Joint Rate Control and Scheduling for Delay-Sensitive Traffic in Multihop Wireless Networks

Soroush Jahromizadeh and Veselin Rakočević
School of Engineering and Mathematical Sciences
City University London
UK
Email: {s.jahromizadeh,v.rakocevic}@city.ac.uk

***Abstract*—The prevalent assumption in the current utility maximisation-based rate control and scheduling solutions for multihop wireless networks has been that bounded packet end-to-end delay can be attained by guaranteeing a pre-specified bandwidth. However, in addition to the inherent complexities in determining the exact bandwidth, this approach requires an admission control strategy. In this paper we present a distributed algorithm that is integrated into the previously proposed rate control and scheduling algorithms and, unlike the previous QoS schemes, regulates packet latency without the need for admission control. The proposed algorithm achieves fairness based on the delay constraints, and enables maximum utilisation of network capacity. We analyse the convergence properties of the proposed algorithm, and demonstrate its performance through numerical examples.**

I. Introduction

Future communication networks rely increasingly on multihop wireless networks to maximise network coverage and capacity specially in areas with limited access to communication infrastructure. Communication in multihop wireless networks are performed over multiple hops and paths through distributed coordination at various network layers. There is an emerging demand for supporting real-time and interactive voice and video applications over such networks. However, delivering the stringent delay and data rate requirement of these applications pose challenging design problems considering the autonomous and distributed nature of adhoc networks.

In this paper, we consider the problem of joint rate control and scheduling for delay-sensitive applications over a multihop wireless network. Delay-sensitive applications are characterised by their bounded end-to-end delay requirements. Although these applications can adjust the signal quality according to the available bandwidth, they require bounded packet end-to-end delay in order to maintain interactivity in real-time communication. We seek an optimisation-based approach which enables a systematic decomposition of the problem into protocol layers and distributed implementation while ensuring full utilisation of network capacity.

The use of optimisation-based approaches in solving network design problems including joint rate control and scheduling has developed extensively (see [1] for a survey). Favourable solutions for the joint rate control and scheduling problem are based on dual decomposition as it allows decomposition of the problem into rate control and scheduling subproblem coupled via link prices (see e.g. [2] and [3]). Furthermore, the structure of the resulting scheduling problem has enabled the development of suboptimal but efficient scheduling algorithms suitable for online implementation [4].

Despite the vast body of research in this area, the issue of controlling packets latency has not yet been thoroughly addressed. Specifically, the prevalent assumption has been that end-to-end delay bounds can be achieved by imposing hard bandwidth constraints in the model (see e.g. [6] and [7]). However, the bandwidth necessary to deliver the required packet end-to-end delay is also dependent on the number packets in the system at the equilibrium, which depends on the transient behaviour of the control algorithm, and is generally difficult to determine in advance. Moreover, this approach requires an admission control policy to avoid network overloading.

In [5], we propose two rate control schemes based on primal decomposition that result in bounded packet delay. The first scheme derives lower bounds on bandwidths and includes an admission control policy, while the second scheme is based on dynamic adjustment of applications utility in order to eliminate the admission control phase. The structure of the resulting scheduling problem in both algorithms is, however, different from those already studied in the literature. Hence, the focus of this paper is rather on the schemes that can regulate packet end-to-end delay and are easily integrated into the previously proposed dual-based algorithms. Furthermore, we seek a best-effort scheme capable of delivering the delay bounds without the need for admission control. As we will show later, within an optimisation framework this is indeed possible through dynamic adjustment of applications utility. The proposed algorithm builds upon the ideas from algorithm 2 in [5], but besides its compatibility with the dual-based solutions, it achieves a fairer bandwidth allocation based on the delay constraints.

The rest of this paper is organised as follows. After formulating the optimisation problem in section II, in section III the proposed algorithm for regulating packets end-to-end delay is presented and its convergence properties are analysed. The numerical results are presented in section IV followed by the conclusion and future work in section V.

