Distributed Fair Resource Allocation in Cellular Networks in the Presence of Heterogeneous Delays

L. Ying, R. Srikant, A. Eryilmaz and G. E. Dullerud
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
{lying,rsrikant,eryilmaz,dullerud}@uiuc.edu

Abstract

We consider the problem of allocating resources at a base station to many competing flows, where each flow is intended for a different receiver. The channel conditions may be time-varying and different for different receivers. It has been shown in [6] that in a delay-free network, a combination of queue-length-based scheduling at the base station and congestion control at the end users can guarantee queue-length stability and fair resource allocation. In this paper, we extend this result to wireless networks where the congestion information from the base station is received with a feedback delay at the transmitters. The delays can be heterogenous (i.e., different users may have different round-trip delays) and time-varying, but are assumed to be upper-bounded, with possibly very large upper bounds. We will show that the joint congestion control-scheduling algorithm continues to be stable and continues to provide a fair allocation of the network resources.

1 Introduction

We study the problem of fair allocation of resources in the downlink of a cellular wireless network consisting of a single base station and many receivers (see Figure 1). The data destined for each receiver is maintained in a separate buffer. The arrivals to the buffers are determined via a congestion control mechanism, which will be described in detail later. We assume that the time is slotted. The channels between the base station and the receivers are assumed to have random time-varying gains which are independent from one time-slot to the next. The independence assumption can be relaxed easily, but we use it here for ease of exposition. The goal is to allocate the network capacity fairly among the users, in accordance with the needs of the users, while exploiting the time-variations in the channel conditions. We associate a utility function with each user that is a concave, increasing function of the mean service that it receives from the network. In an earlier paper [6], it was shown that a combination of Internet-style congestion control at the end-users and queue-length based scheduling at the base station achieves the goal of fair and stabilizing resource allocation. This result is somewhat surprising since the resource constraints in the case of a wireless network are very different from the linear constraints in the case of the Internet [16]. The relative merits of congestion control-based resource allocation scheme as compared to other resource allocation schemes for cellular networks are discussed in [6]. Several other works in the same context are [17, 15, 11]. However, none of these works explicitly

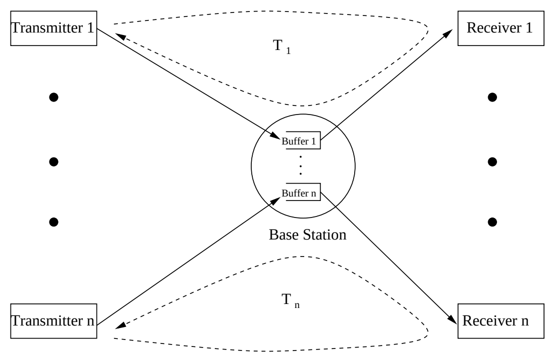

Figure 1. Network with feedback delays. The channel from the base station to the receivers is time-varying.

include the effect of feedback delay in their analysis. In this work, we aim to consider the effect of this essential parameter on the fairness and stability properties of the algorithm presented in [6].

In [6], it is assumed that there are no delays in the transmission of packets from an end-user to the base station and in the transmission of congestion information from the base station back to the end users. In reality, delays exist in both directions: there is a propagation delay τ_i^f from the source i to the base station — we call it the forward delay of user i, and a propagation delay τ_i^b from the base station to the user i — we call it the backward delay. It is well-known that the presence of delays may affect the performance of the network. For example, Internet congestion controllers which are globally stable for the delay-free network may become unstable if the feecback delays are large [16]. So it is interesting to ask whether the conclusions of [6] still hold for wireless networks with heterogeneous delays. We answer this question by showing that for a wireless network with uniformly-bounded delays, which are potentially heterogeneous and time-varying, the algorithm of [6] is stable and can be used to approximate weighted-proportional fair allocation arbitrarily closely. One thing we would like to emphasize is that the results hold not only for networks with small delays, but also for networks with arbitrarily large, but bounded time-varying delays. So even if the end users can only get very old feedback information from the base station, the network is still stable and can eventually reach the fair resource allocation. On the other hand, from the proof, we can also see that when the delays are large, it may take more time for the network to achieve the fair resource allocation. This observation is also supported with simulations.

The rest of paper is organized as follows: in Section 2, we introduce the system model including the congestion controller used by the end users and the scheduler implemented at the base station. In Section 3, we will show that the resulting resource allocation approximates weighted-proportional fairness arbitrarily closely. In Section 4, we further show that the network is stable for any positive K. Section 5 is devoted to simulations that complement the theoretical results. Finally, we will conclude in Section 6.

2 System Model

We consider a cellular network shared by n flows in the downlink and assume that the base station maintains n separate queues, one corresponding to each flow. Each source i has a utility function given by $U_i(\overline{z_i}) = \alpha_i \log \overline{z_i}$, where $\overline{z_i}$ is the average rate at which user i transmits and α_i is a positive weighting factor. The vector $\overline{\mathbf{z}}^*$ that maximizes $\sum_i U_i(\overline{z_i})$ over the set of possible rates yields the *weighted-proportionally fair allocation* [8]. We assume that time is slotted and denote the length of the queue i at the beginning of the time slot t by $x_i[t]$, the number of arrivals to queue i in time slot t by $a_i[t]$, and the amount of service offered to queue i in slot t by $\mu_i[t]$. We assume that each of these parameters can only take non-negative integer values. The evolution of the size of the i^{th} queue is given by

$$x_i[t+1] = x_i[t] + a_i[t] - \mu_i[t] + u_i[t],$$

where $u_i[t]$ is a non-negative quantity which denotes the wasted service in queue i at time slot t and it guarantees that $x_i[t] \geq 0$. We also assume that the channel between the base station and the receivers can be in one of J states in a given slot. We use $s[t] \in \{1, \cdots, J\}$ to denote the state in time slot t. The channel state is assumed to be fixed within a time slot, but may vary from one slot to another, thus capturing the time-varying characteristics of a fading environment. Corresponding to each channel state, say j, is an achievable rate region, C_j, that is defined to be the convex hull of the feasible rate vectors, $\eta[t] := (\eta_1[t], \ldots, \eta_n[t])$, that can be offered to the queues. We assume that each C_j is a bounded region and use the symbol $\hat{\eta} < \infty$ to denote the upper bound on the achievable rates for all channel states. The channel state process is assumed to be independent and identically distributed in each time slot, but we do not require that the statistics be known at the base station. Furthermore, we define the mean achievable rate region as

$$\bar{C} := \left\{ \eta : \eta = \sum_{j=1}^{J} \pi_j^{ch} \eta^{(j)}, \eta^{(j)} \in C_j \right\},$$

where π_j^{ch} stands for the stationary probability of the channel state process being in state j. The scheduler will use the following algorithm:

