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 problem of optimal data rate allocation to a group of users in a multihop wireless network and simultaneously finding a stabilizing scheduling policy. 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.

1 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 the 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 bounded). The DRPC policy generalizes the result of [5], and is similar to the *maximum weighted matching* policy in switch scheduling [6]. 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* such as DRPC can ensure bounded a 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 [7] for a good survey), these results cannot be directly applied to multihop wireless networks.

*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.

In wireline networks, the capacity region is of a simple form, i.e., the sum of the data rates at each link should be less than the known link capacity. In multihop wireless networks, the capacity of each radio link depends on the signal and interference levels, and is thus correlated with the capacities of other links. Therefore, the capacity region of the network is usually a complex function of the underlying scheduling policy and cannot be characterized in a simple form. Past works on rate control in wireless networks either consider only single-hop flows [8, 9, 10], or impose simplified assumptions on a restrictive set of scheduling policies [11, 12, 13, 14, 15]. 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.

The rest of the paper is structured as follows. The system model and the problem formulation are presented in Section 2, followed by our solution in Section 3. We then study two special cases of the problem in Section 4 and compare our solution with existing works. Simulation results are presented in Section 5, and the conclusion is given in Section 6.

2 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 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 rate with which data are 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 its data rate is x_s. We assume that $U_s(\cdot)$ is strictly concave, non-decreasing and continuous differentiable on $[0, M_s]$. The concavity assumption models the "principle of dimishing 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

$$\begin{aligned} \max_{x_s \le M_s} \quad & \sum_s U_s(x_s) \\ \text{subject to} \quad & \vec{x} \in \Lambda. \end{aligned} \tag{1}$$

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 beforehand. 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:

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

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

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

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.

3 The Solution

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

3.1 The Route-Independent Case

We first study the route-independent case. We can assign a Lagrange multiplier q_i^d $(i \neq d)$for each constraint in (2). Let $q_d^d = 0$. The Lagrangian is then:

$$
\begin{aligned}
& L(\vec{x}, \vec{R}, \vec{q}) \\
= & \sum_s U_s(x_s) + \sum_d \sum_{i \neq d} q_i^d \left[\sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d \right. \\
& \qquad \left. - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d - \sum_{s:f_s=i,d_s=d} x_s \right] \\
= & \sum_s \left[U_s(x_s) - x_s q_{f_s}^{d_s} \right] + \sum_d \sum_{(i,j)\in\mathcal{L}} r_{ij}^d (q_i^d - q_j^d).
\end{aligned}
$$

The dual objective function is

$$
\begin{aligned}
D(\vec{q}) &= \max_{x_s \le M_s, \vec{r}^d \ge 0, \sum_d \vec{r}^d \in \mathrm{Co}(\mathcal{R})} L(\vec{x}, \vec{R}, \vec{q}) \\
&= \sum_s B_s(q_{f_s}^{d_s}) + V(\vec{q}),
\end{aligned}
$$

where

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

and

$$V(\vec{q}) = \max_{\vec{r}^d \geq 0, \sum_d \vec{r}^d \in \mathrm{Co}(\mathcal{R})} \sum_d \sum_{(i,j)\in\mathcal{L}} r_{ij}^d (q_i^d - q_j^d). \tag{4}$$

When we maximize the objective in (4), it is easy to verify that, for each link $(i,j) \in \mathcal{L}$, $r_{ij}^d > 0$ only if $q_i^d - q_j^d = \max_{d'} q_i^{d'} - q_j^{d'}$. Hence, letting $r_{ij} = \sum_d r_{ij}^d$ and $\vec{r} = [r_{ij}, (i,j) \in \mathcal{L}]$, we have,

$$V(\vec{q}) = \max_{\vec{r} \geq 0, \vec{r} \in \mathrm{Co}(\mathcal{R})} \sum_{(i,j)\in\mathcal{L}} r_{ij} \max_d (q_i^d - q_j^d). \tag{5}$$

Further, because the objective function in (5) is a linear function of $\vec{r}$, the optimal point must lie in the set $\mathcal{R}$, i.e.,

$$V(\vec{q}) = \max_{\vec{r} \geq 0, \vec{r} \in \mathcal{R}} \sum_{(i,j)\in\mathcal{L}} r_{ij} \max_d (q_i^d - q_j^d). \tag{6}$$

We can then solve $\vec{R} = [\vec{r}^d]$ by picking a $d^*(i,j) = \mathrm{argmax}_d \, q_i^d - q_j^d$ for each link (i,j), and letting $r_{ij}^d = r_{ij}$ if $d = d^*(i,j)$ and $r_{ij}^d = 0$ otherwise.