II. System Model

We consider a multihop wireless network with L links and S sources of delay-sensitive traffic. Each source s can transmit data over I_s alternative paths, at the rate of x_i^s over path $i \in I_s$, and the overall rate of $x_s = \sum_i x_i^s$. However, each source s is constrained by the maximum packet end-to-end delay d_s. Let $I = \sum_s I_s$. The set of links used by paths are given by the $L \times I$ matrix R with elements

$$R_l^i = \begin{cases} 1 & \text{if path } i \text{ uses link } l \\ 0 & \text{otherwise} \end{cases}$$

With a slight abuse of notation, sets are denoted the same as their cardinality. Links data rates c are assumed to be a function of global power assignment p, i.e. $c = u(p)$. Let Π denote the set of feasible power assignments, and let $C = \{u(p), p \in \Pi\}$ denote the set of feasible data rates. Define the capacity region Λ as the largest set of path rates x such that for any $x \in \Lambda$ there exists some scheduling policy that can stabilise the system under the offered load x. Each source s gains a utility $w_s f_s(x_s)$ when it transmits at the total rate of x_s. The functions f_s are assumed to be continuously differentiable, increasing and strictly concave. Furthermore, their curvatures are bounded away from zero, i.e. $-f_s''(x_s) \geq \frac{1}{\bar{\alpha}} > 0$. In this paper, the parameters w are used as a means to provide service differentiation based on the required packet end-to-end delay, and assumed to be primarily unknown.

The joint rate control and scheduling problem is to find path rates x and link data rates c such that

$$\begin{array}{ll} \max_{x \geq 0} & \sum_s w_s f_s(\sum_i x_i^s) \\ \text{s.t.} & x \in \Lambda \end{array} \quad (1)$$

where the capacity region Λ is defined as the set of path rate vectors x for which there exists a link data rate vector c such that

$$Rx \leq c \quad (2)$$

$$c \in \text{Co}(C) \quad (3)$$

As mentioned previously, we assume the parameters w guarantee the required packet end-to-end delay, but are primarily unknown. Note that the objective function in (1) is not strictly concave with respect to (x, c) and as a result may have several global optimal solutions.

III. Proposed Algorithm

We first consider the case where each source has only one available path. The following algorithm is the continuous-time counterpart of the dual-based algorithm presented in [2] for solving (1)

$$\dot{\lambda}_l = \beta(R_l x - c_l)^+_{\lambda_l} \quad \forall l \quad (4)$$

where λ denotes the vector of Lagrange variables associated with constraint (2) or link prices, $\beta > 0$, and $[g(x)]^+_x$ denotes the following

$$[g(x)]^+_x = \begin{cases} g(x) & x > 0 \\ \max(g(x), 0) & x = 0 \end{cases}$$

Source rates x_s and link data rates c are determined by solving the following problems

$$x_s = \arg\max_x (w_s f_s(x_s) - x_s q_s) \quad \forall s \quad (5)$$

$$c = \arg\max_{c \in C} \lambda^T c \quad (6)$$

where $q = R^T \lambda$ is the vector of path prices, and q_s is the price of the path available to source s. Note that since the objective function in (6) is linear, optimal points must lie on C.

Algorithm (4) has an interesting property in that the Lagrange variables λ evolve in proportion to the queueing length at links. This feature is central to our proposed scheme, as we will show later in theorems 2 and 3. First, we show the convergence properties of (4)-(6). Let $\lambda^*(w)$ be the optimal Lagrange variables associated with constraint (2), $x^*(w)$ be the optimal path rates, and $c^*(w)$ be the optimal link data rates for the problem (1) with parameters w. Also, let $q^*(w) = R^T \lambda^*(w)$.

Theorem 1: Given w, algorithms (4)-(6) converge exponentially to an optimal solution of (1).

