Joint Rate Control and Scheduling in Multihop Wireless Networks

Xiaojun Lin and Ness B. Shroff
School of Electrical and Computer Engineering, Purdue University
West Lafayette, IN 47907, U.S.A.
{linx, shroff}@ecn.purdue.edu

***Abstract*— We study the joint problem of allocating data rates and finding a stabilizing scheduling policy in a multihop wireless network. We propose a dual optimization based approach through which the rate control problem and the scheduling problem can be decomposed. We demonstrate via both analytical and numerical results that the proposed mechanism can fully utilize the capacity of the network, maintain fairness, and improve the quality of service to the users.**

I. Introduction

Future wireless networks are expected to support applications with high data rate requirements. Since the wireless spectrum is scarce, it is important to fully utilize the potential capacity of the network. One approach to improve the capacity of a wireless network is to use multi-hop instead of traditional single-hop communication [1], [2]. Another approach is to jointly control multiple layers of the network, including adaptive coding, link scheduling, power control, and routing. The capacity of wireless multihop networks with joint control over multiple layers has been studied in [3], [4]. In [3], the authors characterize the capacity region of a multihop wireless network and propose a Dynamic Routing and Power Control (DRPC) policy. As long as the exogenous data rates fall within the capacity region, the DRPC policy can schedule radio transmissions such that the system is stable (i.e., all queues inside the network remain finite). The DRPC policy generalizes the result of [5]. A related scheduling policy that also minimizes power consumption is studied in [4].

An issue that has not been treated thoroughly in the literature is how to control the data rates of the applications so that they fall within the capacity region. In future wireless networks, more and more applications will be data-oriented. Such applications are *elastic*, i.e., they can transmit data over a wide range of data rates. A network without an appropriate rate control mechanism could perform poorly in practice. Although, in theory, a dynamic scheduling policy[1] such as DRPC can ensure a bounded queue length whenever the *long term* exogenous demand is within the capacity region, the queue length can still be very large if the data rates chosen by the applications are bursty. Large queue lengths will either result in large delays, or, when the buffer size is limited, lead to a large amount of packet losses in the network. Both will negatively affect the quality of service experienced by the users. Further, in a network without rate control, a user could potentially monopolize the network resources by pouring large amounts of data into the network, which not only causes congestion but also unfairness towards the other users. Hence, developing a solid rate control strategy is important for the efficient management of future wireless networks. The objective of rate control is two-fold: to fully utilize the available capacity of the network, and to ensure fairness and good quality of service for the users.

Although rate control (or congestion control) has been studied extensively for *wireline* networks (see [6] for a good survey), these results cannot be directly applied to multihop wireless networks. In wireline networks, the capacity of each link is fixed. In wireless networks, however, the capacity of each link is a function of the underlying schedule used at each time. Past works on rate control in wireless networks either consider only single-hop flows [7], [8], [9], or impose simplified assumptions on a restrictive set of scheduling policies [10], [11], [12], [13], [14]. Hence, these works have not fully exploited the benefit of multihop communication and joint multi-layer control.

In this work, we present a unified framework for joint rate control and scheduling in multihop wireless networks. Our solution to the joint rate control and scheduling problem has an attractive *decomposition* property. Using a dual approach, we show that the rate control problem and the scheduling problem can be decomposed and solved individually. The two problems are then coupled by the implicit cost associated with each queue to solve the joint problem. We show via both analytical and numerical results that our joint rate control and scheduling algorithm can significantly reduce the queue length inside the network and improve the quality of service to the users. Due to space constraints, we will omit the proofs for most results. They are available in [15].

II. The System Model

We consider a wireless network with N nodes. Let $\mathcal{L}$ denote the set of node pairs (i, j) such that transmission from node i to node j is allowed. Due to the shared nature of the wireless media, the data rate r_{ij} of a link (i, j) depends

[1] From now on, we will use the term *scheduling* to refer to the joint control of layers other than rate control, including adaptive coding, link scheduling, power control and routing.

