A simple fluid model to analyze resource sharing in Internet routers

A. Lakshmikantha and R. Srikant
Coordinated Science Lab and
Dept. of Electrical and Computer Engineering
University of Illinois Urbana-Champaign
{lkshmknt,rsrikant}@uiuc.edu

C. L. Beck
Coordinated Science Lab and
Dept. of General Engineering
University of Illinois Urbana-Champaign
beck3@uiuc.edu

Abstract—**It has been widely recognized that providing a simple priority to short flows in Internet routers can dramatically reduce their mean delay while having little impact on the long flows which carry the bulk of Internet traffic. In this paper, we present simple fluid models which can be used to quantify the above observation. Further, we justify these fluid models by showing that stochastic models of resource-sharing among TCP flows converge to the fluid models when the router capacity and the number of users are large, which is the typical scenario of interest.**

I. INTRODUCTION

Numerous studies have shown that most of the files transferred over the Internet are small in size (of the order of a few tens of kilobytes) and only a small fraction of the files are large. It is estimated that 10-20% of the flows (that correspond to large files) carry about 80-90% of the Internet traffic. These rough statistical observations are a consequence of the fact that the file size distribution in the Internet is *heavy tailed* (see, for example, [7]).

As the short-flows have a small amount of data to be transmitted, they never leave the slow-start phase of the TCP. Any packet loss seen by these flows can lead to a significant loss in the throughput since the window size is reset to 1. A natural solution to the above problem is to provide priority to the short-flows in accessing the available bandwidth in the network. It has been shown in [12] that the Shortest Remaining Processing Time (SRPT) policy can dramatically improve the performance of short flows while having little impact on the long flows in the network.

However, in the current Internet, SRPT cannot be implemented very easily since the routers do not have access to per-flow information. It is easier to implement a simple priority scheme whereby packets from short flows are enqueued in a higher priority queue which the router serves whenever it has any packets. When the high priority queue does not have any packets, then the packets from long flows which are stored in a low priority queue are served. Such schemes have been studied through simulations in [1], [6].

The above priority scheme requires the router to estimate whether a flow is a short flow or a long flow. It, it has been shown that by using simple sampling techniques [9], one can identify with a high degree of confidence whether a flow is a long-flow or a short-flow.

In this paper, we evaluate the performance of such priority-based schemes in the Internet via a fluid model and then strengthen our observations via simulations. Without priorities, the nature of bandwidth sharing favors the long-flows in the current Internet. Such a sharing discipline can be approximated by a discriminatory processor sharing (DPS). Stochastic analysis of DPS is extremely hard and closed-form solutions exist only for exponentially distributed service times [14], [10], [19]. However, in a system which has a large number of files and a large server capacity (such as the Internet), some form of law of large numbers holds and the resulting stochastic system can be approximated by a deterministic system which can be modeled by a set of differential equations. In the earlier papers [16], [17] we had considered less accurate fluid models to model the resource sharing of the TCP flows. In this paper, we consider more accurate fluid models which consider the impact of the access bandwidth constraints of TCP flows. More importantly, for the first time, we justify the fluid approximation by showing that a stochastic model of DPS converges to the fluid limit in a large system limit. Finally we also characterize the speed of convergence of the fluid limit to its equilibrium value.

The main contributions of this paper can be summed up as follows:

- It has been shown earlier in [12] that under SRPT short-flows have significant gains compared to priority sharing without significant increase in the delays for the longer flows. In this paper, we show that we can achieve similar gains (qualitatively) with a much simpler priority scheme.
- The analysis carried out in [12] compares SRPT with processor sharing. However, in the current Internet, the nature of bandwidth sharing is "unfair" in the sense that it favors longer flows. Despite this, our analysis seems to indicate that, the longer flows do not suffer significantly when priority is given to the short-flows.
- We show, under appropriate scaling, that the stochastic process describing the DPS queue converges almost surely (abbreviated henceforth as a.s.) to the fluid model over large intervals of time.

Research supported by AFOSR URI grant F49620-01-1-0365

We note that we do not address the problem of identifying short and long-flows. We assume that such mechanisms exist and they can be implemented without causing significant errors [9], [1].

