

# Optimal Scheduling for Fair Resource Allocation in Ad Hoc Networks with Elastic and Inelastic Traffic

Juan José Jaramillo, *Member, IEEE*, and R. Srikant, *Fellow, IEEE*

**Abstract**—This paper studies the problem of congestion control and scheduling in ad hoc wireless networks that have to support a mixture of best-effort and real-time traffic. Optimization and stochastic network theory have been successful in designing architectures for fair resource allocation to meet long-term throughput demands. However, to the best of our knowledge, strict packet delay deadlines were not considered in this framework previously. In this paper, we propose a model for incorporating the quality of service (QoS) requirements of packets with deadlines in the optimization framework. The solution to the problem results in a joint congestion control and scheduling algorithm which fairly allocates resources to meet the fairness objectives of both elastic and inelastic flows, and per-packet delay requirements of inelastic flows.

**Index Terms**—Wireless networks, ad hoc networks, quality of service, congestion control, scheduling, best-effort traffic, real-time traffic.

## I. INTRODUCTION

AS wireless networks become more prevalent, they will be expected to support a wide variety of services, including best-effort and real-time traffic. Such networks will have to serve flows that require quality of service requirements, such as minimum bandwidth and maximum delay constraints, while at the same time keeping the network queues stable for data traffic and guaranteeing throughput optimality. For the case of wireless networks with best-effort traffic only, optimization-based algorithms which naturally map into different layers of the protocol stack have been proposed in the last few years [2]–[7]; see [8] for a survey. However, these models do not take into account strict per-packet delay bounds.

Scheduling packets with strict deadlines has been studied in [9]–[12], but all of these papers provide approximate solutions. The model that we study in this paper builds upon the recent work in [13]–[15] on admission control and scheduling for inelastic flows in collocated wireless networks, i.e., networks where all links interfere with each other. Among the many contributions in these papers is a key modeling innovation whereby the network is studied in frames, where a frame is

a contiguous set of time-slots of fixed duration. Packets with deadlines are assumed to arrive at the beginning of a frame and have to be served by the end of the frame. In this paper, we explore this modeling paradigm further to study the design of resource allocation algorithms for ad hoc networks. The frame-based model allows us to incorporate delay deadlines in the optimization framework for very general network models, and somewhat surprisingly, allows us to design a common framework for handling both elastic and inelastic flows.

The main contributions of the paper are as follows:

- 1) We present an optimization framework for resource allocation in a wireless network consisting of both best-effort flows and flows that generate traffic with per-packet delay constraints. The framework allows for very general interference, channel and arrival models.
- 2) We characterize the capacity region for wireless networks when there is inelastic traffic. Specifically, we show that the capacity region not only depends on the statistics of the channel as is the case when there is only elastic traffic, but it also depends on the probability distribution of the inelastic arrivals.
- 3) Using a dual decomposition approach, we derive an optimal scheduling and congestion control algorithm that fairly allocates resources and ensures that a required fraction of each inelastic flow’s packets are delivered on time by appealing to connections between Lagrange multipliers, queues, and service deficits. The scheduling algorithm seamlessly integrates inelastic and elastic traffic into a unified max-weight scheduling framework, extending the well-known results in [16].
- 4) The convergence of the above algorithm in an appropriate stochastic sense is proved and it is also shown that the network is stable.

The rest of the paper is organized as follows. In Section II we present the network model, which is used in Section III to present a static optimization formulation of our problem. Section IV presents a dual decomposition of the optimization that give us the intuition to formulate our dynamic solution and prove its convergence results in Section V. Section VI summarizes the differences in formulation and solution for an alternate model of the wireless channel. Section VII presents a heuristic for distributed admission control, and Section VIII presents a simulation-based study of our algorithm. Finally, Section IX presents the conclusions and future work.

## II. NETWORK MODEL

The network is represented by a directed graph  $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ , where  $\mathcal{N}$  is the set of nodes and  $\mathcal{L}$  is the set of directional

Research supported by NSF Grants 07-21286, 05-19691, 03-25673, ARO MURI Subcontracts, AFOSR Grant FA-9550-08-1-0432 and DTRA Grant HDTRA1-08-1-0016.

This work was completed during J. J. Jaramillo’s graduate studies at the University of Illinois at Urbana-Champaign. He is currently with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA. Email: jjjarami@iastate.edu

R. Srikant is with the Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801 USA. Email: rsrikant@illinois.edu

This paper is a revised version of an earlier paper that appeared in [1]. The main differences between this version and the previous one are that the proofs have been rewritten to make the results more general and simulations have been extensively expanded.

links such that for all  $n_1, n_2 \in \mathcal{N}$  if  $(n_1, n_2) \in \mathcal{L}$  then node  $n_1$  can transmit to node  $n_2$ . The links are numbered 1 through  $|\mathcal{L}|$ , and by abusing notation, we sometimes use  $l \in \mathcal{L}$  to mean  $l \in \{1, 2, \dots, |\mathcal{L}|\}$ .

Traffic is assumed to be a mixture of elastic and inelastic flows, where an inelastic flow is one that has maximum per-packet delay requirements. In contrast, elastic flows do not have such requirements.

Time is divided in slots, where a set of  $T$  consecutive time slots makes a *frame*. We assume that packet arrivals only occur at the beginning of a frame, and every inelastic packet has a deadline of  $T$  time slots. If a packet misses its deadline it is discarded, and it is required that the loss probability at link  $l \in \mathcal{L}$  due to deadline expiry must be no more than  $p_l$ . For elastic traffic we associate a utility function  $U_l(x_{el})$  which is a function of the mean elastic arrival rate per frame  $x_{el}$ . We assume that  $U_l(\cdot)$  is a concave function.

For a given frame, we denote by the vector  $a_i = (a_{il})_{l \in \mathcal{L}}$  the number of inelastic packet arrivals at every link, where  $a_{il}$  is a random variable with mean  $x_{il}$  and variance  $\sigma_{il}^2$ . We further assume that arrivals are independent between different frames and that  $Pr(a_{il} = 0) > 0$  and  $Pr(a_{il} = 1) > 0$ . The last two assumptions are used to guarantee that the Markov chain we define later is both irreducible and aperiodic, although these can be replaced by other similar assumptions. Similarly, we define  $a_e = (a_{el})_{l \in \mathcal{L}}$  to be the number of elastic packet arrivals at every link in a given frame.

The channel state is assumed to be constant in a given frame, independent between different frames, and independent of arrivals. The vector  $c = (c_l)_{l \in \mathcal{L}}$  denotes the number of packets link  $l$  can successfully transmit on a time slot in a given frame.

Depending on the wireless technology used, we can have some channel feedback before or after a transmission occurs. If channel estimation is performed before transmitting, we can determine the optimal rate at which we can successfully transmit. Alternatively, feedback from the receiver after the transmission can be used to detect if a transmission is successful or not. In this paper we try to capture both scenarios in the following cases:

- 1) Known channel state: It is assumed that  $c_l$  is a *non-negative* random variable with mean  $\bar{c}_l$  and variance  $\sigma_{c_l}^2$ , and we get to know the channel state at the *beginning* of the frame.
- 2) Unknown channel state: It is assumed that  $c_l$  is a *Bernoulli* random variable with mean  $\bar{c}_l$  and we only get to know the channel state at the *end* of the frame.

In the known channel state case where we do channel estimation to determine the optimal transmission rate, we can potentially send more than one packet in a time slot at higher rates. This is captured by the fact that we make no assumptions on the values  $c_l$  can take since it will be determined by the particular wireless technology used. In the case of unknown channel state we assume that we only get the binary feedback of acknowledgments, which is reflected in the Bernoulli assumption on  $c_l$ . In this case, and without any loss of generality, we assume only one packet can be transmitted per time slot per link.

The reader is referred to [17] for an alternative method to define more general channel and interference models, where the channel is modeled by a reliability function that captures the probability of successful transmission as a function of the channel state and the physical layer transmissions parameters. It must be noted though that having more general models does not change the nature of the solution to our problem, so for the sake of simplicity we will use a channel and interference model that allow us to present the main ideas behind our algorithm without the need for more complex notation.

In the rest of the paper we will consider the known channel case first and then in Section VI we will highlight the differences in the analysis for the unknown channel case.

### III. PROBLEM FORMULATION

We first formulate the problem as a static optimization problem. Using decomposition theory, we will then obtain a dynamic solution to this problem and prove its stability using stochastic Lyapunov techniques.

A feasible schedule  $s = (s_{il,t}, s_{el,t})$  is such that  $s_{il,t}, s_{el,t}$  respectively denote the number of inelastic and elastic packets that can be scheduled for transmission at link  $l \in \mathcal{L}$  and time slot  $t \in \{1, 2, \dots, T\}$ ; thus,  $s_{il,t} + s_{el,t} > 0$  means that link  $l$  is scheduled to transmit in time slot  $t$  of the frame. Furthermore, for any  $t$ , if  $s_{i_1,t} + s_{e_1,t} > 0$  and  $s_{i_2,t} + s_{e_2,t} > 0$  then links  $l_1$  and  $l_2$  can be scheduled to simultaneously transmit without interfering with each other. Assuming the inelastic arrivals and the channel state are given by  $a_i$  and  $c$  respectively, we have the following constraints:

$$\sum_{t=1}^T s_{il,t} \leq a_{il} \text{ for all } l \in \mathcal{L} \text{ and} \quad (1)$$

$$s_{il,t} + s_{el,t} \leq c_l \text{ for all } l \in \mathcal{L} \text{ and } t \in \{1, 2, \dots, T\}. \quad (2)$$

Since an inelastic packet that misses its deadline is discarded, (1) states that the number of successful inelastic transmissions within a frame is bounded by the number of packet arrivals. In other words, there is no concept of queueing inelastic packets after deadline expiration. Similarly, (2) states that the number of successful transmissions in any slot cannot exceed the state of the channel. We denote by  $\mathcal{S}(a_i, c)$  the set of all feasible schedules when the arrival state is  $a_i$  and the channel state is  $c$ ; thus,  $\mathcal{S}(a_i, c)$  captures any interference constraints we have on our network and satisfies (1) and (2).