Proof: Consider the dual problem

$$\min_{\lambda \geq 0} D(\lambda) \quad (7)$$

where

$$D(\lambda) = \max_{x \geq 0, c \geq 0} \sum_s w_s f_s(x_s) - \lambda^T (Rx - c) \quad \text{s.t.} \quad (3) \quad (8)$$

Note that since (1) is not strictly concave in (x, c), (8) is non-differentiable. It can be shown that $c - Rx$, where x and c are given by (5) and (6) respectively, is a subgradient of $D(\lambda)$ [8]. Let $\hat{\lambda}$ denote an equilibrium point of (4). It is easy to see that $\hat{\lambda}$ satisfies the Karush-Kuhn-Tucker (KKT) optimality condition [9], and thus is an optimal solution for (7), i.e. $\hat{\lambda} = \lambda^*$.

Now consider the candidate Lyapunov function

$$\begin{aligned} V(\lambda) &= \frac{1}{2} \|\lambda^* - \lambda\|_2^2 \\ &> 0 \qquad \lambda \neq \lambda^* \end{aligned}$$

We have

$$\begin{aligned} \dot{V}(\lambda) &= \beta(c - Rx)^T (\lambda^* - \lambda) \\ &\leq \beta(D(\lambda^*) - D(\lambda)) \\ &< 0 \qquad \lambda \neq \lambda^* \end{aligned}$$

where the first inequality follows from the definition of subgradient [8] and the second inequality follows from the fact that λ^* is optimal. Consequently, there exists $k(\beta) > 0$ such that

$$\dot{V}(\lambda) \leq -k(\beta) \|\lambda^* - \lambda\|_2^2 \quad (9)$$

Hence, the conditions for exponential stability of the equilibrium point of (4)-(6) hold [11]. ■

Packet latency at the equilibrium is a function of the queueing length $q^*(w)$ as well as path rates $x^*(w)$. Interestingly, the equilibrium values of both variables are linked to the choice of parameters w, as the following sensitivity result shows.

Theorem 2: (Bounds on the sensitivity of $q_s^*(w)$ and $x_s^*(w)$ to the variation of parameters w_s). Suppose f is continuously differentiable, increasing and strictly concave. Furthermore, $-f_s''(x_s) \geq \frac{1}{\bar{\alpha}} > 0$, for all $s \in S$. Then

$$0 < \frac{\partial q_s^*}{\partial w_s} < \frac{q_s^*}{w_s} \tag{10}$$

$$0 < \frac{\partial x_s^*}{\partial w_s} < \frac{q_s^*}{w_s^2}\bar{\alpha} \tag{11}$$

for all $s \in S$.

Proof: The optimality condition for the optimisation problem (1) with parameters w yield

$$\sum_s w_s f_s'(x_s^*(w))(x_s - x_s^*(w)) \leq 0, \quad \forall x \in \Lambda$$

Thus

$$\sum_s w_s f_s'(x_s^*(w))(x_s^*(\tilde{w}) - x_s^*(w)) \leq 0$$

Similarly, for problem (1) with parameters $\tilde{w}$ we obtain

$$\sum_s \tilde{w}_s f_s'(x_s^*(\tilde{w}))(x_s^*(w) - x_s^*(\tilde{w})) \leq 0$$

Define $\tilde{w}$ as

$$\tilde{w}_s = \begin{cases} w_s & s \neq r \\ w_s + dw_r & s = r \end{cases}$$

where $dw_r > 0$. Using (5) and adding both inequalities results in

$$\sum_s \frac{(q_s^*(\tilde{w}) - q_s^*(w))}{dw_r}\frac{(x_s^*(\tilde{w}) - x_s^*(w))}{dw_r} \geq 0$$

Taking the limit $dw_r \to 0$ yields

$$\sum_s \frac{\partial x_s^*}{\partial w_r}\frac{\partial q_s^*}{\partial w_r} \geq 0 \tag{12}$$