SCHEDULER: Given the current queue-length $\mathbf{x}[t] := (x_1[t], \ldots, x_n[t])$ and current channel state $s[t]$, the scheduler chooses a service rate vector $\mu[t] := (\mu_1[t], \ldots, \mu_n[t]) \in C_{s[t]}$ that satisfies

$$\mu[t] \in \arg \max_{\eta \in C_{s[t]}} \sum_{i=1}^{n} x_i[t] \eta_i.$$

◇

This scheduling rule was introduced in the context of fixed arrival rates (i.e., where the arrival rates are not adjusted by a congestion controller) in [18], where it was also shown that it is a stabilizing rule, i.e., the mean queue-lengths are upper-bounded. This result was extended in many different directions in [2, 14, 13, 7, 3, 9, 5, 12].

In our model, the packet arrival rate into each queue is assumed to be controlled according to the well-known dual controller that has been studied extensively in the context of Internet congestion control [8, 10, 19, 16]. In the context of Internet congestion control, a dual controller chooses the transmission rate z_i such that

$$\frac{\alpha_i K}{x_i} = z_i$$

for some constant $K > 0$. Next, we describe the operation of our congestion controller followed by some assumptions.

CONGESTION CONTROLLER In our wireless network model, accounting for the forward delay τ_i^f and the backward delay τ_i^b, the mean of the number of arrivals into queue i at time slot t is:

$$E[a_i[t]|x_i[t-T_i]] = \min\left\{\frac{\alpha_i K}{x_i[t-T_i]}, M\right\}, \tag{1}$$

where $T_i = \tau_i^f + \tau_i^b$ and $M > 2\hat{\eta}$ is a positive constant which ensures that the arrival rate into the queue is upper-bounded when the queue-length is small. We assume $a_i[t]$ is independent across time slots and that the variance of $a_i[t]$ given $x_i[t-T_i]$ is bounded:

$$E\left[a_i^2[t]|x_i[t-T_i]\right] \leq V < \infty \ \text{ for all } x_i[t-T_i]. \tag{2}$$

Furthermore, we assume that there exist positive numbers $\theta, A > T\hat{\eta}$ and $h > 2$ such that for any $N > A$,

$$p(\sum_{j=1}^{T} a_i[t-j] = N) < \frac{\theta}{N^h} \text{ for all } i. \tag{3}$$

$\diamond$

We will later show that K plays a critical role in the context of a cellular network: it is important to choose K to be large enough to ensure weighted proportionally-fair resource allocation. The parameter K does not play such a role in the context of wireline networks.

In summary, the combined Scheduler-Congestion Controller Algorithm can be defined as follows:

$$x_i[t+1] = x_i[t] + a_i[t] - \mu_i[t] + u_i[t] \tag{4}$$

$$\mu[t] \in \arg\max_{\mu \in C_{s[t]}} \sum_{i=1}^{n} x_i[t]\eta_i, \tag{5}$$

where $a_i[t]$ is a random variable satisfying the conditions in (1), (2) and (3).

We now present the following theorem, which will be useful later. This theorem is similar to Proposition 1 of [6].

Theorem 1 *There exists a unique pair of vectors* $(\mathbf{x}^\star, \mu^\star)$ *which satisfy following conditions*

- $\mu^\star \in \arg\max_{\eta \in C} \sum_{i=1}^{n} x_i^\star \eta_i$; $x_i^\star = \frac{\alpha_i K}{\mu_i^\star}$ *for all i, and*
- $\mu^\star$ *is the optimal solution to* $\max_{\mu \in C} \sum_{i=1}^{n} K\alpha_i \log \mu_i$.

$\square$

From the above theorem, we can see that $\mu^\star$ is weighted-proportionally fair. For the stochastic model, we will show that $\mathbf{x}[t]$ converges to $\mathbf{x}^\star$, defined in Theorem 1, in a probabilistic sense. This then implies that the network reaches a fair operating point.

In the rest of the paper, we will first show that fair resource allocation can be achieved when the linear gain, K, used in the congestion controller goes to infinity. The proof is quite complicated, but the idea behind the proof is relatively simple. From [6], it has been known that the result holds for a delay-free network. In that paper the authors define a quadratic Lyapunov function, which we denote as $W_0(t)$, and prove that for large K,

$$E[W_0(t+1) - W_0(t)|W_0(t)] \leq G(K) < 0,$$

where $G(\cdot)$ is a properly defined function. Here, we will use a similar Lyapunov function that incorporates the feedback delay. Let us call this Lyapunov function $W(t)$. We show that

$$E[W(t+1) - W(t)|W(t)] \leq G(K) + H(K),$$

where $H(K)$ is an additional term due to the delays. We further prove that $H(K)/G(K) \to 0$ as $K \to \infty$, which implies that for large K,

$$E[W(t+1) - W(t)|W(t)] \leq G(K)/2 < 0.$$

Thus, following the arguments of [6], we can show that the fair resource allocation can be achieved for large K.

Secondly, we will show the network is stable not only for large K, but for any positive K. The proof uses a standard Foster- Lyapunov argument. It will be shown that when the sum of queue-lengths is much larger than the equilibrium value, the Foster Lyapunov function will have a negative drift.