The dual approach thus results in an elegant decomposition of the original problem. Given the Lagrange multipliers q_i^d, the rate control problem B_s and the scheduling problem $V(\vec{q})$ are decomposed. 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 DRPC policy in [3].

The dual problem is then

$$\min_{\vec{q} \geq 0} D(\vec{q}). \tag{7}$$

The dual objective function $D(\vec{q})$ is convex. We can show that 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 $\vec{x} = [x_s]$ and $\vec{R} = [\vec{r}^d]$ solve (3) and (4), respectively. 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:

$$q_i^d(t+1) = \left\{ q_i^d(t) - h_t \left[\sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d(t) \right.\right.$$
$$\left.\left. - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d(t) - \sum_{s:f_s=i,d_s=d} x_s(t) \right] \right\}^+, \tag{8}$$

where $h_t, t = 1, 2, ...$ is a sequence of positive stepsizes, $\vec{x}(t)$ and $\vec{r}^d(t)$ solve (3) and (4), respectively, with $\vec{q} = \vec{q}(t)$. The following proposition is a consequence of Theorem 2.3 in [16, p26]. The details of the proof is in Appendix A.

Proposition 1 *a) There is no duality gap, i.e., the minimal value of (7) 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 (8) 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.

3.1.1 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 (8). 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 $\vec{r}$ of (6). For each link $(i,j) \in \mathcal{L}$, we then pick the queue $Q_i^{d^*}$ at node i with $q_i^{d^*} - q_j^{d^*} = \max_d q_i^d - q_j^d$, 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

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

Comparing (9) with (8), 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. For details, see Appendix B.

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

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

3.1.2 Fairness and Stability

Fairness among users can be controlled by appropriately choosing the utility functions [17]. For example, utility functions of the form

$$U_s(x_s) = w_s \log x_s \tag{10}$$

correspond to *weighted proportional fairness*, where $w_s, s = 1, ..., S$ are the weights. A general form of utility function is

$$U_s(x_s) = w_s \frac{x_s^{1-\gamma}}{1-\gamma}, \gamma > 0. \tag{11}$$

Maximizing the total utility will correspond to *maximizing weighted throughput* as $\gamma \to 0$, *weighted proportional delay* as $\gamma \to 2$, and *weighted max-min faineness* 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 (8) 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:

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

Proposition 3 *Assume that the rate allocation perfectly solves (1) at each time and the utility function is of the forms (10) or (11) for some $\gamma > 0$. If $[\rho_s]$ resides strictly inside the capacity region Λ, then the Markov process $[n_s]$ is ergodic.*

Proof: The proof follows that of Theorem 1 in [17]. We just need to replace the link capacity constraint there (Inequality (6) in [17]) by the constraint $\vec{x} \in \Lambda$. *Q.E.D.*

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 5, 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.

3.1.3 Virtual Queues

As we have pointed out earlier, the queue size Q_i^d and the implicit cost q_i^d are tightly coupled by iteration (8). Hence, the more congested the network, the higher the implicit costs, and the larger the delay at each node. This undesirable coupling can be broken by using the *virtual queue* concept as in wireline networks [18, 19]. We only need to modify the iteration (8) as:

The Node-Centric Solution with Virtual Queue:

$$q_i^d(t+1) = \left\{ q_i^d(t) - h \left[\sum_{j:(i,j)\in\mathcal{L}} \delta r_{ij}^d(t) - \sum_{j:(j,i)\in\mathcal{L}} \delta r_{ji}^d(t) - \sum_{s:f_s=i,d_s=d} x_s(t) \right] \right\}^+, \tag{12}$$

where δ is a positive factor slightly smaller than 1. The iteration (12) 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 (9). 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 5, we will show 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.

3.2 The Route-Dependent Case

The route-independent formulation in Section 3.1 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. In the worst case, the number of queues at each node can be as many as the number of flows. 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:

$$\max_{[x_{sv}]\geq 0} \quad \sum_s U_s(\sum_v x_{sv}) \tag{13}$$

$$\text{subject to} \quad \sum_{s,v} H_{ij}^{sv} x_{sv} \leq c_{ij} \text{ for all } (i,j) \in \mathcal{L} \tag{14}$$

$$\text{and} \quad [c_{ij}] \in \text{Co}(\mathcal{R}), \sum_v x_{sv} \leq M_s. \tag{15}$$