As an example of a feasible schedule, consider Figure 1. We consider a network with three links such that only one link can be scheduled at any given time, and assume that a frame is composed of five time slots. Furthermore, we assume that the channel is always ON and only one packet can be transmitted at any given time, so  $c = (1, 1, 1)$ , and that the inelastic arrivals are given by the vector  $a_i = (1, 0, 1)$ . Since only one link can be scheduled at any given time, we know that a feasible schedule must satisfy  $\sum_{l=1}^3 s_{il,t} + s_{el,t} \leq 1$  for all time slots in a frame. Thus, it is easy to see that our schedule meets the interference constraints and (2). In the case of link 2, we have that  $\sum_{t=1}^5 s_{i_2,t} = 0 \leq a_{i_2} = 0$ , meeting constraint (1). On the other hand, since elastic packets can be queued during multiple

Fig. 1. An example of a feasible schedule when  $a_i = (1, 0, 1)$  and  $c = (1, 1, 1)$

		Time Slot				
		1	2	3	4	5
Inelastic	Link 1	1	0	0	0	0
	Link 2	0	0	0	0	0
	Link 3	0	1	0	0	0
Elastic	Link 1	0	0	1	0	0
	Link 2	0	0	0	1	1
	Link 3	0	0	0	0	0

frames, we can schedule their transmission independent of the current queue length. Therefore  $\sum_{t=1}^5 s_{e2,t} = 2$  is a valid schedule. A similar analysis can be done to verify that (1) is also satisfied for links 1 and 3.

At the beginning of any frame we must choose a feasible schedule to serve all links and decide how many elastic packets are allowed to be injected in the network. Therefore, our goal is to find a function  $Pr(s|a_i, c)$  which is the probability of using schedule  $s \in \mathcal{S}(a_i, c)$  when the inelastic arrivals are given by  $a_i$  and the channel state is  $c$ , subject to the constraint that the loss probability at link  $l \in \mathcal{L}$  due to deadline expiry cannot exceed  $p_l$ . For elastic traffic, we want to select the vector  $a_e$  such that we maximize the network utility while keeping the queues stable.

To properly formulate the problem, let us first define  $\mu_{il}(a_i, c)$  to be the expected number of inelastic packets served if the number of packet arrivals is given by  $a_i$  and the channel state is  $c$ . Similarly,  $\mu_{el}(a_i, c)$  denotes the expected number of elastic packets that can be served. Therefore, we have the following constraints:

$$\mu_{il}(a_i, c) \leq \sum_{s \in \mathcal{S}(a_i, c)} \sum_{t=1}^T s_{il,t} Pr(s|a_i, c)$$

$$\mu_{el}(a_i, c) \leq \sum_{s \in \mathcal{S}(a_i, c)} \sum_{t=1}^T s_{el,t} Pr(s|a_i, c).$$

The expected service for mixed traffic at link  $l$  is then given by

$$\mu_{il} \stackrel{def}{=} \sum_{a_i} \sum_c \mu_{il}(a_i, c) Pr(c) Pr(a_i)$$

$$\mu_{el} \stackrel{def}{=} \sum_{a_i} \sum_c \mu_{el}(a_i, c) Pr(c) Pr(a_i)$$

and due to QoS requirements and capacity constraints, we require that

$$\mu_{il} \geq x_{il}(1 - p_l) \text{ and } x_{el} \leq \mu_{el}.$$

Furthermore, to avoid trivialities, we will assume that  $x_{il}(1 - p_l) > 0$  for all  $l \in \mathcal{L}$ .

If we define the capacity region for fixed arrival and channel states  $a_i$  and  $c$  as follows:

$$\mathcal{C}(a_i, c) \stackrel{def}{=} \left\{ (\bar{\mu}_{il}, \bar{\mu}_{el})_{l \in \mathcal{L}} : \text{there exists } \bar{s} \in \mathcal{S}(a_i, c) \mathcal{CH}, \right. \\ \left. \bar{\mu}_{il} \leq \sum_{t=1}^T \bar{s}_{il,t} \text{ and } \bar{\mu}_{el} \leq \sum_{t=1}^T \bar{s}_{el,t} \right\},$$

it is easy to see that  $\mu(a_i, c) \stackrel{def}{=} (\mu_{il}(a_i, c), \mu_{el}(a_i, c))_{l \in \mathcal{L}} \in \mathcal{C}(a_i, c)$ , where  $\mathcal{S}(a_i, c) \mathcal{CH}$  is the convex hull of  $\mathcal{S}(a_i, c)$ .

Similarly, if the overall capacity of the network is defined as  $\mathcal{C} \stackrel{def}{=}$

$$\left\{ \begin{array}{l} (\mu_{il}, \mu_{el})_{l \in \mathcal{L}} : \text{there exists } (\bar{\mu}_{il}(a_i, c), \bar{\mu}_{el}(a_i, c))_{l \in \mathcal{L}} \in \\ \mathcal{C}(a_i, c) \text{ for all } a_i, c \text{ and } \mu_{il} = E[\bar{\mu}_{il}(a_i, c)], \\ \mu_{el} = E[\bar{\mu}_{el}(a_i, c)] \text{ for all } l \in \mathcal{L} \end{array} \right\},$$

we have that  $\mu \stackrel{def}{=} (\mu_{il}, \mu_{el})_{l \in \mathcal{L}} \in \mathcal{C}$ .

We will focus on maximizing the following objective for some given vector  $w \in \mathbb{R}_+^{|\mathcal{L}|}$ :

$$\max_{\mu \in \mathcal{C}, x_e} \sum_{l \in \mathcal{L}} U_l(x_{el}) + w_l \mu_{il} \quad (3)$$

subject to

$$\mu_{il} \geq x_{il}(1 - p_l) \text{ for all } l \in \mathcal{L}$$

$$0 \leq x_{el} \leq \mu_{el} \text{ for all } l \in \mathcal{L}.$$

The vector  $w$  can be used to allocate additional bandwidth fairly to inelastic flows beyond what is required to meet their QoS needs. Other uses for  $w$  will be explored in the simulations section. We will assume that the inelastic arrivals and loss probability requirements are feasible and thus the optimization problem has a solution  $(\mu^*, x_e^*)$ .

#### IV. SOLUTION USING DUAL DECOMPOSITION

Using the definition of the dual function [18], we have that  $D(\delta_i, \delta_e) =$

$$\max_{\mu \in \mathcal{C}, x_e} \sum_{l \in \mathcal{L}} \left\{ \begin{array}{l} U_l(x_{el}) + w_l \mu_{il} - \delta_{el}[x_{el} - \mu_{el}] \\ - \delta_{il}[x_{il}(1 - p_l) - \mu_{il}] \end{array} \right\}$$

subject to

$$x_{el} \geq 0 \text{ for all } l \in \mathcal{L}.$$

Slater's condition [19] states that, since the objective is concave and the constraints are affine functions, the duality gap is zero and therefore  $D(\delta_i^*, \delta_e^*) = \sum_{l \in \mathcal{L}} U_l(x_{el}^*) + w_l \mu_{il}^*$ , where

$$(\delta_i^*, \delta_e^*) \in \arg \min_{\delta_{il} \geq 0, \delta_{el} \geq 0} D(\delta_i, \delta_e).$$

We are interested in finding  $(\mu^*, x_e^*)$  but not the value  $D(\delta_i^*, \delta_e^*)$ , so if we rewrite the objective in the dual function as

$$\max_{\mu \in \mathcal{C}, x_e} \left\{ \begin{array}{l} \sum_{l \in \mathcal{L}} U_l(x_{el}) - \delta_{el} x_{el} \\ + \sum_{l \in \mathcal{L}} (w_l + \delta_{il}) \mu_{il} + \delta_{el} \mu_{el} \\ - \sum_{l \in \mathcal{L}} \delta_{il} x_{il}(1 - p_l) \end{array} \right\}$$

we notice that the problem can be decomposed into the following subproblems:

$$\max_{x_{el} \geq 0} U_l(x_{el}) - \delta_{el} x_{el}$$

and

$$\max_{\mu \in \mathcal{C}} \sum_{l \in \mathcal{L}} (w_l + \delta_{il}) \mu_{il} + \delta_{el} \mu_{el}. \quad (4)$$

Furthermore, since we are interested in solving the problem for non-negative values of  $\delta_{il}$  and  $\delta_{el}$ , it must be the case

that  $\mu_i^*$  and  $\mu_c^*$  are as large as the constraints allow, and since the upper bounds for  $\mu_{il}^*(a_i, c)$  and  $\mu_{el}^*(a_i, c)$  are expressed as a convex combination, and the objective function in (4) is linear, the problem can be decomposed into the following subproblems for fixed  $a_i$  and  $c$ :

$$\max_{s \in \mathcal{S}(a_i, c)} \sum_{l \in \mathcal{L}} \left\{ (w_l + \delta_{il}) \sum_{t=1}^T s_{il,t} + \delta_{el} \sum_{t=1}^T s_{el,t} \right\}.$$

This suggests the following iterative algorithm to find the solution to our optimization problem, where  $k$  is the step index and  $X_{max} > 0$  is a large enough parameter:

$$\tilde{x}_{el}^*(k) \in \arg \max_{0 \leq x_{el} \leq X_{max}} U_l(x_{el}) - \delta_{el}(k)x_{el}$$

$$\tilde{s}^*(a_i, c, k) \in \arg \max_{s \in \mathcal{S}(a_i, c)} \sum_{l \in \mathcal{L}} \left\{ [w_l + \delta_{il}(k)] \sum_{t=1}^T s_{il,t} + \delta_{el}(k) \sum_{t=1}^T s_{el,t} \right\}$$

$$\tilde{\mu}_{il}^*(k) = \sum_{a_i} \sum_c \sum_{t=1}^T \tilde{s}_{il,t}^*(a_i, c, k) Pr(c) Pr(a_i)$$

$$\tilde{\mu}_{el}^*(k) = \sum_{a_i} \sum_c \sum_{t=1}^T \tilde{s}_{el,t}^*(a_i, c, k) Pr(c) Pr(a_i).$$

We update the Lagrange multipliers  $\delta_i(k)$ ,  $\delta_e(k)$  at every step according to the following equations:

$$\delta_{il}(k+1) = \{\delta_{il}(k) + \epsilon[x_{il}(1-p_l) - \tilde{\mu}_{il}^*(k)]\}^+$$

and

$$\delta_{el}(k+1) = \{\delta_{el}(k) + \epsilon[\tilde{x}_{el}^*(k) - \tilde{\mu}_{el}^*(k)]\}^+$$

where  $\epsilon > 0$  is a fixed step-size parameter, and for any  $\alpha \in \mathbb{R}$ ,  $\alpha^+ \stackrel{\text{def}}{=} \max\{\alpha, 0\}$ .

