is of the order of milliseconds, then one could treat it in practice as a quasi-static system and employ the results obtained here.

through the network, paying prices to nodes every time it requests transmission, but is compensated with a reward if it reaches its destination prior to the hard deadline. The optimal prices paid to nodes for energy can be tractably computed offline and stored. Some packets may need to randomize between nonunique optimal actions. The overall network is optimally scheduled by each packet simply following its own optimal actions.

To compute at one shot the optimal randomized solutions for all packets of all flows, one only needs to solve offline a tractable LP with number of variables bounded by the product of the square of the number of nodes, the number of flows, the maximum relative deadline, and the number of transmit power levels. This is a dramatic reduction from solving the gigantic MDP that governs the evolution of the complete network state, since that has an exponentially large number of states, rather than polynomial as above. The reduction in the number of variables is because of the central structural result that packet decisions are taken based only on their own state, and not on network state. We also show that the optimal policy can have a threshold structure that can further simplify implementation.

If the nodes instead have peak-power constraints, then the above decentralized policy can be simply truncated to yield a policy that is quantifiably near optimal as link capacities increase in a proportional way. This pursues Whittle's relaxation approach for multiarmed bandits [9] for networks with peak-power constraints.

A preliminary announcement of some of these results was presented in the conference paper [10]. The decoupling of the problem into packets is mentioned there, as is the possibility of using Whittle's relaxation. Except for strong duality, no proofs are presented. The need for an appropriate level of randomization to satisfy complementary slackness, the exploitation of decentralization to obtain a tractable LP solution of the overall problem with polynomial number of variables and constraints, and the threshold structure are not mentioned there. This paper contains a comprehensive treatment, including the treatment of the tractable LP that solves the problem, and its proof, the proof of threshold structure, the proof of asymptotic optimality for the peak-power case, and explicitly solved examples to illustrate the theory as well as other simulation examples.

This paper is organized as follows. In Section III, we summarize some notable previous work. In Section IV, we describe the system model. In Sections V–VIII, we address the case where nodes have average-power constraints. In Section V, we show that there is an optimal solution that is fully decentralized, and determine its structure. In Section VI, we show how the decentralized optimal policy can be computed through a tractable low-dimensional LP. In Section VII, we show that there can be a further simplifying threshold structure for the optimal policy. In Section VIII, we provide other iterative online methods for determining the optimal prices. In Section IX, we determine an asymptotically optimal policy based on truncation when there are link-capacity constraints, and in Section X for peak-power constraints at nodes. In Section XI, we address situations where there are both real-time and non-real-time flows. In Section XII, we address fading wireless channels. In Section XIII, we provide examples showing how the theory can be used to calculate optimal decentralized policies, and present a comparative simulation-based performance study of the truncated policy for link-capacity constraints. Finally, conclusion is drawn in Section XIV.

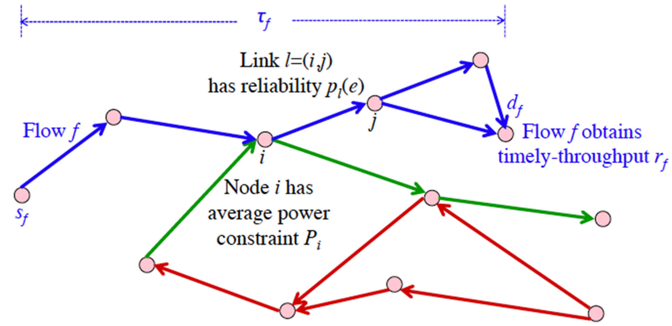


Fig. 3. Multihop network serving F flows. Flow f, with source s_f and destination d_f, has several feasible routes. Its end-to-end relative deadline is τ_f. Node i has an average-power constraint P_i. A packet transmitted on link $\ell = (i, j)$ at power level e has a probability $p_\ell(e)$ of being successfully received by node j.

III. Previous Works

In addition to the work on max-weight and backpressure noted in Section I, Kelly *et al.* [11] have shown in another seminal contribution that the problem of congestion control of the Internet can be formulated as a convex programming problem and have provided a quantitative framework for design based on primal or dual approaches. In another breakthrough, Jiang and Walrand [12] have designed a novel adaptive carrier sensing multiple access algorithm for a general interference model that achieves maximal throughput through completely decentralized scheduling under slow adaptation, without slot synchrony, if packet collisions are ignored. Combined with end-to-end control, it also achieves fairness among the multiple flows.

Concerning scheduling with hard delay constraints, there has been considerable progress on the problem of scheduling an access point, in which multiple one-hop flows with hard relative deadlines share a wireless channel. The Pareto optimal frontier of timely throughput vectors has been characterized, and simple optimal policies have been determined [13]–[26].

IV. System Model

We consider networks in which the data packets have a hard delay constraint on the time within which they should be delivered to their destination nodes if they are to be counted in the throughput.

The communication network of interest is described by a directed graph $G = (V, \mathcal{L})$, as shown in Fig. 3, where $V = \{1, 2, \ldots, |V|\}$ is the set of nodes that are connected via communication links. A directed edge $(i, j) \in \mathcal{L}$ signifies that node i can transmit data packets to node j. We will call this link $\ell = (i, j)$.

We assume that time is discrete, and evolves over slots numbered $1, 2, \ldots$. One time slot is the time taken to attempt a packet transmission over any link in the network.

There are a finite set $\mathcal{E}$ of possible transmit power levels at which a packet can be transmitted. For convenience, we normalize each time slot to 1 s so that power and energy of a transmission are interchangeable.

The outcome of a transmission over a link between any two nodes is allowed to be random, which enables us to model unreliable channels. If a packet transmission occurs on the link $\ell \in \mathcal{L}$ at a certain power level that consumes energy $e \in \mathcal{E}$, then the transmission is successful with probability $p_\ell(e)$, which is monotone increasing in e. We can model the phenomena of wireless fading by allowing the success probability $p_\ell(t,e)$ to also be a function of time that can be assumed to be governed by a finite-state Markov process, whose state is known at the transmitting node. However, for simplicity of exposition, we consider time invariant $p_\ell(e)$'s only. In this paper, we do not consider contention for the transmission medium.

The network is shared by F flows. Packets of flow f have source node s_f and destination node d_f. They may traverse any of several alternative routes.

The packet arrivals of a flow at its source node are independent identically distributed (i.i.d.) across time slots, although the distribution can vary from flow to flow. For simplicity of exposition, we suppose that these distributions have bounded support, i.e., the number of arrivals in a time slot is bounded, although we can relax this to merely assuming they are finite valued. Packet arrivals across flows are independent. The analysis in the following carries over to the case when the arrivals and relative deadlines (detailed in the following) are governed by a finite-state Markov process. We will denote the average arrival rate of flow f in packets/time slot by A_f.

Each packet of flow f has a "relative deadline," or "allowable delay" τ_f. If a packet of flow f arrives to the network at time t, then it needs to be either delivered to its destination node by time slot $t+\tau_f$, or else it is discarded from the network at time $t+\tau_f$ if it has not yet reached its destination d_f. We suppose that all relative deadlines of packets are bounded by a quantity Δ. We can allow packets of a flow to have random i.i.d. relative deadlines, independent across flows; however, both for simplicity of exposition as well as its importance, we will suppose that all packets of flow f have the same relative deadline τ_f.

We assume that no matter the scheduling policy, there is a positive probability for the network to reach an empty state after some time, for example, by no arrivals for a period greater than Δ. This ensures that the Markov chain has a single closed communicating class.

The "timely throughput" r_f attained by a flow f under a policy is the expected value of the average number of packets delivered prior to deadline expiry per unit time

$$r_f := \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\delta_f(t) \tag{1}$$

where the random variable $\delta_f(t)$ is the number of packets of a packet of flow f that are delivered in time to their destination at time t, with the expectation taken under the policy being applied.

The vector $\boldsymbol{r} := (r_1, r_2, \ldots, r_F)$ is called the "timely throughput vector." A timely throughput vector $\boldsymbol{r}$ that can be achieved via some scheduling policy will be called an "achievable timely throughput vector." The set of all achievable timely throughput vectors constitutes the "rate region," denoted by $\boldsymbol{\Lambda}$. It is a compact convex set. The maximal (i.e., undominated) vectors of this set constitute its Pareto frontier. In Sections V–VIII, we consider an *average-power constraint on each node* $i \in V$. If the total energy consumed by all the concurrent packet transmissions on link ℓ at time t is $e_\ell(t)$ units of energy, then the nodal average-power constraints are required to satisfy

$$\limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\sum_{\ell:\ell=(i,\cdot)} e_\ell(t) \le P_i \quad \forall i \in \{1,2,\ldots,V\}. \tag{2}$$

The second summation mentioned above is taken over all links ℓ, where $\ell=(i,j)$ for some node j. We note that the above constraint on the average power allows a node to transmit packets simultaneously over several outgoing links, which can be achieved by employing various techniques such as time-division/frequency-division multiple access, code division multiple access, etc., [27]–[29]. We suppose that nodes can simultaneously receive any number of packets while they are transmitting.

Given a weight $\beta_f \ge 0$ for the timely throughput of each flow f, we define the "weighted timely throughput" as $\beta^T\boldsymbol{r}$, where $\beta=(\beta_1,\beta_2,\ldots,\beta_F)$. We will derive completely decentralized scheduling policies that maximize the weighted timely throughput for any given weight vector β. We will also show how they can be tractably computed.

In Section IX, as an alternative to (2), or in addition to it, we will consider *peak-power constraints on each link*

$$e_\ell(t) \le C_\ell \quad \forall \ell \in \mathcal{L}, \text{ and } t = 1,2,\ldots. \tag{3}$$

Alternatively, we can constrain the number of concurrent packets that can be transmitted on a link ℓ at each time t. For either of these situations, we will obtain quantifiably near-optimal decentralized scheduling policies.

V. Optimality and Structure of a Decentralized Scheduling Policy for Maximizing Weighted Timely Throughput

In this section, we show that a decentralized policy is optimal for the problem of maximizing the weighted timely throughput $\sum_{i=1}^{F}\beta_i r_i$ for a given weight vector $\beta=(\beta_1,\beta_2,\ldots,\beta_F)$ with $\beta_i \ge 0$.