not only on the power P_{ij} assigned to the link, but also on the interference due to the power assignments on other links. Let $\vec{P} = [P_{ij}, (i,j) \in \mathcal{L}]$ denote the power assignments and let $\vec{r} = [r_{ij}, (i,j) \in \mathcal{L}]$ denote the data rates. We assume that $\vec{r} = u(\vec{P})$, i.e., the data rates are completely determined by the global power assignment. (Channel variation, e.g., due to fading, is not considered.) The function $u(\cdot)$ is called the *rate-power function* of the system. There may be constraints on the feasible power assignment. For example, if each node has a total power constraint $P_{i,\max}$, then $\sum_{j:(i,j)\in\mathcal{L}} P_{ij} \leq P_{i,\max}$. Let Π denote the set of feasible power assignments, and let $\mathcal{R} = \{u(\vec{P}), \vec{P} \in \Pi\}$. We assume that $\mathrm{Co}(\mathcal{R})$, the convex hull of $\mathcal{R}$, is closed and bounded.

In our system, there are S users and each user s is associated with a source node f_s and a destination node d_s. Let x_s be the rate with which data is sent from f_s to d_s, over possibly multiple paths and multiple hops. We assume that x_s is bounded in $[0, M_s]$. Each user is associated with a utility function $U_s(x_s)$, which reflects the "utility" to the user s when it can transmit at data rate x_s. We assume that $U_s(\cdot)$ is strictly concave, non-decreasing and continuously differentiable on $[0, M_s]$. The concavity assumption models the "principle of diminishing returns" for elastic applications. We assume that time is divided into slots. At each time slot, the scheduling policy will select a power assignment vector $\vec{P}$ and select data to be forwarded on each link. Given a user rate vector $\vec{x} = [x_s, s = 1, ..., S]$, we say that a system is *stable* under a scheduling policy if the queue length at each node remains finite. In this paper, we are interested in the following joint rate control and scheduling problem:

- Find the user rate vector $\vec{x}$ that maximizes the sum of the utilities of all users $\sum_s U_s(x_s)$ subject to the constraint that the system is stable under some scheduling policy.
- Find the associated scheduling policy that stabilizes the system.

We define the *capacity region* Λ of the system as the largest set of rate vectors $\vec{x}$ such that $\vec{x} \in \Lambda$ is a necessary condition for network stability under *any* scheduling policy. Hence, our rate control problem is simply

$$\max_{x_s \leq M_s} \quad \sum_s U_s(x_s) \qquad \text{subject to} \quad \vec{x} \in \Lambda. \tag{1}$$

The capacity region can be determined by the rate-power function $u(\cdot)$ and the power constraint set Π [2], [3], [4]. In this paper, we consider two cases: the *route-independent* case and the *route-dependent* case.. In the *route-independent* case, the routes for each user are not specified before-hand. Let $\mathcal{D} = \{d_s, s = 1, ..., S\}$ denote the set of destination nodes. The capacity region is determined as in [3]:

The Route-Independent Case: The capacity region Λ is the set of user rates $\vec{x}$ such that there exists a link rate vector $\vec{r}^d$ associated with each destination node d and the vector $\vec{R} = [\vec{r}^d, d \in \mathcal{D}]$ satisfies:

$$\begin{aligned} & r_{ij}^d \geq 0 \text{ for all } (i,j) \in L \text{ and for all } d \in \mathcal{D} \\ & \sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d - \sum_{s:f_s=i,d_s=d} x_s \geq 0 \\ & \qquad \text{for all } d \text{ and for all } i \neq d \\ & [\sum_d r_{ij}^d] \in \mathrm{Co}(\mathcal{R}), \end{aligned} \tag{2}$$

where r_{ij}^d can be interpreted as the amount of capacity on link (i,j) that is allocated for data towards destination d.

In the *route-dependent* case, each user s has $\theta(s)$ alternate routes. Let $H = [H_{ij}^{sv}]$ denote the routing matrix, i.e., $H_{ij}^{sv} = 1$ if path v of user s uses link (i,j), and $H_{ij}^{sv} = 0$ otherwise. The capacity region is determined as in [4]:

The Route-Dependent Case: The capacity region Λ is the set of user rates $\vec{x}$ such that there exists x_{sv} for each s, v and the vector $[x_{sv}, s = 1, ..., S, v = 1, ...\theta(s)]$ satisfies:

$$x_s = \sum_v x_{sv} \text{ for all } s, \text{ and } [\sum_{s,v} H_{ij}^{sv} x_{sv}] \in \mathrm{Co}(\mathcal{R}),$$

where x_{sv} can be interpreted as the data rate on path v of user s, and $\sum_{s,v} H_{ij}^{sv} x_{sv}$ is the total rate on link (i,j).

