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 transmitters 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.

I. 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,

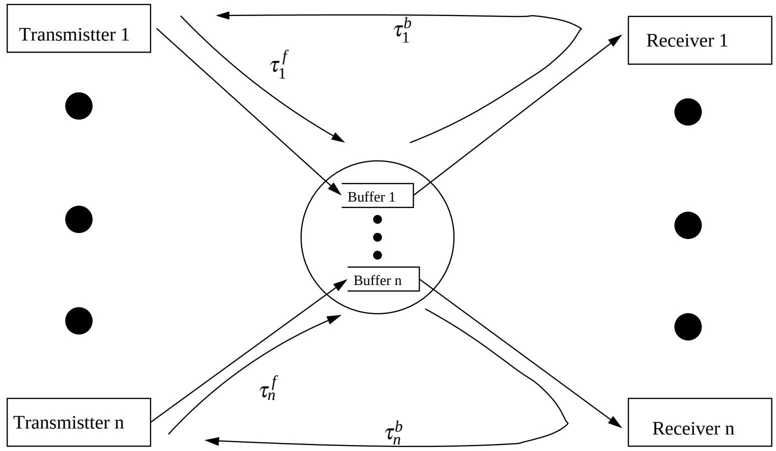

Fig. 1. Network with feedback delays. The channel from the base station to the receivers is time-varying.

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 [17]. 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 [18], [10], [13]. However, none of these works explicitly include the effect of feedback delay in their analysis. One of the reasons that delay is not important in these other works is that a specific scheduling algorithm is used in the network which allows the congestion control to be based only on the queue length at the entry node of each source. However, here we consider a situation where such scheduling is not used and where the bottleneck is at the cellular network while the sources may be located far away from the base station. An example of such a situation is a file transfer from a remote host over the downlink of a cellular network. 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 (transmitter) to the base station and in the transmission of congestion information from the base station back to the end users. But if we consider the case where the end users are connected to the base station through the Internet, then delays exist in both directions: there is a propagation delay τ_i^f from the end user i to the base station — we call it the forward delay of the end user i, and a propagation delay τ_i^b from the base station to the end 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 feedback delays are large [17]. In our problem, when delays exist, the information the end users obtain will be "outdated" information. So the congestion information the users obtain at time t does not reflect the queue status at the base station at time t. So it is interesting to study a wireless network with delays and ask whether the conclusions of [6] still hold for wireless networks with heterogeneous delays. We answer this question by showing that for a 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-m fair allocation arbitrarily closely. We emphasize that the results hold 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 by simulations, not shown here due to page limitations, which are presented in [21].

The rest of the paper is organized as follows: in Section II, we introduce the system model including the congestion controller used by the end users and the scheduler implemented at the base station. In Section III, we show that the resulting resource allocation approximates weighted-m fairness arbitrarily closely. Finally, we conclude in Section IV.

II. 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. We study the fair resource allocation problem in this paper. Specifically we consider weighted-m fairness. It means that each source i has a utility function given by $U_i(\overline{z_i}) = \alpha_i \frac{\overline{z_i}^{1-m}}{1-m}$, where $\overline{z_i}$ is the average rate at which user i transmits and α_i is a positive weighting factor [12]. Here, $m = 1$, $m = 2$ and $m \to \infty$ correspond to respectively proportional fair, minimum potential delay fair and max-min fair resource allocations. Note that in the case, $m = 1$, the above utility function is not well-defined; however, it can be shown that as $m \to 1$, the limiting resource allocation corresponds to the case where $U_i(\bar{z_i}) = \alpha_i \log \bar{z_i}$ [12]. We assume that the 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 ith 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 given to 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 $s[t]$ 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 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 let $\hat{\eta} < \infty$ 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 distribution of the channel state process being in state j. The scheduler will use 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 at the base station 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 [19], 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], [16], [15], [7], [3], [9], [5], [14].

In our model, the packet arrival rate into the 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], [11], [20], [17]. 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 any constant $K > 0$. Next, we describe the operation of our congestion controller followed by some assumptions.