Given the strict concavity of f, $\{x_s^*(w)\}$ and $\{x_s^*(\tilde{w})\}$, are the unique maximisers for the problem (1) with parameters w and $\tilde{w}$, respectively. So

$$\sum_s w_s f_s(x_s^*(w)) > \sum_s w_s f_s(x_s^*(\tilde{w}))$$

and

$$\sum_s \tilde{w}_s f_s(x_s^*(\tilde{w})) > \sum_s \tilde{w}_s f_s(x_s^*(w))$$

Adding both inequalities results in

$$\sum_s (\tilde{w}_s - w_s)(f_s(x_s^*(\tilde{w})) - f_s(x_s^*(w))) > 0$$

Except for $s = r$, all the elements in the above summation are zero. Since $\tilde{w}_r - w_r = dw_r > 0$, $f_r(x_r^*(\tilde{w})) > f_r(x_r^*(w))$, and consequently, since f is an increasing function, $x_r^*(\tilde{w}) > x_r^*(w)$. Thus

$$\frac{x_r^*(\tilde{w}) - x_r^*(w)}{dw_r} > 0$$

Taking the limit $dw_r \to 0$ yields the lower bound of (11).

Let $S_d = \{s \in S, s \neq r | x_s^*(\tilde{w}) < x_s^*(w)\}$. Clearly S_d is non-empty (otherwise $\{x_s^*(w)\}$ would not be optimal). Since we assumed $f_s''(x_s) < 0$, for all $s \in S_d$ we have $f_s'(x_s^*(\tilde{w})) > f_s'(x_s^*(w))$ and thus from (5), $q_s^*(\tilde{w}) > q_s^*(w)$. Taking the limit $dw_r \to 0$ we obtain $\frac{\partial x_s^*}{\partial w_r} < 0$ and $\frac{\partial q_s^*}{\partial w_r} > 0$, so

$$\sum_{s \in S_d} \frac{\partial x_s^*}{\partial w_r}\frac{\partial q_s^*}{\partial w_r} < 0$$

Let $S_i = \{s \in S, s \neq r | x_s^*(\tilde{w}) \geq x_s^*(w)\}$. Using a similar argument, $q_s^*(\tilde{w}) \leq q_s^*(w)$, for all $s \in S_i$ and thus we have

$$\sum_{s \in S_i} \frac{\partial x_s^*}{\partial w_r}\frac{\partial q_s^*}{\partial w_r} \leq 0$$

Hence, since $\frac{\partial x_s^*}{\partial w_s} > 0$, in order for (12) to hold we must have $\frac{\partial q_s^*}{\partial w_s} > 0$, for all $s \in S$.

To derive an upper bound for $\frac{\partial q_s^*}{\partial w_s}$, note that from (5) we have

$$q_s^* = w_s f'(x_s^*) \tag{13}$$

Hence

$$\frac{\partial q_s^*}{\partial w_s} = w_s f_s''(x_s^*)\frac{\partial x_s^*}{\partial w_s} + \frac{q_s^*}{w_s}$$

Since $f_s''(x_s^*) < 0$, the first and second terms on the right side of the equation are negative and positive, respectively. The term on the left side of the equation has already been shown to be positive. Therefore

$$0 < -w_s f_s''(x_s^*)\frac{\partial x_s^*}{\partial w_s} < \frac{q_s^*}{w_s}$$

from which the upper bounds in (10) and (11) follow immediately. ■

Motivated by the theorem 2, we propose the following algorithm which determines the utility weight parameters that attain the required packet end-to-end delay at the equilibrium point

$$\dot{w}_s = \alpha[d_s \beta x_s(w) - q_s(w)]_{w_s}^+ \quad \forall s \tag{14}$$

where $\alpha > 0$. The following theorem, presents the conditions for the stability of algorithms (4)-(6) and (18).

Theorem 3: Algorithms (4)-(6) and (18) are exponentially stable, given α is sufficiently small compared to β. Furthermore, at equilibrium, utility weights guarantee the required end-to-end delays, and path and data rates maximise the aggregate utility of the optimisation problem (1) with equilibrium utility weights w^*.

