Resource Allocation for Multi-hop Wireless Networks

Atilla Eryilmaz
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA, USA
Email: eryilmaz@mit.edu

R. Srikant
Coordinated Science Laboratory
Electrical and Computer Engineering
University of Illinois at Urbana, Champaign
Urbana, IL, USA
Email: rsrikant@uiuc.edu

***Abstract*— In this work, we describe and analyze a joint scheduling, routing and congestion control mechanism for wireless networks, that asymptotically guarantees stability of the buffers and fair allocation of the network resources. The queue lengths serve as common information to different layers of the network protocol stack, which are otherwise decoupled. Our main contribution is to prove the asymptotic optimality of a *primal-dual congestion controller*, which is known to model different versions of TCP well.**

I. Introduction

Consider a set of flows that share the resources of a fixed wireless network. Each flow is described by its source-destination node pair, with no apriori established routes. The limited power resources and interference amongst concurrent transmissions necessitate multi-hop transmission. Hence, the nodes that constitute the network must cooperate by forwarding each others' packets towards their destinations. Thus, each node may need to maintain buffers to hold packets of those flows other than its own. For such a system, in this paper, we give an affirmative answer to the following question: is it possible to design a realistic mechanism that finds the optimal routes for the flows, guarantees stability of the buffers at the nodes, and drives the mean flow rates to a system-wide optimum level? Moreover, we show that the analysis is robust to a number of modifications to the mechanism, and extensions to the network model.

The question of stability of wireless networks has first been addressed by Tassiulas and Ephremides[20] with the assumption that the incoming flows are *inelastic*, i.e., the flow rates are fixed as for voice or video traffic. Subsequently, there has been a large body of work that extended the same idea to different scenarios and more general settings[21], [18], [19], [2], [16], [6], [15], [5]. However, these works do not consider the case of traffic whose rate can be adjusted online.

In the context of wireline networks, the idea of a distributed flow control based on a system-wide optimization problem was developed in [7], and followed by others in [12], [8], [1], [**?**]; see [17] for a survey. In these works, the main contribution was the design of a distributed congestion control mechanism to drive the rates of *elastic* flows towards the system-wide optimum.

More recently, the problem of serving elastic traffic over wireless networks has been investigated in [**?**], [**?**], [9], [**?**], [4], [14], [10], [**?**]. Here, the queues and the wireless characteristics of the network are included in the system model. The main idea in these works has been to combine the results on scheduling inelastic traffic in wireless networks and distributed congestion control in wireline networks to design *joint scheduling-congestion control mechanisms* that guarantee optimal routes, stability and optimal rate allocation. These papers prove that a decentralized congestion controller at the transport layer working in conjunction with a queue-length-based scheduler at the medium access control (MAC) layer will asymptotically achieve buffer stability, optimal routing and fair rate allocation. Moreover, these layers are coupled through common queue length information.

In [9], [4], [14], [**?**], authors propose and study rate control algorithms that adapt the flow rates instantaneously as a function of the entry queue lengths. The rate control mechanism studied in all of these works can be categorized as the *Dual Congestion Controller* since it can be interpreted as a gradient algorithm for the dual of an optimization problem. However, the intrinsic assumption of the dual congestion control mechanism that the rate levels can be changed instantaneously may not be reasonable. For example, it is well known that sliding-window based flow control mechanisms such as TCP respond to congestion feedback not instantaneously, but gradually. Such a response is desired by practitioners because the rate fluctuations are small. Thus, the study of another algorithm that modifies the flow rates iteratively is important. To this end, we propose and study the so called *Primal-Dual Congestion Controller* in this work. Primal-dual algorithms are well known in the optimization literature and have been studied extensively in different contexts[1], [**?**], [17], [11]. Since the response of the primal-dual controller is more gradual compared to the dual controller, it is not immediately clear as to whether the buffer stability and rate convergence properties will be maintained.