First, we note that the problem can be formulated as a finite state, finite action, constrained Markov decision Process (CMDP) [30], where the goal is to maximize the average reward subject to several average cost constraints, although the number of states in the system will be exponentially large. The state of an individual packet of flow f present in the network at time t is described by the two tuple (i,s), where i is the node at which it is present, and s is its time-till-deadline. The state of the network at time t, $x(t)$, is described by specifying the state of each packet of each flow present in the network at time t. Since the time spent by a packet in the network is bounded by Δ, and since the number of arrivals in any time slot is also bounded due to the bounded support assumption, the system state $x(t) \in \mathcal{X}$, a finite set, although it is exponentially large. A scheduling policy π has to choose, at each time t, based on the past history, possibly in a randomized way, which packets to transmit at each node from the set of available packets, and over which links and at what powers. The link choice allows routing to be optimized. The choice made at time t will be denoted $U(t) \in \mathcal{U}$, a finite set. Since the probability distribution of the system state $x(t+1)$ at time $t+1$ depends only $x(t)$ and $U(t)$, the problem of maximizing the timely throughput subject to nodal average-power constraints (2) is a CMDP, where a reward of β_f is received when a packet of flow f is delivered to its destination before its deadline expires

$$\underset{\pi}{\text{Maximize}} \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{f}\sum_{t=1}^{T}\beta_f\delta_f(t)$$
$$\text{subject to (2)}$$

where $\delta_f(t)$ is the number of packets of flow f delivered before deadline expiry to d_f at time t.

As written above, the problem has a prohibitively large number of states. We will show in this section that it allows a packet-by-packet decoupling that makes possible an elegant and simple solution.

Our solution will be comprised of two aspects. First, at each node v, we will have a "nodal price" $\lambda_v^\star \geq 0$ for energy. Every packet requesting transmission at node v at a power level that consumes energy e over an outgoing link (v, v') will be charged an amount $\lambda_v^\star e$. It can choose to not get transmitted, which will be regarded as requesting transmission over the self-loop (v, v), and it will not be charged anything. Second, if a packet of flow f reaches its destination d_f in time before its deadline expires, then it obtains a reward β_f.

In the solution, which we show to be optimal, each packet optimizes its scheduling, power control and routing decisions in a completely decentralized way, independent of all other packets and nodes in the network, so as to maximize its own total expected reward. With the state of the packet described by the two tuple (v, s), where v is the node at which it is present, and s its remaining time-till-deadline, it determines its optimal decisions through the following simple dynamic programming problem, which we call the "optimal single-packet transportation problem":

$$\begin{aligned} V^f(v,s) =\ & \max\{V^f(v,s-1), \\ & \max_{v':(v.v')\in\mathcal{L}, e\in\mathcal{E}}\{-\lambda_v^* e + p_{(v,v')}(e)V^f(v',s-1) \\ & + \left(1-p_{(v,v')}(e)\right)V^f(v,(s-1)^+)\}\} \\ V^f(d_f,s) =\ & \beta_f \qquad \text{if } s \geq 0 \end{aligned} \tag{4}$$

where $V^f(v,s)$ is the optimal total expected reward for a packet of flow f starting from state (v, s).

Any maximizer of the right-hand side (RHS) yields an optimal action, i.e., whether to transmit or not, and if so, at what level, and over which link. If there are two actions that achieve the maximum, then the packet can choose either action. In particular, a packet can choose an action randomly from all those that maximize the RHS of above according to a probability distribution. We will suppose that a packet employs an *optimal stationary randomized policy for its transportation problem*, by which it is meant that in each state (v, s) it chooses a probability distribution supported on the set of actions that achieve the maximum of the RHS and chooses an action independently of all other actions according to this probability distribution.

Theorem 1 (Optimality of a fully decentralized policy):

1) Let $\lambda^\star = \left(\lambda_1^\star, \lambda_2^\star, \ldots, \lambda_{|V|}^\star\right)$ with $\lambda_i^\star \geq 0$. Denote by $\pi_f(\lambda^\star)$ an optimal stationary randomized policy for packets of flow f that is optimal for the single-packet optimal transportation problem (4), and by $\pi(\lambda^\star)$ the policy that implements $\pi_f(\lambda^\star)$ for each packet belonging to flow f. Suppose that, at every node i, either the average-power constraint (2) is satisfied with equality by $\pi(\lambda^\star)$, or $\lambda_i^\star = 0$ and the constraint (2) is respected. Then, $\pi(\lambda^\star)$ is an optimal policy for maximizing the weighted timely throughput $\sum_{f=1}^F \beta_f r_f$ subject to the average-power constraints (2).
2) There exists an optimal $\lambda^\star$ and an optimal $\pi(\lambda^\star)$ that satisfies the conditions of 1).

Proof: A history-dependent randomized policy for the CMDP is one which, dependent on past network history $(x(0), u(0), \ldots, x(t))$, chooses $u(t)$ according to a probability distribution on $\mathcal{U}$. A stationary randomized policy is one where the probability distribution only depends on the current state $x(t)$.

A stationary randomized policy is optimal in the class of all history-dependent randomized policies [30]. We provide a self-contained proof of this[2] because it will be used in the following to define an equivalent LP. Denote $\text{Prob}(x(t+1)=j|x(t)=i, u(t)=u)$ by $p_{ij}(u)$, and consider the problem of maximizing a long-term average reward Max $\liminf_{T\to\infty}\frac{1}{T}\mathbb{E}\sum_{t=1}^{T} r(x(t),u(t))$, subject to long-term average cost constraints $\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\sum_{t=1}^{T} c(x(t),u(t)) \leq \bar{C}$. For any history-dependent randomized policy π_{hdr}, consider the induced "occupation measure" until time t, $\psi_{\pi_{\text{hdr}}}(i,u;t) := \frac{1}{t}\left(\sum_{s=1}^{t}\mathbb{P}\left(x(s)=i, u(s)=u\right)\right)$. Similarly define $\psi_{\pi_{\text{hdr}}}(i;t)$. The infinite horizon reward is $\leq \sum_{i\in\mathcal{X},u\in\mathcal{U}} \psi_{\pi_{\text{hdr}}}(i,u)r(i,u)$, and the infinite horizon cost constraints are $\geq \sum_{i\in\mathcal{X},u\in\mathcal{U}}\psi_{\pi_{\text{hdr}}}(i,u)c(i,u)$, for any limit point $\psi_{\pi_{\text{hdr}}}$ of the sequence $\{\psi_{\pi_{\text{hdr}}}(i,u;t)\}_{(i,u)\in\mathcal{X}\times\mathcal{U}}$ along some subsequence t_n. Define a stationary randomized policy π_{sr} as the one that takes the action u when the system is in state i, with a probability $\pi_{\text{sr}}(u|i) = \frac{\psi_{\pi_{\text{hdr}}}(i,u)}{\sum_{\tilde{u}\in\mathcal{U}}\psi_{\pi_{\text{hdr}}}(i,\tilde{u})}$. We now complete the proof by showing that its occupation measures $\psi_{\pi_{\text{sr}}}(i,u;t) \to \psi_{\pi_{\text{hdr}}}$, thus establishing that the reward/cost constraint values achievable by any history-dependent policy are also achievable by a stationary randomized policy. $\psi_{\pi_{\text{hdr}}}(i;t) = \frac{1}{t}(\sum_{s=1}^{t}\mathbb{P}(x(s)=i)) = \frac{1}{t}(\sum_{s=2}^{t}\sum_{(j,\tilde{u})}\mathbb{P}(x(s-1)=j, u(s-1)=\tilde{u})p_{ji}(u)) + \frac{1}{t}\mathbb{P}(x(1)=i) = \sum_j\sum_{\tilde{u}}\psi_{\pi_{\text{hdr}}}(j,\tilde{u};t-1)p_{ji}(\tilde{u}) + \frac{1}{t}\mathbb{P}(x(1)=i) = \sum_j\sum_{\tilde{u}}\psi_{\pi_{\text{hdr}}}(j,\tilde{u};t)p_{ji}(\tilde{u}) + \frac{1}{t}\mathbb{P}(x(1)=i) - \frac{1}{t}\mathbb{P}(x(t+1)=i)$. Taking limits along the subsequence t_n, $\psi_{\pi_{\text{hdr}}}(i) = \lim_{n\to\infty}\sum_j\sum_{\tilde{u}}\psi_{\pi_{\text{hdr}}}(j,\tilde{u};t_n)p_{ji}(\tilde{u}) = \sum_j\sum_{\tilde{u}}\psi_{\pi_{\text{hdr}}}(j,\tilde{u})p_{ji}(\tilde{u}) = \sum_j\psi_{\pi_{\text{hdr}}}(j)\sum_{\tilde{u}}\pi_{\text{sr}}(\tilde{u}|j)p_{ji}(\tilde{u})$, where the last equality follows from the definition of policy π_{sr}. However,

[2] At a reviewer's suggestion.

$\psi_{\pi_{\rm sr}}$ is the unique solution z of $z(i) = \sum_{j\in\mathcal{X}} z(j) \sum_u \pi_{\rm sr}(u|j)p_{ji}(u), \forall i \in \mathcal{X}$. Thus, it follows that $\psi_{\pi_{\rm sr}} = \psi_{\pi_{\rm hdr}}$. By stationarity, the limits also hold almost surely. Although we have only shown that $\psi_{\pi_{\rm sr}}(i) = \psi_{\pi_{\rm hdr}}(i), \forall i \in \mathcal{X}$, the proof of $\psi_{\pi_{\rm sr}}(i,u) = \psi_{\pi_{\rm hdr}}(i,u), \forall i \in \mathcal{X}, u \in \mathcal{U}$ is similar. One could consider the quantity $\psi_{\pi_{\rm sr}}(i,u)$ and repeat the equalities considered above to arrive at the equation $z(i,u) = \pi_{\rm sr}(u|i)(\sum_{j\in\mathcal{X}} z(j) \sum_u \pi_{\rm sr}(u|j)p_{ji}(u)),\ \forall\ (i,u) \in \mathcal{X} \times \mathcal{U}$. This equation is uniquely solved by $\{\psi_{\pi_{\rm sr}}(i,u)\}_{(i,u)\in\mathcal{X}\times\mathcal{U}}$. Thus, it follows that $\psi_{\pi_{\rm sr}} = \psi_{\pi_{\rm hdr}}$.

Since we can restrict ourselves to stationary randomized policies, we can replace $\limsup$ and $\liminf$ by $\lim$ and write the CMDP as follows:

$$\operatorname*{Max}_{\pi} \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_f \sum_{t=1}^{T} \beta_f \delta_f(t), \text{ subject to} \tag{5}$$

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T} \sum_{\ell:\ell=(i,\cdot)} e_\ell(t) \le P_i \quad \forall i \in \{1,2,\ldots,V\}. \tag{6}$$

Moreover, by considering the occupation measure variables of the stationary randomized policy as the decision variables, (5) can be written as a linear objective in the decision variables, and the constraints (6) can be written as linear inequalities in the decision variables. Hence, the CMDP can be written as an LP.