CONGESTION CONTROLLER: Recall that, for user i, the forward delay is τ_i^f and the backward delay is τ_i^b. In our model, the downlink is the only bottleneck of the system, so τ_i^f, which is the propagation delay from user i to the base station, is a constant. On the other hand, packets experience queueing delay at the base station, and the transmissions between the base station and the receivers are over wireless links. Thus, $\tau_i^b[t]$ is time varying. Since users will always use the latest feedback information, we define $\tau_i^b[t]$ such that

$$\tau_i^b[t] = \min\{\tilde{\tau}_i^b[t], \tau_i^b[t-1]+1\},$$

where $\{\tilde{\tau}_i^b[t]\}_t$ are i.i.d. random variables, and $T_{\max} - \tau_i^f \geq \tilde{\tau}_i^b[t] \geq \tau_i^p$. Note that $T_{\max}$ is the upper bound on the round trip delays, and τ_i^p is the propagation delay from the base station to user i via receiver i. Now, denote the amount of data sent out by user i in slot t by $\lambda_i[t]$. The congestion controller at user i regulates the mean of $\lambda_i[t]$ such that

$$E\left[\lambda_i[t] \middle| x_i\left[t - \tau_i^b[t]\right]\right] = \min\left\{\frac{\alpha_i K}{\left(x_i\left[t-\tau_i^b[t]\right]\right)^m}, M\right\}, \tag{1}$$

where $m > 0$, $M > 2\hat{\eta}$ is a positive constant which ensures that the arrival rate into the queue is upper bounded when the queue length is small, and $x_i[t-\tau_i^b[t]]$ is the congestion information measured by the based station and feedback to user i via receiver i. We will later show that K has to be large to approximate weighted-m fair resource allocation.

Since $a_i[t] = \lambda_i\left[t-\tau_i^f\right]$, the mean of the number of arrivals into queue i at time t given by

$$E\left[a_i[t]|x_i[t-T_i[t]]\right] = E\left[\lambda_i\left[t-\tau_i^f\right]\Big|x_i\left[t-\tau_i^f-\tau_i^b\left[t-\tau_i^f\right]\right]\right] = \min\left\{\frac{\alpha_i K}{(x_i[t-T_i[t]])^m}, M\right\}, \quad (2)$$

where $T_i[t] = \tau_i^f + \tau_i^b\left[t-\tau_i^f\right]$. Define $\tilde{T}_i[t] = \tau_i^f + \tilde{\tau}_i^b\left[t-\tau_i^f\right]$, we have $T_i[t] = \min\{\tilde{T}_i[t], T_i[t-1]+1\}$, and $\{\tilde{T}_i[t]\}_t$ are i.i.d. random variables such that $T_{\max} \geq \tilde{T}_i[t] \geq \tau_i^f + \tau_i^p$. We assume $a_i[t]$ is independent across time slots and

$$E\left[a_i^2[t]|x_i[t-T_i(t)]\right] \leq V < \infty \text{ for all } x_i[t-T_i(t)]. \quad (3)$$

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

$$P\left(\sum_{j=1}^{T} a_i[t-j] = N\right) < \frac{\theta}{N^h} \text{ for all } i. \quad (4)$$

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

$$\begin{aligned} x_i[t+1] &= x_i[t] + a_i[t] - \mu_i[t] + u_i[t] \quad (5) \\ \mu[t] &\in \arg\max_{\eta \in C_{s[t]}} \sum_{i=1}^{n} x_i[t]\eta_i, \quad (6) \end{aligned}$$

where $a_i[t]$ is a random variable satisfying the conditions in (2), (3) and (4). Note that the congestion control part of this algorithm is slightly different from the algorithm in [6]. We impose an upper-bound on the source rates in a more natural manner than in [6]. Our results continue to hold for the algorithm in [6] too.