The route-dependent case can then be treated analogously to the route-independent case. We can associate a Lagrange multiplier q_{ij} for each constraint in (14). The Lagrangian is then:

$$\begin{aligned} & L(\vec{x}, \vec{c}, \vec{q}) \\ = \; & \sum_s U_s(\sum_v x_{sv}) - \sum_{(i,j)\in\mathcal{L}} q_{ij} \left[\sum_{s,v} H_{ij}^{sv} x_{sv} - c_{ij} \right] \\ = \; & \sum_s \left[U_s(x_s) - \sum_v \sum_{(i,j)\in\mathcal{L}} H_{ij}^{sv} q_{ij} x_{sv} \right] \qquad (16) \\ & + \sum_{(i,j)\in\mathcal{L}} q_{ij} c_{ij}, \qquad (17) \end{aligned}$$

where $\sum_{(i,j)\in\mathcal{L}} H_{ij}^{sv} q_{ij}$ can be viewed as the implicit cost of path v of user s. The objective function of the dual problem is:

$$\begin{aligned} D(\vec{q}) &= \max_{x_{sv}\geq 0, \sum_v x_{sv}\leq M_s, \vec{c}\in \mathrm{Co}(\mathcal{R})} L(\vec{x},\vec{c},\vec{q}) \\ &= \sum_s B_s(\vec{q}) + V(\vec{q}), \end{aligned}$$

where

$$\begin{aligned} B_s(\vec{q}) &= \max_{x_{sv}\geq 0, \sum_v x_{sv}\leq M_s} \Bigg[U_s(\sum_v x_{sv}) \\ &\qquad - \sum_v \sum_{(i,j)\in\mathcal{L}} H_{ij}^{sv} q_{ij} x_{sv} \Bigg], \end{aligned} \tag{18}$$

and

$$\begin{aligned} V(\vec{q}) &= \max_{\vec{c}\in \mathrm{Co}(\mathcal{R})} q_{ij}c_{ij} \\ &= \max_{\vec{c}\in\mathcal{R}} q_{ij}c_{ij}. \end{aligned} \tag{19}$$

As we can see, the rate control problem B_s and the scheduling problem V are again decomposed. The subgradient of D is given by,

$$\frac{\partial D}{\partial q_{ij}} = -\left(\sum_{s,v} H_{ij}^{sv} x_{sv} - c_{ij} \right).$$

Hence, the following subgradient method can be used to solve the dual. We will refer to this solution as the link-centric solution, since the Lagrange multiplier is associated with each link.

The Link-Centric Solution:

$$q_{ij}(t+1) = \left\{ q_{ij}(t) + h_t \left[\sum_{s,v} H_{ij}^{sv} x_{sv}(t) - c_{ij}(t) \right] \right\}^+ . \tag{20}$$

where h_t is a sequence of stepsizes, $\vec{x}(t)$ and $\vec{c}(t)$ solve (18) and (19) respectively.

There is one problem with the iterations (18)-(20). Although the iteration (20) 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 (18) 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 (13) and (18) is not strictly concave in $[x_{sv}]$. One can overcome this problem of oscillation by using the Proximal Optimization Algorithm [20]. We can introduce an auxiliary variable y_{sv} for each x_{sv}, and modify the objective function to be

$$\begin{aligned} \max_{[x_{sv}]\geq 0} \quad & \sum_s U_s(\sum_v x_{sv}) - \frac{\nu}{2}\sum_{s,v}(x_{sv} - y_{sv})^2 \\ \text{subject to} \quad & (14) \text{ and } (15), \end{aligned} \tag{21}$$

where $\nu > 0$. For any fixed $\vec{y}$, the objective function becomes strictly concave. We can retain the link-centric solution (20) 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 same techniques as in the proof of 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 (20), where the vector* $[x^*_{sv}]$ *is the unique optimal solution to (21) given* $\vec{y}$.

Finally, an optimal solution to the original problem (13) can be obtained by the following iteration [20]:

- Given $\vec{y} = \vec{y}(k)$, solve (21). 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 (21) to converge before we set $\vec{y}$ to the new value of $\vec{x}$. Our simulation results indicate that these modifications converge satisfactorily in practice. Readers can refer to [21] for some related theoretical analysis.

For the route-dependent case, we can derive analogous results to those in Sections 3.1.1 to 3.1.3 regarding joint scheduling, stability and virtual queue algorithms. Further, the implicit costs provide us with guidelines for finding better routes. We can search the minimum cost path between the source f_s and the destination d_s. If the minimum cost path is cheaper than all current alternate paths of user s, we can add this path to the set of alternate paths. This procedure can be used to gear the network towards optimal routing.

4 Special Cases

The most computationally expensive step of our solution is to solve the scheduling subproblems (6) or (19). They are both of the form

$$\max_{\vec{r}\in\mathcal{R}} \sum_{(i,j)\in\mathcal{L}} w_{ij} r_{ij}, \tag{22}$$

where the weights w_{ij} is $\max_d(q_i^d - q_j^d)$ in the route-independent case (6), and is q_{ij} in the route-dependent case (19). We now briefly discuss two special cases: when the scheduling policy is allowed to incorporate power control, and when it is not.