The rest of this paper is organized as follows. In Section II, we develop the fluid model for the case of no priorities and show that the system converges to the equilibrium exponentially fast. In Section III, we evaluate the performance of the system with priorities. The validity of the fluid model for the case of no priorities is discussed in Section IV. In this section, we show that, under appropriate scaling, the stochastic process describing the number of backlogged short-flows and long-flows, converges a.s. to the fluid model over large intervals of time. The proof of validity of the fluid model for the case of priorities is very similar and thus will be omitted in the interest of space. Section V contains simulation results that strengthen the observations made in Section III. Concluding remarks are provided in Section VI.

II. FLUID MODELS WITH NO PRIORITIES

A. Model Description

We consider a single link accessed by many flows. Short-flows (long-flows) arrive into the system according to a Poisson process with mean rate $\lambda_s(\lambda_l)$. We assume that the file-size distributions belong to the class of Coxian distributions. (Note that Coxian distributions form a dense subset of the set of all distributions on the positive real line. We can therefore, approximate any distribution arbitrarily closely by a Coxian distribution) Coxian distributions are defined as follows.

Definition 1: Consider a Markov chain on the states $\{1, 2, ..., m+1\}$, where state $m+1$ is an absorption state and states $\{1, 2..., m\}$ are all transient. The system starts at state 1 w.p. 1. In each state i, the system spends a random amount of time which is exponentially distributed with mean $\frac{1}{\mu_i}$. At the end of this dwell time, it moves to state $m+1$ w.p. $(1-p_i)$ or moves to state $i+1$ w.p. p_i. At the end of the dwell time in state m, the system moves to state $m+1$ w.p. 1. A Coxian distribution of m phases is the distribution of the time to absorption of the above described Markov chain.

The short-flows (respectively long-flows) are modeled by a Coxian distribution with m_s (m_l) phases. In phase i, the flow transmits a random amount of data which is exponentially distributed with mean $\frac{1}{\mu_{s_i}}$ $(\frac{1}{\mu_{l_i}})$. With the exception of the final phase, at the end of phase i, the flow leaves the system with probability $(1-p_{s_i})$ $((1-p_{l_i}))$ and begins a new exponential phase with probability p_{s_i} (p_{l_i}). At the end of the final phase, the flow leaves the system with probability one. To ease the notation, we use $\mathcal{S}$ $(\mathcal{L})$ to denote the index set $\{1, 2..., m_s\}$ $(\{1, 2, ..., m_l\})$. Define $p_{s_0} = 1$ and $p_{l_0} = 1$. Let $\frac{1}{\mu_l}$ and $\frac{1}{\mu_s}$ denote the mean file-size of short-flows and long-flows respectively. Then,

$$\frac{1}{\mu_s} = \sum_{i \in \mathcal{S}} \left(\prod_{j=0}^{i-1} p_{s_j} \right) \frac{1}{\mu_{s_i}}, \tag{1}$$

and

$$\frac{1}{\mu_l} = \sum_{i \in \mathcal{L}} \left(\prod_{j=0}^{i-1} p_{l_j} \right) \frac{1}{\mu_{l_i}}. \tag{2}$$

For any flow accessing the bottleneck link, there are two factors that determine the bandwidth available to that flow. First, the maximum rate at which a flow can transmit is constrained by the rate of its access link and due to TCP's maximum window size constraints. Second, the capacity of the link is shared by many flows and thus, when the number of flows is large, the link capacity will determine the bandwidth received by each flow.

To model the access constraints, we assume that all the short-flows (long-flows) accessing the link have an access constraint of $r_s(r_l)$. In the case of short-flows, this limit occurs mainly due to the inherent nature of TCP itself (more on this in Section III), while in the case of long-flows this will be primarily due to the limitations in the access links.

It is well known that the data rate of a user, can be approximated by window size divided by the RTT (round trip time). Thus, at any given time, the data rate of a flow is roughly proportional to window size. Since the short-flows never leave the slow-start phase of TCP, the window sizes of the short-flows will be very small. Thus, the bandwidth received by short-flows will be small. As observed earlier, short flows never leave the slow-start phase and hence their window sizes are always small, leading to a small bandwidth share for these flows. To model the effects on bandwidth sharing due to sizable differences in the window sizes, we assume that the capacity of the bottleneck link is shared in a proportional manner with weights w_s and w_l $(w_l > w_s)$. Relating these weights to the actual window sizes seems very complicated and we do not attempt to do so. Based on empirical evidence from the simulations, we assume that ratio $\frac{w_l}{w_s}$ is roughly around three. The precise value of this ratio is unimportant for our analysis.