In both cases, it is easy to show that the capacity region Λ is a convex and compact set. Because the utility function is strictly concave, an optimal solution to (1) exists and is unique. We next present a methodology for solving (1) and the associated scheduling policy.

III. The Solution

In this section, we will develop a framework for solving the joint rate control and scheduling problem.

A. The Route-Independent Case

We first study the route-independent case using a duality approach. We can assign a Lagrange multiplier q_i^d ($i \neq d$) for each constraint in (2), construct the Lagrangian and obtain the dual objective function (see [15] for the detail). Let $q_d^d = 0$ and $\vec{q} = [q_i^d]$. The dual objective function is of the following form:

$$D(\vec{q}) = \sum_s B_s(q_{f_s}^{d_s}) + V(\vec{q}),$$

where

$$B_s(q) = \max_{x_s \leq M_s} U_s(x_s) - x_s q, \tag{3}$$

$$V(\vec{q}) = \max_{\vec{r}=u(\vec{P}),\vec{P}\in\Pi} \sum_{(i,j)\in\mathcal{L}} r_{ij} \max_d (q_i^d - q_j^d). \tag{4}$$

The dual approach thus results in an elegant decomposition of the original problem. Given $\vec{q}$, we have decomposed the problem into the rate control problem (3) and the scheduling problem (4). The Lagrange multiplier q_i^d can be interpreted as the implicit cost at node i for destination node d. Each user s solves its own utility maximization problem B_s independently as if the "price" for user s is $q_{f_s}^{d_s}$. The scheduling problem $V(\vec{q})$ is precisely the the DRPC policy in [3].

The dual problem of (1) is then

$$\min_{\vec{q}\geq 0} D(\vec{q}). \tag{5}$$

Let $\vec{r} = u(\vec{P})$ be the solution to the scheduling problem (4). Compute the related set of vectors $\vec{r}^d, d \in \mathcal{D}$ as follows. For each link (i,j), let $d^*(i,j) = \text{argmax}_d\, q_i^d - q_j^d$, and let $r_{ij}^d = r_{ij}$ if $d = d^*(i,j)$ and $r_{ij}^d = 0$ otherwise. We can show that $D(\vec{q})$ is convex and its subgradient is given by,

$$\frac{\partial D}{\partial q_i^d} = \left[\sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d - \sum_{s:f_s=i,d_s=d} x_s\right],$$

where $\vec{x} = [x_s]$ solves (3) and the vectors $\vec{r}^d$ are as defined earlier. We can then use the following subgradient method to solve the dual problem. We will refer to this solution as the node-centric solution, since the Lagrange multipliers (i.e., implicit costs) are associated with the nodes.

The Node-Centric Solution:

$$q_i^d(t+1) = \left\{ q_i^d(t) - h_t \left[\sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d(t) - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d(t) - \sum_{s:f_s=i,d_s=d} x_s(t)\right]\right\}^+, \tag{6}$$

where $h_t, t = 1,2,...$ is a sequence of positive stepsizes, $\vec{x}(t)$ and $\vec{r}^d(t)$ are defined as $\vec{x}$ and $\vec{r}^d$ earlier with $\vec{q} = \vec{q}(t)$. The following proposition is a consequence of Theorem 2.3 in [16, p26].

Proposition 1: a) There is no duality gap, i.e., the minimal value of (5) coincides with the optimal value of (1).

b) Let Φ be the set of $\vec{q}$ that minimizes $D(\vec{q})$. For any $\vec{q} \in \Phi$, let $\vec{x}$ solve (3), then $\vec{x}$ is the unique optimal solution $\vec{x}^*$ of (1).

c) Let $\rho(\vec{q},\Phi) = \min_{\vec{p}\in\Phi} ||\vec{q} - \vec{p}||$. If

$$h_t \to 0 \text{ as } t \to \infty, \text{ and } \textstyle\sum_t h_t = \infty,$$

then $\rho(\vec{q}(t),\Phi) \to 0$ and $\vec{x}(t) \to \vec{x}^*$ as $t \to \infty$.