Making the change of variables  $\hat{d}_l(k) = \delta_i(k)$  and  $\hat{e}_l(k) = \delta_e(k)$ , we have that our iterative algorithm can be rewritten as

$$\tilde{x}_{el}^*(k) \in \arg \max_{0 \leq x_{el} \leq X_{max}} \frac{1}{\epsilon} U_l(x_{el}) - \hat{q}_l(k)x_{el}$$

$$\tilde{s}^*(a_i, c, k) \in \arg \max_{s \in \mathcal{S}(a_i, c)} \sum_{l \in \mathcal{L}} \left\{ \left[ \frac{1}{\epsilon} w_l + \hat{d}_l(k) \right] \sum_{t=1}^T s_{il,t} + \hat{q}_l(k) \sum_{t=1}^T s_{el,t} \right\}$$

$$\tilde{\mu}_{il}^*(k) = \sum_{a_i} \sum_c \sum_{t=1}^T \tilde{s}_{il,t}^*(a_i, c, k) Pr(c) Pr(a_i)$$

$$\tilde{\mu}_{el}^*(k) = \sum_{a_i} \sum_c \sum_{t=1}^T \tilde{s}_{el,t}^*(a_i, c, k) Pr(c) Pr(a_i).$$

with update equations

$$\hat{d}_l(k+1) = [\hat{d}_l(k) + x_{il}(1-p_l) - \tilde{\mu}_{il}^*(k)]^+$$

$$\hat{q}_l(k+1) = [\hat{q}_l(k) + \tilde{x}_{el}^*(k) - \tilde{\mu}_{el}^*(k)]^+.$$

It should be noted that due to the change of variables  $\hat{d}_l(k)$  can be interpreted as a queue that has  $x_{il}(1-p_l)$  arrivals and  $\tilde{\mu}_{il}^*(k)$  departures at step  $k$ ;  $\hat{q}_l(k)$  can have a similar queue

interpretation. The dual decomposition approach only provides an intuition behind the solution, but the real network has stochastic and dynamic arrivals and channel state conditions. In the next section, we present the complete solution which takes into account these dynamics and we also establish its convergence properties.

## V. DYNAMIC ALGORITHM AND ITS CONVERGENCE ANALYSIS

### A. Scheduler and Congestion Controller

To implement the algorithm online, we propose the following congestion control algorithm in frame  $k$ , where the queue length at link  $l$  is given by  $q_l(k)$ :

$$\tilde{x}_{el}^*(k) \in \arg \max_{0 \leq x_{el} \leq X_{max}} \frac{1}{\epsilon} U_l(x_{el}) - q_l(k)x_{el}. \quad (5)$$

We need to convert this elastic arrival rate, which in general is a non-negative real number, into a non-negative integer indicating the number of elastic packets allowed to enter the network in a given frame. This conversion can be made in many different ways: we assume the elastic arrivals at link  $l$ ,  $\tilde{a}_{el}(k)$ , are a random variable with mean  $\tilde{x}_{el}^*(k)$  and variance upper-bounded by  $\sigma_e^2$ , and are such that  $Pr(\tilde{a}_{el}(k) = 0) > 0$  and  $Pr(\tilde{a}_{el}(k) = 1) > 0$  for all  $l \in \mathcal{L}$  and all  $k$ . The last two assumptions are used to guarantee the Markov chain we define below is both irreducible and aperiodic, although these can be replaced by other similar assumptions.

Letting the number of inelastic arrivals be denoted by  $a_i(k)$  and the channel state by  $c(k)$ , we propose the following scheduling algorithm:

$$\tilde{s}^*(a_i(k), c(k), d(k), q(k)) \in \quad (6)$$

$$\arg \max_{s \in \mathcal{S}(a_i(k), c(k))} \sum_{l \in \mathcal{L}} \left\{ \left[ \frac{1}{\epsilon} w_l + d_l(k) \right] \sum_{t=1}^T s_{il,t} + q_l(k) \sum_{t=1}^T s_{el,t} \right\}.$$

The vectors  $d(k)$  and  $q(k)$  are updated from frame to frame as follows:

$$d_l(k+1) = [d_l(k) + \tilde{a}_{il}(k) - I_{il}^*(a_i(k), c(k), d(k), q(k))]^+$$

$$q_l(k+1) = [q_l(k) + \tilde{a}_{el}(k) - I_{el}^*(a_i(k), c(k), d(k), q(k))]^+,$$

where

$$I_{il}^*(a_i(k), c(k), d(k), q(k)) = \sum_{t=1}^T \tilde{s}_{il,t}^*(a_i(k), c(k), d(k), q(k))$$

$$I_{el}^*(a_i(k), c(k), d(k), q(k)) = \sum_{t=1}^T \tilde{s}_{el,t}^*(a_i(k), c(k), d(k), q(k))$$

and  $\tilde{a}_{il}(k)$  is a binomial random variable with parameters  $a_{il}(k)$  and  $1-p_l$ . The quantity  $\tilde{a}_{il}(k)$  can be generated by the network as follows: upon each inelastic packet arrival, toss a coin with probability of *heads* equal to  $1-p_l$ , and if the outcome is *heads*, add a one to the deficit counter.

In our notation we make explicit the fact that for fixed  $\epsilon$  and  $w$ , the optimal scheduler (6) is a function of  $a_i(k)$ ,  $c(k)$ ,  $d(k)$ , and  $q(k)$ . We interpret  $d_l(k)$  as a virtual queue that counts the deficit in service for link  $l$  to achieve a loss probability due to

deadline expiry less than or equal to  $p_l$ . This deficit queue was first used in the inelastic traffic context in [13] for the case of collocated networks; the connection to the dual decomposition approach now provides a Lagrange multiplier interpretation to it and allows the extension to general ad hoc networks. Note that  $q_l(k)$  is just the queue size for elastic packets at link  $l$ .

### B. Convergence Results

For readability, we present the main results in this section, but the proofs are deferred to the appendixes. We start by noting that  $(d(k), q(k))$  defines an irreducible and aperiodic Markov chain. We will first bound the expected drift of  $(d(k), q(k))$  for a suitable Lyapunov function.

*Lemma 1:* Consider the Lyapunov function  $V(d, q) = \frac{1}{2} \sum_{l \in \mathcal{L}} d_l^2 + q_l^2$ . If there exists a point  $\mu(\Delta) \in \mathcal{C}/(1 + \Delta)$  for some  $\Delta > 0$  such that

$$\mu_{il}(\Delta) \geq x_{il}(1 - p_l) \text{ for all } l \in \mathcal{L} \quad (7)$$

and

$$\mu_{el}(\Delta) > 0 \text{ for all } l \in \mathcal{L}, \quad (8)$$

then

$$\begin{aligned} E[V(d(k+1), q(k+1)) | d(k) = d, q(k) = q] - V(d, q) \\ \leq B_1 - B_2 \sum_{l \in \mathcal{L}} d_l - B_3 \sum_{l \in \mathcal{L}} q_l \\ - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \\ - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} \{w_l(1 + \Delta)\mu_{il}(\Delta) - w_l E[I_{il}^*(a_i(k), c(k), d, q)]\} \end{aligned}$$

for some positive constants  $B_1, B_2, B_3$ , any  $\epsilon > 0$ , where  $\tilde{x}_e^*(k)$  is the solution to (5), and  $I_i^*(a_i(k), c(k), d, q)$  is obtained from the solution to (6).  $\diamond$

It is important to note that since the last two terms in the right-hand side of the inequality can be upper-bounded, Lemma 1 implies that  $(d(k), q(k))$  is positive recurrent since the expected drift is negative but for a finite set of values of  $(d(k), q(k))$ . As a direct consequence of this fact, we note that the total service deficit and queue length have a  $O(1/\epsilon)$  bound.

*Corollary 1:* If there exists a point  $\mu(\Delta) \in \mathcal{C}/(1 + \Delta)$  for some  $\Delta > 0$  such that (7) and (8) hold true, then the total expected service deficit and network queue length is upper-bounded by

$$\limsup_{k \rightarrow \infty} E \left[ \sum_{l \in \mathcal{L}} d_l(k) + q_l(k) \right] \leq B_4 + \frac{1}{\epsilon} B_5,$$

where

$$B_4 = \frac{B_1}{\min\{B_2, B_3\}}$$

and

$$B_5 \leq \frac{\sum_{l \in \mathcal{L}} \max_{0 \leq x_{el} \leq X_{max}} 2|U_l(x_{el})| + w_l x_{il}}{\min\{B_2, B_3\}}.$$

$\diamond$

Lemma 1 also implies that the scheduling and congestion control algorithm fulfills all links' inelastic requirements.