3 Positive recurrence and weighted-proportional fairness

Notice that the wireless network we introduced in Section 2 is a discrete-time stochastic system, so we have to show the stochastic stability of this system, i.e., we have to show that a Markov chain describing the evolution of this system is positive recurrent. We define $\mathbf{y}[t] = (\mathbf{x}[t], \ldots, \mathbf{x}[t-T])$, where $T = \max_i T_i$. It is easy to see that the process $\{\mathbf{y}[t]\}_{t\geq 0}$ is a Markov chain because $a_i[t]$ depends only on $x_i[t-T_i]$, so $x_i[t+1]$ and $\mathbf{y}[t+1]$ are determined by $\mathbf{y}[t]$.

Foster's Criterion will be used to prove the main results in this paper. The basic idea is to find a Foster-Lyapunov function $W(\mathbf{y}[t])$ and a finite set S such that if $\mathbf{y}[t] \notin S$, then the drift $E[W(\mathbf{y}[t+1]) - W(\mathbf{y}[t])|\mathbf{y}[t]] < -\varepsilon$. The Foster-Lyapunov function that we will use in this paper is given by

$$W(\mathbf{y}[t]) = \frac{1}{2}\sum_i^n (x_i[t] - x_i^\star)^2 .$$

We will first show that $\{\mathbf{y}[t]\}$ is positive recurrent, which implies that it has a stationary distribution. Then, we can prove that as $K \to \infty$, the i^{th} scaled queue-length, (x_i/K), tends to concentrate around the point $x_i^\star/K$. Thus,

$$E[a_i] = \frac{\alpha_i}{\frac{x_i[t]}{K}} \approx \frac{\alpha_i}{\frac{x_i^\star}{K}}.$$

Therefore, we can approximate the weighted-proportional fair allocation arbitrarily closely by choosing a large K. For a positive constant c and $\sigma < 1$, define

$$S = \left\{\mathbf{y}[t] : \|\mathbf{x}[t] - \mathbf{x}^\star\| \leq c\sqrt{K}\right\}. \tag{6}$$

Note that $\|\mathbf{x}[t] - \mathbf{x}^\star\| \leq c\sqrt{K}$ implies that

$$\sum_i x_i[t-s] \leq \sum_i x_i^\star + nc\sqrt{K} + nT\hat{\eta}, \quad \text{for all} \ \ 0 \leq s \leq T.$$

Thus, S is a finite set. Define

$$E[\Delta W_t(\mathbf{y})] := E\left[W(\mathbf{y}[t+1]) - W(\mathbf{y}[t])|\mathbf{y}[t]\right],$$

the following theorem proves the positive recurrence of $\mathbf{y}[t]$.

Theorem 2 *There exist positive numbers c, ζ and δ^*, such that for large K*

$$E[\Delta W_t(\mathbf{y})] \leq -\frac{\delta^*}{\sqrt{K}}\|\mathbf{x} - \mathbf{x}^\star\| I_{\mathbf{y}\in S^c} + \zeta I_{\mathbf{y}\in S}, \tag{7}$$

where S is defined in (6). And the Markov chain $\{\mathbf{y}[t]\}$ is positive recurrent.

Proof: First, it is easy to see that if $\mathbf{y}[t] \in S$, there exists $0 < \zeta < \infty$ such that $E[\Delta W_t(\mathbf{y})] < \zeta$. Now, consider $\mathbf{y}[t] \notin S$, define the event χ_0^t such that

$$\chi_0^t := \left\{\max_i \sum_{j=1}^T a_i[t-j] \leq A\right\},$$

and events χ_l^t for $l = 1, 2, \ldots$ such that

$$\chi_l^t := \left\{\max_i \sum_{j=1}^T a_i[t-j] = A + l\right\}.$$

Then we can rewrite $E[\Delta W_t(\mathbf{y})]$ as follows:

$$E[\Delta W_t(\mathbf{y})] = \sum_{l=0}^\infty E[\Delta W_t(\mathbf{y})|\chi_l^t] p(\chi_l^t).$$

For convenience, we let $\{y\}^M$ denote $\min\{y, M\}$. Then, along the lines of the proof of Theorem 1 of [6], it can be shown that there exists $B_d < \infty$, which is independent of K and $\mathbf{x}[t]$, such that

$$E[\Delta W_t(\mathbf{y})] \leq \sum_{i=1}^n \Delta x_i[t] \left(\left\{\frac{\alpha_i K}{x_i[t-T_i]}\right\}^M - \mu_i^\star\right) + B_d \tag{8}$$

$$= \sum_{i=1}^n \Delta x_i[t] \left(\left\{\frac{\alpha_i K}{x_i[t-T_i]}\right\}^M - \left\{\frac{\alpha_i K}{x_i[t]}\right\}^M\right) \tag{9}$$

$$+ \sum_{i=1}^n \Delta x_i[t] \left(\left\{\frac{\alpha_i K}{x_i[t]}\right\}^M - \mu_i^\star\right) + B_d, \tag{10}$$

where $\Delta x_i[t] = x_i[t] - x_i^\star$. The only difference between this analysis and the delay-free case is the additional term (9), which we will also refer to as $H(K)$. It has been shown in [6] that $G(K) := (10)$ is negative when $\|\mathbf{x}\|$ is large. Here, we do not know the sign of (9), but we can show that when K increases, $|(9)|$ increases much slower than $|(10)|$. So $|H(K)|/G(K) \to 0$ as $K \to \infty$, and when K is large enough, we will have $(9) + (10) = G(K) + H(K) < G(K)/2 < 0$, which implies $E[\Delta W_t(\mathbf{y})] < 0$.