Here, it must be stressed that even though the congestion control is distributed, the scheduling is still assumed to be centralized in this work. In [10], [22], [3], [**?**], [**?**], the impact of de-centralized implementations of the scheduler is studied. We note that the results of this work can be extended to distributed and asynchronous implementations for a special class of interference models using the approach in [3].

The paper is organized as follows: Section II describes the system model. In Section III, we state the goal as an optimization problem and state the characteristics of the optimum point. Section IV-A introduces the queue-length-based scheduler that is implemented at the MAC layer. We propose and provide an extensive study of the primal-dual congestion controller in Section IV-B. Various modifications and extensions to the system are studied in Section VI. Finally, we give concluding remarks in Section VII.

II. System Model

We assume that the network is represented by a graph, $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of directed links. If a link (n, m) is in $\mathcal{L}$, then it is possible to send packets from node n to node m subject to the interference constraints to be described shortly. We let $\mu = \{\mu_l\}_{l\in\mathcal{L}}$ denote the rate vector at which data can be transfered over each link $l \in \mathcal{L}$. We assume that there is an upper bound, $\hat{\eta} < \infty$, on each μ_l. For ease of presentation we assume that there is no fading in the environment. We will discuss the extension of the model to include time-variations in a later section.

We let $\hat{\Gamma}$ denote a bounded region in the $|\mathcal{L}|$ dimensional real space, representing the set of μ that can be achieved in a given time slot, i.e., it represents the interference constraint. In general, the set need not be convex. In fact, a typical case would be a discrete set of rates that can be achieved, and hence be non-convex. We let $\Gamma := \mathcal{CH}\{\hat{\Gamma}\}$ denote the convex hull of the set $\hat{\Gamma}$. It is well known that by time-sharing between different rate vectors in $\hat{\Gamma}$, any point in Γ can be attained.

We use $\mathcal{F}$ to denote the set of flows that share the network resources. The routes of these flows are not specified apriori, but established by the backpressure scheduling algorithm to be described in Section 2. We use $b(f)$ to denote the beginning node, and $e(f)$ to denote the end node of flow f.

Associated with each flow f is a utility function $U_f(x_f)$, which is a function of the flow rate x_f. The utility function, denoted by $U_f(\cdot)$ for flow f, is assumed to be strictly concave and nondecreasing, and to satisfy a set of other conditions which we omit in this paper due to space constraints. The details can be found in [ref]. However, we note that these conditions are not restrictive and hold for the following class of utility functions.

$$U_f(x) = \beta_f \frac{x^{1-\alpha_f}}{(1-\alpha_f)} \qquad \forall\, \alpha_f > 0. \tag{1}$$

This class of utility functions is known to characterize a large class of fairness concepts [13]. Next, we describe the capacity region of the network as in [15], [9].

Definition 1 (Capacity region): The *capacity region*, Λ, of the network contains the set of flow rates $\mathbf{x} \geq \mathbf{0}$ for which there exists a set $\left\{\mu_l^{(d)}\right\}_{l\in\mathcal{L}}^{d\in\mathcal{N}}$ that satisfies

- $\mu_l^{(d)} \geq 0$ for all $l \in \mathcal{L}$, $d \in \mathcal{N}$.
- $\left[\sum_d \mu_l^{(d)}\right]_l \in \Gamma.$
- $\mu_{into(n)}^{(d)} + \sum_f x_f \mathcal{I}_{\{b(f)=n\}}^{\{e(f)=d\}} \leq \mu_{out(n)}^{(d)}$, $\forall n, d(\neq n) \in \mathcal{N}$,

where $\mu_{into(n)}^{(d)} := \sum_{(k,n)\in\mathcal{L}} \mu_{(k,n)}^{(d)}$ denotes the potential number of packets that are destined for node d, incoming to node n. $\mu_{out(n)}^{(d)} := \sum_{(n,m)\in\mathcal{L}} \mu_{(n,m)}^{(d)}$ has a similar definition for the potential outgoing number of packets. Also, $\mathcal{I}_A^B$ is the indicator function of the event $A \cap B$.