*Corollary 2:* If there exists a point  $\mu(\Delta) \in \mathcal{C}/(1 + \Delta)$  for some  $\Delta > 0$  such that (7) and (8) hold true, then the online algorithm fulfills all the inelastic constraints. That is:

$$\liminf_{K \rightarrow \infty} E \left[ \frac{1}{K} \sum_{k=1}^K I_{il}^*(a_i(k), c(k), d(k), q(k)) \right] \geq x_{il}(1 - p_l)$$

for all  $l \in \mathcal{L}$ .  $\diamond$

The above corollary simply states that the arrival rate into the deficit counter is less than or equal to the departure rate. This result is an obvious consequence of the stability of the deficit counters and so a formal proof is not provided here.

To prove that our dynamic algorithm achieves the optimal solution to the static problem (3) in some average sense and fulfills all links' requirements, we first state a weaker result than Lemma 1:

*Lemma 2:* Consider the Lyapunov function  $V(d, q) = \frac{1}{2} \sum_{l \in \mathcal{L}} d_l^2 + q_l^2$ . Then

$$\begin{aligned} E[V(d(k+1), q(k+1)) | d(k) = d, q(k) = q] - V(d, q) \\ \leq B_1 - B_6 \sum_{l \in \mathcal{L}} d_l - B_7 \sum_{l \in \mathcal{L}} q_l - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) - U_l(\tilde{x}_{el}^*(k))] \\ - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} \{w_l \mu_{il}^* - w_l E[I_{il}^*(a_i(k), c(k), d, q)]\} \end{aligned}$$

for some non-negative constants  $B_1 > 0, B_6, B_7$ , any  $\epsilon > 0$ , where  $(\mu^*, x_e^*)$  is the solution to (3),  $\tilde{x}_e^*(k)$  is the solution to (5), and  $I_i^*(a_i(k), c(k), d, q)$  is obtained from the solution to (6).  $\diamond$

The proof follows the same lines as in Lemma 1 and is thus omitted.

Now we are ready to prove that our online algorithm is within  $O(\epsilon)$  of the optimal value.

*Theorem 1:* For any  $\epsilon > 0$ , and assuming that  $U_l(\cdot)$  is a concave function, we have that

$$\begin{aligned} \limsup_{K \rightarrow \infty} E \left[ \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) + w_l \mu_{il}^*] - \sum_{l \in \mathcal{L}} U_l \left( \frac{1}{K} \sum_{k=1}^K \tilde{x}_{el}^*(k) \right) \right. \\ \left. - \sum_{l \in \mathcal{L}} w_l \frac{1}{K} \sum_{k=1}^K I_{il}^*(a_i(k), c(k), d(k), q(k)) \right] \leq B\epsilon \end{aligned}$$

for some  $B > 0$ , where  $(\mu^*, x_e^*)$  is the solution to (3),  $\tilde{x}^*(k)$  is the solution to (5), and  $I_i^*(a_i(k), c(k), d(k), q(k))$  is obtained from the solution to (6).  $\diamond$

It is interesting to note that the assumption that  $U_l(\cdot)$  is a concave function is only used to prove that our algorithm is close to the optimal value, but it is not necessary to prove any of the other results in this section.

In conclusion, there is a trade-off in choosing the parameter  $\epsilon$ : smaller values will achieve a solution closer to the optimal, but at the same time the deficit in service at the links and the aggregate queue length increase. The statement and the proofs of Lemma 1 and Theorem 1 follow techniques similar to those in [4]. The results can also be derived, in a slightly different form, using the techniques in [5]. A closely related result can be obtained using the methods in [2].  $\diamond$

## VI. UNKNOWN CHANNEL STATE

The analysis for the unknown channel case is similar to the one we presented for the known channel case, so in this section we will only highlight the differences.

A feasible schedule  $s = (s_{il,t}, s_{el,t})$  is such that  $s_{il,t}, s_{el,t}$  respectively denote the number of inelastic and elastic packets that can be scheduled for transmission at link  $l \in \mathcal{L}$  and time  $t \in \{1, 2, \dots, T\}$  without violating any interference constraints. Assuming the inelastic arrivals are given by  $a_i$ , and since we can only schedule at most one packet per link at every time slot, we have the following constraints:

$$\sum_{t=1}^T s_{il,t} \leq a_{il} \text{ for all } l \in \mathcal{L} \text{ and } \quad (9)$$

$$s_{il,t} + s_{el,t} \leq 1 \text{ for all } l \in \mathcal{L} \text{ and } t \in \{1, 2, \dots, T\}. \quad (10)$$

We denote by  $\mathcal{S}(a_i)$  the set of all feasible schedules for fixed arrivals, capturing any interference constraints we have on our network, and satisfying (9) and (10).

Our goal now is to find a function  $Pr(s|a_i)$  which is the probability of using schedule  $s \in \mathcal{S}(a_i)$  when the inelastic arrivals are given by  $a_i$ , subject to the constraint that the loss probability at link  $l \in \mathcal{L}$  due to deadline expiry cannot exceed  $p_l$ . For elastic traffic, we still want to select the vector  $a_e$  such that we maximize the total utility while keeping the queues stable.

For a given distribution  $Pr(s|a_i)$  we have that  $\mu_{il}(a_i)$  is the expected number of attempted inelastic transmissions if arrivals are given by  $a_i$ . Similarly,  $\mu_{el}(a_i)$  denotes the expected number of times link  $l$  is scheduled to serve elastic packets in a given frame. As before, we have the following constraints:

$$\mu_{il}(a_i) \leq \sum_s \sum_{t=1}^T s_{il,t} Pr(s|a_i)$$

$$\mu_{el}(a_i) \leq \sum_s \sum_{t=1}^T s_{el,t} Pr(s|a_i)$$

When the (unknown) channel state is  $c$ , we have that  $c_l \mu_{il}(a_i)$  is the expected number of successful inelastic transmissions per frame at link  $l$  for fixed arrivals, while  $c_l \mu_{el}(a_i)$  is the expected service to link  $l$  for inelastic arrivals. Thus, the expected service for mixed traffic at link  $l$  is given by

$$\mu_{il} \stackrel{def}{=} \sum_{a_i} \sum_c c_l \mu_{il}(a_i) Pr(c) Pr(a_i)$$

$$\mu_{el} \stackrel{def}{=} \sum_{a_i} \sum_c c_l \mu_{el}(a_i) Pr(c) Pr(a_i).$$

Simplifying both expressions we get

$$\mu_{il} = \sum_{a_i} \bar{c}_l \mu_{il}(a_i) Pr(a_i)$$

$$\mu_{el} = \sum_{a_i} \bar{c}_l \mu_{el}(a_i) Pr(a_i).$$

Due to service requirements and capacity constraints we need that

$$\mu_{il} \geq x_{il}(1 - p_l) \text{ and } x_{el} \leq \mu_{el}.$$

To avoid trivialities, we will assume that  $x_{il}(1 - p_l) > 0$  for all  $l \in \mathcal{L}$ .

If we define the capacity region for fixed arrival and channel states  $a_i$  and  $c$  as follows:

$$\mathcal{C}(a_i, c) \stackrel{def}{=} \left\{ \begin{array}{l} (\bar{\mu}_{il}, \bar{\mu}_{el})_{l \in \mathcal{L}} : \text{there exists } \bar{s} \in \mathcal{S}(a_i)_{\mathcal{C}\mathcal{H}}, \\ \bar{\mu}_{il} \leq c_l \sum_{t=1}^T \bar{s}_{il,t} \text{ and } \bar{\mu}_{el} \leq c_l \sum_{t=1}^T \bar{s}_{el,t} \end{array} \right\},$$

it is easy to check that  $\mu(a_i, c) \stackrel{def}{=} (c_l \mu_{il}(a_i), c_l \mu_{el}(a_i))_{l \in \mathcal{L}} \in \mathcal{C}(a_i, c)$ , where  $\mathcal{S}(a_i, c)_{\mathcal{C}\mathcal{H}}$  is the convex hull of  $\mathcal{S}(a_i, c)$ .

Similarly, the overall capacity of the network can be defined as  $\mathcal{C} \stackrel{def}{=} \left\{ \begin{array}{l} (\mu_{il}, \mu_{el})_{l \in \mathcal{L}} : \text{there exists } (\bar{\mu}_{il}(a_i, c), \bar{\mu}_{el}(a_i, c))_{l \in \mathcal{L}} \in \\ \mathcal{C}(a_i, c) \text{ for all } a_i, c \text{ and } \mu_{il} = E[\bar{\mu}_{il}(a_i, c)], \\ \mu_{el} = E[\bar{\mu}_{el}(a_i, c)] \text{ for all } l \in \mathcal{L} \end{array} \right\},$

we have that  $\mu \stackrel{def}{=} (\mu_{il}, \mu_{el})_{l \in \mathcal{L}} \in \mathcal{C}$ .

With the definitions and constraints stated above we can formulate the optimization problem in a similar way as in (3) for some vector  $w \in \mathbb{R}_+^{|\mathcal{L}|}$ :

$$\max_{\mu \in \mathcal{C}, x_e} \sum_{l \in \mathcal{L}} U_l(x_{el}) + w_l \mu_{il}$$

subject to

$$\mu_{il} \geq x_{il}(1 - p_l) \text{ for all } l \in \mathcal{L}$$

$$0 \leq x_{el} \leq \mu_{el} \text{ for all } l \in \mathcal{L}.$$

Using the dual decomposition approach outlined in Section IV, we find that the only difference with the known channel state case is the scheduling algorithm. Assuming inelastic arrivals are given by  $a_i(k)$  the scheduling algorithm is given by

$$\tilde{s}^*(a_i(k), d(k), q(k)) \in$$

$$\arg \max_{s \in \mathcal{S}(a_i(k))} \sum_{l \in \mathcal{L}} \left\{ \left[ \frac{1}{\epsilon} w_l + d_l(k) \right] \bar{c}_l \sum_{t=1}^T s_{il,t} + q_l(k) \bar{c}_l \sum_{t=1}^T s_{el,t} \right\}.$$

The main difference in the scheduling algorithm compared to the known channel state case is that the network now uses the expected channel state in making scheduling decisions. Thus, the network needs to know or estimate  $\bar{c}_l$  as in [13].