Proof: Since λ in (4) evolves in proportion to the queueing length at links, $q^*(w^*)/\beta$ is equal to paths queueing length at equilibrium and (18) ensures that at equilibrium path end-to-end delays $\{q_s^*(w^*)/\beta x_s^*(w^*)\}$ are equal to the required end-to-end delays d. Now consider the following function

$$V(\lambda) = \frac{1}{2}\|\lambda^*(w) - \lambda(w)\|_2^2 \tag{15}$$

Thus

$$\begin{aligned} \dot{V}(\lambda) &= (\lambda^*(w) - \lambda(w))^T(\dot{\lambda}^*(w) - \dot{\lambda}(w)) \\ &= (\lambda^*(w) - \lambda(w))^T(\frac{\partial \lambda^*}{\partial w}\dot{w} + \beta(c - Rx)) \end{aligned}$$

Notice that algorithm (18) can also be written as

$$\dot{w} = \alpha R^T[\tilde{\lambda}(w) - \lambda(w)]_w^+$$

where $R^T\tilde{\lambda}(w) = \{d_s\beta x_s(w)\}$. Let $d(w) = \{\frac{q_s^*(w)}{\beta x_s^*(w)}\}$ be the equilibrium end-to-end delay given w. Suppose $\|\lambda^*(w) - \lambda(w)\|_2 \leq \epsilon$. When w is near w^*, $d(w)$ is near d, and there exists $\gamma(\epsilon) \leq 1$ such that $\gamma(\epsilon)\|\tilde{\lambda}(w) - \lambda(w)\|_2 \leq \|\lambda^*(w) - \lambda(w)\|_2$ (if $\dot{V}(\lambda) < 0$, λ approaches λ^* and therefore this inequality still holds, as $t \to \infty$). Hence

$$\begin{aligned} \dot{V}(\lambda) &= \alpha(\lambda^*(w) - \lambda(w))^T\frac{\partial \lambda^*}{\partial w}R^T(\tilde{\lambda}(w) - \lambda(w)) \\ &\quad +(\lambda^*(w) - \lambda(w))^T\beta(c - Rx) \\ &\leq (\frac{\alpha}{\gamma(\epsilon)}\rho_{\max}(\frac{\partial \lambda^*}{\partial w}R^T) - k(\beta))\|\lambda^*(w) - \lambda(w)\|_2^2 \end{aligned}$$

where $\rho_{\max}$ denotes the maximum eigenvalue, and the inequalities follow from (9) and the definition of matrix Euclidean norm. Suppose $\rho_{\max}(\frac{\partial \lambda^*}{\partial w}R^T)$ is upperbounded by P. If α satisfy

$$\alpha \leq k(\beta)\frac{\gamma(\epsilon)}{P} \tag{16}$$

then the equilibrium of (4)-(6) and (18) is exponentially stable [11]. ■

When each source can transmit over multiple alternative paths, KKT optimality conditions imply that at optimality, paths with positive flow have the minimum price [12]. Let $\underline{q}_s = \min_i q_i^s$. The source rate algorithm (5) is then modified to

$$x_s = \arg\max_x(w_s f_s(x_s) - x_s\underline{q}_s) \text{ and } x_i^s = 0 \text{ if } q_i^s > \underline{q}_s \quad \forall s \tag{17}$$

Note that (17) only computes the aggregate source rate. In order to determine path rates, any load balancing strategy that satisfies (17) and results in a subgradient of (8), such as those proposed for algorithm 2 in [12], can be used. Subsequently, the multipath version of (18) is given by

$$\dot{w}_s = \alpha[d_s\beta\underline{x}_s(w) - \underline{q}_s(w)]_{w_s}^+ \quad \forall s \tag{18}$$

where $\underline{x}_s = \min_i x_i^s$, $x_i^s \neq 0$. The results from theorems (1), (2) and (3) are readily extended to the multipath case and therefore omitted.