The last part of Proposition 1 shows that, when $h_t \downarrow 0$ in an appropriate fashion, iteration (6) converges and solves the optimal rate assignment. In practice, we usually use a constant stepsize. As long as the stepsize is small, we can still show that $\vec{q}(t)$ will converge to a small neighborhood around Φ. Because the mapping from $\vec{q}(t)$ to $\vec{x}(t)$ is continuous, we can then conclude that the user rates $\vec{x}(t)$ will also converge to a small neighborhood of the optimal rate assignment $\vec{x}^*$. Further, using constant stepsize also allows the algorithm to track the non-stationarity in the network conditions.

1) The Joint Scheduling Policy : We now show that a stabilizing scheduling policy can be obtained as a by-product of solving (4) at each iteration (6). We have each node i maintain a queue for each destination node $d \neq i$, and let Q_i^d denote its queue length. At each time slot, we use the power assignment that generates the optimal vector $\vec{r}$ in (4) (and subsequently the vectors $\vec{r}^d$). For each link $(i,j) \in \mathcal{L}$, we then pick the queue $Q_i^{d^*}$ at node i with $r_{ij}^{d^*} = r_{ij}$ and transmit data from queue $Q_i^{d^*}$ to node j at rate r_{ij}. The evolution of Q_i^d is then determined by

$$Q_i^d(t+1) = \left\{ Q_i^d(t) - \left[\sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d(t) - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d(t) - \sum_{s:f_s=i,d_s=d} x_s(t)\right]\right\}^+. \tag{7}$$

Comparing (7) with (6), we can see that $Q_i^d(t) = q_i^d(t)/h$ when $h_t = h > 0$. Since $q_i^d(t)$ is bounded, we conclude that $Q_i^d(t)$ is bounded as well.

Proposition 2: Assume $h_t = h > 0$ for all t. If h is small enough, then using the above scheduling policy, we have,

$$\sup_t Q_i^d(t) < +\infty \text{ for all } i \neq d.$$

2) Fairness and Stability : Fairness among users can be controlled by appropriately choosing the utility functions [17]. A general form of the utility function is

$$U_s(x_s) = w_s \frac{x_s^{1-\gamma}}{1-\gamma}, \gamma > 0, \tag{8}$$

where $w_s, s = 1,...,S$ are the weights. Maximizing the total utility will correspond to *weighted proportional fairness* as $\gamma \to 1$, and *weighted max-min fainess* as $\gamma \to \infty$.

An interesting question that we need to address is: although fairness and quality of service are improved by applying rate control, does this reduce the *stability region* (to be defined below) of the system? To be precise, we replace each user s by a dynamic group of users with the same utility function and the same source and destination pair. We assume that users of group s arrive according to a Poisson process with rate λ_s and each user brings a file for transfer whose size is exponentially distributed with mean $1/\mu_s$. The load brought by group s is then $\rho_s = \lambda_s/\mu_s$. The *stability region* Θ of the system under a given rate-control and scheduling policy is the set of load vectors $[\rho_s, s = 1,...,S]$ such that the number of users and the queue lengths remain finite when $[\rho_s] \in \Theta$. Note that in the framework of [3], there is no rate control, i.e., each user can send the entire file into the system immediately upon its arrival. The authors of [3] show that as long as $[\rho_s]$ resides strictly inside the capacity region Λ, the DRPC policy (without rate control) will stabilize the system. Hence, the stability region coincides with the capacity region Λ. We now study whether the same conclusion holds when rate control is applied. Let n_s denote the number of users from group s that are currently in the system. We assume that the rate allocation $\vec{x}$ at each time is perfectly determined by the solution to the utility maximization problem (1) given the current set of users. Because the iteration (6) takes time to converge, we can expect that this assumption will better capture systems transferring "longer" files [17]. The transition of n_s is then given by:

$$n_s \to n_s + 1, \qquad \text{with rate } \lambda_s$$
$$n_s \to n_s - 1, \qquad \text{with rate } x_s n_s \mu_s.$$

Proposition 3: Assume that the rate allocation perfectly solves (1) at each time and the utility function is of the form (8) for some $\gamma > 0$. If $[\rho_s]$ resides strictly inside the capacity region Λ, then the Markov process $[n_s]$ is ergodic.