Similar results can be proved for this algorithm using the techniques developed in Section V-B, whereby one can show that the algorithm meets all the inelastic QoS constraints, the total expected service deficits and the queue lengths have a  $O(1/\epsilon)$  bound, and the mean value of the objective is within  $O(\epsilon)$  of the optimal value.

## VII. DISTRIBUTED ADMISSION CONTROL

In this section we present a heuristic for distributed admission control based on the fact that the deficit counters  $d(k)$  are a good indicator of stability in the network.

So far, we have proved that if the inelastic arrivals and loss probability requirements are feasible, then our solution stabilizes the network and achieves an optimal allocation of resources without explicitly needing to characterize the capacity region. In the case that the inelastic arrivals and loss probability requirements are not necessarily feasible, we need

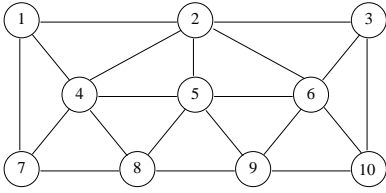


Fig. 2. Interference graph used in the simulations

to develop an admission control mechanism that guarantees that the set of admitted flows can be supported by the network. Since the congestion controller (5) guarantees that the rate of elastic packets injected in the network are always within the capacity region, we only need to decide on the admission of inelastic flows. A possible solution would be to characterize the capacity region analytically and use an admission controller which exploits this characterization to admit or reject new flows. However, a simple characterization of the capacity region appears to be difficult for ad hoc networks, and further any controller that uses such a characterization will be centralized in any case, and is thus not very useful.

In Corollary 1 we proved that if the requirements can be supported by the network using the service vector  $\mu(\Delta)$ , then the expected service deficit is upper bounded. Thus, if we see that the deficit counter of an inelastic flow does not stabilize over time, it means that such flow cannot be supported by the network. This observation can be exploited to perform admission control as follows. When a new flow arrives at a link, it can ask its neighbors for the values of their deficit queues. Admit the new flow only if the neighbors' deficit counters are all below a threshold  $\delta$ . The proper choice of such a  $\delta$  and whether one can prove desirable properties of such an admission controller are topics for further research.

### VIII. SIMULATIONS

The purpose of this simulation study is to understand how the parameter  $\epsilon$  and the link weights  $w_l$  impact the performance of the algorithm, how a greedy heuristic can be used to implement the optimal scheduler, and how channel information helps improve performance. We simulate a 10-link network with an interference graph given by Fig. 2, where each node represents a link and each edge means that the two adjacent links cannot be scheduled simultaneously. For example, if link 1 is scheduled, then links 2, 4, and 7 cannot be activated. The required loss probability due to deadline expiry of inelastic packets is set to 0.1, the link arrivals are assumed to have a Bernoulli distribution with mean 0.6 packets/frame, and there are 3 time slots per frame. The channel for every link is assumed to have a Bernoulli distribution with mean 0.96, and for the first part of the simulations we assume we get to know the channel state at the beginning of the frame. We set  $U_l(x_{el}) = \log(x_{el})$  for all links. The simulation time was  $10^6$  frames.

As can be noted from (6), the max-weight scheduler requires that we do an exhaustive search to find the optimal schedule at every frame. For large networks this can become a burden due to the large search space; thus we explore a greedy heuristic

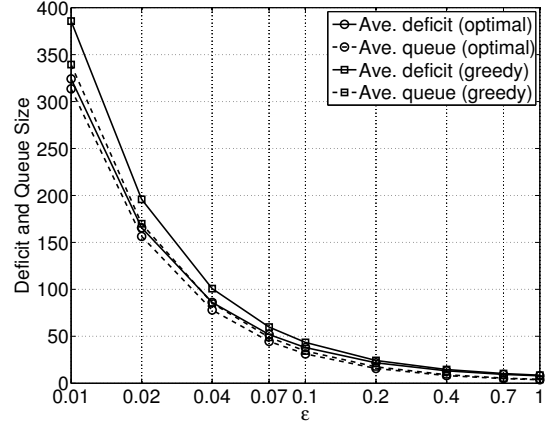


Fig. 3. Deficit size and queue length when  $w_l = 0$

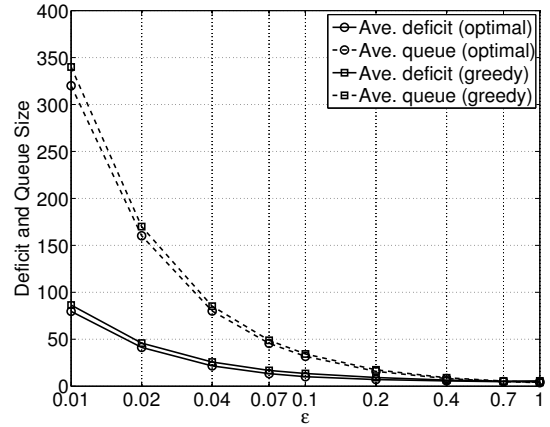
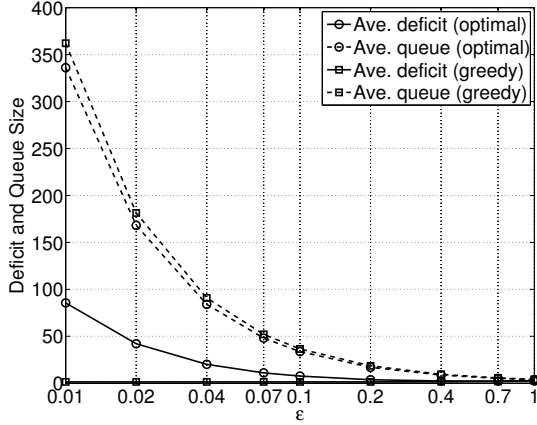
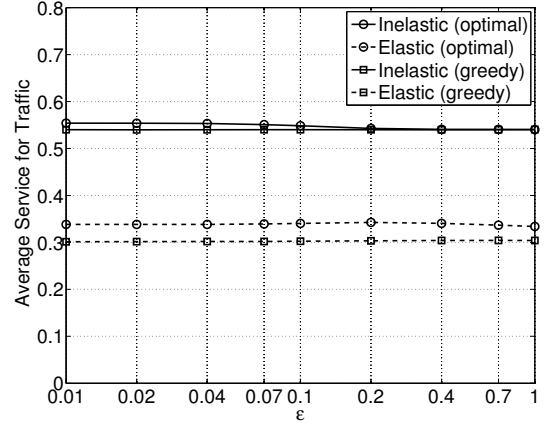
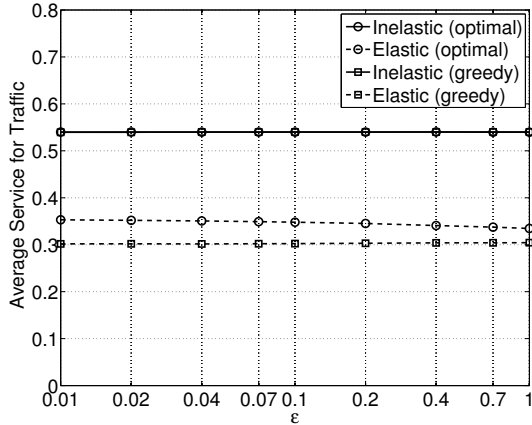
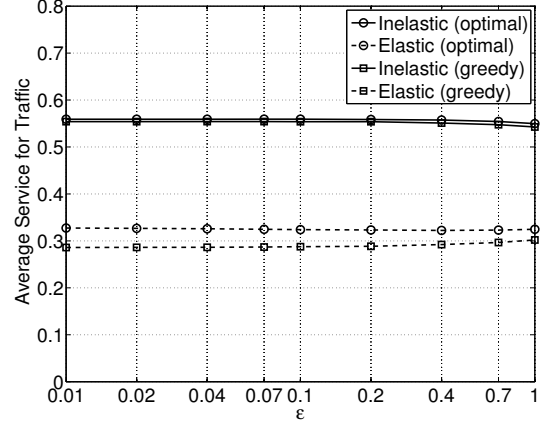


Fig. 4. Deficit size and queue length when  $w_l = 3$

and check how close it is to the optimal solution: at any given time slot, the greedy scheduler orders all links according to their weights. The greedy scheduler adds one of the links with the largest weight to the schedule, then removes all links that interfere with this link from the graph, then schedules a link with the largest weight among the remaining links, and so on. This procedure continues until no more links can be scheduled.

In Figs. 3, 4, and 5, we plot the expected values of the deficit counters and queues per link for various values of  $w_l$ , and compare their evolution for both the scheduler with optimal decisions and the greedy scheduler.

We see that as  $w_l$  increases, the deficit counters become small. The upper bound in Corollary 1 only suggests that the sum of the deficit counters and queues is  $O(1/\epsilon)$ . Thus, it is interesting to note that by changing  $w_l$ , one can nearly eliminate the backlog in deficit for inelastic traffic while maintaining the same order of queue sizes. The reason for this can be understood by examining the scheduling algorithm (6). Note that the algorithm gives priority to elastic traffic if queues are larger than counters. When  $w_l$  is small compared to  $\epsilon$ , the effect of  $w_l$  is negligible in the scheduling algorithm. On the other hand, when  $w_l$  is  $O(1)$ ,  $w_l/\epsilon$  is  $O(1/\epsilon)$  which

Fig. 5. Deficit size and queue length when  $w_l = 6$ Fig. 7. Average service when  $w_l = 3$ Fig. 6. Average service when  $w_l = 0$ Fig. 8. Average service when  $w_l = 6$ 

is comparable to the queue lengths and hence, the deficit does not have to be large to provide service to inelastic traffic under algorithm (6).