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 $(\mathbf{x}^*, \mu^*)$ which satisfy following conditions

- $\mu^* \in \arg\max_{\eta \in \bar{C}} \sum_{i=1}^n x_i^* \eta_i$; $\quad x_i^* = \left(\frac{\alpha_i K}{\mu_i^*}\right)^{\frac{1}{m}}$ for all i, and
- μ^* is the optimal solution to $\max_{\mu \in \bar{C}} \sum_{i=1}^n K\alpha_i \frac{\mu_i^{1-m}}{1-m}$.

$\diamond$

From the above theorem, we can see that μ^* is weighted-m fair. For the stochastic model, we will show that $\mu[t]$ converges to μ^*, 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 show that fair resource allocation can be achieved when the linear gain, K, used in the congestion controller goes to infinity.

III. Weighted-m Fairness and Stability

Use $\mu[\infty]$ to denote the steady state of μ, we will show in our main theorem — Theorem 3, that for any $\varepsilon > 0$,

$$\lim_{K\to\infty} P\left(|\mu[\infty]-\mu^*| \geq \varepsilon\right) = 0,$$

which implies that the network can achieve weighted-m fairness when K is chosen to be large. To prove this, we first need following lemma which characterizes the mean distance between $\mathbf{x}^*$ and the steady state of $\mathbf{x}[t]$.

Lemma 2: There exists a positive constant $\sigma < 1/m$, and a positive constant $\bar{c}$ that depends on the mean achievable rate region, the algorithm parameters $\{\alpha_i\}$, and the moments of the channel and arrival process, such that

$$E\left[\|\mathbf{x}[\infty]-\mathbf{x}^*\|\right] \leq \bar{c}K^{\frac{1}{m}-\sigma} \text{ for large } K,$$

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

Proof: Define $\mathbf{y}[t] = (\mathbf{x}[t],\ldots,\mathbf{x}[t-T_{\max}],\mathbf{T}[t])$, where $T_{\max} \geq T_i[t]$. 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 $x_i[t-T_i[t]]$ and $T_i[t]$ depends only on $T_i[t-1]$, so $x_i[t+1]$, $T_i[t+1]$ and $\mathbf{y}[t+1]$ are determined by $\mathbf{y}[t]$. Further, define the Foster-Lyapunov function

$$W(\mathbf{y}[t]) = \frac{1}{2}\sum_{i}^{n}(x_i[t]-x_i^*)^2,$$

and

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

Then following an argument similar to Theorem 2 of [6], the lemma will hold if there exist a finite set S_σ, positive numbers $\sigma < 1/m$, δ^*, and ζ such that for large K

$$E[\Delta W_t(\mathbf{y})] \leq -\frac{\delta^*}{K^{\frac{1}{m}-\sigma}}\|\mathbf{x}-\mathbf{x}^*\| I_{\mathbf{y}\in S_\sigma^c} + \zeta I_{\mathbf{y}\in S_\sigma}. \tag{7}$$

Thus we only need to show inequality (7). For a positive constant c and $\sigma < 1/m$ define

$$S_\sigma = \{\mathbf{y}[t] : \|\mathbf{x}[t] - \mathbf{x}^*\| \leq cK^\sigma \ \}. \tag{8}$$

Note that $T_i[t] \leq T_{\max}$, and $\|\mathbf{x}[t] - \mathbf{x}^*\| \leq cK^\sigma$ implies that

$$\sum_i x_i[t-s] \leq \sum_i x_i^* + ncK^\sigma + nT_{\max}\hat{\eta}, \quad \text{for all} \quad 0 \leq s \leq T_{\max}.$$

Thus, S_σ is a finite set. Also, it is easy to see that if $\mathbf{y}[t] \in S_\sigma$, there exists $0 < \zeta < \infty$ such that $E[\Delta W_t(\mathbf{y})] < \zeta$. Now, consider $\mathbf{y}[t] \notin S_\sigma$, define the event χ_0^t such that