It is assumed that each node maintains a separate queue for those flows that have the same destination. We use $q_{n,d}[t]$ to denote the number of packets that are destined for node d, waiting for service at node n at time t. Then, for each $n \in \mathcal{N}$, and $d \in \mathcal{N}\backslash\{n\}$, the evolution of $q_{n,d}$ is given by

$$\begin{aligned} q_{n,d}[t+1] = (q_{n,d}[t] + \sum_f x_f[t]\, \mathcal{I}_{\{b(f)=n\}}^{\{e(f)=d\}} \\ + s_{into(n)}^{(d)}[t] - s_{out(n)}^{(d)}[t]), \end{aligned} \tag{2}$$

where $s_{into(n)}^{(d)}[t] := \sum_{(k,n)\in\mathcal{L}} s_{(k,n)}^{(d)}[t]$ and $s_{out(n)}^{(d)}[t] := \sum_{(n,m)\in\mathcal{L}} s_{(n,m)}^{(d)}[t]$, and $s_{(n,m)}^{(d)}[t]$ denotes the rate provided to d-destined packets over link (n, m) at slot t. Notice that, this is the actual amount of packets served over the link, not the potential amount denoted by $\mu_{(n,m)}^{(d)}[t]$.

III. Problem Statement and Characterization of the Optimal Point

Our goal is to design a congestion control/scheduling mechanism such that the flow rate vector, $\mathbf{x}$, solves the following optimization problem:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \sum_{f\in\mathcal{F}} U_f(x_f) \\ s.t. \quad & \mathbf{x} \in \Lambda \end{aligned} \tag{3}$$

which refer to (3) as the *primal problem.* Due to the strict concavity assumption of $U_f(\cdot)$ and the convexity of the capacity region Λ, there exists a unique optimizer of the primal problem, which we refer to as $\mathbf{x}^\star$. We call this the *fair rate allocation*. One can use duality theory by defining $\lambda_{n,d}$ to be the Lagrange multiplier associated with the constraint $\mu_{into(n)}^{(d)} + \sum_f x_f \mathcal{I}_{\{b(f)=n\}}^{\{e(f)=d\}} \leq \mu_{out(n)}^{(d)}$ to get the dual function, which we denote by $D(\lambda)$. Here, $\lambda_{n,d}$ can be interpreted as the price of transferring a unit amount of data from node n to node d. Thus, $\lambda_{b(f),e(f)}$ is nothing but the price of transferring a unit amount of data from the source of flow f to its destination. Such an approach was first taken in [9] where it was shown that for this problem, the duality gap vanishes. Hence, there exists a nonempty set $\Psi^\star$ of optimal Lagrange multipliers that satisfy $\sum_{f\in\mathcal{F}} U_f(x_f^\star) = D(\lambda^\star)$, for all $\lambda^\star \in \Psi^\star$.

IV. Joint Scheduling, Routing and Flow Control Mechanism

We describe the mechanism in two parts. The first part deal with the routing and scheduling component, and the second part is the congestion control algorithm.

A. Routing and Scheduling Algorithm

We use a queue-length-based scheduler known as the *back-pressure scheduler* introduced by Tassiulas [18], which uses the differential backlog at the two end nodes of a link to determine the rate of that link.

Definition 2 (Back-pressure Scheduler): At slot t, for each $(n,m) \in \mathcal{L}$, we define the differential backlog for destination node d as $W_{(n,m),d}[t] := (q_{n,d}[t] - q_{m,d}[t])$. Also, we let $W_{(n,m)}[t] = \max_d \left\{W_{(n,m),d}[t]\right\}$ and $d_{(n,m)}[t] = arg\,max_d \left\{W_{(n,m),d}[t]\right\}$. Then, choose the rate vector $\mu[t] \in \hat{\Gamma}$ that satisfies

$$\mu[t] \in arg\,max_{\{\eta\in\hat{\Gamma}\}} \sum_{\{(n,m)\in\mathcal{L}\}} \eta_{(n,m)} W_{(n,m)}[t], \qquad (4)$$

then serve the queue holding packets destined for node $d_{(n,m)}[t]$ over link (n,m) at rate $\mu_{(n,m)}[t]$. That is, we set $\mu_{(n,m)}^{(d_{(n,m)}[t])}[t] = \mu_{(n,m)}[t]$. The rest of the queues at node n are not served at slot t.