It must be noted that small deficit counters mean that there is a small backlog in providing acceptable service to inelastic arrivals. For the case of real-time traffic this is a desirable property, since we do not want to have large variations in the service provided that could affect the perceived quality. Thus, even if fair allocation of bandwidth beyond the minimum is not required for inelastic flows, choosing  $w_l$  an order of magnitude larger than  $\epsilon$  is desirable to maintain small deficits.

As can be noted, the greedy scheduler seems to give lower deficit values than the optimal scheduler for larger values of  $w_l$ . We believe that the reason is that weights given to inelastic flows increase with increasing  $w_l$  and therefore, the greedy scheduler picks them first. However, our optimality goal is given by (3) which is determined by the rates received by the various flows. The rates achieved by the two schedulers are quite close in the simulations, as seen in Figs. 6, 7, and 8, while keeping the dropping probabilities below the requirement, as shown in Figs. 9, 10, and 11.

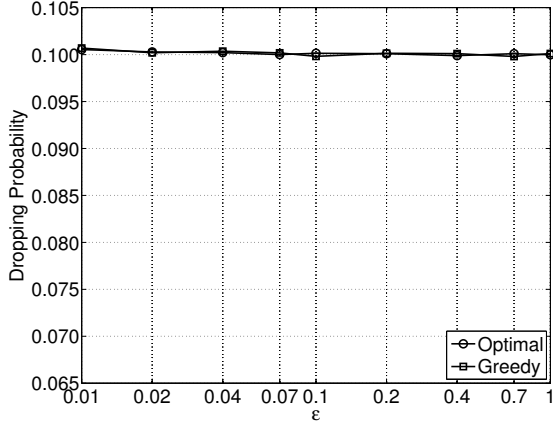
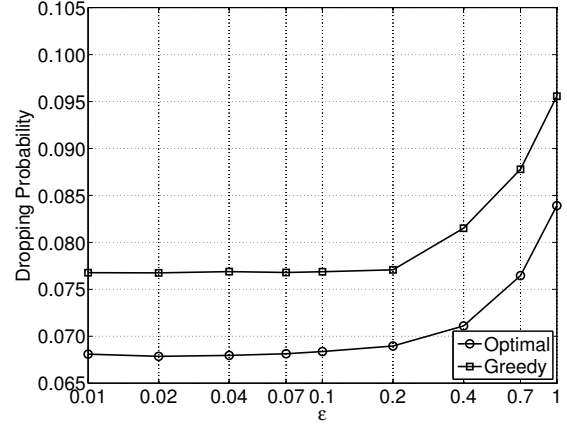
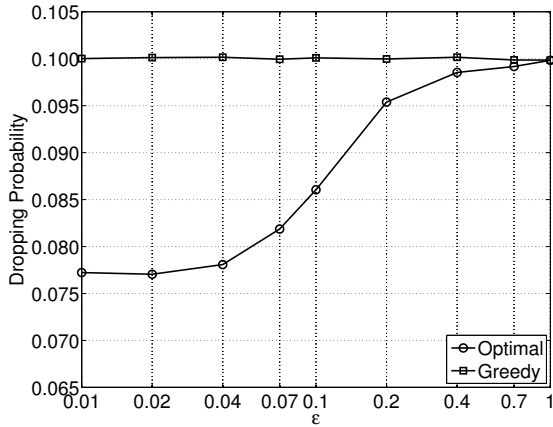
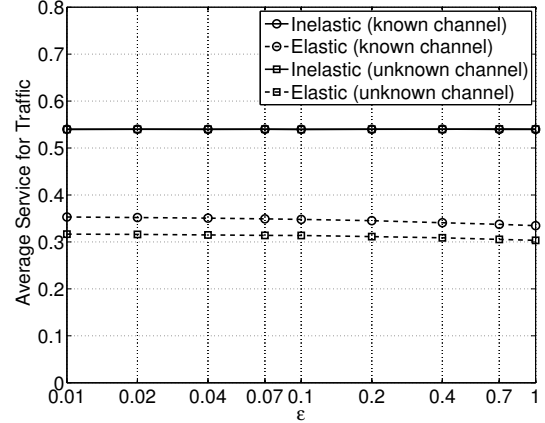
Finally, we study the loss of performance when we get to

know the channel state at the *end* of the frame compared to knowing the channel at the *beginning* of the frame. In Figs. 12, 13, and 14 we see that the rates achieved for inelastic traffic are identical, while elastic traffic has a performance loss no greater than 10% in any case. The reason for this comes from the fact that inelastic traffic has to meet throughput constraints, while elastic traffic uses the excess capacity. Thus, any loss in capacity is mostly absorbed by elastic traffic.

## IX. CONCLUSIONS AND FUTURE WORK

In this paper we have presented an optimization framework for the problem of congestion control and scheduling of elastic and inelastic traffic in ad hoc wireless networks. The model was developed for general interference graphs, general arrivals and time-varying channels. Using a dual function approach we presented a decomposition of the problem into an online algorithm that is able to make optimal decisions while keeping the network stable and fulfilling the inelastic flow's QoS constraints. A key result is that, through the use of deficit counters, one can treat the scheduling problem for elastic and inelastic flows in a common framework. It is also interesting to note that the deficit counters introduced in [13]–[15] have the



Fig. 9. Dropping probability when  $w_l = 0$ Fig. 11. Dropping probability when  $w_l = 6$ Fig. 10. Dropping probability when  $w_l = 3$ Fig. 12. Average service when  $w_l = 0$ 

interpretation of Lagrange multipliers. Simulations corroborate our results and show the dependency of the performance of the algorithm on the auxiliary parameter  $\epsilon$  and its role into assigning resources to both elastic and inelastic traffic.

It must be noted that in the unknown channel state case we assume that the channel remains constant during the entire frame. The case when the channel state varies from time slot to time slot is more challenging to solve since it involves a dynamic programming solution. Similarly, our traffic model for inelastic packets assumes that packets arrive at the beginning of the frame and all have the same delay. Possible extensions include heterogeneous delays and arrivals at any time slot. These topics are left for future work.

#### APPENDIX A PROOF OF LEMMA 1

To prove Lemma 1, we start by first proving two auxiliary lemmas and then stating a fact.

*Lemma 3:* Given that at frame  $k$  we have the event  $d(k) = d$  and  $q(k) = q$  and if there exists a point  $\mu(\Delta) \in \mathcal{C}/(1 + \Delta)$

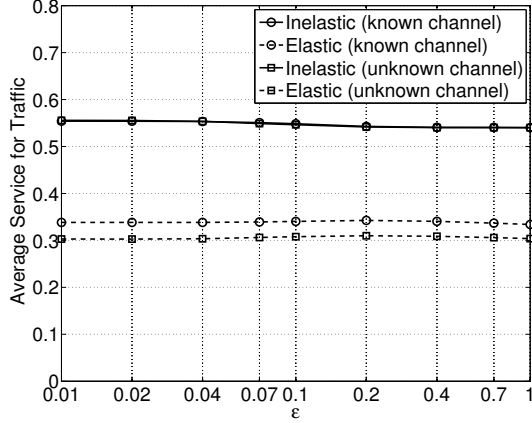
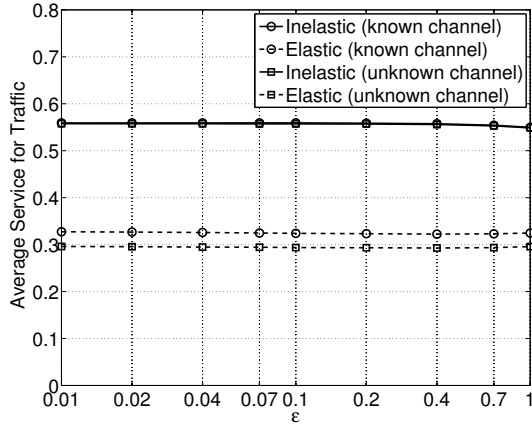
for some  $\Delta > 0$  such that (7) and (8) hold true, then

$$\begin{aligned} & E \left[ \frac{1}{2} \sum_{l \in \mathcal{L}} \{ [d_l + \tilde{a}_{il}(k) - I_{il}^*(a_i(k), c(k), d, q)]^+ \}^2 \right] - \sum_{l \in \mathcal{L}} \frac{d_l^2}{2} \\ & \leq B_8 + \sum_{l \in \mathcal{L}} d_l \mu_{il}(\Delta) \\ & \quad - E \left[ \sum_{l \in \mathcal{L}} \left( \frac{1}{\epsilon} w_l + d_l \right) I_{il}^*(a_i(k), c(k), d, q) \right. \\ & \quad \left. - \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} w_l I_{il}^*(a_i(k), c(k), d, q) \right] \end{aligned}$$

for some non-negative constant  $B_8$ , and where  $I_{il}^*(a_i(k), c(k), d, q)$  is given by the solution to (6).  $\diamond$

*Proof:*

$$\begin{aligned} & E \left[ \frac{1}{2} \sum_{l \in \mathcal{L}} \{ [d_l + \tilde{a}_{il}(k) - I_{il}^*(a_i(k), c(k), d, q)]^+ \}^2 \right] - \sum_{l \in \mathcal{L}} \frac{d_l^2}{2} \\ & \leq E \left[ \frac{1}{2} \sum_{l \in \mathcal{L}} [d_l + \tilde{a}_{il}(k) - I_{il}^*(a_i(k), c(k), d, q)]^2 \right] - \sum_{l \in \mathcal{L}} \frac{d_l^2}{2} \end{aligned}$$