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

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

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

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 also 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 $B_d > 0$, which is independent on K and $\mathbf{x}[t]$, such that

$$\begin{aligned}
E[\Delta W_t(\mathbf{y})] &\leq \sum_{i=1}^{n} \Delta x_i[t] \left(\left\{ \frac{\alpha_i K}{(x_i[t-T_i])^m} \right\}^M - \mu_i^* \right) + B_d & (9)\\
&= \sum_{i=1}^{n} \Delta x_i[t] \left(\left\{ \frac{\alpha_i K}{(x_i[t-T_i])^m} \right\}^M - \left\{ \frac{\alpha_i K}{(x_i[t])^m} \right\}^M \right) & (10)\\
&\quad + \sum_{i=1}^{n} \Delta x_i[t] \left(\left\{ \frac{\alpha_i K}{(x_i[t])^m} \right\}^M - \mu_i^* \right) + B_d, & (11)
\end{aligned}$$

where $\Delta x_i[t] = x_i[t] - x_i^*$. Define $G(K) := (11)$ and $H(K) := (10)$, to prove inequality (7), we will show the following three facts. The first one is that there exists $\delta_d > 0$ such that for all events

χ_l^t,

$$G(K) \leq -\frac{\delta_d}{K^{\frac{1}{m}-\sigma}}\|\mathbf{x}[t]-\mathbf{x}^*\|. \tag{12}$$

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

$$p(\chi_0^t)\,|H(K)| \leq p(\chi_0^t)\frac{\delta_0}{K^{\xi}}\|\mathbf{x}[t]-\mathbf{x}^*\|. \tag{13}$$

The last one is that there exists $\delta_1 > 0$ such that

$$\sum_{l=1}^{\infty} p(\chi_l^t)\,|H(K)| \leq \frac{\delta_1}{K^{\xi}}\|\mathbf{x}[t]-\mathbf{x}^*\|. \tag{14}$$

If all three inequalities — (12), (13) and (14) — hold and $\xi > \frac{1}{m}-\sigma$, we will have

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

Then, when $K > ((2\delta_0+\delta_1)/\delta_d)^{1/(\xi+\sigma-1/m)}$, we have

$$\frac{\delta_d}{2K^{\frac{1}{m}-\sigma}}-\frac{\delta_0+\delta_1}{K^{\xi}} > 0,$$

which implies

$$E[\Delta W_t(\mathbf{y})] \leq -\left(\frac{\delta^*}{K^{\frac{1}{m}-\sigma}}\right)\|\mathbf{x}[t]-\mathbf{x}^*\|$$

and the inequality (7) holds with $\delta^* = \delta_d/2$.

Now we prove (12), (13) and (14). We will first show (12). The proof is similar to the proof of Theorem 1 of [6]. But here we consider a general m instead of $m=1$. Define σ as follows:

$$\sigma = \begin{cases} \lfloor \frac{1}{m} \rfloor, & \text{if } m \leq 1 \text{ and } \frac{1}{m} \text{ is not an integer;} \\ \frac{1}{m}-\frac{1}{2}, & \text{if } m \leq 1 \text{ and } \frac{1}{m} \text{ is an integer;} \\ \frac{1}{2m}, & \text{if } m > 1. \end{cases}$$

From above definition, we have that

$$0 < \frac{1}{m} - \sigma < \min\{\sigma, 1\}.$$

Notice that we have

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

Letting $i_0 = \arg\max_i |x_i[t] - x_i^*|$, then

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

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

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

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

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

Because

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

and

$$\left|\left(\frac{x_{i_0}^*}{x_{i_0}^* - |x_{i_0}[t] - x_{i_0}^*|}\right)^m - 1\right| \geq \left|\left(\frac{x_{i_0}^*}{x_{i_0}^* + |x_{i_0}[t] - x_{i_0}^*|}\right)^m - 1\right|,$$

we can show that

$$\mu_{i_0}^* \left|\left(\frac{x_{i_0}^*}{x_{i_0}[t]}\right)^m - 1\right| \geq \mu_{i_0}^* \left|\left(\frac{x_{i_0}^*}{x_{i_0}^* + |x_{i_0}[t] - x_{i_0}^*|}\right)^m - 1\right| \geq \mu_{i_0}^* \left|1 - \frac{1}{(1+\varepsilon)^m}\right|,$$