Fact 1: The maximization in (4) can be performed over Γ instead of $\hat{\Gamma}$, because the optimal rate vector must always contain at least one element from $\hat{\Gamma}$. This follows from the linearity of the objective function and the fact that $\Gamma = \mathcal{CH}\{\hat{\Gamma}\}$.

B. Primal-Dual Congestion Controller

The primal-dual congestion controller changes the flow rate vector in an iterative manner. This is a realistic model for window based flow control mechanisms implemented in many versions of TCP, because for such mechanisms the flow rates are gradually increased or decreased depending on the congestion feedback from the network.

Definition 3 (Primal-Dual Congestion Controller): At the beginning of time slot t, each flow, say f, has access to the queue length of its first node, i.e. $q_{b(f),e(f)}[t]$. Then the data rate $x_f[t]$ of flow f is an independently distributed random variable that satisfies

$$x_f[t+1] = \min\left\{x_f[t] + \alpha\left(KU_f'\left(x_f[t]\right) - q_{b(f),e(f)}[t]\right)\right\}_m^M,$$

where the notation $\{y\}_a^b$ projects the value of y onto the interval $[a,b]$. We assume that m is a fixed positive valued quantity that can be arbitrarily small, $M > 2\hat{\eta}$. These parameters guarantee that the amount of data generated at every time slot is within realistic boundaries. Also, K is a multiplicative constant that will be used to guarantee convergence of the achieved rates to the fair allocation. In particular, we are interested in the performance of the system for large K.

V. Analysis of the System

In this section, we first introduce a heuristic fluid model of the joint scheduler-congestion control mechanism, and prove its stability and convergence properties, and then show the convergence properties of the discrete-time primal-dual algorithm.

1) Continuous-time Fluid model Analysis: We assume that the time is continuous and the evolution of each queue is governed by the differential equation: for each $n \in \mathcal{N}$, and $d \neq n$,

$$\dot{q}_{n,d}(t) = \left(\sum_f x_f(t)\,\mathcal{I}_{b(f)=n}^{e(f)=d} + \mu_{into(n)}^{(d)}(t) - \mu_{out(n)}^{(d)}(t)\right)_{q_{n,d}}^+$$

where $(y)_z^+ := (y)_{z\geq 0}$ and $(y)_{z\geq a}$ is equal to y when $z > a$ and is equal to $\max(y,0)$ when $z = a$. Here, (t) is used instead of $[t]$ to signify that we are working in continuous-time. The back-pressure algorithm computes the link schedules and rates at every instant of time as described in Section 2. Finally, the congestion controller is assumed to determine the instantaneous flow rates such that

$$\dot{x}_f(t) = \alpha\left(KU_f'\left(x_f(t)\right) - q_{b(f),e(f)}(t)\right)_{x_f(t)\geq m} \qquad \forall\, f \in \mathcal{F},$$

where we take $m > 0$ such that it satisfies: $m < x_f^\star$ for all $f \in \mathcal{F}$. Then, the following global asymptotic stability result holds.