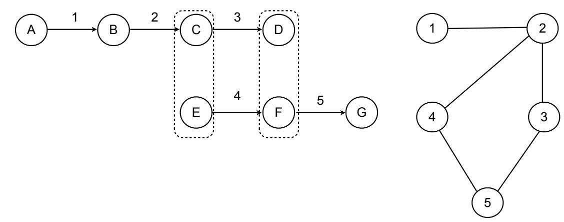

Fig. 1. Example network topology (left) and its contention graph (right)

IV. Numerical Results

To demonstrate how the proposed algorithm regulates packet end-to-end delay, we consider the network in Figure 1 with two flows $A \to D$ and $E \to G$. We assume that links within the interference range of each other (shown in dashed lines) contend for channel access and the successful link transmits at its fixed capacity $c_l^0 = 5$ packets/msec. The contention relations between the links are given by the contention graph in Figure 1. So in this example link power assignment p represent link activation/inactivation, and in the scheduling subproblem (6), the constraint (3) is equivalent to

$$Fc \leq \mathbf{1}, \quad c_l = 0 \quad \text{or} \quad c_l^0, \quad l \in L \tag{19}$$

where F is the contention matrix (see [3] for further details of this model). We assume the utility functions for both flows are logarithmic with initial weights $w_1 = w_2 = 1$. Figures 2 and 3 show the evolution of packet end-to-end delay and source transmission rates when the joint rate control and scheduling algorithms (4)-(6) are simulated. Both end-to-end delay and source transmission rates tend to the equilibrium quickly with appropriate selection of parameter β. The small oscillations are caused by the scheduling process as explained in [3]. The equilibrium end-to-end delay are approximately 154 and 160 milliseconds for source 1 and 2, respectively. We set the target end-to-end delay to $d_1 = d_2 = 140$ msec. Figures 4 and 5 show the evolution of packet end-to-end delay and source transmission rates when algorithms (4)-(6) are simulated jointly with the proposed algorithm (18). As predicted by the theoretical results, the end-to-end delays converge to their target value, although with a slower convergence speed. It will be the subject of our future research to determine the conditions under which the convergence speed of proposed algorithms can be improved.

V. Conclusion

We show that in multihop wireless networks, bounded packet end-to-end delay can be theoretically achieved without the need for admission control, using a network utility maximisation framework. The proposed algorithm is integrated