To show this, we will show the following three facts. First, there exists a $\delta_d > 0$ such that for all events χ_l^t,

$$G(K) \leq -\frac{\delta_d}{\sqrt{K}}\|\mathbf{x}[t] - \mathbf{x}^\star\|. \tag{11}$$

Second, when χ_0^t happens, there exists a $\delta_0 > 0$ such that

$$p(\chi_0^t)|H(K)| \leq p(\chi_0^t)\frac{\delta_0}{K}\|\mathbf{x}[t] - \mathbf{x}^\star\|. \tag{12}$$

And third, there exists a $\delta_1 > 0$ such that

$$\sum_{l=1}^{\infty} p(\chi_l^t)\,|H(K)| \leq \frac{\delta_1}{K}\|\mathbf{x}[t]-\mathbf{x}^\star\|. \tag{13}$$

If all three inequalities — (11), (12) and (13) — hold and $\xi > 1-\sigma$, we will have

$$\begin{aligned}
E[\Delta W_t(\mathbf{y})] &\leq G(K)+p(\chi_0^t)H(K)+\sum_{l=1}^{\infty} p(\chi_l^t)H(K) \\
&\leq -\left(\frac{\delta_d}{\sqrt{K}}-\frac{\delta_0 p(\chi_0^t)+\delta_1}{K}\right)\|\mathbf{x}[t]-\mathbf{x}^\star\| \\
&\leq -\left(\frac{\delta_d}{\sqrt{K}}-\frac{\delta_0+\delta_1}{K}\right)\|\mathbf{x}[t]-\mathbf{x}^\star\|.
\end{aligned}$$

Thus, if we have $K > 4(\delta_0+\delta_1)^2/\delta_d^2$, then we also have

$$\frac{\delta_d}{2\sqrt{K}}-\frac{\delta_0+\delta_1}{K} > 0,$$

which implies

$$E[\Delta W_t(\mathbf{y})] \leq -\left(\frac{\delta^*}{\sqrt{K}}\right)\|\mathbf{x}[t]-\mathbf{x}^\star\|$$

and the theorem is proved with $\delta^* := \delta_d/2$.

Now we prove (11), (12) and (13). We will first show (11). The proof is similar to the proof of Theorem 1 of [6]. Notice that we have

$$(x_i[t]-x_i^\star)\left(\left\{\frac{\alpha_i K}{x_i[t]}\right\}^M-\mu_i^\star\right) \leq 0 \text{ for all } i.$$

Letting $i_0 = \arg\max_i |x_i[t]-x_i^\star|$, we have

$$(10) \leq -|x_{i_0}[t]-x_{i_0}^\star|\left|\left\{\frac{\alpha_{i_0}K}{x_{i_0}[t]}\right\}^M-\mu_{i_0}^\star\right|+B_d.$$

Now, if $\left\{\frac{\alpha_i K}{x_{i_0}[t]}\right\}^M = M$, then from the definition of M, we have $M > 2\hat{\eta}$, so

$$\left|\left\{\frac{\alpha_{i_0}K}{x_{i_0}[t]}\right\}^M-\mu_{i_0}^\star\right| = M-\mu_{i_0}^\star > \hat{\eta}.$$

Otherwise if $\left\{\frac{\alpha_i K}{x_{i_0}[t]}\right\}^M < M$, then

$$\left\{\frac{\alpha_i K}{x_{i_0}[t]}\right\}^M = \frac{\alpha_i K}{x_{i_0}[t]}$$

and

$$\left|\left\{\frac{\alpha_{i_0}K}{x_{i_0}[t]}\right\}^M-\mu_{i_0}^\star\right| = \mu_{i_0}^\star\left|\frac{x_{i_0}^\star}{x_{i_0}[t]}-1\right|.$$

Furthermore, because

$$x_{i_0}[t] = \begin{cases} x_{i_0}^\star-|x_{i_0}[t]-x_{i_0}^\star| \geq 0, & \text{if } x_{i_0}[t]-x_{i_0}^\star \leq 0; \\ x_{i_0}^\star+|x_{i_0}[t]-x_{i_0}^\star| \geq 0, & \text{if } x_{i_0}[t]-x_{i_0}^\star \geq 0, \end{cases}$$

and

$$\left|\frac{x_{i_0}^\star}{x_{i_0}^\star-|x_{i_0}[t]-x_{i_0}^\star|}-1\right| \geq \left|\frac{x_{i_0}^\star}{x_{i_0}^\star+|x_{i_0}[t]-x_{i_0}^\star|}-1\right|,$$

we have

$$\left|\left\{\frac{\alpha_{i_0}K}{x_{i_0}[t]}\right\}^M-\mu_{i_0}^\star\right| \geq \mu_{i_0}^\star\left|\frac{x_{i_0}^\star}{x_{i_0}^\star+|x_{i_0}[t]-x_{i_0}^\star|}-1\right|.$$

Now using the fact that when $\mathbf{x}[t] \notin S$, we have

$$c\sqrt{K} \leq \|\mathbf{x}[t]-\mathbf{x}^\star\| \leq \sqrt{n}\left|x_{i_0}[t]-x_{i_0}^\star\right|,$$

we can write

$$\begin{aligned}
\left|\left\{\frac{\alpha_{i_0}K}{x_{i_0}[t]}\right\}^M-\mu_{i_0}^\star\right| &\geq \mu_{i_0}^\star\left|\frac{x_{i_0}^\star}{x_{i_0}^\star+\frac{c}{\sqrt{n}}\sqrt{K}}-1\right| \\
&= \mu_{i_0}^\star\frac{\frac{c}{\sqrt{n}}\sqrt{K}}{\frac{\alpha_{i_0}}{\mu_{i_0}^\star}K+\frac{c}{\sqrt{n}}\sqrt{K}}.
\end{aligned}$$

It is easy to see that as $K \to \infty$, we have

$$\mu_{i_0}^\star\frac{\frac{c}{\sqrt{n}}\sqrt{K}}{\frac{\alpha_i}{\mu_{i_0}^\star}K+\frac{c}{\sqrt{n}}\sqrt{K}} \to 0$$

Hence, for large K, we will have $\mu_{i_0}^\star\frac{\frac{c}{\sqrt{n}}\sqrt{K}}{\frac{\alpha_i}{\mu_{i_0}^\star}K+c\sqrt{K}} < \hat{\eta}$, which implies for large K,

$$G(K) \leq -|x_{i_0}[t]-x_{i_0}^\star|\left(\mu_{i_0}^\star\frac{\frac{c}{\sqrt{n}}\sqrt{K}}{\frac{\alpha_i}{\mu_{i_0}^\star}K+\frac{c}{\sqrt{n}}\sqrt{K}}+\frac{B_d}{|x_{i_0}[t]-x_{i_0}^\star|}\right).$$

So by choosing a sufficiently large c that is independent of K, we can find a positive constant $\hat{\delta}$ and $\hat{K}$ such that for any $K \geq \hat{K}$

$$G(K) \leq -\frac{\hat{\delta}}{\sqrt{K}}|x_{i_0}[t]-x_{i_0}^\star| \leq -\frac{\delta_d}{\sqrt{K}}\|\mathbf{x}[t]-\mathbf{x}^\star\|,$$

where $\delta_d = \hat{\delta}/\sqrt{n}$.