4.1 With Joint Power Control

We first investigate the following system model that allows the scheduling policy to incorporate power control. This model has been used in [4]. The path loss $G(i,j)$ from any node i to node j is known and fixed. Let $G(i,i) = \infty$. Let P_{ij} be the transmission power used on link $(i,j) \in \mathcal{L}$. Then the signal-to-interference ratio (SIR) for the signal from node i to node j is

$$\mathrm{SIR}_{ij} = \frac{G(i,j)P_{ij}}{N_0 + \sum_{(k,h)\in\mathcal{L},(k,h)\neq(i,j)} G(k,h)P_{k,h}},$$

where N_0 is the ambient noise. We assume that the transmission rate at link (i, j) is proportional to its SIR, i.e.,

$$r_{ij} = W \times \mathrm{SIR}_{i,j}.$$

This assumption is suitable for CDMA systems with a moderate processing gain [4]. Each node i has a power constraint $P_{i,\max}$, i.e., the power allocation must satisfy:

$$\sum_{j:(i,j)\in\mathcal{L}} P_{ij} \le P_{i,\max} \text{ for all } i. \tag{23}$$

The problem (22) can be rewritten as

$$\begin{aligned} V(\vec{q}) &= \max g(\vec{P}) \\ &\quad \text{subject to (23),} \end{aligned} \tag{24}$$

where

$$g(\vec{P}) = \sum_{(i,j)\in\mathcal{L}} \mathrm{SIR}_{ij} w_{ij}. \tag{25}$$

We now show that the maximum point of $g(\cdot)$ in (23) must satisfy the property that at most one $P_{ij} > 0$ for any node i. We show by contradiction: assume that this property does not hold, i.e., there exists an optimal point $\vec{P}^0$ and a node i such that $P^0_{ij} > 0$ and $P^0_{ik} > 0$. Let $P_{ij} = x$ and $P_{ik} = P^0_{ij} + P^0_{ik} - x$. Fix all other power levels and write the function $g(\cdot)$ as a function of x. It is easy to verify that the resulting function $g(x)$ is strictly convex in x, which implies that its value will be strictly larger at either $x = 0$ or $x = P^0_{ij} + P^0_{ik}$, i.e., the function $g(\vec{P})$ will be larger by setting either P_{ij} or P_{ik} to zero. This contradicts with our earlier assumption that $\vec{P}^0$ is optimal. Hence, at most one $P_{ij} > 0$ for any node i.

Further, although the function $g(\vec{P})$ is not convex in $\vec{P}$, it *is* convex in each variable P_{ij}. Combining this property with the fact that only one $P_{ij} > 0$ for any node i, we thus conclude that, in order to maximize g, each node i should either transmit at full power $P_{i,\max}$ or shut off, and each node should only transmit to one other node at any time. These properties substantially simplify the search for the optimal power allocation [4]. Firstly, we only need to search within a finite set of extremal nodal power allocations $\bigotimes_{i=1}^{N}\{0, P_{i,\max}\}$, and compare the optimal values of $g(\cdot)$ given each nodal power assignment. Secondly, given the transmission power at all nodes, each node can decide its own optimal transmission schedule locally. To see this, let P^i be the transmission power of node i. By the above observations, each node i only needs to decide which node j should be the *only* receiver that node i transmits to. Given the transmission power at all other nodes, the SIR from node i to any other node is known. Hence, each node i can locally select the receiver j^* by solving

$$j^* = \operatorname*{argmax}_{j} r_{ij} w_{ij},$$

where r_{ij} is calculated as if the node i uses energy P^i to transmit to node j.

The complexity of the above procedure is $O(2^N)$. When the number of nodes N in the system is large, a clustering heuristics as in [22] can be used to find approximate solutions to (22).

4.2 Without Joint Power Control

We next consider an alternate system model that does not allow the scheduling policy to incorporate power control. This model has been used in earlier studies of rate control (without integrated joint scheduling) in wireless multihop networks [11, 12]. The capacity of each link (i, j) is fixed at c_{ij}. The scheduling policy can only select a link to be active or inactive, but cannot use power control or adaptive coding. The scheduling constraint is that each node can only transmit to or receive from one other node at any time. The problem (22) then becomes a *maximum weighted matching* (MWM) problem where the weights for each link is $c_{ij}w_{ij}$. A polynomial-time algorithm can be found in [23].