where

$$\varepsilon = \frac{c\mu_{i_0}^{*\,1/m}}{\sqrt{n}\alpha_{i_0}{}^{1/m}} K^{\sigma - \frac{1}{m}} > 0,$$

and the last inequality holds because $x_{i_0}^* = \left(\frac{\alpha_{i_0}K}{\mu_{i_0}^*}\right)^{1/m}$ and $cK^\sigma \leq \|\mathbf{x}[t]-\mathbf{x}^*\| \leq \sqrt{n}|x_{i_0}[t]-x_{i_0}^*|$. Further, since $(1+\varepsilon)^m \geq 1+m\varepsilon$, we have

$$\mu_{i_0}^*\left|1-\frac{1}{1+m\varepsilon}\right| = \frac{\mu_{i_0}^* m\varepsilon}{1+m\varepsilon}.$$

It is easy to see that for large K, we will have $\frac{\mu_{i_0}^* m\varepsilon}{1+m\varepsilon} < \hat{\eta}$. Thus, for large K, we have

$$\left|\left\{\frac{\alpha_{i_0}K}{(x_{i_0}[t])^m}\right\}^M - \mu_{i_0}^*\right| \geq \frac{\mu_{i_0}^* m\varepsilon}{1+m\varepsilon},$$

and

$$G(K) \leq -|x_{i_0}[t]-x_{i_0}^*|\left(\frac{\mu_{i_0}^* m\varepsilon}{1+m\varepsilon} - \frac{B_d}{|x_{i_0}[t]-x_{i_0}^*|}\right) \leq -|x_{i_0}[t]-x_{i_0}^*|\left(\frac{\mu_{i_0}^*}{\frac{\sqrt{n}}{mc}\left(\frac{\alpha_{i_0}}{\mu_{i_0}^*}\right)^{\frac{1}{m}} K^{\frac{1}{m}-\sigma}+1} - \frac{B_d}{\frac{c}{\sqrt{n}}K^\sigma}\right).$$

Because $\frac{1}{m}-\sigma \leq \sigma$, by choosing sufficiently large c, we can find a positive constant $\hat{\delta}$ and $\hat{K}$ such that for any $K \geq \hat{K}$

$$G(K) \leq -\frac{\hat{\delta}}{K^{\frac{1}{m}-\sigma}}|x_{i_0}[t]-x_{i_0}^*| \leq -\frac{\delta_d}{K^{\frac{1}{m}-\sigma}}\|\mathbf{x}[t]-\mathbf{x}^*\|,$$

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

Next, we consider (13). It is the case that the arrivals are upper bounded by A. We will show that as K increases, $\left|\left\{\frac{\alpha_i K}{(x_i[t-T_i])^m}\right\}^M - \left\{\frac{\alpha_i K}{(x_i[t])^m}\right\}^M\right|$ decreases.

We use Figure 2 to prove our result. Suppose that $m>0$, then $\alpha_i K/y^m$ is convex and

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

is as Figure 2. Now suppose $c-a = d-b = A$ and $f(a) = M$. Then for any b such that $a<b$, we can see from the figure that $f(a)-f(c) > f(b)-f(d)$. It means that