Next, we consider (12). It is the case where the arrivals are upper-bounded by A. We will show that as K increases,

$$\left| \left\{ \frac{\alpha_i K}{x_i[t-T_i]} \right\}^M - \left\{ \frac{\alpha_i K}{x_i[t]} \right\}^M \right|$$

will decrease.

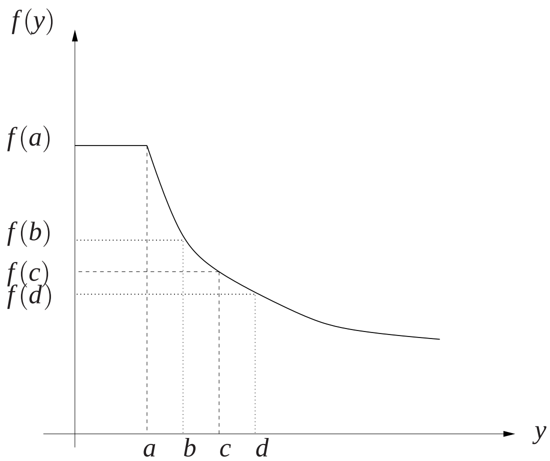

Figure 2. $f(y) = \left\{ \frac{\alpha_i K}{y} \right\}^M$

We will use Figure 2 to prove our result. We know that $\alpha_i K/y$ is convex in y, so

$$f(y) = \left\{ \frac{\alpha_i K}{y} \right\}^M$$

is as shown in Figure 2, where $f(a) = M$ and $a = \alpha_i K/M$. Now suppose $c-a = d-b = A$ and $f(a) = M$. Then for any b satisfying $a < b$, we can see from the figure that $f(a) - f(c) > f(b) - f(d)$. This means that

$$\max_{\{x_1,x_2:x_2-x_1\le A\}} \{f(x_1) - f(x_2)\} = f(a) - f(c).$$

Furthermore, by assumption we have $A \geq T\hat{\eta}$, hence $|x_i[t] - x_i[t-T]| \leq A$ and

$$\begin{aligned} |f(x_i[t-T_i]) - f(x_i[t])| &\leq f(a) - f(c) = M - \frac{\alpha_i K}{\frac{\alpha_i K}{M} + A} \\ &= \frac{M^2 A}{\alpha_i K + MA} < \frac{M^2 A}{\alpha_i} \frac{1}{K}. \end{aligned}$$

This allows us to conclude that

$$\left| \left\{ \frac{\alpha_i K}{x_i[t-T_i]} \right\}^M - \left\{ \frac{\alpha_i K}{x_i[t]} \right\}^M \right| < \frac{M^2 A}{\alpha_i} \frac{1}{K}. \tag{14}$$

Letting $\alpha_{\min} := \min_i \alpha_i$, we define $\delta_0 = M^2 An/\alpha_{\min}$, which satisfies

$$|(9)| \quad \leq \quad \frac{\delta_0}{K} \|\mathbf{x}[t] - \mathbf{x}^\star\|.$$

Finally, we consider the complement of χ_0^t, which is denoted as $\chi_0^{t\,c}$, and derive inequality (13). In this case the arrivals are not upper-bounded and can be arbitrarily large. But from assumption (3), the probability of this event is very small. So we can still obtain an upper bound for $\sum_{i=1}^n (x_i[t] - x_i^\star) \left| \left\{ \frac{\alpha_i K}{x_i[t-T_i]} \right\}^M - \left\{ \frac{\alpha_i K}{x_i[t]} \right\}^M \right|$. Suppose χ_l^t occurs ($l \geq 1$), similar to the inequality (14), we can get

$$\left| \left\{ \frac{\alpha_i K}{x_i[t-T_i]} \right\}^M - \left\{ \frac{\alpha_i K}{x_i[t]} \right\}^M \right| < \frac{M^2(A+l)}{\alpha_i} \frac{1}{K}$$

and

$$\sum_{l=1}^{\infty} p(\chi_l^t)\, |(9)| \leq \|\mathbf{x}[t] - \mathbf{x}^\star\| \sum_{l=1}^{\infty} \frac{nM^2(A+l)}{\alpha_{\min}} \frac{1}{K} p(\chi_l^t).$$

Now we use assumption (3), which yields

$$\begin{aligned} \sum_{l=1}^{\infty} \frac{nM^2(A+l)}{\alpha_{\min}} \frac{1}{K} p(\chi_l^t) &\leq \sum_{l=1}^{\infty} \frac{nM^2(A+l)}{\alpha_{\min}} \frac{1}{K} \frac{\theta}{(A+l)^h} \\ &\leq \frac{nM^2\theta(h-2)}{\alpha_{\min} A^{h-2}} \frac{1}{K} \\ &= \frac{\delta_1}{K}, \end{aligned}$$

where $\delta_1 := \frac{nM^2\theta(h-2)}{\alpha_{\min}A^{h-2}}$.

Thus we have proved that inequalities (11), (12) and (13) hold. Also, it is easy to see that $\sigma < \xi$. Thus, when $K > 4(\delta_0+\delta_1)^2/\delta_d^2$, we have (7) with $\delta^* := \delta_d/2$. Furthermore, by invoking Foster's Criterion [4, Proposition 5.3], we have that the Markov chain $\{\mathbf{y}[t]\}$ is positive recurrent. □

Following an argument similar to that of Theorem 2 in [6], we can show the following result.