Defining λ_i as the Lagrange multiplier associated with the power constraint on node i, and $\lambda := (\lambda_1, \lambda_2, \ldots, \lambda_{|V|})$, we can write the Lagrangian for (6) as

$$L(\pi,\lambda) = \lim_{T\to\infty}\frac{1}{T}\left\{\mathbb{E}\sum_f\sum_{t=1}^{T}\beta_f\delta_f(t) - \sum_i \lambda_i\left(\mathbb{E}\sum_{t=1}^{T}\sum_{\ell:\ell=(i,\cdot)} e_\ell(t)\right)\right\} + \sum_i \lambda_i P_i \tag{7}$$

where the expectation is w.r.t. the policy π that is being used, the random packet transmission outcomes, and the randomness of the packet arrivals and relative deadlines, if the latter are random.

Denoting by $e_{\ell,f,n}(t)$, the amount of energy spent on transmitting the nth packet of flow f at time t on link ℓ, we have

$$e_\ell(t) = \sum_{f,n} e_{\ell,f,n}(t).$$

The Lagrangian (7) reduces to

$$L(\pi,\lambda) = \sum_i \lambda_i P_i + \lim_{T\to\infty}\frac{1}{T}\sum_{f,n}\mathbb{E}\sum_{t=1}^{T}\left(\beta_f\delta_f(t) - \sum_i \lambda_i \sum_{\ell:\ell=i(i,\cdot)} e_{\ell,f,n}(t)\right). \tag{8}$$

The key observation is that this can be decoupled completely on a packet-by-packet basis for any fixed value of the vector λ, as follows.

Let Packets(f) := set of all packets of flow f. We will denote a packet by σ. Let Packets(f,t) := subset of packets of flow f that arrive before time t. Let $e(\sigma,i)$:= total energy consumed by packet σ at node i, $\phi(\sigma)$:= flow that σ belongs to, and $\delta(\sigma)$ be the random variable that assumes value 1 if packet σ reaches its destination before its deadline, and 0 otherwise.

Since the relative deadlines of packets are bounded, $L(\pi,\lambda)$ can be rewritten as a sum over packets

$$L(\pi,\lambda) = \sum_i \lambda_i P_i + \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\sum_f \sum_{\sigma\in\text{Packets}(f,T)}\left(\beta_f\delta(\sigma) - \sum_i \lambda_i e(\sigma,i)\right). \tag{9}$$

The term corresponding to packet σ of flow f

$$\mathbb{E}\left(\beta_f\delta(\sigma) - \sum_i \lambda_i e(\sigma,i)\right) \tag{10}$$

can be interpreted as follows. The packet incurs a payment of λ_i per unit energy used for transmission by node i, and accrues a reward of β_f if it reaches its destination before its deadline expires. This is therefore the *optimal single-packet transportation problem* whose solution is given by (4). Let $R(f)$ denote its optimal expected cost.

Due to the decomposition of (9) over packets, we can optimize packet by packet. Hence, we obtain

$$\operatorname*{Max}_{\pi} L(\pi,\lambda) = \sum_f A_f R_f(\lambda) + \sum_i \lambda_i P_i$$

since A_f is the arrival rate of packets of flow f.

The dual function is

$$D(\lambda) = \operatorname*{Max}_{\pi} L(\pi,\lambda). \tag{11}$$

The dual problem is

$$\operatorname*{Max}_{\lambda\ge 0} D(\lambda). \tag{12}$$

There is no duality gap since the CMDP (6) can equivalently be posed as an LP, in which the variables to be optimized are the *occupation measures* defined above [30]–[33], induced by the policy π on the joint state-action space. The dual function (11) is therefore

$$D(\lambda^\star) = L(\pi(\lambda^\star),\lambda^\star). \tag{13}$$

The result then simply follows from complementary slackness [34] since the primal problem can be written as an LP over variables that are occupation measures, as shown above. The existence of $\lambda^\star$ and $\pi(\lambda^\star)$ follows from the linear programming solution and the dual. ■

A. Features of Our Solution

The following features may be noted.

1) In order to solve the primal problem (5), (6) in its original form, the network is required to make decisions based on the knowledge of the network state $x(t)$. With $|V|$ denoting the number of nodes, Δ an upper bound on

relative deadlines of flows, and F the number of flows, the size of the state space in which the network state $x(t)$ resides is exponential, $(V\Delta)^{F\Delta}$, since there can be $F\Delta$ packets in the network, with each being in one of $V\Delta$ states. It is therefore prohibitive to directly attempt to optimize the control of the network. Thus, an approach based on directly solving the constrained network-wide MDP (6) is computationally intractable.

2) Moreover, the optimal policy requires the entire network state information to be instantaneously known at each node at each slot. Indeed, one of the key reasons why optimal policies for communication networks (and other distributed systems) are generally intractable is that every decision requires instantaneous knowledge of the complete network state, which is something that cannot be obtained since the entire purpose of determining the optimal policy is to communicate information with deadlines. Thus, an approach based on implementing the solution of the constrained network-wide MDP (5), (6) would also have been implementationally futile.
3) In contrast, the solution we obtain is a completely decentralized solution. Each packet makes its own decision at each node on whether it wants to be transmitted, and if so, at what power level, and over what link. No network state knowledge is needed by a packet to determine its optimal decision. Each packet's actions are independent of the actions chosen by all other packets in the network.
4) The MDP for the "optimal single-packet transportation problem" governs the behavior of each packet, oblivious to all other traffic or network state. In this standalone problem, a single packet optimizes its progress through the network, paying prices to nodes, depending on the energy consumption for transmission incurred by the node, every time it requests transmission at a node at a power level. The packet is compensated with a reward if it reaches its destination prior to its hard deadline. The reward it receives is equal to the weight of its flow in the weighted timely throughput.
5) The optimal single-packet transportation problem has a small-sized state space: $|V|\Delta$. Thus, one only needs to solve a small dynamic programming problem over a time horizon of Δ, with $|V|\Delta$ states. This is much smaller than the exponentially large number $(V\Delta)^{F\Delta}$ of network states.
6) Thus, we have reduced the computational complexity from *exponential* to low degree polynomial. Moreover, the resulting solution can be implemented locally at each node. It is highly decentralized with no coupling between flows or nodes or even packets.
7) Our solution exploits the Lagrangian dual of the CMDP. The Lagrange multipliers associated with the average-power constraints are interpreted as prices paid by a packet for utilizing energy every time its transmission is attempted by a node.
8) Complementary slackness requires that at each node, either all the available power is fully used, or the price of energy at that node is zero. In order to fully use up all available power at a node, packets may need to randomize. Suppose that 1.5 W of power is available at a node, but packets can only get transmitted at power levels 1 or 2 W. Then, to use up all the node's power, 50% of the packets will need to be transmitted at power level 1 W and 50% at 2 W. This is the reason for randomization.
9) The only manner in which this optimal single-packet transportation problem is coupled to the overall network, nodal power constraints, other flows, and other packets, is through predetermined prices for nodal energy. The optimal prices can be tractably computed offline and stored, as will be shown in Section VIII.
10) The key to these results is to pursue a fundamentally stochastic approach that considers the Lagrangian of the constrained network-wide MDP governing the entire network, and showing how it decomposes into packet-by-packet decisions. This decomposition approach allows treatment of all variability related aspects, since they affect delay. Through this decomposition of the stochastic system, we can address timely throughput optimality of packets that meet hard per-packet delay deadlines, rather than just throughput optimality.
11) This is in sharp contrast to the backpressure approach that considers the decomposition of the Lagrangian of the fluid model, and can only guarantee throughput optimality. The stochastic decomposition approach of this paper is able to address delay rather than just throughput.

VI. Tractable LP for Computing the Decentralized Randomized Optimal Policy

We will now show that we can tractably compute the decentralized randomized optimal policies for all flows by taking advantage of the decoupling of packet behaviors. We do this by considering the associated LP involving "state-action probabilities" for solving CMDPs [35]. (In this approach, we do not need to first explicitly solve for the optimal prices. They could be computed through the dual of this LP.)

This LP does *not* have an exponentially large number of variables equal to the product of the number of states of the network and the number of possible network-wide actions, which is what it would be if we were directly solving the MDP for the complete network state. Instead, having determined that packets only make decisions based on the node they are at and their time-till-deadline, we see that for each flow, we only need to determine for each of its packet's states what that packet should do. Thus, the number of decision variables is no more than the product of the number of flows, the number of nodes that packets could be at, the maximum relative deadline over all flows, the number of potential next nodes that a packet could be transmitted to, and the power level of that transmission. This is what allows us to determine a tractable solution.

Theorem 2 (Tractable LP for determining decentralized randomized optimal policy): Let $\{\xi^f(i,j,s,e)\}$ be the optimal decision variables of the following LP:

$$\text{Max} \sum_{f\in F}\sum_{s=0}^{\tau_f}\sum_{i\in V}\sum_{e\in\mathcal{E}} \beta_f A_f \xi^f(i,d_f,s,e)p_{i,d_f}(e) \tag{14}$$

Subject to

$$\sum_{f\in F}\sum_{s=0}^{\tau_f}\sum_{j\neq i}\sum_{e\in\mathcal{E}} A_f\xi^f(i,j,s,e)e \le P_i \quad \forall i, \tag{15}$$

$$\sum_{j\in V, j\neq d_f}\ \sum_{e\in\mathcal{E}} \xi^f(j,i,s,e)p_{j,i}(e)$$

$$+\sum_{m\in V}\sum_{e\in\mathcal{E}}\xi^f(i,m,s,e)(1-p_{i,m}(e))$$

$$=\sum_{k\in V}\sum_{e\in\mathcal{E}}\xi^f(i,k,s-1,e) \quad \forall f, i\neq d_f, 1\le s\le\tau_f \tag{16}$$

$$\sum_{j\in V}\sum_{e\in\mathcal{E}}\xi^f(s_f,j,\tau_f,e)=1 \quad \forall f \tag{17}$$

$$\xi^f(i,j,s,e)\ge 0 \quad \forall i\neq d_f, j, f, s, e \tag{18}$$

$$\sum_{j\in V}\sum_{e\in\mathcal{E}}\xi^f(s_f,j,s,e)=1 \quad \forall f, s. \tag{19}$$

Then, a decentralized, yet optimal, policy is for a packet of flow f in node v with time-till-deadline s to transmit to node j at power level e with probability $\frac{\xi^f(i,j,s,e)}{\sum_j'\sum_e'\xi^f(i,j',s,e')}$.