Proposition 3 shows that the stability region is not reduced by applying rate control. On the other hand, the benefit of applying rate control is that the queue length Q_i^d inside the network can be tightly controlled. In Section IV, we will show via simulation that the queue length inside the network can indeed be significantly reduced by applying our joint rate control and scheduling algorithm.

3) Virtual Queues : The fact that the queue size Q_i^d and the implicit cost q_i^d are tightly coupled by iteration (6) may have the following undesirable consequence. The more congested the network, the higher the implicit costs, and the larger the delay at each node. Using the *virtual queue* concept as in wireline networks [18], we can couple the implicit cost with a virtual queue instead of the real queue Q_i^d. We only need to modify the iteration (6) as:

The Node-Centric Solution with Virtual Queue:

$$q_i^d(t+1) = \left\{ q_i^d(t) - h \left[\sum_{j:(i,j)\in\mathcal{L}} \delta r_{ij}^d(t) - \sum_{j:(j,i)\in\mathcal{L}} \delta r_{ji}^d(t) - \sum_{s:f_s=i,d_s=d} x_s(t) \right] \right\}^+, \quad (9)$$

where δ is a positive factor slightly smaller than 1. The iteration (9) corresponds to shrinking the capacity region to $\delta\Lambda$. We can imagine a *virtual queue* VQ_i^d at each node i associated with each destination d such that $VQ_i^d = q_i^d/h$, while the *real queue* still evolves as in (7). If the number of the users is fixed, while the implicit costs (and the virtual queue length) will converge to positive values, the real queue length will eventually go to zero. In Section IV, we will demonstrate that using the virtual queue algorithm can further reduce the queue length inside the network with a minimal cost to the capacity of the system.

B. The Route-Dependent Case

The route-independent formulation in Section III-A is convenient for systems with a small number of destination nodes. For example, for traffic from wireless terminals to the (single) base-station, each node only needs to maintain one queue. No per-flow information needs to be maintained. If the number of destinations is large, each node then needs to maintain many queues, each of which corresponds to one destination node. As we will see next, the route-dependent formulation is more convenient in such scenarios. With the route-dependent formulation, each link again only needs to maintain one queue (or implicit cost) and no per-flow information needs to be maintained.

We introduce an auxiliary variable $c_{ij} \geq 0$ for each link $(i,j) \in \mathcal{L}$, and rewrite the primal problem (1) as:

$$\max_{[x_{sv}]\geq 0} \quad \sum_s U_s(\sum_v x_{sv}) \quad (10)$$

$$\text{subject to} \quad \sum_{s,v} H_{ij}^{sv} x_{sv} \leq c_{ij} \text{ for all } (i,j) \in \mathcal{L} \quad (11)$$

$$\text{and} \quad [c_{ij}] \in \text{Co}(\mathcal{R}), \sum_v x_{sv} \leq M_s. \quad (12)$$

The route-dependent case can then be treated analogously to the route-independent case. (For the detail, please refer to [15].) We can associate a Lagrange multiplier q_{ij} for each constraint in (11), and obtain the following iterative solution. We will refer to this solution as the link-centric solution, since the Lagrange multiplier is associated with each link.

The Link-Centric Solution:

$$q_{ij}(t+1) = \left\{ q_{ij}(t) + h_t \left[\sum_{s,v} H_{ij}^{sv} x_{sv}(t) - c_{ij}(t) \right] \right\}^+. \quad (13)$$

Here h_t is a sequence of stepsizes. The vector $[x_{sv}(t)]$ solves

$$B_s(\vec{q}) = \max_{x_{sv}\geq 0, \sum_v x_{sv} \leq M_s} \left[U_s(\sum_v x_{sv}) - \sum_v \sum_{(i,j)\in\mathcal{L}} H_{ij}^{sv} q_{ij} x_{sv} \right] \quad (14)$$

for each user s and the vector $\vec{c}(t) = [c_{ij}(t)]$ solves

$$V(\vec{q}) = \max_{[c_{ij}]=u(\vec{P}), \vec{P}\in\Pi} \sum_{(i,j)\in\mathcal{L}} q_{ij} c_{ij}. \quad (15)$$

Note that $\sum_{(i,j)\in\mathcal{L}} H_{ij}^{sv} q_{ij}$ in (14) can be viewed as the implicit cost of path v of user s. As we can see, the rate control problem B_s and the scheduling problem V are again decomposed.