The link-centric solution in Section 3.2 combined with the above MWM algorithm can be compared with the rate control algorithm proposed in [11] (although the authors in the latter work address max-min fairness). The rate control algorithm in [11] uses a token-allocation scheme: each node distributes tokens evenly to each flow passing through it. Hence, each node needs to maintain per-flow information. On the other hand, in our link-centric solution, each node only needs to maintain one implicit cost variable for each out-going link. No per-flow information is required. The scheduling algorithm in [11] also uses MWM. However, the scheduling algorithm and the rate control algorithm in [11] are not integrated, i.e., the weights in MWM (equivalent to queue lengths) are decoupled from the token allocation process. On the other hand, in our link-centric solution, the rate control, the MWM and the queue lengths are all tightly coupled by the implicit cost at each link. Finally, our link-centric solution can utility the full capacity of the system, while the algorithm in [11] (and [12]) can utility only 2/3 of the capacity of the system in many cases.

Heuristic solutions to the MWM problem are attractive when computation complexity becomes a main concern. For example, the following procedure can compute an approximate solution to the MWM problem [24]. From all possible links $(i, j) \in \mathcal{L}$, pick the link with the largest $c_{ij}w_{ij}$. Add this link to the schedule. Remove all links that are incident with nodes i or j. Pick the link with the largest $c_{ij}w_{ij}$ from the remaining links, and add to the schedule. Continue until there are no links left. This procedure produces a *maximal* weighted matching. Computing maximal weighted matching is much easier than computing MWM. However, our simulation results indicate that using the above procedure will usually produce a rate allocation that is close to the one obtained from perfect MWM.

5 Simulation Results

In this section, we present simulation results for our joint rate control and scheduling solution. We will first simulate the node-centric solution in Section 3.1 combined with the joint power control model in Section 4.1. We use the "grid" topology in Fig. 1. There are 8 terminals and 1 base station. 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

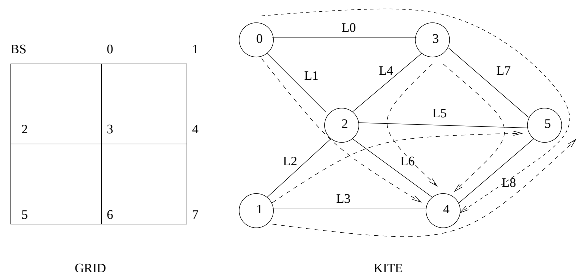

Figure 1: Network Topologies

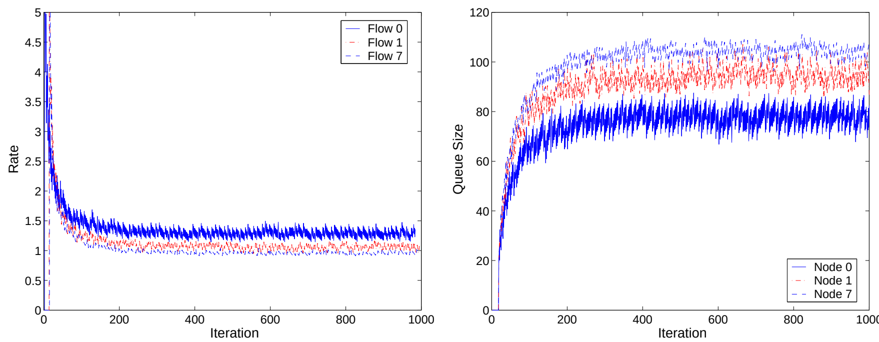

Figure 2: The evolution of the user rates (left) and the queue length (right)

distance from the transmitter to the receiver. The rate of each link is proportional to the SIR, with $r_{ij} = 10.0 \times \mathrm{SIR}_{ij}$. The ambient noise level is $N_0 = 1.0$ unit.

Fig. 2 shows the evolution of the data rates for three users at nodes 0, 1 and 7 respectively, and the evolution of the queue length at these 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 then simulate the virtual queue algorithm in Section 3.1.3. The virtual queue parameter $\delta = 0.95$. The left figure in Fig. 3 shows the evolution of the data rates for the same set of users as before. We find that their values differ only slightly from those in Fig. 2, so does the evolution of the virtual queue length (not shown). However, the real queues are eventually driven close to zero, as shown in the right figure in Fig. 3.

We next simulate the case when there are dynamic arrivals and departures of the users. Users

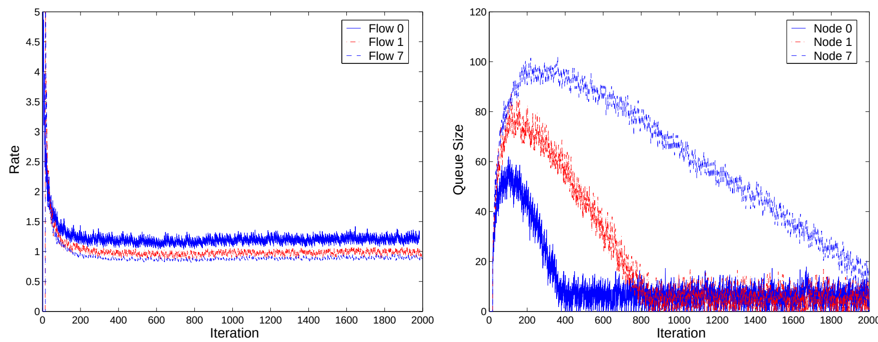

Figure 3: The evolution of the user rates (left) and the evolution of the real queue length (right) when the virtual queue algorithm is used.