Theorem 1: Starting from any $\mathbf{x}(0)$ and $\mathbf{q}(0)$, the rate vector $\mathbf{x}(t)$ converges to $\mathbf{x}^\star$ as t goes to infinity. Moreover, the queue length vector $\mathbf{q}(t)$ approaches the bounded set $\tilde{\mathcal{S}}$ described by

$$\tilde{\mathcal{S}} = \mathcal{S} \cap \{\mathbf{q} \geq \mathbf{0} \,:\, \mathbf{q_{b(f),e(f)}} = \lambda^\star_{\mathbf{b(f),e(f)}} \text{ for all } \mathbf{f} \in \mathcal{F}\},$$

where $\mathcal{S} = \Bigg\{\mathbf{q} \geq \mathbf{0} \,:\, \sum_{\mathbf{n,d}} (\mathbf{q_{n,d}(t)} - \lambda^\star_{\mathbf{n,d}})$
$\times \left(\mu_{out(n)}^{(d)}(t) - \mu_{into(n)}^{(d)}(t) - \sum_f x_f^\star \mathcal{I}_{\{b(f)=n\}}^{\{e(f)=d\}}\right) = 0\Bigg\}$.

Proof: The proof follows from the application of LaSalle's invariance principle applied to the following Lyapunov function:

$$W(\mathbf{x},\mathbf{q};\lambda^\star) := \sum_{f\in\mathcal{F}} \frac{(x_f - x_f^\star)^2}{2\alpha} + \sum_{n\in\mathcal{N}}\sum_{d\in\mathcal{N}} \frac{(q_{n,d} - \lambda^\star_{n,d})^2}{2},$$

which is introduced in [17]. For details, please refer to [ref]. ∎

2) Discrete-time Model Analysis: The evolution of each flow rate is governed by the primal-dual iteration as described in Definition 3. Also, recall that the queue-lengths evolve according to (2). We define μ_{sym} to be the maximum flow rate that can be provided to all the flows, i.e., $\mu_{sym} := \max\{\eta \geq 0 : (\eta, \cdots, \eta) \in \Lambda\}$. We assume that $\mu_{sym} > m > 0$, and that $x_f^\star > m, \forall f$. These are reasonable assumptions given that we are free to choose m as small as necessary to satisfy them.

The following proposition establishes the asymptotic boundedness, and hence the stability of the queue lengths. The proof is provided in [ref].

Proposition 1: For $\alpha = \frac{1}{K^2}$, and for some finite constant c, we have: $\limsup_{t\to\infty} \sum_{f\in\mathcal{F}} q^2_{b(f),e(f)}[t] \leq cK^2$.

Next we state the main theorem which shows that the average rate obtained by each user is close to its *fair* share as defined by the resource allocation problem (3).

Theorem 2: For $\alpha = \frac{1}{K^2}$, and for some finite $B > 0$, we have: for all $f \in \mathcal{F}$,

$$x_f^\star - \frac{B}{\sqrt{K}} \leq \liminf_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} x_f[t] \leq \limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} x_f[t] \leq x_f^\star + \frac{B}{\sqrt{K}}.$$

Proof: The proof of this theorem uses the results of Theorem 1 and drift analysis. We refer the interested reader to [ref] for the whole proof. ■

Note that Theorem 2 directly implies that $\lim_{K\to\infty} \bar{x}_f = x_f^\star$, for all $f \in \mathcal{F}$, where $\bar{x}_f$ denotes the average rate of flow f, i.e., $\bar{x}_f := \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} x_f[t]$. Hence, the fair allocation is asymptotically attained by the joint scheduling, routing, and primal-dual congestion control mechanism.

VI. Extensions and Variations

In this section, we discuss possible extensions and variations to the joint mechanism that we studied up to this point. The goal is two-fold. First, we emphasize that the analysis can be extended to consider more realistic models. Towards this end, we discuss the inclusion of time-variations in the channel conditions into the model. Second, we consider different versions of congestion controllers that may be of interest, and show that their stability and convergence properties can be proved using the same techniques introduced earlier. In the interest of space, we only provide an outline of the arguments and do not provide complete proofs.

A. Stochastic Channel Models

Recall that our wireless network model assumes time-invariant channel conditions. However, in reality, channel conditions will fluctuate due to the change in the environment. Also, the arrivals can be random to model various implementation details. To accommodate these situations, the network model can be appropriately extended[ref]. Then, the arguments of Section IV-B can easily be modified to argue the stability and convergence properties of the system.