Fig. 13. Average service when  $w_l = 3$ Fig. 14. Average service when  $w_l = 6$ 

$$\begin{aligned}
&= E \left[ \sum_{l \in \mathcal{L}} d_l [\tilde{a}_{il}(k) - I_{il}^*(a_i(k), c(k), d, q)] \right. \\
&\quad \left. + \frac{1}{2} \sum_{l \in \mathcal{L}} [\tilde{a}_{il}(k) - I_{il}^*(a_i(k), c(k), d, q)]^2 \right] \\
&\leq E \left[ \sum_{l \in \mathcal{L}} d_l \tilde{a}_{il}(k) - d_l I_{il}^*(a_i(k), c(k), d, q) \right. \\
&\quad \left. + \frac{1}{2} \sum_{l \in \mathcal{L}} \tilde{a}_{il}^2(k) + a_{il}^2(k) \right] \\
&\leq B_8 + \sum_{l \in \mathcal{L}} d_l x_{il}(1 - p_l) \\
&\quad - E \left[ \sum_{l \in \mathcal{L}} \left( \frac{1}{\epsilon} w_l + d_l \right) I_{il}^*(a_i(k), c(k), d, q) \right. \\
&\quad \left. - \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} w_l I_{il}^*(a_i(k), c(k), d, q) \right] \\
&\leq B_8 + \sum_{l \in \mathcal{L}} d_l \mu_{il}(\Delta)
\end{aligned}$$

$$\begin{aligned}
&- E \left[ \sum_{l \in \mathcal{L}} \left( \frac{1}{\epsilon} w_l + d_l \right) I_{il}^*(a_i(k), c(k), d, q) \right. \\
&\quad \left. - \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} w_l I_{il}^*(a_i(k), c(k), d, q) \right]
\end{aligned}$$

where (11) follows from the definition of  $I_{il}^*(a_i(k), c(k), d, q)$ , (12) follows from (7), and

$$B_8 = \frac{1}{2} \sum_{l \in \mathcal{L}} (x_{il}^2 + \sigma_{il}^2) [1 + (1 - p_l)^2] + x_{il} p_l (1 - p_l).$$

**Lemma 4:** Given that at frame  $k$  we have the event  $d(k) = d$  and  $q(k) = q$  and if there exists a point  $\mu(\Delta) \in \mathcal{C}/(1 + \Delta)$  for some  $\Delta > 0$  such that (7) and (8) hold true, then

$$\begin{aligned}
&E \left[ \frac{1}{2} \sum_{l \in \mathcal{L}} \{ [q_l + \tilde{a}_{el}(k) - I_{el}^*(a_i(k), c(k), d, q)]^+ \}^2 \right] - \sum_{l \in \mathcal{L}} \frac{q_l^2}{2} \\
&\leq B_9 - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \\
&\quad - \sum_{l \in \mathcal{L}} q_l \{ E [I_{el}^*(a_i(k), c(k), d, q)] - \mu_{el}(\Delta) \}
\end{aligned}$$

for some constant  $B_9 > 0$ , where  $x_e^*$  and  $\tilde{x}^*(k)$  are the solutions to (3) and (5) respectively.  $\diamond$

*Proof:*

$$\begin{aligned}
&E \left[ \frac{1}{2} \sum_{l \in \mathcal{L}} \{ [q_l + \tilde{a}_{el}(k) - I_{el}^*(a_i(k), c(k), d, q)]^+ \}^2 \right] - \sum_{l \in \mathcal{L}} \frac{q_l^2}{2} \\
&\leq E \left[ \frac{1}{2} \sum_{l \in \mathcal{L}} [q_l + \tilde{a}_{el}(k) - I_{el}^*(a_i(k), c(k), d, q)]^2 \right] - \sum_{l \in \mathcal{L}} \frac{q_l^2}{2} \\
&= E \left[ \sum_{l \in \mathcal{L}} q_l [\tilde{a}_{el}(k) - I_{el}^*(a_i(k), c(k), d, q)] \right. \\
&\quad \left. + \frac{1}{2} \sum_{l \in \mathcal{L}} [\tilde{a}_{el}(k) - I_{el}^*(a_i(k), c(k), d, q)]^2 \right] \\
&\leq E \left[ \sum_{l \in \mathcal{L}} q_l \tilde{a}_{el}(k) - q_l I_{el}^*(a_i(k), c(k), d, q) \right. \\
&\quad \left. + \frac{1}{2} \sum_{l \in \mathcal{L}} (\tilde{a}_{el}^2(k) + c_l^2 T^2) \right]
\end{aligned} \tag{13}$$

$$\begin{aligned}
&\leq B_9 + \sum_{l \in \mathcal{L}} - \left[ \frac{1}{\epsilon} U_l(\tilde{x}_{el}^*(k)) - q_l \tilde{x}_{el}^*(k) \right] \\
&\quad + \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} U_l(\tilde{x}_{el}^*(k)) - \sum_{l \in \mathcal{L}} q_l E [I_{el}^*(a_i(k), c(k), d, q)] \\
&\leq B_9 + \sum_{l \in \mathcal{L}} - \left[ \frac{1}{\epsilon} U_l(\mu_{el}(\Delta)) - q_l \mu_{el}(\Delta) \right] \\
&\quad + \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} U_l(\tilde{x}_{el}^*(k)) - \sum_{l \in \mathcal{L}} q_l E [I_{el}^*(a_i(k), c(k), d, q)]
\end{aligned} \tag{14}$$

$$\begin{aligned}
&= B_9 - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \\
&\quad - \sum_{l \in \mathcal{L}} q_l \{ E [I_{el}^*(a_i(k), c(k), d, q)] - \mu_{el}(\Delta) \}
\end{aligned} \tag{12}$$

where (13) follows from the definition of  $I_{el}^*(a_i(k), c(k), d, q)$ ,

$$B_9 = \frac{1}{2} \sum_{l \in \mathcal{L}} [X_{max}^2 + \sigma_e^2 + (\bar{c}_l^2 + \sigma_{cl}^2) T^2],$$

and (14) follows from the fact that  $\tilde{x}_e^*(k)$  is the optimal point of (5). ■

*Fact 1:* The optimization in (6) can be performed over  $\mathcal{S}(a_i(k), c(k))_{\mathcal{CH}}$ , the convex hull of  $\mathcal{S}(a_i(k), c(k))$ ; that is,

$$\begin{aligned} & \max_{s \in \mathcal{S}(a_i(k), c(k))} \sum_{l \in \mathcal{L}} \left[ \frac{1}{\epsilon} w_l + d_l(k) \right] \sum_{t=1}^T s_{il,t} + q_l(k) \sum_{t=1}^T s_{el,t} = \\ & \max_{s \in \mathcal{S}(a_i(k), c(k))_{\mathcal{CH}}} \sum_{l \in \mathcal{L}} \left[ \frac{1}{\epsilon} w_l + d_l(k) \right] \sum_{t=1}^T s_{il,t} + q_l(k) \sum_{t=1}^T s_{el,t}. \end{aligned}$$

The reason for this comes from the fact that the objective function is linear and therefore there must be an optimal point  $s^*(a_i(k), c(k), d(k), q(k)) \in \mathcal{S}(a_i(k), c(k))$ . ◊

*Proof of Lemma 1:* From the definition of  $\mathcal{C}$ , we know that  $\mu(\Delta) \in \mathcal{C}/(1 + \Delta)$  means that there exist  $(\bar{\mu}_{il}(a_i, c), \bar{\mu}_{el}(a_i, c))_{l \in \mathcal{L}} \in \mathcal{C}(a_i, c)$  for all  $a_i, c$  and  $(1 + \Delta)\mu_{il}(\Delta) = E[\bar{\mu}_{il}(a_i, c)]$ ,  $(1 + \Delta)\mu_{el}(\Delta) = E[\bar{\mu}_{el}(a_i, c)]$  for all  $l \in \mathcal{L}$ . For the rest of the proof we define  $\bar{\mu}_{il}(a_i, c)$ ,  $\bar{\mu}_{el}(a_i, c)$  to be such set of values associated to  $\mu(\Delta)$ .

From Lemmas 3 and 4 we have:

$$\begin{aligned} & E[V(d(k+1), q(k+1)) | d(k) = d, q(k) = q] - V(d, q) \\ & \leq B_1 + \sum_{l \in \mathcal{L}} d_l \mu_{il}(\Delta) + \sum_{l \in \mathcal{L}} q_l \mu_{el}(\Delta) \\ & \quad - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \\ & \quad - E \left[ \sum_{l \in \mathcal{L}} \left( \frac{1}{\epsilon} w_l + d_l \right) I_{il}^*(a_i(k), c(k), d, q) \right. \\ & \quad \left. + \sum_{l \in \mathcal{L}} q_l I_{el}^*(a_i(k), c(k), d, q) \right] \\ & \quad + \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} w_l E[I_{il}^*(a_i(k), c(k), d, q)] \\ & \leq B_1 + \sum_{l \in \mathcal{L}} d_l \mu_{il}(\Delta) + \sum_{l \in \mathcal{L}} q_l \mu_{el}(\Delta) \quad (15) \\ & \quad - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \\ & \quad - E \left[ \sum_{l \in \mathcal{L}} \left( \frac{1}{\epsilon} w_l + d_l \right) \bar{\mu}_{il}(a_i(k), c(k)) \right. \\ & \quad \left. + \sum_{l \in \mathcal{L}} q_l \bar{\mu}_{el}(a_i(k), c(k)) \right] \\ & \quad + \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} w_l E[I_{il}^*(a_i(k), c(k), d, q)] \\ & = B_1 + \sum_{l \in \mathcal{L}} d_l \mu_{il}(\Delta) + \sum_{l \in \mathcal{L}} q_l \mu_{el}(\Delta) \\ & \quad - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \end{aligned}$$