Theorem 3 *There exists a positive constant $\bar{c}$, that depends on the mean achievable rate region, the algorithm parameters $\{\alpha_i\}$, and the first and second moments of the channel and arrival processes, such that*

$$E\left[\|\mathbf{x}[\infty] - \mathbf{x}^\star\|\right] \leq \bar{c}\sqrt{K} \text{ for large } K,$$

where $\mathbf{x}[\infty]$ is an informal notation for the steady state of $\mathbf{x}$ and $\|\cdot\|$ denotes the Euclidean distance in the $\Re^n$. □

Now using the Markov inequality, Theorem 3 yields

$$P\left(\frac{1}{K}|x_i[\infty]-x_i^\star|>\varepsilon\right)\le\frac{\bar{c}}{\varepsilon\sqrt{K}},$$

which implies that $\frac{\mathbf{x}[\infty]}{K}\approx\frac{\mathbf{x}^\star}{K}$ for large K and

$$E[a_i]=\frac{\alpha_i}{\frac{x_i}{K}}\approx\frac{\alpha_i}{\frac{x_i^\star}{K}}.$$

So $\mu[t]$ converges to $\mu^\star$ for large t in a probabilistic sense and the network is weighted proportionally fair according to Theorem 1.

These results allow us to conclude that even in the presence of delays, the network will achieve weighted-proportional fairness.

4 Stability for Heterogeneous Delays

In Section 3, we have seen that when K is large the network is stable. This raises the concern that the network may not be stable for small K, creating the risk of operating the network in an unstable regime. But actually condition (7) is much stronger than what is necessary to prove the stability. In this section, we will show that for any $K>0$, the Markov chain is positive recurrent. Define

$$S_{\bar{X}} := \left\{\mathbf{y}[t]:\sum_i x_i[t]\le\bar{X}\right\}, \tag{15}$$

$$\chi_0^t = \left\{\max_i\sum_{l=1}^{T}a_i[t+l]\le A_s,\ \text{for all } i\right\}. \tag{16}$$

Clearly, $S_{\bar{X}}$ is a finite set. From (1), we know that the mean of the arrivals is upper-bounded. Hence for any $\varepsilon>0$, there exists an N such that for all $A_s\ge N$, we have $P(\chi_0^t)>1-\varepsilon$.

Theorem 4 *For any $K>0$, there exists positive numbers ζ, $\bar{X}$ and δ such that*

$$E[\Delta W_t(\mathbf{y})]\le-\delta\sum_{i=1}^{n}x_i[t]I_{\mathbf{y}\in S_{\bar{X}}^c}+\zeta I_{\mathbf{y}\in S_{\bar{X}}},$$

where $S_{\bar{X}}$ is defined in (15). And the Markov chain $\{\mathbf{y}[t]\}$ is positive recurrent.

Proof: In this proof, we assume that K is fixed. First the finiteness of $S_{\bar{X}}$ implies that if $\mathbf{y}[t]\in S_{\bar{X}}$, there exists $0<\zeta<\infty$ such that $E[\Delta W_t(\mathbf{y})]<\zeta$. Now, consider $\mathbf{y}[t]\notin S_{\bar{X}}$, which means that

$$\sum_i x_i[t]>\bar{X}.$$

We can rewrite $E[\Delta W_t(\mathbf{y})]$ as

$$E[\Delta W_t(\mathbf{y})|\chi_0^t]p(\chi_0)+E[\Delta W_t(\mathbf{y})|\chi_0^{t\,c}](1-p(\chi_0)).$$

Let $P(\chi_0^t)=p$. In the following, we will show that

$$E[\Delta W_t(\mathbf{y})|\chi_0^t]p\le\hat{C}-pK\frac{\alpha_{\min}}{x_{\max}^\star}\sum_{i=1}^{n}x_i[t]. \tag{17}$$

where $\hat{C}$ is a positive constant which depends on A_s, $\alpha_{\min}=\min_i\alpha_i$ and $x_{\max}^\star=\max_i x_i^\star$. Then, we will show that

$$E[\Delta W_t(\mathbf{y})|\chi_0^{t\,c}](1-p)\le(1-p)\left(M\sum_{i=1}^{n}x_i[t]+B\right). \tag{18}$$

If these inequalities hold, we have

$$\begin{aligned}E[\Delta W_t(\mathbf{y})] &\le -\left(pK\frac{\alpha_{\min}}{x_{\max}^\star}-(1-p)M\right)\sum_{i=1}^{n}x_i[t]\\ &\quad +(1-p)B+\hat{C}\end{aligned}$$

Then, from (16), we know that $P(\chi_0^t)$ is strictly increasing with A_s. Thus, by choosing A_s large enough, we can obtain a small $(1-p)$ such that

$$\frac{1}{3}pK\frac{\alpha_{\min}}{x_{\max}^\star}-(1-p)M>0.$$

We fix A_s so that $(1-p)B+\hat{C}$ is a constant. Thus, there exists an $\bar{X}$ such that for $\sum_i x_i[t]>\bar{X}$, we have

$$\tfrac{1}{3}pK\tfrac{\alpha_{\min}}{x_{\max}^\star}\bar{X}-(1-p)B-p\hat{C}\ge0.$$

Hence there exist A_s and $\bar{X}$ such that

$$E[\Delta W_t(\mathbf{y})]\le-\delta\sum_{i=1}^{n}x_i[t]I_{\mathbf{y}\in S_{\bar{X}}^c}+\zeta I_{\mathbf{y}\in S_{\bar{X}}},$$

where $\delta=\frac{1}{3}pK\frac{\alpha_{\min}}{(x_{\max}^\star)^m}$. Then, by invoking the Foster's criterion, we can conclude that the Markov Chain is positive recurrent.