There is one problem with the iterations (13)-(15). Although the iteration (13) may converge to the optimal value of $\vec{q}$, the user rates x_{sv} will not converge if some users have multiple paths. In fact, when we solve (14) for a user s that has multiple paths, only paths that have the minimum cost will have positive data rates. If the costs of several paths are close to each other, the data rates on these paths will oscillate as the implicit cost $\vec{q}(t)$ is being updated. This difficulty arises because the objective function in (10) and (14) is not strictly concave in $[x_{sv}]$. One can overcome this problem of oscillation by using the Proximal Optimization

Algorithm [19]. We can introduce an auxiliary variable y_{sv} for each x_{sv}, and modify the objective function to be

$$\max_{[x_{sv}]\geq 0} \sum_s U_s(\sum_v x_{sv}) - \frac{\nu}{2}\sum_{s,v}(x_{sv}-y_{sv})^2 \quad (16)$$
$$\text{subject to } \ (11) \text{ and } (12),$$

where $\nu > 0$. For any fixed $\vec{y} = [y_{sv}]$, the objective function becomes strictly concave. We can retain the link-centric solution (13) except that now the vector $[x_{sv}(t)]$ should solve

$$B_s(\vec{q}|\vec{y}) = \max_{x_{sv}\geq 0, \sum_v x_{sv}\leq M_s} \Bigg[U_s(\sum_v x_{sv}) - \frac{\nu}{2}\sum_v (x_{sv}-y_{sv})^2 - \sum_v \sum_{(i,j)\in\mathcal{L}} H_{ij}^{sv} q_{ij} x_{sv} \Bigg].$$

No oscillation will occur because the mapping from $\vec{q}$ to $\vec{x}$ is now continuous. Using similar techniques as in Proposition 1, we can show the following result.

Proposition 4: If

$$h_t \to 0 \text{ as } t \to \infty, \text{ and } \textstyle\sum_t h_t = \infty,$$

then $[x_{sv}(t)] \to [x_{sv}^*]$ by the iteration (13), where the vector $[x_{sv}^*]$ is the unique optimal solution to (16) given $\vec{y}$.

Finally, an optimal solution to the original problem (10) can be obtained by the following iteration [19]:

- Given $\vec{y} = \vec{y}(k)$, solve (16). Let $\vec{x}$ be its optimal solution.
- Let $\vec{y}(k+1) = \vec{x}\beta + \vec{y}(k)(1-\beta)$ where $0 < \beta < 1$.

In practice, we usually use constant stepsize $h_t = h$, and we do not wait for the proximal problem (16) to converge before we set $\vec{y}$ to the new value of $\vec{x}$. Readers can refer to [20] for related theoretical analysis. We can also derive analogous results to those in Sections III-A.1 to III-A.3 for the route-dependent case.

IV. Simulation Results

Due to space constraints, we will only report simulations with the node-centric solution in Section III-A. Readers can refer to [15] for additional simulation results for the link-centric solution. Note that the most computationally expensive step of our solution is to solve the scheduling subproblems (4) or (15). The complexity depends on the form of the rate-power function $u(\cdot)$. For our simulation, we assume that the data rate at each link is proportional to its Signal-to-Interference ratio (SIR). This assumption is suitable for CDMA systems with a moderate processing gain [4]. Each node i also has a maximum power constraint $P_{i,\max}$. With a rate-power function of this form, the scheduling problem can be solved by applying a similar procedure as the one used in [4]. For details, please refer to [15]. The complexity of the procedure is $O(2^N)$. When the number of nodes N in the system is large, a clustering heuristic as in [21] can be used to find approximate solutions.