$$\begin{aligned} & - \sum_{l \in \mathcal{L}} \left( \frac{1}{\epsilon} w_l + d_l \right) (1 + \Delta) \mu_{il}(\Delta) \\ & - \sum_{l \in \mathcal{L}} q_l (1 + \Delta) \mu_{el}(\Delta) \\ & + \sum_{l \in \mathcal{L}} \frac{1}{\epsilon} w_l E[I_{il}^*(a_i(k), c(k), d, q)] \\ & = B_1 - \Delta \sum_{l \in \mathcal{L}} d_l \mu_{il}(\Delta) - \Delta \sum_{l \in \mathcal{L}} q_l \mu_{el}(\Delta) \\ & \quad - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \\ & \quad - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} \{w_l(1 + \Delta) \mu_{il}(\Delta) - w_l E[I_{il}^*(a_i(k), c(k), d, q)]\} \end{aligned}$$

where  $B_1 = B_8 + B_9$ ; (15) follows from the definition of  $\bar{\mu}_{il}(a_i, c)$ ,  $\bar{\mu}_{el}(a_i, c)$ ,  $\mathcal{C}(a_i, c)$ , and Fact 1. Thus we have the following:

$$\begin{aligned} & E[V(d(k+1), q(k+1)) | d(k) = d, q(k) = q] - V(d, q) \\ & \leq B_1 - B_2 \sum_{l \in \mathcal{L}} d_l - B_3 \sum_{l \in \mathcal{L}} q_l \\ & \quad - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(\mu_{el}(\Delta)) - U_l(\tilde{x}_{el}^*(k))] \\ & \quad - \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} \{w_l(1 + \Delta) \mu_{il}(\Delta) - w_l E[I_{il}^*(a_i(k), c(k), d, q)]\} \end{aligned}$$

where

$$B_2 = \Delta \min_{l \in \mathcal{L}} \{\mu_{il}(\Delta)\}$$

and

$$B_3 = \Delta \min_{l \in \mathcal{L}} \{\mu_{el}(\Delta)\}. \quad \blacksquare$$

## APPENDIX B PROOF OF THEOREM 1

From Lemma 2 we know that

$$\begin{aligned} & \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) - U_l(\tilde{x}_{el}^*(k))] \\ & + \frac{1}{\epsilon} \sum_{l \in \mathcal{L}} \{w_l \mu_{il}^* - w_l E[I_{il}^*(a_i(k), c(k), d, q)]\} \\ & \leq B_1 - B_6 \sum_{l \in \mathcal{L}} d_l - B_7 \sum_{l \in \mathcal{L}} q_l + V(d, q) \end{aligned}$$

$$- E[V(d(k+1), q(k+1)) | d(k) = d, q(k) = q]$$

$$\leq B_1 + V(d, q)$$

$$- E[V(d(k+1), q(k+1)) | d(k) = d, q(k) = q]$$

since  $B_6 \sum_{l \in \mathcal{L}} d_l + B_7 \sum_{l \in \mathcal{L}} q_l \geq 0$ . Taking expectations:

$$\begin{aligned} & \frac{1}{\epsilon} E \left[ \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) - U_l(\tilde{x}_{el}^*(k))] \right. \\ & \quad \left. + \sum_{l \in \mathcal{L}} \{w_l \mu_{il}^* - w_l I_{il}^*(a_i(k), c(k), d(k), q(k))\} \right] \\ & \leq B_1 - E[V(d(k+1), q(k+1))] + E[V(d(k), q(k))]. \end{aligned}$$

Adding the terms for  $k = \{1, \dots, K\}$  and dividing by  $K$  we get:

$$\begin{aligned} & \frac{1}{\epsilon} E \left[ \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) + w_l \mu_{il}^*] \right. \\ & \quad \left. - \sum_{l \in \mathcal{L}} \frac{1}{K} \sum_{k=1}^K \{U_l(\tilde{x}_{el}^*(k)) + w_l I_{il}^*(a_i(k), c(k), d(k), q(k))\} \right] \\ & \leq B_1 - \frac{E[V(d(K+1), q(K+1))]}{K} + \frac{E[V(d(1), q(1))]}{K} \\ & \leq B_1 + \frac{E[V(d(1), q(1))]}{K} \end{aligned} \quad (16)$$

where (16) follows from the fact that the Lyapunov function  $V$  is non-negative. Using Jensen's inequality [19] we get:

$$\begin{aligned} & \frac{1}{\epsilon} E \left[ \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) + w_l \mu_{il}^*] - \sum_{l \in \mathcal{L}} U_l \left( \frac{1}{K} \sum_{k=1}^K \tilde{x}_{el}^*(k) \right) \right. \\ & \quad \left. - \sum_{l \in \mathcal{L}} w_l \frac{1}{K} \sum_{k=1}^K I_{il}^*(a_i(k), c(k), d(k), q(k)) \right] \\ & \leq \frac{1}{\epsilon} E \left[ \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) + w_l \mu_{il}^*] \right. \\ & \quad \left. - \sum_{l \in \mathcal{L}} \frac{1}{K} \sum_{k=1}^K \{U_l(\tilde{x}_{el}^*(k)) + w_l I_{il}^*(a_i(k), c(k), d(k), q(k))\} \right] \\ & \leq B_1 + \frac{E[V(d(1), q(1))]}{K} \end{aligned}$$

Assuming  $E[V(d(1), q(1))] < \infty$  we get the following limit expression:

$$\limsup_{K \rightarrow \infty} E \left[ \sum_{l \in \mathcal{L}} [U_l(x_{el}^*) + w_l \mu_{il}^*] - \sum_{l \in \mathcal{L}} U_l \left( \frac{1}{K} \sum_{k=1}^K \tilde{x}_{el}^*(k) \right) \right. \\ \left. - \sum_{l \in \mathcal{L}} w_l \frac{1}{K} \sum_{k=1}^K I_{il}^*(a_i(k), c(k), d(k), q(k)) \right] \leq B\epsilon$$

where  $B = B_1$ . ■

## REFERENCES

- [1] J. J. Jaramillo and R. Srikant, "Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic," in *IEEE INFOCOM*, San Diego, CA, USA, Mar. 15–19, 2010.
- [2] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," in *IEEE INFOCOM*, vol. 3, Miami, FL, USA, Mar. 13–17, 2005, pp. 1794–1803.
- [3] X. Lin and N. B. Shroff, "Joint rate control and scheduling in multihop wireless networks," in *43rd IEEE Conference on Decision and Control (CDC)*, vol. 2, Atlantis, Paradise Island, Bahamas, Dec. 14–17, 2004, pp. 1484–1489.
- [4] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," in *IEEE INFOCOM*, vol. 3, Miami, FL, USA, Mar. 13–17, 2005, pp. 1723–1734.
- [5] A. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Systems*, vol. 50, no. 4, pp. 401–457, Aug. 2005.
- [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] L. Chen, S. H. Low, M. Chiang, and J. C. Doyle, "Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks," in *IEEE INFOCOM*, Barcelona, Catalunya, Spain, Apr. 23–29, 2006.

- [8] X. Lin, N. B. 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, Jun. 2006.
- [9] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *Wireless Networks*, vol. 8, no. 1, pp. 13–26, Jan. 2002.
- [10] V. Raghunathan, V. Borkar, M. Cao, and P. R. Kumar, "Index policies for real-time multicast scheduling for wireless broadcast systems," in *IEEE INFOCOM*, Phoenix, AZ, USA, Apr. 13–18, 2008, pp. 1570–1578.
- [11] A. Dua and N. Bambos, "Downlink wireless packet scheduling with deadlines," *IEEE Trans. Mobile Comput.*, vol. 6, no. 12, pp. 1410–1425, Dec. 2007.
- [12] Q. Liu, X. Wang, and G. B. Giannakis, "A cross-layer scheduling algorithm with QoS support in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 839–847, May 2006.
- [13] I.-H. Hou, V. Borkar, and P. R. Kumar, "A theory of QoS for wireless," in *IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 19–25, 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 *10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, New Orleans, LA, USA, May 18–21, 2009, pp. 175–184.
- [15] —, "Scheduling heterogeneous real-time traffic over fading wireless channels," in *IEEE INFOCOM*, San Diego, CA, USA, Mar. 15–19, 2010.
- [16] 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.
- [17] M. J. Neely, "Delay-based network utility maximization," in *IEEE INFOCOM*, San Diego, CA, USA, Mar. 15–19, 2010.
- [18] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Norwell, MA: Kluwer Academic Publishers, 2003.
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. New York, NY: Cambridge University Press, Mar. 2004.



**Juan José Jaramillo** (S '06, M '11) received his B.S. degree (summa cum laude) from Universidad Pontificia Bolivariana, Colombia, in 1998, his M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign in 2005 and 2010, respectively, all in electrical engineering. He is the recipient of a Fulbright fellowship. From 1999 to 2003 he worked at Empresas Publicas de Medellin in Colombia. He is currently with Iowa State University, where he is a Postdoctoral Research Associate at the Department of Electrical and Computer Engineering.

His research interests include communication networks and game theory.



**R. Srikant** (S '90-M '91-SM '01-F '06) received his B.Tech. from the Indian Institute of Technology, Madras in 1985, his M.S. and Ph.D. from the University of Illinois in 1988 and 1991, respectively, all in Electrical Engineering. He was a Member of Technical Staff at AT&T Bell Laboratories from 1991 to 1995. He is currently with the University of Illinois at Urbana-Champaign, where he is the Fredric G. and Elizabeth H. Nearing Endowed Professor in the Department of Electrical and Computer Engineering, and a Research Professor in the Coordinated Science

Lab.

He was an associate editor of *Automatica* and the *IEEE Transactions on Automatic Control*, and is currently an associate editor of the *IEEE/ACM Transactions on Networking*. He has also served on the editorial boards of special issues of the *IEEE Journal on Selected Areas in Communications* and *IEEE Transactions on Information Theory*. He was the chair of the 2002 IEEE Computer Communications Workshop in Santa Fe, NM and was a program co-chair of IEEE INFOCOM, 2007. His research interests include communication networks, stochastic processes, queueing theory, information theory, and game theory.