First, we will show (17). Suppose χ_0^t occurs. Then along the lines of the proof of Theorem 1 of [6], it can be shown that there exists a $B_d<\infty$, which is independent on K and $\mathbf{x}[t]$, such that

$$E[\Delta W_t(\mathbf{y})|\chi_0^t]p(\chi_0^t)\le pB_d \tag{19}$$

$$+\quad p\sum_{i=1}^{n}\Delta x_i[t]\left(\left\{\frac{\alpha_iK}{x_i[t]}\right\}^M-\frac{\alpha_iK}{x_i^\star}\right) \tag{20}$$

$$+\quad p\sum_{i=1}^{n}\Delta x_i[t]\left(\left\{\frac{\alpha_iK}{x_i[t-T_i]}\right\}^M-\left\{\frac{\alpha_iK}{x_i[t]}\right\}^M\right). \tag{21}$$

Here, we need two inequalities to show that (17) holds. The first one is

$$(20) \leq -pK \sum_{i=1}^{n} \alpha_i \left(\frac{x_i[t]}{x_i^\star} - 2 \right). \tag{22}$$

And the second one is

$$(21) \leq pK \sum_{i=1}^{n} \alpha_i + pM \sum_{i=1}^{n} x_i^\star + pnMA_s. \tag{23}$$

If both of them hold, we can define

$$\hat{C} = pB_d + 3pK \sum_{i=1}^{n} \alpha_i + pM \sum_{i=1}^{n} x_i^\star + pnMA_s,$$

and $pB_d + (22) + (23)$ yields (17). To prove (22), we rearrange the index such that for users $i = 1, \ldots, \hat{n}$, $x_i[t] > x_i^\star$, which implies $\left\{ \frac{\alpha_i K}{x_i[t]} \right\}^M = \frac{\alpha_i K}{x_i[t]}$. Notice that for $i = 1, \cdots, \hat{n}$,

$$(x_i[t] - x_i^\star) \left(\left\{ \frac{\alpha_i K}{x_i[t]} \right\}^M - \frac{\alpha_i K}{x_i^\star} \right) \leq 0,$$

so we have

$$\begin{aligned} (20) \quad & \leq \quad p \sum_{i=1}^{\hat{n}} (x_i[t] - x_i^\star) \left(\frac{\alpha_i K}{x_i[t]} - \frac{\alpha_i K}{x_i^\star} \right) \\ & \leq \quad -pK \sum_{i=1}^{\hat{n}} \alpha_i \left(\frac{x_i[t]}{x_i^\star} - 2 \right). \end{aligned}$$

Since $1 - \frac{x_i[t]}{x_i^\star} \geq 0$ for any $i = \hat{n}+1, \ldots, n$, we have

$$(20) \leq -pK \sum_{i=1}^{n} \alpha_i \left(\frac{x_i[t]}{x_i^\star} - 2 \right).$$

Next, we show (23). Define

$$\Delta X_i = (x_i[t] - x_i^\star) \left(\left\{ \frac{\alpha_i K}{x_i[t - T_i]} \right\}^M - \left\{ \frac{\alpha_i K}{x_i[t]} \right\}^M \right).$$

If $x_i[t] \leq x_i^\star$, then

$$|\Delta X_i| \leq M x_i^\star. \tag{24}$$

Otherwise, if $x_i[t] > x_i^\star$, and if $x_i[t - T_i] > x_i[t]$, then

$$\Delta X_i \leq 0. \tag{25}$$

If instead $x_i[t] > x_i[t - T_i]$ and $x_i[t] > x_i^\star$, then

$$\begin{aligned} |\Delta X_i| \quad & \leq \quad (x_i[t - T_i] + A_s) \left\{ \frac{\alpha_i K}{x_i[t - T_i]} \right\}^M \\ & = \quad x_i[t - T_i] \left\{ \frac{\alpha_i K}{x_i[t - T_i]} \right\}^M + A_s \left\{ \frac{\alpha_i K}{x_i[t - T_i]} \right\}^M, \end{aligned}$$

which implies

$$|\Delta X_i| \leq \alpha_i K + MA_s. \tag{26}$$

From (24), (25) and (26), we can conclude that (23) holds. Combining this with (22), we managed to show that (17) holds.

Finally, we consider the event $\chi_0^{t\,c}$ and show (18). In this case we can write

$$\begin{aligned} E[\Delta W_t(\mathbf{y}) | \chi_0^{t\,c}](1 - p(\chi_0)) \quad & \leq \quad (1-p)B_d + (1-p) \sum_{i=1}^{n} \Delta X_i \\ & \leq \quad (1-p) \left(M \sum_{i=1}^{n} x_i[t] + B_d \right), \end{aligned}$$

which proves that (18) holds.

Now that we have shown that both (17) and (18) holds, we can conclude that the Markov chain $\{\mathbf{y}[t]\}$ is positive recurrent for any $K > 0$. $\square$

5 Simulations

In this section, we study the performance properties of the system in the presence of feedback delays through simulations. For this purpose, we consider 10 users with asymmetric channel conditions. The number of packets that can be served for user i in a given slot is assumed to be Poisson distributed with mean $(0.4 + i \times 0.1)$ for $i = 1, \cdots, 10$. The distribution of the arrivals to each buffer is also assumed to be Poisson distributed with the time-varying mean determined by the congestion control mechanism described in Section 2.

We take $M = 100$ and $\alpha_i = 1$ for all i throughout the simulations. As a representative case, we will be studying the queue-length evolution of the first queue with varying delay and K parameters. Recall that, the larger the K parameter, the closer the corresponding resource allocation to the weighted-proportionally fair allocation. In Figure 3, we take $K = 100$ and the buffers are assumed to be empty at the beginning of time; i.e., $x_i[0] = 0$ for all i. We plot the $x_1[t]$ for different values of the delay parameter T_1. We observe that in this scenario, the delay causes a significant shift in the transients of the evolution. However, as our theoretical analysis suggests, the equilibrium point is eventually reached in each case. The convergence takes longer as the delay increases.