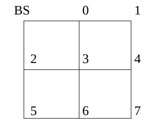

Fig. 1. The "grid" topology

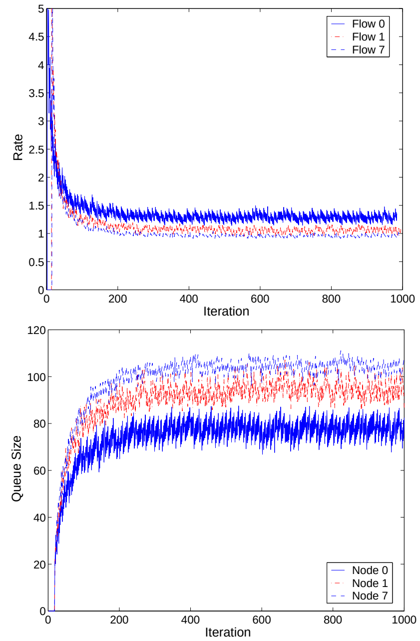

Fig. 2. The evolution of the user rates (top) and the queue length (bottom)

Our simulation uses the "grid" topology in Fig. 1. There are 8 terminals and 1 base-station (BS). The horizontal and vertical distance between neighboring nodes is 1.0 unit. We first assume that each terminal has one user sending data towards the base-station. The utility function for each user is $U(x) = \ln x$. Each node can communicate directly with any other node. The power constraint at each node is $P_{i,\max} = 1.0$ unit. The path loss is d^{-4} where d is the distance from the transmitter to the receiver. The rate of each link is proportional to the SIR, with $r_{ij} = 10.0 \times \text{SIR}_{ij}$. The ambient noise level is $N_0 = 1.0$ unit.

In Fig. 2, we report the evolution of the data rates of three selected users at nodes 0, 1 and 7, respectively, and the evolution of the queue lengths at these three nodes. The stepsize $h = 0.01$. (The implicit cost at each node is simply h times the respectively queue length.) As we can see, all quantities of interest converge to a small neighborhood.

We have also simulated the virtual queue algorithm in Section III-A.3 with $\delta = 0.95$. The result is not reported here due to space constraints and is available in [15]. We find that both the data rates of the users and the virtual queue length converge to values that are only slightly different from those in Fig. 2. However, the real queues are eventually driven close to zero, as predicted by our theoretical result.

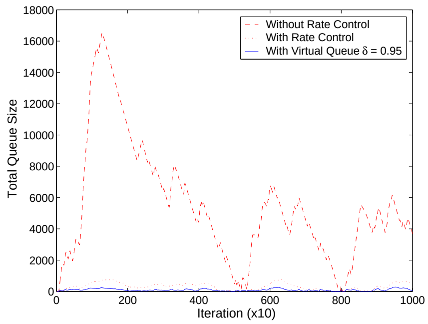

Fig. 3. The evolution of the total queue length summed over all nodes

We next simulate the case when there are dynamic arrivals and departures of the users. Users arrive at each node according to a Poisson process with rate $\lambda_s = 0.009$. Each user needs to send to the base-station a file whose size is exponentially distributed with mean $1/\mu_s = 100$ unit. The utility function for each user and the radio transmission model are the same as before. We also simulate the case when there is no rate control, i.e., when a user arrives, it pumps all data into the source node at a high data rate of 20.0. In Fig. 3, we compare the total queue length summed over all nodes with rate control (the dotted line) and without rate control (the dashed line). Obviously, the queue length is reduced significantly when rate control is employed.

We also simulate the virtual queue algorithm when there are dynamic arrivals and departures of the users. The (lowest) solid line in Fig. 3 shows the total queue length summed over all nodes when the virtual queue algorithm is used. Although the virtual queue parameter $\delta = 0.95$ is close to 1 (which means that we only incur a small cost at the capacity of the system), the virtual queue algorithm is still very effective in further reducing the queue length inside the network. The peak queue length is reduced from 770 (without virtual queue) to 284 (with virtual queue).

V. Conclusion

In this paper, we have presented a framework for joint rate-control and scheduling in multihop wireless networks. We proposed a dual approach through which the rate control problem and the scheduling problem can be decomposed. Our solution not only fully utilizes the capacity of the network, but also ensures fairness and good quality of service to the users. We demonstrate via both analytical and numerical results that the proposed mechanism can effectively reduce the queue length and the packet delay inside the network.