Proof: From the proof of Theorem 1, we can restrict attention to randomized Markov policies where a packet of flow f is transmitted with a certain probability over a certain outgoing link at a certain power level, or not transmitted at all, with the probabilities depending only on the state (i,s) of the packet. Let $\xi^f(i,j,s,e)$ denote the resulting "state-action probability" that a packet of flow f is transmitted over link (i,j) at power level e when its time-till-deadline is s. We let $\sum_{e\in\mathcal{E}}\xi^f(i,i,s,e)$ denote the probability that it is not transmitted, and also define $p_{ii}(e)\equiv 1$. The probabilities $\{\xi^f(i,j,s,e)\}$ are characterized by the balance constraints (16), nonnegativity (18), normalization (17), and the initial condition (19). The power constraint is captured by (15), and the accrued reward by (14). Thus, the solution of the LP is an upper bound on accruable reward. However, it is also achievable through the policy indicated. ■

The LP exploits the structure of the optimal solution determined in Section V, that packet actions are determined only by their own states, to dramatically reduce the complexity of the LP to only $|V|^2F\Delta|\mathcal{E}|$ variables and $|V|+|V|F\Delta+F+|V|^2F\Delta|\mathcal{E}|+F\Delta$ constraints. This is a dramatic reduction of the exponential complexity of the network-wide optimal control problem. Moreover, being an LP, it is eminently tractable.

VII. Optimality of Threshold Policies

In fact, there is even more structure in the optimal single-packet optimal transportation problem that can simplify its implementation. Here, we suppose that optimal prices are given and consider only the solution of (4) given the prices. We do not address the need for randomization to satisfy complementary slackness of the power constraint. For simplicity, we illustrate this structure when there is only one transmit power level that corresponds to a fixed energy usage e for any transmission. The structure we show is that each packet's decision is simply governed by a threshold on time-till-deadline.

Theorem 3 (Threshold policy): For each flow f, and node i, there is a threshold $\tau_f(i)$, such that the optimal decision for a packet of flow f at node i with a time-till-deadline s is to be transmitted/not transmitted according to whether the time-till-deadline s is strictly greater than/equal to or lesser than the threshold $\tau_f(i)$.

Proof: In a state where the decisions to transmit/not transmit are both optimal (i.e., the maximizer of the RHS of the dynamic programming equation (4) is not unique), we choose "not to transmit," so that we thereby obtain an optimal policy that uniquely assigns an optimal action to each state. (Note that we are not addressing here the need for appropriate randomization to satisfy complementary slackness.) We will prove the following property (P) of this optimal policy, from which the theorem readily follows: (P) If the optimal decision is to not transmit a packet at a node, then it is optimal to never again transmit that packet at that node.

The reason is that one can then simply define $\tau_f(i)$ as the maximum value of s at which the decision to not transmit is the optimal action. Now we prove property (P) by using stochastic coupling. Suppose that for a packet of flow f at a node i, it is optimal to not transmit it at time-till-deadlines equal to $\sigma, \sigma-1, \ldots \sigma-k$, but it is optimal to transmit it when its time-till-deadline is $\sigma-k-1$. Consider a packet, called Packet-1, that follows this optimal policy. It waits for k slots at node i, and then gets transmitted when its time-till-deadline is $\sigma-k$. Now consider another packet, called Packet-2, that waits no time at node i, and is transmitted when its time-till-deadline is σ. We will couple the subsequent experiences of Packet-1 and Packet-2, i.e., whether a transmission at a link is failure or success, after that transmission. Then, if Packet-1 reaches the destination d in time, then so does Packet-2. Hence, the reward accrued by Packet-2 is no less than the reward accrued by Packet-1, while its costs are the same. Hence, the decision of Packet-2 to immediately get transmitted at time-till-deadline σ is optimal. ■

However, it is not advocated to search for an optimal policy by trying to find the thresholds. It is preferable to search for optimal *prices* since they are the *same* for all packets of all flows, whereas the optimal threshold is only for a specific flow. That is, the prices provide the right tradeoff between packets of different flows.

VIII. Determining the Optimal Prices

Now we return to the problem of determining the correct energy prices $\lambda^\star := \{\lambda_i^\star : i\in V\}$ to be charged by the nodes. One method is to just obtain them simply as the sensitivities of the power constraints in the above LP. This however requires a model of the network, its reliabilities, and the requirements of all the flows. If one does not know the system model, then one can employ "tatonnement" over a running system. Such price discovery is based on the dual function (11). First, we discuss a hybrid of optimization and simulation, and subsequently a purely learning approach.

A. Repeated Simulations

Price discovery can be performed offline by repeated simulation. Since the Dual Problem is convex, each node can employ subgradient descent on the optimal price vector $\lambda^\star$. The subgradient iteration is simply Walrasian tatonnement [36], which attempts to drives the "excess power consumption" at each node toward zero. Specifically, at the nth iterate of the price vector, node i chooses

$$\lambda_i^{n+1} = \lambda_i^n + \epsilon\,[\text{Power consumed by node } i - P_i]$$

which simply amounts to a standard subgradient iteration with a step size $\epsilon > 0$. There are issues that need to be take into consideration with respect to the subgradient method, such as choice of step sizes, and the fact that it may not be a descent method [37]. For any fixed price vector λ, one can solve the dynamic programming equations for the optimal packet policy $\pi(\lambda)$.

B. Employing Online Learning

Instead of using simulation-based optimization to determine the optimal prices, one can determine both the optimal policy as well as the optimal policy contingent on that price, by using two time-scale stochastic approximation [38]–[40]. The faster time-scale stochastic approximation for the policy is

$$\begin{aligned} V_{n+1}(i,s) &= 1_n(i,s)\,\{V_n(i,s)\,(1-a_n) \\ &+ a_n \max\{V_n(i,(s-1)^+), X\}\} + (1-1_n(i,s))\,V_n(i,s) \end{aligned}$$

where X is the maximum, over all links $\ell = (i,j)$, of

$$\lambda_{i,n} + p_\ell(e) V_n(j,(s-1)^+) + (1-p_\ell(e))\,V_n(i,(s-1)^+).$$

In the above equation, $1_n(i,s)$ assumes the value 1 if the packet state at iteration n is (i,s). $\{a_n\}$ is a positive sequence that satisfies $\sum_n a_n = \infty, \sum_n a_n^2 < \infty$.

We can use a slower time-scale stochastic approximation for the prices

$$\lambda_{n+1} = \lambda_n(1-b_n) + b_n(P - \bar{P}(\pi(\lambda_n))) \tag{20}$$

where P is the vector consisting of nodal power bounds, $\bar{P}_i(\pi(\lambda))$ is the average-power utilization at node i under $\pi(\lambda)$, and the sequence b_n satisfies $\sum_n b_n = \infty, \sum_n b_n^2 < \infty$, as well as $\sum_n (\frac{b(n)}{a(n)})^\gamma < \infty$, where $\gamma \geq 1$ [39]. The iterations converge to the optimal prices $\lambda^\star$ [41].

When network parameters are not known, one can both solve for the optimal policy $\pi(\lambda^\star)$ as well as the optimal nodal prices $\lambda^\star$ in a decentralized manner. One way to achieve this task is to perform the value iterations using reinforcement learning for each price vector λ until convergence, and then to update the price λ using a gradient descent method.

IX. Link-Capacity Constraints: The Near Optimality of a Truncation Policy

So far we have focused on systems where nodes have *average*-power constraints (2) or constraints on *average* number of packets concurrently transmitted. In this section, we consider *peak link-capacity constraints* on the number of concurrent packets that can be scheduled at any given time slot t. The quantity $e_\ell(t)$, now redefined as the number of packets transmitted on link ℓ at time t, has to satisfy

$$e_\ell(t) \leq C_\ell \quad \forall \text{ links } \ell \in \mathcal{L}, \text{ and } t = 1,2,\ldots. \tag{21}$$

The more stringent problem that results is

$$\begin{aligned} &\underset{\pi}{\text{Maximize}} \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_f \sum_{t=1}^T \beta_f d_f(t) \\ &\text{subject to (21).} \end{aligned} \tag{22}$$

Our approach to addressing this problem is similar to that developed by Whittle [9] for multiarmed bandits. Since there is no simple index policy [42] that is optimal when one is allowed to pull n arms concurrently, if $n > 1$, Whittle has suggested relaxing this constraint for *each* time t to a constraint that the *average* number of arms concurrently pulled is n. This relaxed problem has a tractable solution under an "indexability" condition [43]. Importantly, it is near optimal when the number of arms available goes to infinity, with the proportion of arms of each type held constant [44]. Our approach pursues this idea for multihop networks.

We now proceed to construct a decentralized policy with a provably close approximation to optimality. We begin with the policy $\pi^\star$ that is optimal for the relaxed version of the problem (5)–(6) which involves *average linkwise* power constraints

$$\begin{aligned} &\max_\pi \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_f \sum_{t=1}^T \beta_f d_f(t), \text{ subject to} \\ &\lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^T e_{(i,j)}(t) \leq C_{(i,j)} \quad \forall \text{ links } (i,j) \in \mathcal{L}. \end{aligned} \tag{23}$$

This optimal policy can be obtained in exactly the same fashion as for the problem (5)–(6) with average *nodal* power constraints, except that now there are *link-based prices* $\lambda_{(i,j)}$, instead of nodal prices λ_i. This is similarly a decentralized policy, as we have shown in the preceding sections, and is moreover tractable to compute.

However, $\pi^\star$ only ensures that the constraint (21) is met on average, and *not* at each time t. On the occasions that the number of packets that it prescribes for concurrent transmission does not exceed the constraint (21), we just transmit all the packets specified by that policy. However, on the occasions that it specifies an excessive number of transmissions exceeding the RHS of (21), we simply truncate the list of packets by picking any $C_{(i,j)}$ of these packets and transmitting them. Clearly, this leads to a policy that does satisfy the constraint (21) at each time instant. Moreover, we eject from the network those packets that $\pi^\star$ dictated to be scheduled, but were not picked for transmission. (Discarding the packets is not strictly required, but it simplifies the discussion.) Let us denote this modified policy by $\tilde{\pi}^\star$. It may be noted that under this truncated policy, the evolution of the network is not independent across different packets, as was the case with $\pi^\star$. We will show that this policy is nearly optimal in a precise asymptotic sense quantified in the following.

Theorem 4 (Asymptotic optimality of truncation policy): Consider the sequence of systems described in problem (23) parameterized by N, in which the arrivals for the N-th system are i.i.d. with binomial parameters $(N, A_f/N)$.[3] The deviation from optimality of the Nth system in the sequence operating under the policy $\tilde{\pi}^\star$ is $O(\frac{1}{\sqrt{N}})$, and hence the policy $\tilde{\pi}^\star$ is asymptotically optimal for the joint routing-scheduling problem (22) under hard link-capacity constraints.