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 the left figure in Fig. 4, 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. In the right figure, we plot the total amount of data that are pending to be sent at the user level, when rate control is employed. We can observe that, our rate control algorithm maintains small queue lengths inside the network by preventing an excessive amount of pending data from entering the network.

We also simulate the virtual queue algorithm when there are dynamic arrivals and departures of the users. The solid line in the left figure in Fig. 4 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 can further reduce the queue length inside the network. The peak queue length is reduced from 770 (without virtual queue) to 284 (with virtual queue).

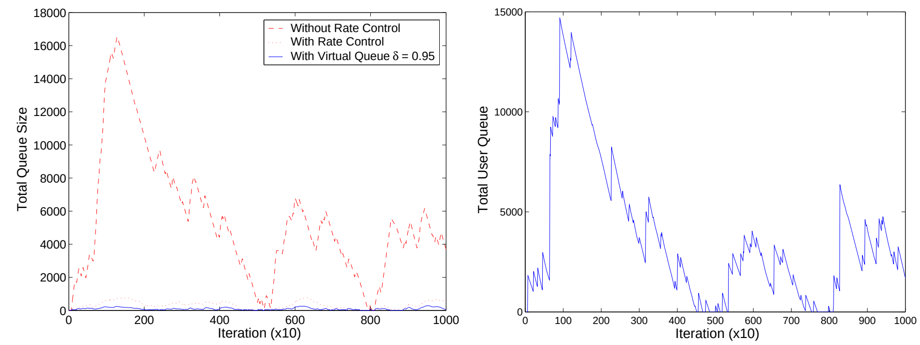

Figure 4: The evolution of the total queue length summed over all nodes (left), and the evolution of the total amount of pending data at the user level (right).

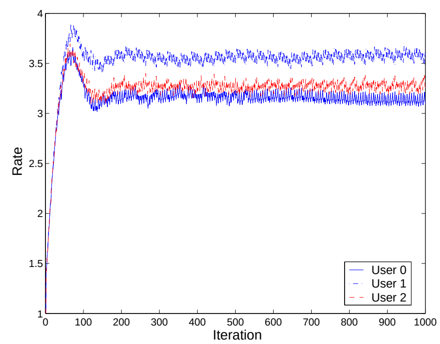

Figure 5: The evolution of each user's total data rate using perfect maximum weighted matching.

We next simulate the link-centric solution combined with the model in Section 4.2 (without joint power control). We use the "kite" topology is in Fig. 1. There are three users. User 0 sends data from node 0 to node 4, user 1 from node 1 to node 5, and user 2 from node 3 to node 4. Each user has two alternate routes as shown by the dashed lines in Fig. 1. The utility function for each user is again $U(x) = \ln x$. The capacity of each link is 10 units. We implement a precise algorithm for solving the maximum weighted matching (MWM) problem and plot in Fig. 5 the evolution of each user's total data rate summed over all alternate paths. The stepsizes are $h = 0.1, \beta = 0.1$. We also implement the simple heuristics in Section 4.2 and plot the same figure again (Fig. 6). As we can use, using a much simpler heuristic for MWM does not change the data rates significantly.

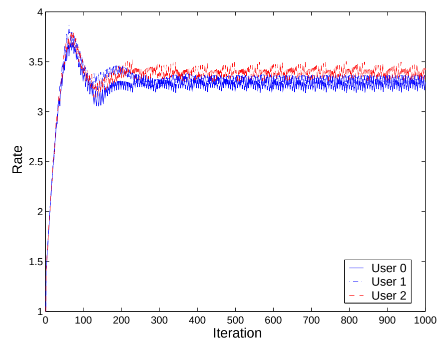

Figure 6: The evolution of each user's total data rate using heuristic matching.

6 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 are 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. Our simulation result using a lower complexity heuristic for the MWM problem suggests that such an approach can be quite attractive. 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.