If n_{l_i} denotes the number of long-flows in phase i and n_{s_i} denotes the number of short-flows in phase i, then the data rate of a short-flow is given by

$$x_s = \min \left(r_s, \frac{w_s c}{\sum_{i \in \mathcal{S}} n_{s_i} w_s + \sum_{j \in \mathcal{L}} n_{l_j} w_l} \right), \tag{3}$$

where c denotes the capacity of the link. Similarly, the data rate of a long flow is given by

$$x_l = \min \left(r_l, \frac{w_l c}{\sum_{i \in \mathcal{S}} n_{s_i} w_s + \sum_{j \in \mathcal{L}} n_{l_j} w_l} \right). \tag{4}$$

Without loss of generality we assume $c = 1$. In this paper, we refer to the above sharing policy as modified discriminatory processor sharing (DPS). Such a policy tends to be unfair to short-flows. Under heavy loading (when there are large number of flows in the network, the access rate constraint becomes redundant), it is easy to see that $\frac{x_s}{x_l} = \frac{w_s}{w_l}$. As $w_s < w_l$, the bandwidth seen by short-flows is smaller than the bandwidth seen by the large-flows.

Let us denote the total load on the system by

$$\rho = \frac{\lambda_l}{\mu_l} + \frac{\lambda_s}{\mu_s} = \rho_l + \rho_s. \tag{5}$$

Throughout this paper, we will assume that the system is critically loaded, i.e. $\rho = 1$.

The number of bits in the system at any given time t, is denoted by $W(t)$. If there is a short-flow in phase i, then after the completion of phase i, it moves to phase $i+1$ with a probability p_{s_i}, and from phase $i+1$ it moves to phase $i+2$ with probability $p_{s_{i+1}}$ and so on. Thus if X_{s_i} denotes the amount of data this file transmits before it exits the system, then,

$$E[X_{s_i}] = \frac{1}{\mu_{s_i}} + \sum_{k>i, i\in\mathcal{S}} \frac{1}{\mu_{s_k}} \prod_{j=i}^{j=k-1} p_{s_j} \overset{def}{\equiv} \frac{1}{\nu_{s_i}}.$$

Thus, if there are n_{s_i} files in phase i, the average amount of data that these files will transmit before exiting the system will be $\frac{n_{s_i}}{\nu_{s_i}}$. Summing over all phases of both short and long flows , the average number of bits still remaining in the system is given by,

$$W(t) = \sum_{i\in\mathcal{S}} \frac{n_{s_i}}{\nu_{s_i}} + \sum_{i\in\mathcal{L}} \frac{n_{l_i}}{\nu_{l_i}}.$$

B. Fluid Model Analysis

We will now consider a fluid model that appropriately describes DPS system. It will be later shown (Section IV) that the fluid model is valid over large intervals of time. The evolution of the number of files in phase i is governed by the following set of differential equations.

$$\begin{aligned} \dot{n}_{l_i} &= \begin{cases} \lambda_l - \mu_{l_i} n_{l_i} x_l & i = 1, \\ p_{l_{i-1}} \mu_{l_{i-1}} n_{l_{i-1}} x_l - \mu_{l_i} n_{l_i} x_l & i \in \mathcal{L},\, i \neq 1, \end{cases} \\ \dot{n}_{s_i} &= \begin{cases} \lambda_s - \mu_{s_i} n_{s_i} x_s & i = 1, \\ p_{s_{i-1}} \mu_{s_{i-1}} n_{s_{i-1}} x_s - \mu_{l_s} n_{l_s} x_s & i \in \mathcal{S},\, i \neq 1. \end{cases} \end{aligned} \tag{6}$$