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

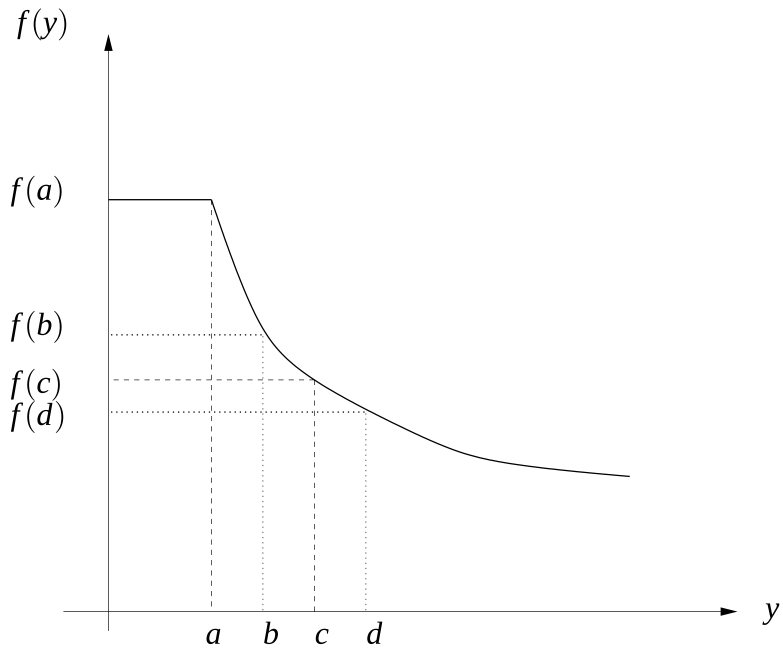

Fig. 2. A plot of $f(y) = \{\alpha_i K/y^m\}^M$

Further, from the assumption we have $A \geq T_{\max}\hat{\eta}$, so $|x_i[t] - x_i[t - T_{\max}]| \leq A$ and

$$|f(x_i[t-T_i]) - f(x_i[t])| \leq M - \frac{\alpha_i K}{\left(\left(\frac{\alpha_i K}{M}\right)^{\frac{1}{m}} + A\right)^m} = M - \frac{\alpha_i K}{\frac{\alpha_i K}{M}\left(1 + \left(\frac{A^m M}{\alpha_i K}\right)^{\frac{1}{m}}\right)^m}.$$

Since $\left(\frac{A^m M}{\alpha_i K}\right)^{\frac{1}{m}}$ is small when K is large, and $(1+\varepsilon)^m \leq 1 + 2m\varepsilon$ for small enough ε, we can conclude that for sufficiently large K

$$|f(x_i[t-T_i]) - f(x_i[t])| \leq M - \frac{M}{1 + 2m\left(\frac{A^m M}{\alpha_i K}\right)^{\frac{1}{m}}} < \frac{2mAM^{\frac{1+m}{m}}}{\alpha_i^{\frac{1}{m}}} \frac{1}{K^{\frac{1}{m}}}. \tag{15}$$

Let $\alpha_{\min} = \min_i \alpha_i$, then there exists $\delta_0 = 2n\alpha_{\min}^{-\frac{1}{m}} mAM^{\frac{1+m}{m}}$ and $\xi = \frac{1}{m}$ such that

$$|H(K)| \leq \frac{\delta_0}{K^{\xi}} \|\mathbf{x}[t] - \mathbf{x}^*\|.$$

Finally, we consider the complement of χ_0^t, which is denoted as $\chi_0^{t\,c}$, and derive inequality (14). Now the arrivals are not upper bounded and can be arbitrarily large. But from assumption (4), the probabilities of these events is very small. So we can still obtain a upper bound for

$\sum_{i=1}^{n}(x_i[t]-x_i^*)\left|\left\{\frac{\alpha_i K}{x_i[t-T_i]}\right\}^M-\left\{\frac{\alpha_i K}{x_i[t]}\right\}^M\right|$. Now suppose χ_l^t occurs ($l \geq 1$), similar to inequality (15), we can get

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

and

$$\sum_{l=1}^{\infty} p(\chi_l^t)\,|H(K)| \leq \|\mathbf{x}[t]-\mathbf{x}^*\| \sum_{l=1}^{\infty} 2n\alpha_{\min}^{-\frac{1}{m}} mM^{\frac{1+m}{m}}\frac{A+l}{K^{\xi}}p(\chi_l^t).$$

Under assumption (4), we can further obtain

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

where $\delta_1 = 2n\alpha_{\min}^{-\frac{1}{m}} mM^{\frac{1+m}{m}}(h-2)\frac{1}{A^{h-2}}$.