Appendix

A Proof of Proposition 1

The proof of part a) is quite standard (see, for example, Theorem 3.2.8 in [25, p44]). In fact, let $\vec{x}^*$ denote the optimal solution of the primal problem (1), and let $\vec{R}^* = [\vec{r}^{d,*}]$ denote the corresponding vector of link rates that satisfies (2). It is easy to verify that

$$\max_{x_s \le M_s, \sum_d \vec{r}^d \in \mathrm{Co}(\mathcal{R})} L(\vec{x}, \vec{R}, \vec{q}) \ge \sum_x U_s(x_s^*) \text{ for all } \vec{q} \ge 0 \ .$$

To proof part a), we only need to find a $\vec{q} \ge 0$ such that

$$\max_{x_s \le M_s, \sum_d \vec{r}^d \in \mathrm{Co}(\mathcal{R})} L(\vec{x}, \vec{R}, \vec{q}) = \sum_x U_s(x_s^*).$$

Towards this end, let $\vec{b} = [b_i^d, d \in \mathcal{D}, i \ne d]$ and let

$$G(\vec{b}) = \max_{x_s \le M_s} \quad \sum_s U_s(x_s) \tag{26}$$

$$\text{subject to} \quad r_{ij}^d \ge 0 \text{ for all } (i,j) \in L \text{ and for all } d \in \mathcal{D}$$

$$\sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d - \sum_{s:f_s=i,d_s=d} x_s \ge -b_i^d \quad \text{for all } d \text{ and for all } i \ne d \tag{27}$$

$$[\sum_d r_{ij}^d] \in \mathrm{Co}(\mathcal{R}),$$

Note that the right hand side of the constraint (27) is $-b_i^d$. Then the original problem (1) corresponds to $\vec{b} = 0$ and $G(0) = \sum_s U_s(x_s^*)$. It is easy to show that $G(\vec{b})$ is a concave function of $\vec{b}$. Hence, by Theorem 3.1.8 of [25, p36], there exists a subgradient $\vec{q}_0$ of $G(\vec{b})$ at $\vec{b} = 0$. We now show that $\vec{q}_0$ is the desired dual vector. For any $\vec{b} \ge 0$, by the concavity of $G(\vec{b})$, we have

$$G(\vec{b}) \le G(0) + \vec{q}_0^T \vec{b}.$$

Further, by the definition of the problem (27),

$$G(0) \le G(\vec{b}) \text{ for all } \vec{b} \ge 0.$$

Hence, for any $\vec{b} \geq 0$,

$$G(0) \geq G(\vec{b}) - \vec{q}_0^T \vec{b} \geq G(0) - \vec{q}_0^T \vec{b},$$

and we have,

$$\vec{q}_0^T \vec{b} \geq 0 \text{ for all } \vec{b} \geq 0.$$

Therefore, $\vec{q}_0 \geq 0$. Next, for any $\vec{x}$ such that $x_s \leq M_s$ for all s, and for any $\vec{R}$ such that $\sum_d \vec{r}^d \in \text{Co}(\mathcal{R})$, if we let $\vec{g}(\vec{x}, \vec{R}) = [g_i^d(\vec{x}, \vec{R})]$, where

$$g_i^d(\vec{x}, \vec{R}) = -\left[\sum_{j:(i,j)\in\mathcal{L}} r_{ij}^d - \sum_{j:(j,i)\in\mathcal{L}} r_{ji}^d - \sum_{s:f_s=i,d_s=d} x_s \right],$$

then $(\vec{x}, \vec{R})$ is a feasible point in the problem (27) with $\vec{b} = \vec{g}(\vec{x}, \vec{R})$. Hence, using the concavity of $G(\vec{b})$ again, we have

$$\begin{aligned} \sum_s U_s(x_s) &\leq G(\vec{g}(\vec{x}, \vec{R})) \\ &\leq G(0) + \vec{q}_0^T \vec{g}(\vec{x}, \vec{R}) \\ &= \sum_s U_s(x_s^*) + \vec{q}_0^T \vec{g}(\vec{x}, \vec{R}). \end{aligned} \tag{28}$$

Choosing $\vec{x} = \vec{x}^*, \vec{R} = \vec{R}^*$, we have

$$\sum_s U_s(x_s^*) \leq \sum_s U_s(x_s^*) + \vec{q}_0^T \vec{g}(\vec{x}^*, \vec{R}^*),$$

i.e.,

$$\vec{q}_0^T \vec{g}(\vec{x}^*, \vec{R}^*) \geq 0.$$

However, since $\vec{g}(\vec{x}^*, \vec{R}^*) \leq 0$ and $\vec{q}_0 \geq 0$, we must have

$$\vec{q}_0^T \vec{g}(\vec{x}^*, \vec{R}^*) = 0.$$

Finally, using (28) again, we obtain

$$\begin{aligned} L(\vec{x}, \vec{R}, \vec{q}_0) &= \sum_s U_s(x_s) - \vec{q}_0^T \vec{g}(\vec{x}, \vec{R}) \\ &\leq \sum_s U_s(x_s^*) = \sum_s U_s(x_s^*) - \vec{q}_0^T \vec{g}(\vec{x}^*, \vec{R}^*) \\ &= L(\vec{x}^*, \vec{R}^*, \vec{q}_0). \end{aligned}$$

Hence

$$\max_{x_s \le M_s, \sum_d \vec{r}^d \in \mathrm{Co}(\mathcal{R})} L(\vec{x}, \vec{R}, \vec{q}_0) = \sum_s U_s(x_s^*),$$

i.e., there is no duality gap.