B. Dual Congestion Controller

A dual congestion controller aims to change the end-to-end flow rates in a direction so as to minimize the dual objective of (3). For this controller, the data rate $x_f[t]$ of flow f is an independently distributed random variable that satisfies

$$E[x_f[t] \mid q_{b(f),e(f)}[t]] = \min\left\{ U_f'^{-1}\left(\frac{q_{b(f),e(f)}[t]}{K}\right), M\right\}.$$

The heuristic fluid model of this controller is given by $x_f(t) = U_f'^{-1}\left(\frac{q_{b(f),e(f)}(t)}{K}\right)$ forall $f \in \mathcal{F}$. For this model, the global asymptotic stability of the joint scheduling and congestion control mechanism can be proved using LaSalle's invariance principle by studying the Lyapunov function: $V(\mathbf{q};\lambda^\star) = \frac{1}{2}\sum_{n\in\mathcal{N}}\sum_{d\in\mathcal{N}} (q_{n,d} - \lambda^\star_{n,d})^2$, for any given $\lambda^\star \in K\Psi^\star$. Hence, for the fluid model, system evolves toward the optimal rate allocation, and the queue lengths change in such a way that the backlogs converge to the optimal Lagrange multiplier set, $K\Psi^\star$. Then, we can return to the original discrete-time, stochastic system model described above and study its performance as we did in Section V-.2.

Here, due to the direct relationship between the queue lengths and mean flow rates, we can also give bounds for the queue length vector's proximity to $K\Psi^\star$ by adding a new condition on the utility functions, which is again satisfied by the class of utility functions described by (1). Then, we can state the following two results:

Result 1: There exists positive constants $\bar{c} < \infty$ and $\sigma \in (0, 1)$, that depend on Λ, the utility function set $\{U_f(\cdot)\}$, and the moments of the arrival processes, such that for each $\lambda^\star \in K\Psi^\star$,

$$E\left[\|\mathbf{q}^\infty - \lambda^\star\|\right] \leq \bar{c}(K)^\sigma, \qquad \text{for large } K \qquad (5)$$

where $\mathbf{q}^\infty$ is a notation used to denote the state of the Markov chain in steady-state and $\|\cdot\|$ denotes the Euclidean distance.

Result 2: The mean of the stationary rate vector converges to $\mathbf{x}^\star$ as K increases, i.e.,
$E[\mathbf{x}^\infty] \to \mathbf{x}^\star$ as $K \to \infty$, where $\mathbf{x}^\infty$ is defined by $x_f^\infty = \min\left\{U_f'^{-1}\left(\frac{q^\infty_{b(f),e(f)}}{K}\right), M\right\}$ for all $f \in \mathcal{F}$.

Thus, Result 1 establishes the stability of the buffers, and Result 2 proves the asymptotic optimality of the rates achieved by the dual controller, and as an alternate proof of the results in [9], [**?**], [14].

C. Capturing TCP behavior

It is shown in [17] that different versions of TCP can be modeled by the following rate evolution: for all $f \in \mathcal{F}$

$$\dot{x}_f(t) = \kappa_f(x_f(t))\left(KU_f'(x_f(t)) - q_{b(f),e(f)}(t)\right)_{x_f(t)\geq m},$$

where $\kappa_f(\cdot)$ is any non-decreasing, continuous function with $\kappa_f(y) > 0$ for any $y > 0$. For this rate evolution, the approach used to prove the primal-dual algorithm of Section IV-B can be applied. However, the Lyapunov function must be modified. Then, convergence of the model can be proved using the same line of reasoning as in Section IV-B (see [ref] for details).

VII. Conclusions

In this work, we study a cross-layer scheduling-congestion control mechanism for wireless networks. We model many existing versions of TCP/AQM schemes using the primal-dual congestion controller. It is shown that this controller, along with suitable MAC/routing protocol achieves fairness and stability. Architecturally, we maintain the traditional protocol stack, but couple them through the use of queue length information.