Proof: Let us denote by $p_f(\tau, \ell)$ the probability that under $\pi^\star$ a packet (of flow f) with age τ time slots would be attempted on link ℓ. Since the arrival rate of flow f packets is A_f, then, on account of the fact that the policy $\pi^\star$ satisfies the average-power constraints C_ℓ imposed by the network, we have

$$\sum_f \sum_{\tau=1}^{\Delta} A_f p_f(\tau, \ell) \le C_\ell \quad \forall \ell \in \mathcal{E}. \tag{24}$$

We will now obtain lower bounds on the performance of the policy $\tilde{\pi}^\star$. The following arguments are based on analysis of the evolutions of policies on an appropriately constructed probability space. Let us denote by r_0 the (average) reward earned by policy $\pi^\star$. First note that the reward collected by $\tilde{\pi}^\star$ (denoted by r_1) does not increase if it were to, instead of dropping a packet because of capacity constraint violation, schedule it as dictated by $\pi^\star$, but no reward is given to it if this packet is delivered to its destination node (denoted by r_2). However, r_2 is more than the reward if now a penalty of β_f units per packet was imposed for scheduling a packet via utilizing "excess capacity" at some link $l \in \mathcal{E}$, but it were given a reward in case this packet reaches the destination node (denoted by r_3). r_3 is certainly more than the reward, which $\pi^\star$ earns if it is penalized an amount equal to the sum of the excess bandwidths that its links utilize (denoted by r_4) multiplied by β_f, since any individual packet may be scheduled multiple times by utilizing excess bandwidth. Let $E_{f,\ell,\tau}(t)$ denote the number of packets of flow f that have an age of τ time slots, and are served on link ℓ at time t under the policy $\pi^\star$, and let $\mathrm{MAD}(X) := \mathbb{E}|X - \bar{X}|$ denote the mean absolute deviation of X with respect to its mean $\bar{X}$. The deviation from optimality satisfies

$$\begin{aligned}
r_0 - r_4 &\le \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\sum_{\ell\in\mathcal{E}}\left(\sum_{f,\tau}\beta_f\left(E_{f,\ell,\tau}(t) - C_\ell\right)\right)^+ \\
&= \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\sum_{\ell\in\mathcal{E}}\left\{\sum_{f,\tau}\beta_f\left(E_{f,\ell,\tau}(t) - \bar{E}_{f,l}(t)\right)\right. \\
&\qquad \left. + \sum_f \beta_f\left(\bar{E}_{f,l}(t) - C_\ell\right)\right\}^+ \\
&\le \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\sum_{\ell\in\mathcal{E}}\left\{\sum_{f,\tau}\beta_f\left(E_{f,\ell,\tau}(t) - \bar{E}_{f,l}(t)\right)\right\}^+ \\
&\le \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\sum_{\ell\in\mathcal{E}}\sum_{f,\tau}\beta_f\left\{E_{f,\ell,\tau}(t) - \bar{E}_{f,l}(t)\right\}^+ \\
&\le \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\sum_{\ell\in\mathcal{E}}\sum_{f,\tau}\beta_f\left(\left|E_{f,\ell,\tau}(t) - \bar{E}_{f,l}(t)\right)\right| \\
&= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{\ell\in\mathcal{E}}\sum_{f,\tau}\beta_f \mathrm{MAD}\left(E_{f,\ell,\tau}(t)\right) \\
&= O\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

where the the last equality follows from [45]. ∎

[3] An equivalent formulation is to keep the size of packets fixed, while the link capacities for the Nth system are scaled as NC_ℓ, with the arrivals being i.i.d. with binomial parameters (N, A_f). A similar analysis can be performed for this case.

X. Near-Optimal Scheduling Under Peak-Power Constraints

We can similarly address problems where there are *peak-power* constraints on each node. The problem is formally stated as

$$\underset{\pi}{\text{Maximize}} \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_f\sum_{t=1}^{T}\beta_f d_f(t), \text{ subject to} \tag{25}$$

$$\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\sum_{t=1}^{T}\sum_{\ell:\ell=(i,j)} e_\ell(t) \le P_i \quad \forall i \in \{1,\ldots,V\}$$

$$\text{and} \sum_{\ell:\ell=(i,j)} e_\ell(t) \le P_i^{\max} \quad \forall i \in \{1,\ldots,V\} \text{ and } t \tag{26}$$

where $P_i^{\max}$ (with value greater than P_i) is the peak-power constraint on node i.

Theorem 5 (Asymptotically optimal policy for peak-power constraints): Consider the sequence of networks operating under the policy $\tilde{\pi}^\star$, in which the arrivals for the Nth system are i.i.d. with binomial parameters $(N, A_f/N)$. The deviation from optimality of the Nth system in the sequence operating under the policy $\tilde{\pi}^\star$ is $O(\frac{1}{\sqrt{N}})$, and hence the policy $\tilde{\pi}$ is asymptotically optimal for the peak-power problem (25).

Proof: Denote the policy that maximizes the objective function (25) under the average-power constraints (26) by $\pi^\star$. Now we modify it to a policy $\tilde{\pi}^\star$ described as follows: At each time t, each node $i \in V$ looks up the decision rule $\pi^\star$ and obtains the optimal power levels at which $\pi^\star$ would have carried out transmissions of the packets available with it. For this purpose, each node i only needs to have knowledge of the age of packets present with it. Node i then chooses a maximal subset of the packets present with it, such that the transmission power levels assigned to them by $\pi^\star$ sum to less than the bound $P_i^{\max}$. One way to choose such a set of packets and associated power levels is as follows: Arrange the packets in decreasing order of the transmission power assigned by $\pi^\star$, and label them. Then, $\tilde{\pi}^\star$

schedules the largest index packet such that the energy of all packets upto that index sum to less than $P_i^{\max}$. The asymptotic optimality of $\tilde{\pi}^\star$ follows as in Theorem 4. ■

XI. Jointly Serving Real-Time and Non-Real-Time Flows

In the previous sections, we have considered networks exclusively serving real-time flows for which the utility of a packet arriving after its deadline is zero. Often one is interested in networks that serve a mix of real and non-real-time flows [46], [47]. The system model can be easily extended. To incorporate this case, we simply set the relative deadlines of the packets belonging to the non-real-time flows as $+\infty$ so that they are never dropped.

XII. Wireless Fading

We can also incorporate fading of the wireless channels as follows. We model the channel state as a finite-state Markov process $Y(t)$, with the link transmission success probabilities $p_\ell(Y(t), e)$, a function of the channel state $Y(t)$, and the transmit power level. As before, we assume that the probabilities are monotone increasing in e.

The network state is then described by 1) the state of each packet, and 2) the channel state $Y(t)$. The optimal policy can be determined along similar lines as before, by augmenting the system state with the channel state $Y(t)$. The optimal policy will be of the following form: the decision to be taken by a node i at time t will depend on the state of the packet and the channel state $Y(t)$.

The above assumes that the channel condition is known to each transmitter. A simplification is possible if we assume that the process $Y(t)$ is i.i.d., which would eliminate the need for communicating $Y(t)$. Alternately, it could be deterministically time-varying. A common model that can be approximated is block fading [48], [49], under which the channel state needs only to be communicated periodically.

XIII. Illustrative Examples and Simulations

We first illustrate how the theory developed here can be used to explicitly hand compute the optimal decentralized policy in two examples. In the second example, the deadlines are slightly more relaxed than in the first example, and we can see how the prices change as a consequence, and how the optimal policy reacts to this.

Subsequently, we consider a more complex example and provide a comparative simulation illustrating the performance of the asymptotically optimal policy for the case of link-capacity constraints, comparing it with the well known backpressure, shortest path, and earliest deadline first (EDF) policies.

A. Two Illustrative Examples

Example 1: Consider the system shown in Fig. 4. It consists of two flows traversing the nodes 1, 2, and 3 in opposite directions. Flow 1, with source node $s_1 = 1$ and destination node $d_1 = 3$, has an end-to-end relative deadline τ_1 of two slots. Flow 2, with source node $s_2 = 3$ and destination node $d_2 = 1$, also has an end-to-end relative deadline τ_2 of two slots. Packets cannot afford even one failure on any transmission if they are to reach their destinations in time, since the relative deadlines for the flows are exactly equal to the total number of hops to be traversed. One packet of each flow arrives in every time slot, so $A_1 = A_2 = 1$. Each packet transmission at any node is at 1 W, so $e = 1$ since all time slots are 1 s. Nodes 1–3 have average-power constraints $P_1 = 0.5, P_2 = 0.4$, and $P_3 = 0.5$ W, respectively. Links (1, 2), (2, 3), (2, 1), and (3, 2) have reliabilities of $p_{(1,2)} = 0.4, p_{(2,3)} = 0.3, p_{(2,1)} = 0.7$, and $p_{(3,2)} = 0.6$, respectively. Denoting by r_1 and r_2 the timely throughputs of flows 1 and 2, we wish to maximize $5r_1 + 2r_2$, i.e., packets of flow 1 are 2.5 times more valuable than packets of flow 2. So, $\beta_1 = 5$ and $\beta_2 = 2$.

The dynamic programming equations for the optimal single-packet transportation problem for flow 1 yield

$$V^1(1,2) = \text{Max}\{0, -\lambda_1 + 0.4V^1(2,1)\}$$
$$V^1(2,1) = \text{Max}\{0, -\lambda_2 + (0.3)\cdot 5\}.$$

So $V^1(2,1) = (1.5 - \lambda_2)^+$ and $V^1(1,2) = [(0.6 - 0.4\lambda_2)^+ - \lambda_1]^+$. Similarly, for flow 2, $V^2(3,2) = [(0.84 - 0.6\lambda_2)^+ - \lambda_3]^+$ and $V^2(2,1) = (1.4 - \lambda_2)^+$.

Packets of flow 1 at node 2 are more valuable than packets of flow 2 at node 2, since packets of flow 1 have expected reward of $(0.3)5 = 1.5$, whereas packets of flow 2 have expected reward of $(0.7)2 = 1.4$. So, we will push as many packets of flow 1 as possible to node 2.