Proof of part b): For any $\vec{q} \in \Phi$, we have

$$\begin{aligned} \sum_s U_s(x_s^*) &= D(\vec{q}) = \max_{x_s \le M_s, \sum_d \vec{r}^d \in \mathrm{Co}(\mathcal{R})} \sum_s U_s(x_s) - \vec{q}^T \vec{g}(\vec{x}, \vec{R}) \\ &\ge \sum_s U_s(x_s^*) - \vec{q}^T \vec{g}(\vec{x}^*, \vec{R}^*) \end{aligned} \tag{29}$$

Hence,

$$\vec{q}^T \vec{g}(\vec{x}^*, \vec{R}^*) \ge 0.$$

However, since $\vec{g}(\vec{x}^*, \vec{R}^*) \le 0$ and $\vec{q} \ge 0$, we must have

$$\vec{q}^T \vec{g}(\vec{x}^*, \vec{R}^*) = 0.$$

Using (29) again, we have,

$$\max_{x_s \le M_s, \sum_d \vec{r}^d \in \mathrm{Co}(\mathcal{R})} \sum_s U_s(x_s) - \vec{q}^T \vec{g}(\vec{x}, \vec{R}) = \sum_s U_s(x_s^*) - \vec{q}^T \vec{g}(\vec{x}*, \vec{R}^*).$$

However, given $\vec{q}$, the point $(\vec{x}, \vec{R})$ that maximizes

$$L(\vec{x}, \vec{R}, \vec{q}) = \sum_s U_s(x_s) - \vec{q}^T \vec{g}(\vec{x}, \vec{R})$$

must satisfy the property that $\vec{x}$ is the optimal solution in (3). Since $U_s(x_s)$ is strictly concave, the optimal solution of (3) is unique. Therefore, it must be equal to $\vec{x}^*$.

Proof of part c): Given any $\vec{q}$, let

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

where $\vec{x} = [x_s]$ and $\vec{R} = [\vec{r}^d, d \in \mathcal{D}]$ solve (3) and (4), respectively. Let

$$\partial D(\vec{q}) = [\frac{\partial D}{\partial q_i^d}, d \in \mathcal{D}, i \ne d].$$

We first verify that $\partial D(\vec{q})$ is a subgradient of $D(\cdot)$ at $\vec{q}$. To see this, for any other vector $\vec{q}_1$, we have,

$$
\begin{aligned}
D(\vec{q}') \geq L(\vec{x}, \vec{R}, \vec{q}_1) &= L(\vec{x}, \vec{R}, \vec{q}) + (\vec{q}_1 - \vec{q})^T \partial D(\vec{q}) \\
&= D(\vec{q}) + (\vec{q}_1 - \vec{q})^T \partial D(\vec{q}).
\end{aligned}
$$

Hence, $\partial D(\vec{q})$ is a subgradient of $D(\cdot)$ at $\vec{q}$. Further, it is easy to verify that $\partial D(\vec{q})$ is bounded. Therefore, part c) follows from Theorem 2.3 of [16, p26].

B Proof of Proposition 2

Let $\delta = 1$. Following the proof of Theorems 2.2 and 2.3 in [16, p25-26], we can show that, there exists a number $h^* > 0$ such that when $h_t = h > 0$ and $h < h^*$, we can find a t_0 such that

$$\rho(\vec{q}(t), \Phi) \leq \delta \text{ for } t \geq t_0.$$

This implies that $\vec{q}(t)$ is bounded. Hence, Q_i^d is bounded for all $i \neq d$.

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] N. Mckeown, V. Anantharan, and J. Walrand, "Achieving 100% Throughput in an Input-Queued Switch," in *Proceedings of IEEE INFOCOM*, vol. 1, San Francisco, CA, March 1996, pp. 296–302.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[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. Qiu and N. B. Shroff, "A New Predictive Flow Control Scheme for Efficient Network Utilization and QoS," in *Proceedings of ACM Sigmetrics*, Boston, MA, 2001.

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

[21] 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.

[22] 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.

[23] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs, New Jersey: Prentice-Hall, 1982.

[24] E. Leonardi, M. Mellia, F. Neri, and M. A. Marsan, "On the Stability of Input-Queued Switches with Speed-Up," *IEEE/ACM Transactions on Networking*, vol. 9, no. 1, pp. 104–118, February 2001.

[25] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*. New York: Springer, 2000.