We have proved that inequalities (12), (13) and (14) hold, and it is easy to see that $\xi > 1/m-\sigma$. Thus, when $K > 4(\delta_0+\delta_1)^2/\delta_d^2$, inequality (7) holds with $\delta^* = \delta_d/2$, and the lemma follows. ■

Recall that $x_i^* = \left(\frac{\alpha_i K}{\mu_i^*}\right)^{\frac{1}{m}}$, so from the lemma above, we have

$$E\left[\frac{|x_i[\infty]-x_i^*|}{x_i^*}\right] \leq E\left[\frac{\|\mathbf{x}[\infty]-\mathbf{x}^*\|}{x_i^*}\right] \leq \frac{\bar{c}(\mu^*)^{1/m}}{\alpha_i^{1/m}K^{\sigma}},$$

which converges to zero when K goes to infinity. Thus, the mean of $x_i[\infty]$ concentrates around x_i^* for large K, from which we can show that weighted-m fairness can be achieved. This is stated in the next theorem which is the main result of this paper.

Theorem 3: Consider the combined Scheduling-Congestion Control Algorithm defined by (2)-(6). Then the steady-state service rate vector $\mu[\infty]$ satisfies the following: for any $\varepsilon > 0$,

$$\lim_{K\to\infty} P(|\mu[\infty]-\mu^*| \geq \varepsilon) = 0.$$

Proof: Using the Markov inequality, Lemma 2 yields for any $\varepsilon > 0$,

$$P\left(\frac{1}{K^{\frac{1}{m}}}|x_i[\infty]-x_i^*| > \varepsilon\right) \leq \frac{\bar{c}}{\varepsilon K^{\sigma}}.$$

Further, since

$$\mu[t] \in \arg\max_{\mu \in C_{s[t]}} \sum_{i=1}^{n} x_i[t]\eta_i = \arg\max_{\mu \in C_{s[t]}} \sum_{i=1}^{n} \frac{x_i[t]}{K^{1/m}}\eta_i,$$

we can conclude

$$\lim_{K\to\infty} P\left(|\mu[\infty] - \mu^*| \geq \varepsilon\right) = 0,$$

and the network is weighted m-fair according to Theorem 1. ■

Theorem 3 allows us to conclude that even in the presence of delays, the network will achieve weighted-m fairness. Note that, from inequality (7), when K is large, $\{\mathbf{y}[t]\}$ is positive recurrent and the system is stable. Actually if we are only concerned with the stability of the system, inequality (7) is much stronger than what is necessary to prove the stability. In fact, we can show that for any $K > 0$, the Markov chain is positive recurrent. Define the $S_{\bar{X}}$:

$$S_{\bar{X}} = \left\{ \mathbf{y}[t] : \sum_i x_i[t] \leq \bar{X} \right\}. \tag{16}$$

Clearly, $S_{\bar{X}}$ is a finite set. Stability of the system is established by following theorem.

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

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

where $S_{\bar{X}}$ is defined as (16). Hence, the Markov chain $\{\mathbf{y}[t]\}$ is positive recurrent.

Proof: We omit the proof here because of lack of the space. Please refer Theorem 4 of [21] for the proof of $m = 1$. The case of general weighted-m fairness is similar. ■

From Theorem 3, we see that fairness can only be achieved when $K \to \infty$. However, Theorem 4 assures that we are guaranteed at least stability for all K.

IV. Conclusion

In this paper, we have shown that the algorithm (5) and (6) is stable even in the presence of heterogeneous delays and when K is large, the network will achieve weighted-m fairness. When delays are not negligible in some situations, our result reinforces the result that the combination of queue-length-based scheduling and congestion control is a good distributed fair 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*, Vol. 49, No. 11, pp 1954 - 1962, November, 2004.

[6] A. Eryilmaz and R. Srikant. Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. *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] X. Lin and N. Shroff. The impact of imperfect scheduling on cross-layer rate control in wireless networks. In *Proceedings of IEEE Infocom 2005*, Miami, Florida, March 2005.

[11] 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.

[12] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 5:556–567, 2000.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

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

[19] 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.

[20] 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.

[21] L. Ying, R. Srikant, A. Eryilmaz and G. Dullerud. Distributed Fair Resource Allocation in Cellular Networks in the Presence of Heterogeneous Delays. *Proc. of WiOpt 2005*.