In order for a packet of flow 1 to choose to be transmitted at node 2, however, the price λ_2 that it pays needs to be less than the expected reward (0.3)5 that it can obtain in the future. Hence,

$$\lambda_2 \leq 1.5.$$

Similarly, in order for a packet of flow 1 to choose to be transmitted at node 1, the total expected price it expects to pay, $\lambda_1 + 0.4\lambda_2$ (since λ_1 is the price it pays at node 1, and if it succeeds to reach node 2, which happens with probability 0.4, it then pays a price λ_2) must be less than the expected reward, which is (0.4)(0.3)5. Hence,

$$\lambda_1 + 0.4\lambda_2 \leq 0.6. \tag{27}$$

But flow 1 can only push $(0.5)(0.4) = 0.2$ of its packets to node 2. So, there is spare capacity at node 2 that flow 2 can use. For flow 2 to use that, we need $\lambda_2 \leq (0.7)2 = 1.4$. Now, flow 2 needs to utilize the spare capacity of 0.2 left at node 2. So, it needs to ensure a flow of 0.2 reaches node 2. To do that, it needs

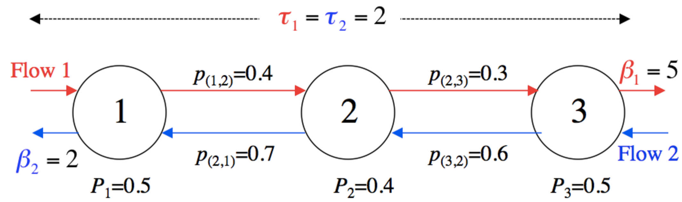


Fig. 4. System considered in Example 1.

to transmit $\frac{1}{3}$ of the packets that arrive since $\frac{1}{3}(0.6) = 0.2$. So, it needs to randomize at node 3. By complementary slackness, this can only happen if packets at node 3 are indifferent to being transmitted or not. So

$$\lambda_3 + 0.6\lambda_2 = (0.6)(0.7)(2) = 0.84.$$

Since we want to maximize $D(\lambda)$, we choose $\lambda_2 = 1.4, \lambda_3 = 0$, and, from (27), $\lambda_1 = 0.04$.

Therefore, we arrive at the following solution, where we denote by $\pi^f(i, s)$ the probability with which a packet of flow f is transmitted at node i when the time-till-deadline is s:

$$\lambda^\star = (0.04, 1.4, 0)$$

$$\pi^1(1,2) = 0.5, \pi^1(2,1) = 1$$

$$\pi^2(3,2) = \frac{1}{3}, \pi^2(2,1) = 1.$$

Now we verify that this policy is optimal. $\lambda_2 = 1.4$ implies $\pi^1(2,1) = 1$ since $1.4 \le (0.3)5$. Now $\lambda_1 + 0.4\lambda_2 = 0.6$ implies $\pi^1(1,2) = 1$ and $\pi^1(2,1) = 0$ are both optimal, i.e., a packet is indifferent to them, and so one may randomize between them to satisfy the average-power constraint. Similarly, $\lambda_2 = 1.4$ implies that both decisions $\pi^2(2,1) = 1$ and $\pi^2(2,1) = 0$ are optimal. Also, $\lambda_3 + 0.6\lambda_2 = 0.84$ implies both $\pi^2(3,2) = 1$ and $\pi^2(3,2) = 0$ are both optimal. So, we can randomize the transmission of packets of flow 2 in state $(3,2)$. The average-power usages are 0.5 W at node 1, 0.4 W at node 2, and 1/3 W at node 3. The average-power constraints of $P_1 = 0.5$ and $P_2 = 0.4$ at nodes 1 and 2, respectively, are met with equality. The average-power constraint at node 3 is slack but $\lambda_3 = 0$. So, complementary slackness holds. Hence, the policy is optimal.

Example 2: We now consider the same system as in Example 1, except that we relax the relative deadlines to $\tau^1 = \tau^2 = 3$, so that every packet can afford to have one hop that is retransmitted and still make it to its destination in time.

Consider a packet that has just arrived at node 1. It can either make it to its destination in two hops if both transmissions are successful the first time they are attempted, or it can fail once at node 1 and then be successful on subsequent transmissions at nodes 1 and 2, or it can succeed the first time at node 1, fail once at node 2, and then succeed at node 2 on the second attempt. If it does so reach its destination, it obtains a reward of 5. Hence, taking these possibilities into account, if a packet of flow 1 gets transmitted at every available opportunity, then the expected reward for a packet of flow 1 at its first visit to node $1 = [(0.4)(0.3) + (0.6)(0.4)(0.3) + (0.4)(0.7)(0.3)]5 = 1.38$. Similarly, expected reward for a packet of flow 2 at its first visit to node $3 = 1.428$, expected reward for a packet of flow 1 at its second visit to node $1 = 0.6$, expected reward for a packet of flow 2 at its second visit to node $3 = 0.84$, expected reward for a packet of flow 1 at its first visit to node $2 = 2.55$, expected reward for a packet of flow 2 at its first visit to node $2 = 1.82$, expected reward for a packet of flow 1 at its second visit to node $2 = 1.5$, and expected reward for a packet of flow 2 at its second visit to node $2 = 1.4$.

Packets of flow 1 are more valuable at node 2 than flow 2. So, we want to maximize the throughput of packets of 1 to node 2. If we transmit with probability 0.5 on the first attempt at node 1, then all power is used up. The maximum power that can be consumed by packets of flow 1 at node $2 = (0.5)(0.4) + (0.5)(0.4)(0.7) = 0.34$ W. So, there is still 0.06 W left at node 2 that can be used by packets of flow 2. After arriving at node 2 for the first time, a packet of flow 2 can use a maximum power of 1.3 W. So, flow 2 at node 3 needs to make $\frac{0.06}{(1.3)(0.6)}$ attempts, which amounts to randomization with probability $1/13$. In order to transmit a packet of flow 2 on its second visit to node 2, the price λ_2 cannot be any more than the expected reward $(0.7)2 = 1.4$. So, we could attempt some of the packets of flow 2 that arrive at node 3, and transmit some packets of flow 2 that arrive at node 2.

With $\lambda_2 = 1.4$, λ_1 needs to satisfy $\lambda_1 + (0.4)\lambda_2 + (0.4)(0.7)\lambda_2 = [(0.4)(0.7) + (0.4)(0.7)(0.3)]5$, so $\lambda_1 = 0.068$. Similarly, λ_3 needs to satisfy, $\lambda_3 + (0.6)\lambda_2 + (0.6)(0.3)\lambda_2 = [(0.6)(0.7) + (0.6)(0.7)(0.3)]2$, which yields $\lambda_3 = 0$. The power constraint at node 3 is slack, but $\lambda_3 = 0$. So, the price vector is $\lambda = (0.068, 1.4, 0)$. The corresponding probabilities of transmission are as follows:

$$\pi^1(1,3) = 0.5, \pi^1(1,2) = 0$$

$$\pi^1(2,2) = 1, \pi^1(2,1) = 1$$

$$\pi^2(3,3) = 1/13, \pi^2(3,2) = 0$$

$$\pi^2(2,2) = 1, \pi^2(2,1) = 1.$$

The optimal single-packet transportation dynamic programming equations yield

$$\begin{aligned} V^1(1,3) &= \text{Max}\left\{0, -0.068 + (0.6)V^1(1,2) + 0.4V^1(2,2)\right\} \\ &= 0, \text{ so both choices are optimal,} \\ &\quad \text{permitting randomization} \\ V^1(1,2) &= \text{Max}\left\{0, -0.68 + (0.6)(0) + 0.4V^1(2,1)\right\} = 0, \\ &\quad \text{again, both choices are optimal,} \\ &\quad \text{permitting randomization} \\ V^1(2,2) &= \text{Max}\left\{0, -1.4 + (0.7)V^1(2,1) + (0.3)5\right\} = 0.17 \\ V^1(2,1) &= \text{Max}\left\{0, -1.4 + (0.3)5\right\} = 0.1. \end{aligned}$$

In all of the following, both choices are again optimal

$$\begin{aligned} V^2(3,3) = \text{Max}\left\{0, 0 + (0.6)V^2(2,2) + 0.4V^2(3,2)\right\} = 0 \\ V^2(3,2) = \text{Max}\left\{0, 0 + 0.6V^2(2,1)\right\} = 0 \\ V^2(2,2) = \text{Max}\left\{0, -1.4 + (0.3)V^2(2,1) + (0.7)2\right\} = 0 \\ V^2(2,1) = \text{Max}\left\{0, -1.4 + (0.7)2\right\} = 0. \end{aligned}$$

Note that the power consumptions are as follows:

$$\begin{aligned} P_1 &= (1)(0.5) = 0.5, \text{ so it is tight} \\ P_2 &= (0.5)(0.4)\left[1 + (0.7)1\right] + \frac{1}{13}(0.6)\left[1 + (0.3)1\right] \\ &= 0.4, \text{ tight} \\ P_3 &= \frac{1}{13}. \end{aligned}$$

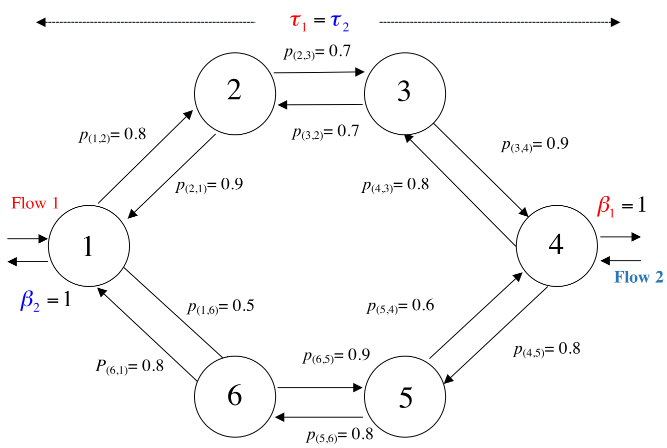


Fig. 5. Network with two source–destination pairs $(s_1 = 1, d_1 = 4)$ and $(s_2 = 4, d_2 = 1)$. Arrivals are deterministic with rates $A_1 = A_2 = 1$ per time slot. Link capacities are $C_{(i,j)} \equiv 1$ packet/time slot for all links (i, j) shown.

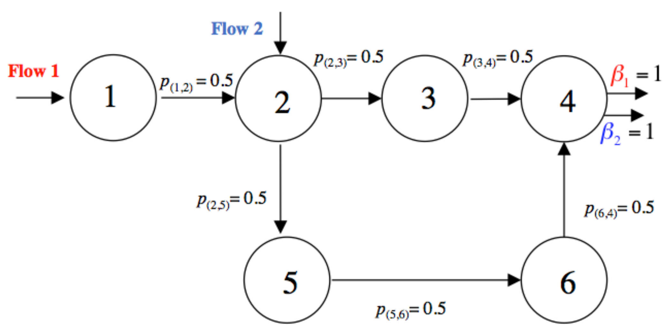


Fig. 6. Network with two source–destination pairs $(s_1 = 1, d_1 = 4), (s_2 = 2, d_2 = 4)$. The arrivals are deterministic with rates $A_1 = A_2 = 1$ per time slot. Link capacities are $C_{(i,j)} \equiv 1$ packet/time slot for all links (i, j) shown.