Next, we change the initial buffer occupancy levels to 1000 for each queue while keeping K intact. The resulting

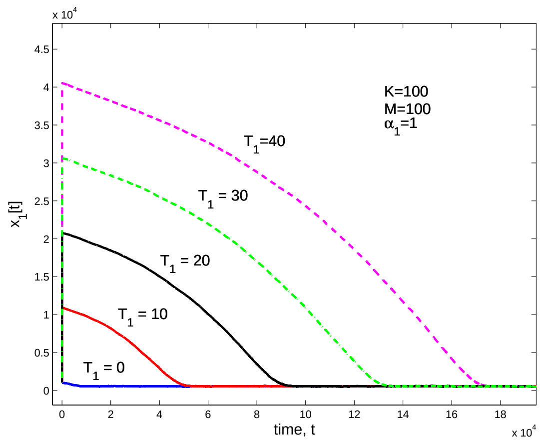

Figure 3. The queue-length evolution of flow 1 **with zero initial buffer occupancies and** $K = 100$**.**

queue-length level of the first buffer is plotted in Figure 4. In this case, we observe that even the transients do not differ by much when T_1 varies between 0 and 1000. This is due to the fact that the delayed and delay-free version of the queue-length will not cause a huge difference in the arrival rates to the queue.

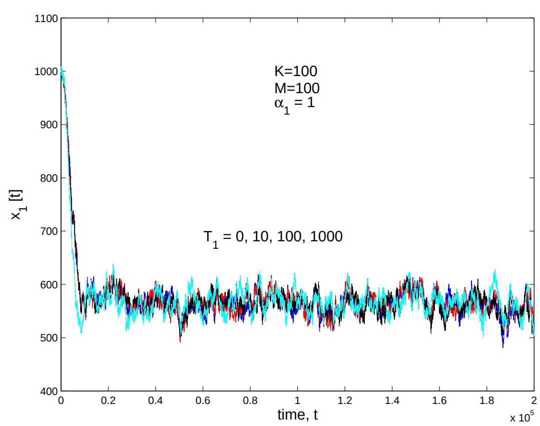

Figure 4. The queue-length evolution of flow 1 **with** $x_1[0] = 1000$ **and** $K = 100$**.**

Finally, we note that the choice of K is also a key parameter in the transient behavior. To show this, we keep the initial queue-length levels the same with the previous simulation (i.e., $x_i[0] = 1000$ for all i), but change K to 1000 from 100. Notice that this modification will also shift the equilibrium point up. We observe in Figure 5 that the queue-length evolution is inseparable for T_1 values between 0 and 100, however, there is a discernable difference when we increase T_1 to 1000. The difference is much more drastic when we further increase it to 10000.

We note that for a typical cellular network, the duration of a time slot is roughly 1 msec. And the Round-Trip-Time experienced in the Internet varies between 1 msec and 10000 msecs, with its bulk being centered around 100 msecs [1]. Thus, a typical value for T_i would be 100. Our simulations suggest that if the initial buffer occupancies are not very small, the effect of feedback delay on a typical internet session will be very insignificant. However, if the delay exceeds a certain point, the delay will start to influence the transient behavior.

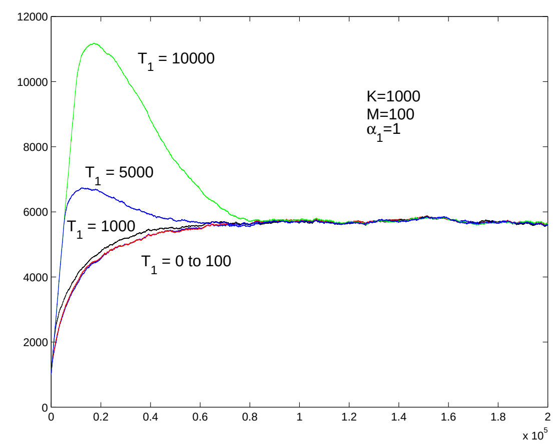

Figure 5. The queue-length evolution of flow 1 **when** $x_1[0] = 1000$ **and** $K = 1000$**.**

6 Conclusion

In this paper, we have shown that the algorithm (4) and (5) is stable even in the existence of heterogeneous delays. Furthermore, we proved that for large K, the network will achieve weighted-proportional fair allocation. When delays are not negligible, our result reaffirms the conclusion in [6] that the combination of queue-length-based scheduling and congestion control is a good distributed fair and stabilizing resource allocation scheme.

References

[1] J. Aikat, J. Kaur, F. D. Smith and K. Jeffay. Variability of TCP Round-trip Times *Proceedings of ACM Sigcomm, Internet Measurement Conference*, pages 279–284, October, 2003.

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

[3] M. Armony and N. Bambos. Queueing dynamics and maximal throughput scheduling in switched processing systems. Technical Report Netlab-2001-09/01, Stanford University.

[4] S. Asmussen. *Applied Probability and Queues*. Springer-Verlag, New York, 2003.

[5] R. Buche and H. J. Kushner. Control of mobile communication systems with time-varying channels via stability methods. *IEEE Transactions on Automatic Control*, 2004. To appear.

[6] A. Eryilmaz and R. Srikant. Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. To appear in the *Proceedings of IEEE Infocom*, Miami, Florida, March 2005.

[7] A. Eryilmaz, R. Srikant, and J. Perkins. Stable scheduling policies for fading wireless channels. In *Proceedings of IEEE International Symposium on Information Theory*, 2003. To appear in the *IEEE/ACM Transactions on Networking*, February, 2005.

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

[9] R. Leelahakriengkrai and R. Agrawal. Scheduling in multimedia wireless networks. In *Proceedings of ITC*, Brazil, 2001.

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

[11] M. Neely, E. Modiano, and C. Li. Fairness and optimal stochastic control for heterogeneous networks. In *Proceedings of IEEE Infocom 2005*, Miami, Florida, March 2005.

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

[13] S. Shakkottai, R. Srikant, and A. Stolyar. Pathwise optimality of the exponential scheduling rule for wireless channels. In *Proceedings of ISIT*, Lausanne, Switzerland, July 2002. To appear in the *Advances in Applied Probability*, 2004.

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

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

[16] R. Srikant. *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.

[17] A. Stolyar. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. Submitted.

[18] 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*, pages 1936–1948, December 1992.

[19] H. Yaiche, R. R. Mazumdar, and C. Rosenberg. A game-theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Transactions on Networking*, 8(5):667–678, October 2000.