Acknowledgment

The LaTeX style file for IZS 2006 is essentially identical with `IEEEtran.cls` (version V1.6b), which is maintained by Michael Shell.

References

[1] T. Alpcan and T. Başar. A utility-based congestion control scheme for internet-style networks with delay. In *Proceedings of IEEE Infocom*, San Francisco, California, March-April 2003.

[2] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting. Scheduling in a queueing system with asynchronously varying service rates, 2000. Bell Laboratories Technical Report.

[3] L. Bui, A. Eryilmaz, R. Srikant, and X. Wu. Joint asynchronous congestion control and distributed scheduling for wireless networks. Submitted to IEEE Infocom 2006.

[4] A. Eryilmaz and R. Srikant. Fair resource allocation in wireless networks using queue-length based scheduling and congestion control. In *Proceedings of IEEE Infocom*, Miami, FL, March 2005.

[5] A. Eryilmaz, R. Srikant, and J. Perkins. Stable scheduling policies for fading wireless channels. In *Proceedings of IEEE International Symposium on Information Theory*, 2003. To appear in the *IEEE/ACM Transactions on Networking, 2004*.

[6] A. Eryilmaz, R. Srikant, and J. R. Perkins. Throughput optimal scheduling for broadcast channels. In *Modelling and Design of Wireless Networks, Proceedings of SPIE, E. K. P. Chong (editor)*, volume 4531, pages 70–78, Denver, CO, 2001.

[7] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[8] S. Kunniyur and R. Srikant. A time-scale decomposition approach to adaptive ECN marking. *IEEE Transactions on Automatic Control*, 47(6):882–894, June 2002.

[9] X. Lin and N. Shroff. Joint rate control and scheduling in multihop wireless networks. In *Proceedings of IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004.

[10] X. Lin and N. Shroff. The impact of imperfect scheduling on cross-layer rate control in multihop wireless networks. In *Proceedings of IEEE Infocom*, Miami, FL, March 2005.

[11] S. Liu, T. Başar, and R. Srikant. Exponential RED: A stabilizing AQM scheme for low and high speed TCP, 2003. An earlier version appeared in the Proceedings of IEEE Conference on Decision and Control, December 2003, under the title "Controlling the Internet: A survey and some new results".

[12] S. H. Low and D. E. Lapsley. Optimization flow control, I: Basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7:861–875, December 1999.

[13] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, October 2000.

[14] M.J. Neely, E. Modiano, and C. Li. Fairness and optimal stochastic control for heterogeneous networks. In *Proceedings of IEEE Infocom*, pages 1723–1734, Miami, FL, March 2005.

[15] M.J. Neely, E. Modiano, and C.E. Rohrs. Dynamic power allocation and routing for time varying wireless networks. In *Proceedings of IEEE Infocom*, pages 745–755, April 2003.

[16] S. Shakkottai and A. Stolyar. Scheduling for multiple flows sharing a time-varying channel: The exponential rule. *Translations of the AMS, Series 2,* A volume in memory of F. Karpelevich, 207:185–202, 2002.

[17] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhäuser, Boston, MA, 2004.

[18] L. Tassiulas. Scheduling and performance limits of networks with constantly varying topology. *IEEE Transactions on Information Theory*, 43:1067–1073, May 1997.

[19] L. Tassiulas. Linear complexity algorithms for maximum throughput in radio networks and input queued switches. In *Proceedings of IEEE Infocom*, pages 533–539, 1998.

[20] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 36:1936–1948, December 1992.

[21] L. Tassiulas and A. Ephremides. Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, 39:466–478, March 1993.

[22] X. Wu and R. Srikant. Regulated maximal matching: A distributed scheduling algorithm for multi-hop wireless networks with node-exclusive spectrum sharing, 2005. Submitted to *IEEE Conference on Decision and Control*.