	Links (i,j)					
(τ_1,τ_2)	(1,2)	(2,3)	(3,4)	(2,5)	(5,6)	(6,4)
(4,3)	0.0	24.078	0.0	-0.5	22.7674	0.0
(5,4)	0.0107	48.3409	0.0518	0.0	48.7170	0.0
(6,5)	0.0037	40.3655	0.0276	0.0	39.7392	0.0

Fig. 7. Prices of links for the network in Fig. 6.

The last constraint is loose, but then $\lambda_3 = 0$, and we still have complementary slackness. So, the solution is optimal.

B. Simulations

Now we consider the case of link-capacity constraints (or equivalently peak-power constraints). We present a comparative simulation study of the asymptotically optimal policy with respect to the following two policies: 1) EDF scheduling combined with backpressure routing (EDF-BP), and 2) EDF scheduling combined with shortest path routing (EDF-SP) that routes packets along the shortest path from source to destination with ties broken randomly. We consider the systems shown in Figs. 5 and 6. All link capacities are just 1, so the asymptotically optimal policy is noteworthy for its excellent performance seen

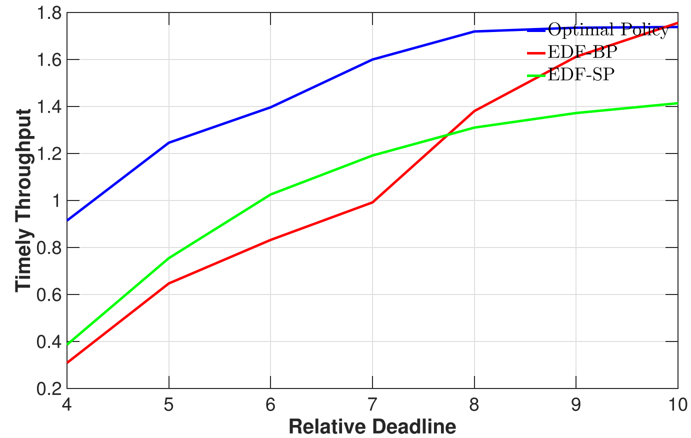


Fig. 8. Timely throughputs of the asymptotically optimal (labeled "Optimal"), EDF-BP and EDF-SP policies for the network in Fig. 5 as the relative deadlines of both flows are increased.

in the following, even in the very much nonasymptotic regime. The prices of the links are indicated in Fig. 7.

We compare the performance of the asymptotically optimal policy $\tilde{\pi}^\star$ of Theorem 4, with the following EDF-SP policy.

1) The link $\ell = (i, j)$ chosen for scheduling packet transmissions for flow f lies on the shortest path that connects the source and destination nodes of flow f.
2) Thereafter, on each link (i, j), it gives higher priority to packets having earlier deadlines. It then serves a maximum of $C_{(i,j)}$ packets in decreasing order of priority.

We also compare the performance with the EDF-BP policy. Under the EDF-BP policy, each node i maintains queues for each flow f and possible age s. Denoting by $Q_{i,f}(t, s)$, the queue length at node i at time t, and by $Q_{f,i}(t) = \sum_s Q_{i,f}(t, s)$ the total number of packets of flow f at node i at time t, the policy functions as follows.

1) For each outgoing link $\ell = (i, j)$, EDF-BP calculates the backlogs $Q_{f,i}(t) - Q_{f,j}(t)$ of flow f.
2) On each link $\ell = (i, j)$, it prioritizes packets on the basis of the backlogs associated with their flows. For packets belonging to the same flow, higher priority is given to packets having earlier deadlines.
3) It then serves a maximum of $C_{(i,j)}$ highest priority packets from among the packets whose flows have a positive backlog $Q_{f,i}(t) - Q_{f,j}(t)$.

Both EDF-SP and EDF-BP eject packets that have crossed their deadlines.

Figs. 8 and 10 show the comparative performances of the policies for the networks in Figs. 5 and 6 as the relative deadlines of the flows are varied. The performance of the asymptotically optimal policy is superior even in the nonasymptotic regime. Figs. 9 and 11 show the comparative performance as network capacities are increased.

Observe that for the network in Fig. 6, a shortcoming of EDF-SP is that it is unable to utilize the path $1 \to 2 \to 5 \to 6 \to 4$, and therefore performs worse than EDF-BP. Although it seems that in a general network the EDF-BP should be able to utilize all source–destination paths, it will neither be able to efficiently

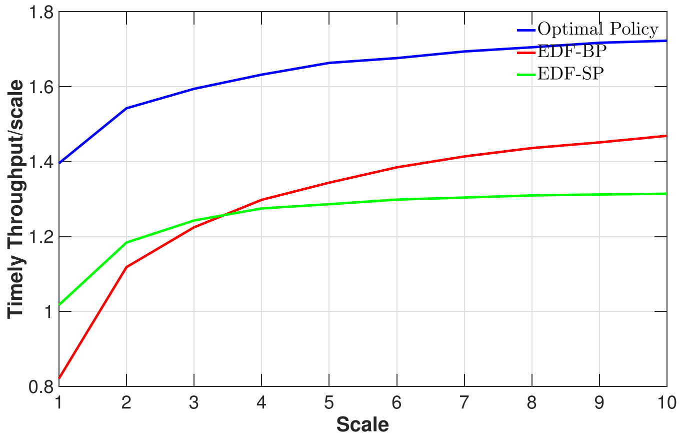


Fig. 9. Timely throughputs for the network in Fig. 5 as the relative deadlines of flows are increased. Relative deadline of flow 1 is one more than that of flow 2.

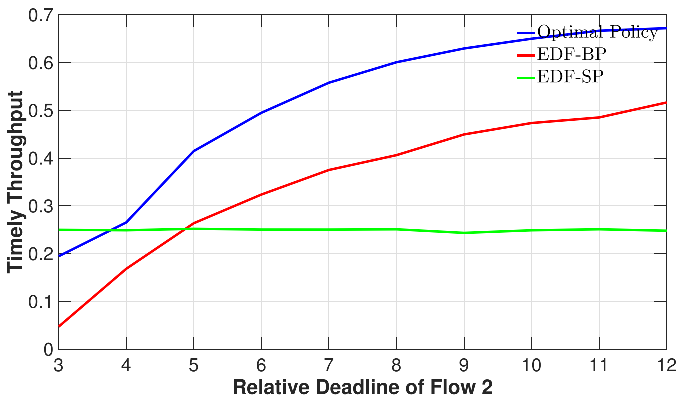


Fig. 10. Timely throughputs of the policies for the network in Fig. 6 as link capacities and arrival rate are scaled. The relative deadlines for both flows are set at six time slots.

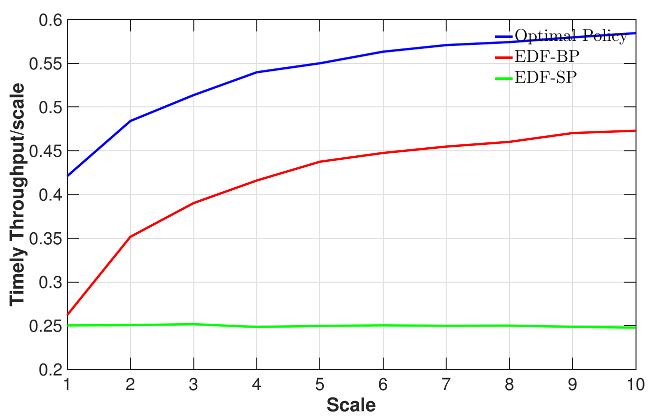


Fig. 11. Timely throughputs for the network in Fig. 6 as link capacities and arrival rates are scaled. Relative deadlines for flows 1 and 2 are 6 and 5, respectively.

prioritize packets based on their age, nor discover which paths are more efficient at delivering packets within their deadlines.

XIV. Concluding Remarks

We have addressed the problem of designing optimal decentralized policies that maximize the timely throughput of multihop wireless networks with average nodal power constraints and unreliable links, in which packets are useful only when they are delivered by their deadline. The key to our results is the observation that if the nodes are subject to average-power constraints, then the optimal solution is decoupled not only along nodes and flows, but also along packets within the same flow at a node. Each packet can be treated exclusively in terms of its time-till-deadline at a node. The decision to transmit a packet is governed by a "transmission price" that packets pay at each node, weighed against the reward that it collects at the destination if it reaches it before the deadline expires. The nodes need not share any information such as queue compositions, etc., among themselves in order to schedule packets. This result may be of interest since obtaining optimal decentralized policies for networks has long been considered an intractable problem. The overall solution is eminently tractable, being completely determined by an LP that can be solved offline, with the number of variables equal to the product of the square of the number of nodes, the number of flows, the maximum relative deadline, and the number of transmit power levels, rather than exponential in problem size. Thus, this paper fills two important gaps in the existing literature of policies for multihop networks: 1) hard per-packet end-to-end delay guarantees, and 2) optimal decentralized policies.

This result should be contrasted with the backpressure-based approach that has been developed over the past quarter century. It essentially considers the Lagrangian of the fluid model, and interprets the queue lengths as prices. It allows the design of throughput optimal policies, but not delay optimality, as one would expect from any fluid model-based analysis. In contrast, our approach studies the problem of joint routing, scheduling, and power control of packets under deadline constraints over a multihop network for the complete stochastic system involving all uncertainties, and consider its Lagrangian and the dual. This captures delay performance fully, and allows us to address the timely throughput optimality of packets that meet hard end-to-end deadlines. The Lagrange multipliers are the prices for energy usage paid by a packet to a node for transmitting its packet. They are very different from the queue length-based prices used by backpressure policies. We also consider the case of peak-power constraints at each node, which may be present in addition to, or as a replacement of, average-power constraints. A minor modification of the optimal policy for the case of average-power constraints, based on truncation, is asymptotically optimal as the network capacity is scaled.

This approach of dualizing the stochastic problem has broad ramifications, as has been explored in [50] and [51] for problems such as video transmission and energy storage.

In addition to millimeter wave networks, these results are also applicable to other networks where transmissions are directional, such as networks consisting of microwave repeaters, networks with directed antennas, or even unreliable wireline links.

References

[1] "FCC adopts rules to facilitate next generation wireless technologies," 2016. [Online]. Available: https://www.fcc.gov/document/fcc-adopts-rules-facilitate-next-generation-wireless-technologies

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] 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. 37, no. 12, pp. 1936–1948, Dec. 1992.