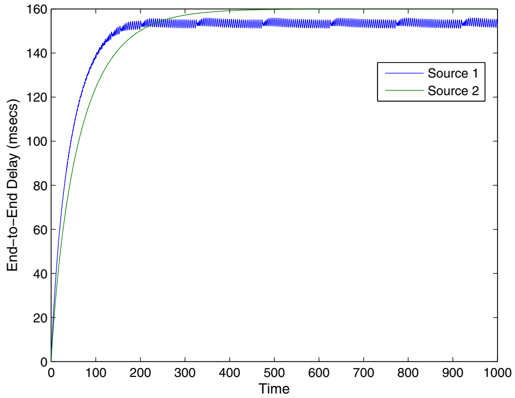

Fig. 2. Evolution of end-to-end delay when algorithms (4)-(6) are simulated

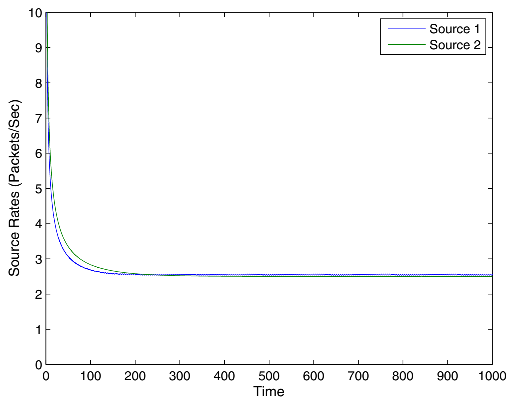

Fig. 3. Evolution of source rates when algorithms (4)-(6) are simulated

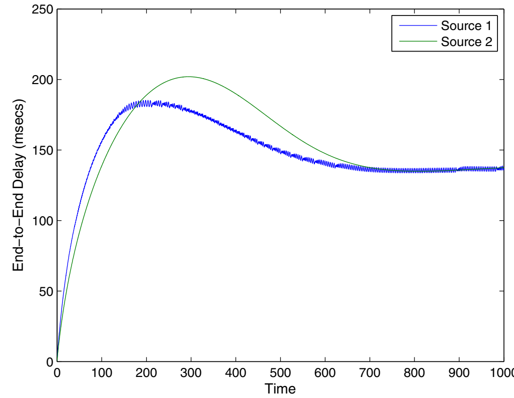

Fig. 4. Evolution of end-to-end delay when algorithms (4)-(6) are simulated jointly with the proposed algorithm (18)

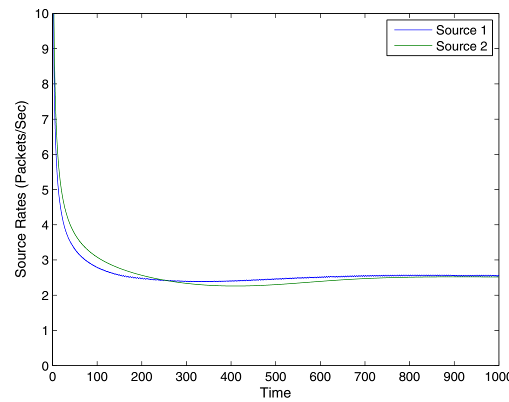

Fig. 5. Evolution of source rates when algorithms (4)-(6) are simulated jointly with the proposed algorithm (18)

into the previously proposed joint rate control and scheduling algorithms, and simply works by dynamically adjusting the applications utility according to the deviation from the delay bounds. The main focus of this paper is on showing the stability of proposed algorithm. Our future work will study the means to improve other performance characteristics essential for practical implementation.

References

[1] M. Chiang, S. H. Low, A. R. Calderbank and J. C. Doyle, Layering as Optimization Decomposition: A Mathematical Theory of Network Architectures, in *Proceedings of the IEEE*, Vol. 95, No. 1, Jan 2007.

[2] X. Lin and N. B. Shroff, A Tutorial on Cross-layer Optimisation in Wireless Networks, *IEEE J. Sel. Areas Commun.*, Vol 24, No. 8, Aug 2006.

[3] L. Chen, S. H. Low and J. C. Doyle, Joint Congestion Control and Media Access Control Design for Ad Hoc Wireless Networks, in *Proc. IEEE INFOCOM*, Miami, FL, Mar 2005.

[4] X. Lin and N. B. Shroff, The Impact of Imperfect Scheduling on Cross-Layer Rate Control in Wireless Networks, *IEEE/ACM Trans. Netw.*, Vol. 14, No. 2, Apr 2006.

[5] S. Jahromizadeh and V. Rakočević, Rate Control for Delay-Sensitive Traffic in Multihop Wireless Networks, in *Proc. 4th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks*, Tenerife, Canary Islands, Oct 2009.

[6] S. Shenker, Fundamental Design Issues for the Future Internet, *IEEE J. Sel. Areas Commun.*, Vol 13, No. 7, Sep 1995.

[7] B. Wydrowski and M. Zukerman, QoS in Best-Effort Networks, *IEEE Communications Magazine*, Dec 2002

[8] D. P. Bertsekas, *Nonlinear Programming*. Belmont, Massachusetts: Athena Scientific, 1995.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[10] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods* Prentice Hall, 1989.

[11] H. K. Khalil, *Nonlinear Systems*, 3rd ed Upper Saddle River, NJ: Prentice Hall, 2002.

[12] W. Wang, M. Palaniswami and S. H. Low, Optimal Flow Control and Routing in Multi-path Networks, *Performance Evaluation*, Vol 52, No. 2-3, Apr 2003.