The most computationally expensive part of the solution is to find the schedule that maximizes the total weighted link capacity at each iteration. For future work, we plan to study simple heuristics that can approximate the optimal schedule. We are particularly interested in heuristic solutions that are easy to implement in a distributed fashion. We will also study how the rate allocation will be affected by the use of these heuristics.

References

[1] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[2] S. Toumpis and A. J. Goldsmith, "Capacity Regions for Wireless Ad Hoc Networks," *IEEE Transcations on Wireless Communications*, vol. 2, no. 4, pp. 736–748, July 2003.

[3] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic Power Allocation and Routing for Time Varying Wireless Networks," in *Proceedings of IEEE INFOCOM*, San Francisco, April 2003.

[4] R. L. Cruz and A. V. Santhanam, "Optimal Routing, Link Scheduling and Power Control in Multi-hop Wireless Networks," in *Proceedings of IEEE INFOCOM*, San Francisco, April 2003.

[5] 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*, vol. 37, no. 12, pp. 1936–1948, December 1992.

[6] S. H. Low and R. Srikant, "A Mathematical Framework for Designing a Low-Loss Low-Delay Internet," *Network and Spatial Economics*, vol. 4, no. 1, pp. 75–102, March 2004.

[7] L. Tassiulas and S. Sarkar, "Maxmin Fair Scheduling in Wireless Networks," in *Proceedings of IEEE INFOCOM*, New York, NY, June 2002, pp. 763–772.

[8] T. Nandagopal, T. Kim, X. Gao, and V. Bharghavan, "Achieving MAC Layer Fairness in Wireless Packet Networks," in *Proceedings of ACM Mobicom*, Boston, MA, August 2000, pp. 87–98.

[9] X. Huang and B. Bensaou, "On Max-min Fairness and Scheduling in Wireless Ad-Hoc Networks: Analytical Framework and Implementation," in *Proceedings of IEEE/ACM MobiHoc*, Long Beach, CA, October 2001, pp. 221–231.

[10] S. Sarkar and L. Tassiulas, "End-to-end Bandwidth Guarantees Through Fair Local Spectrum Share in Wireless Ad-hoc Networks," in *Proceedings of the IEEE Conference on Decision and Control*, Maui, Hawaii, December 2003.

[11] Y. Yi and S. Shakkottai, "Hop-by-hop Congestion Control over a Wireless Multi-hop Network," in *Proceedings of IEEE INFOCOM*, Hong Kong, March 2004.

[12] Y. Qiu and P. Marbach, "Bandwith Allocation in Ad-Hoc Networks: A Price-Based Approach," in *Proceedings of IEEE INFOCOM*, San Francisco, CA, April 2003.

[13] Y. Xue, B. Li, and K. Nahrstedt, "Price-based Resource Allocation in Wireless Ad Hoc Networks," in *Proceedings of the Eleventh International Workshop on Quality of Service (IWQoS 2003), also Lecture Notes in Computer Science, ACM Springer-Verlag*, vol. 2707, Monterey, CA, June 2003, pp. 79–96.

[14] M. Chiang, "To Layer or Not to Layer: Balancing Transport and Physical Layers in Wireless Multihop Networks," in *Proceedings of IEEE INFOCOM*, Hong Kong, March 2004.

[15] X. Lin and N. B. Shroff, "Joint Rate Control and Scheduling in Multihop Wireless Networks," *Technical Report, Purdue University, http://min.ecn.purdue.edu/∼linx/papers.html*, 2004.

[16] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Berlin: Springer-Verlag, 1985.

[17] T. Bonald and L. Massoulie, "Impact of Fairness on Internet Performance," in *Proceedings of ACM Sigmetrics*, Cambridge, MA, June 2001, pp. 82–91.

[18] A. Lakshmikantha, C. Beck, and R. Srikant, "Robustness of Real and Virtual Queue Based Active Queue Management Schemes," in *Proc. American Control Conference*, June 2003, also available at http://comm.csl.uiuc.edu/∼srikant/pub.html.

[19] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. New Jersey: Prentice-Hall, 1989.

[20] X. Lin and N. B. Shroff, "The Multi-path Utility Maximization Problem," in *41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2003.

[21] R. L. Cruz and A. V. Santhanam, "Hierarchical Link Scheduling and Power Control in Multihop Wireless Networks," in *Proceedings of the 40th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2002.