[4] X. Lin and N. Shroff, "Joint rate control and scheduling in multihop wireless networks," in *Proc. 43rd IEEE Conf. Decis. Control*, Dec. 14–17, 2004, pp. 1484–1489.

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

[6] A. Eryilmaz and R. Srikant, "Joint congestion control, routing, and MAC for stability and fairness in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1514–1524, Aug. 2006.

[7] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," in *Proc. IEEE Conf. Comput. Commun.*, Mar. 2003, vol. 1, pp. 745–755.

[8] L. Ying, S. Shakkottai, A. Reddy, and S. Liu, "On combining shortest-path and back-pressure routing over multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 841–854, Jun. 2011.

[9] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probab.*, vol. 25, pp. 287–298, 1988.

[10] R. Singh and P. R. Kumar, "Decentralized throughput maximizing policies for deadline-constrained wireless networks," in *Proc. IEEE Conf. Decis. Control*, 2015, pp. 3759–3766.

[11] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.

[12] L. Jiang and J. Walrand, "A distributed CSMA algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 960–972, Jun. 2010.

[13] I.-H. Hou, V. Borkar, and P. R. Kumar, "A theory of QoS for wireless," in *Proc. 28th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, Apr. 2009, pp. 486–494.

[14] I.-H. Hou and P. R. Kumar, "Admission control and scheduling for QoS guarantees for variable-bit-rate applications on wireless channels," in *Proc. 10th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, May 2009, pp. 175–184.

[15] I.-H. Hou and P. R. Kumar, "Scheduling heterogeneous real-time traffic over fading wireless channels," in *Proc. 29th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, San Diego, CA, USA, Mar. 2010, pp. 1631–1644.

[16] I.-H. Hou and P. R. Kumar, "Utility maximization for delay constrained QoS in wireless," in *Proc. 29th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, Mar. 2010, pp. 1–9.

[17] I.-H. Hou and P. R. Kumar, "Utility-optimal scheduling in time-varying wireless networks with delay constraints," in *Proc. 11th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2010, pp. 31–40.

[18] I.-H. Hou and P. R. Kumar, "Broadcasting delay-constrained traffic over unreliable wireless links with network coding," in *Proc. 12th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2011, pp. 33–42.

[19] I.-H. Hou, A. Truong, S. Chakraborty, and P. R. Kumar, "Optimality of periodwise static priority policies in real-time communications," in *Proc. 50th IEEE Conf. Decis. Control Eur. Control Conf.*, 2011, pp. 5047–5051.

[20] I.-H. Hou and P. R. Kumar, "Scheduling periodic real-time tasks with heterogeneous reward requirements," in *Proc. 32nd IEEE Real-Time Syst. Symp.*, 2011, pp. 282–291.

[21] I.-H. Hou and P. R. Kumar, "Real-time communication over unreliable wireless links: A theory and its applications," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 48–59, Feb. 2012.

[22] I.-H. Hou and P. R. Kumar, "A survey of recent results on real-time wireless networking," *Proc. Real-Time Wireless Ind. Appl.*, pp. 1–6, Apr. 2011.

[23] I.-H. Hou and P. R. Kumar, "Queueing systems with hard delay constraints: A framework and solutions for real-time communication over unreliable wireless channels," *Queueing Syst., Theory Appl.*, vol. 71, pp. 151–177, 2012.

[24] I.-H. Hou and P. R. Kumar, *Packets With Deadlines: A Framework for Real-Time Wireless Networks* (Synthesis Lectures on Communication Networks). San Rafael, CA, USA: Morgan & Claypool, 2013.

[25] R. Singh, I.-H. Hou, and P. R. Kumar, "Pathwise performance of debt based policies for wireless networks with hard delay constraints," in *Proc. 52nd IEEE Conf. Decis. Control*, Dec. 10–13, 2013, pp. 7838–7843.

[26] R. Singh, I.-H. Hou, and P. R. Kumar, "Fluctuation analysis of debt based policies for wireless networks with hard delay constraints," in *Proc. 33rd Annu. Joint Conf. IEEE Comput. Commun. Soc.*, Apr. 2014, pp. 2400–2408.

[27] S. Zhang and V. K. N. Lau, "Resource allocation for OFDMA system with orthogonal relay using rateless code," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4534–4540, Nov. 2008.

[28] M. Duarte and A. Sabharwal, "Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2010, pp. 1558–1562.

[29] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2010, pp. 1–12.

[30] E. Altman, "Constrained Markov decision processes with total cost criteria: Occupation measures and primal LP," *Math. Methods Oper. Res.*, vol. 43, no. 1, pp. 45–72, 1996.

[31] E. Altman, *Constrained Markov Decision Processes*. London, U.K.: Chapman & Hall/CRC Press, Mar. 1999.

[32] A. S. Manne, "Linear programming and sequential decisions," *Manage. Sci.*, vol. 6, pp. 259–267, 1960.

[33] V. S. Borkar, "Control of Markov chains with long-run average cost criterion," in *The IMA Volumes in Mathematics and Its Applications*, W. Fleming and P. Lions, Eds. New York, NY, USA: Springer, 1988, pp. 57–77.

[34] D. P. Bertsekas, A. E. Ozdaglar, and A. Nedic, *Convex Analysis and Optimization* (Athena Scientific Optimization and Computation Series). Belmont, MA, USA: Athena Scientific, 2003.

[35] S. M. Ross, *Applied Probability Models With Optimization Applications*. San Francisco, CA, USA: Holden–Day, 1970.

[36] M. Morishima, *Walras' Economics : A Pure Theory of Capital and Money*. Cambridge, U.K.: Cambridge Univ. Press, 1977.

[37] S. Boyd, 2003. [Online]. Available: https://web.stanford.edu/class/ee392o/sub-grad_method.pdf

[38] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.

[39] V. S. Borkar, *Stochastic Approximation : A Dynamical Systems Viewpoint*. New Delhi, India: Cambridge Univ. Press, 2008.

[40] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 1997.

[41] V. S. Borkar, "Stochastic approximation with two time scales," *Syst. Control Lett.*, vol. 29, no. 5, pp. 291–294, 1997.

[42] J. Gittins and D. Jones, "A dynamic allocation index for the sequential design of experiments," in *Progress in Statistics*, J. Gani, Ed., Amsterdam, The Netherlands: North Holland, 1974, pp. 241–266.

[43] J. Gittins, K. Glazebrook, and R. Weber, *Multi-Armed Bandit Allocation Indices.* Hoboken, NJ, USA: Wiley, 2011.

[44] R. R. Weber and G. Weiss, "On an index policy for restless bandits," *J. Appl. Probab.*, vol. 27, no. 3, pp. 637–648, 1990.

[45] P. Diaconis and Y. Zabell, "Closed form summation for classical distributions: Variations on theme," *Statist. Sci.*, vol. 6, no. 3, pp. 284–302, 1991.

[46] J. Jaramillo and R. Srikant, "Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic," in *Proc. IEEE Conf. Comput. Commun.*, 2010, pp. 1–9.

[47] S. Shakkottai and A. L. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," in *Teletraffic Engineering in the Internet Era: Proceedings of the International Teletraffic Congress* (Teletraffic Science and Engineering), J. M. de Souza, N. L. S. da Fonseca, and E. de Souza e Silva, Eds., New York, NY, USA: Elsevier, 2001, vol. 4, pp. 793–804.

[48] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.

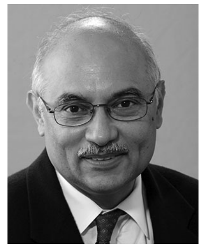
[49] T. Rappaport, *Wireless Communications: Principles and Practice*. New York, NY, USA: Prentice-Hall, 2002.

[50] R. Singh and P. R. Kumar, "Optimizing quality of experience of dynamic video streaming over fading wireless networks," in *Proc. 54th IEEE Conf. Decis. Control*, Dec. 15–18, 2015, pp. 7195 –7200.

[51] R. Singh and P. R. Kumar, "On storage and renewables: A theory of sizing and uncertainty," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 26–30, 2015, pp. 7195–7200.

Rahul Singh (S'16–M'18) received the B.E. degree in electrical engineering from Indian Institute of Technology Kanpur, Kanpur, India, in 2009, the M.Sc. degree in electrical engineering from the University of Notre Dame, South Bend, IN, USA, in 2011, and the Ph.D. degree in electrical and computer engineering from the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, in 2015.

In 2015, he joined, as a Postdoctoral Associate, the Laboratory for Information Decision Systems, Massachusetts Institute of Technology, where he worked on several stochastic and adversarial decision-making problems that arise in network control. He then worked with Encoredtech as a Data Scientist on machine learning problems arising in time-series modeling. He is currently with Intel Santa Clara, CA, USA, as Deep Learning Engineer. His research interests include decentralized control of large-scale complex cyber-physical systems, operation of electricity markets with renewable energy, machine learning, deep learning, and scheduling of stochastic networks serving real-time traffic.



P. R. Kumar (F'88) received the B.Tech. degree in electronics engineering from Indian Institute of Technology (IIT) Madras, Chennai, India, in 1973, and the D.Sc. degree in systems science and mathematics from Washington University in St. Louis, St. Louis, MO, USA, in 1977.

He is currently with Texas A&M University, College Station, TX, USA. He was a faculty member with the University of Maryland, Baltimore County (1977–1984) and the University of Illinois at Urbana-Champaign (1985–2011). He was the Leader of the Guest Chair Professor Group on Wireless Communication and Networking with Tsinghua University. He is a D. J. Gandhi Distinguished Visiting Professor with IIT Bombay, and an Honorary Professor with IIT Hyderabad. His research interests include cyber-physical systems, cybersecurity, privacy, wireless networks, renewable energy, smart grid, autonomous vehicles, and unmanned air vehicle systems.

Prof. Kumar is a member of the U.S. National Academy of Engineering, The World Academy of Sciences, and the Indian National Academy of Engineering. He was awarded a Doctor Honoris Causa by ETH Zurich. He was the recipient of the IEEE Field Award for Control Systems, the Donald P. Eckman Award of the AACC, Fred W. Ellersick Prize of the IEEE Communications Society, the Outstanding Contribution Award of ACM SIGMOBILE, the INFOCOM Achievement Award, and the SIGMOBILE Test-of-Time Paper Award. He is a Fellow ACM. He was also the recipient of the Distinguished Alumnus Award from IIT Madras, the Alumni Achievement Award from Washington University in St. Louis, and the Daniel Drucker Eminent Faculty Award from the College of Engineering, University of Illinois at Urbana-Champaign.