The stationary point of these equations (assuming $\rho = 1$) is given by

$$\begin{aligned} n_{l_i}{}^* x_l^* &= \left(\prod_{j=0}^{j=i-1} p_{l_j} \right) \frac{\lambda_l}{\mu_{l_i}} \quad \forall i \in \mathcal{L}, \\ n_{s_i}{}^* x_s^* &= \left(\prod_{j=0}^{j=i-1} p_{s_j} \right) \frac{\lambda_s}{\mu_{s_i}} \quad \forall i \in \mathcal{S}, \end{aligned} \tag{7}$$

where $n_{s_i}^*$ and $n_{l_i}^*$ denote the equilibrium values. The above set of equations are however not enough to characterize the equilibrium point completely. As the system is critically loaded, the actual equilibrium point depends on the average work in the system at equilibrium. It will be shown later (Remark 1) that if the initial work load is high enough, then the average work in the system is invariant with time. Under this assumption, given the load at time $t = 0$, which we denote by $W(0)$, one can explicitly characterize the unique equilibrium point. After some algebraic manipulations, we get the following closed form expressions for the equilibrium point:

$$\begin{aligned} n_{l_i}^* &= W(0) \frac{\rho_{l_i}}{\sum_{i\in\mathcal{S}} \frac{\rho_{s_i}}{\nu_{s_i}} \frac{w_l}{w_s} + \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}}}, \\ n_{s_i}^* &= W(0) \frac{\rho_{s_i}}{\sum_{i\in\mathcal{S}} \frac{\rho_{s_i}}{\nu_{s_i}} + \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}} \frac{w_s}{w_l}}, \end{aligned} \tag{8}$$

where

$$\rho_{s_i} = \frac{\lambda_s}{\mu_{s_i}} \prod_{j=1}^{i-1} p_{s_i}, \quad \rho_{l_i} = \frac{\lambda_l}{\mu_{l_i}} \prod_{j=1}^{i-1} p_{l_i}.$$

The bandwidth received by short flows and long flows at equilibrium is given by

$$\begin{aligned} x_s^* &= \frac{1}{W(0)} \left(\sum_{i\in\mathcal{S}} \frac{\rho_{s_i}}{\nu_{s_i}} + \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}} \frac{w_s}{w_l} \right), \\ x_l^* &= \frac{1}{W(0)} \left(\sum_{i\in\mathcal{S}} \frac{\rho_{s_i}}{\nu_{s_i}} \frac{w_l}{w_s} + \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}} \right). \end{aligned} \tag{9}$$

Given the initial conditions $(n_{s_i}(0) i \in \mathcal{S}, n_{l_i}(0) i \in \mathcal{L})$ the amount of work left in the system $W(0)$ can be determined. Thus, the unique equilibrium point is determined by these initial conditions. However, for the equilibrium analysis to be valid, the system has to be stable. In the rest of this section, we show that for a large class of initial conditions that are lower bounded, the system will converge to the equilibrium value. We start with the following lemma.

Lemma 2: In the modified DPS which is critically loaded, the amount of work left in the system is nondecreasing. i.e. $W(t) \geq W(0)\ \forall t \geq 0$.

Proof: The proof follows from the fact that the system is critically loaded and thus at any time t, the rate at which bits enter the system is more than or equal to the maximum rate at which the server can serve. More formally, to show that

$W(t) \geq W(0)$, it suffices to show that $\dot{W}(t) \geq 0$, $\forall t \geq 0$.

$$\begin{aligned}
\dot{W}(t) &= \frac{d}{dt}\left[\sum_{i\in\mathcal{S}} \frac{n_{s_i}}{\nu_{s_i}} + \sum_{i\in\mathcal{L}} \frac{n_{l_i}}{\nu_{l_i}}\right] \\
&= \lambda_s \sum_{i\in\mathcal{S}} \left(\prod_{j=0}^{i-1} p_{s_j}\right) \frac{1}{\mu_{s_i}} + \lambda_l \sum_{i\in\mathcal{L}} \left(\prod_{j=0}^{i-1} p_{l_j}\right) \frac{1}{\mu_{l_i}} \\
&\quad -\left(\sum_{i\in\mathcal{S}} n_{s_i} x_s + \sum_{i\in\mathcal{L}} n_{l_i} x_l\right) \\
&= (\frac{\lambda_s}{\mu_s} + \frac{\lambda_l}{\mu_l}) - \min\left(1, (\sum_j n_{s_j} r_s + \sum_j n_{l_j} r_l)\right) \\
&\geq 0
\end{aligned} \tag{10}$$

■

Now we would like to develop conditions under which the system will behave like a DPS queue i.e. $\forall t \geq 0$ $\min(\sum_i n_{s_i} r_s + \sum_i n_{l_i} r_l, 1) = 1$. Suppose,

$$\begin{aligned}
\textstyle\sum_j n_{s_j}(0) &> \tfrac{1}{r_s} \\
\textstyle\sum_j n_{l_j}(0) &> \tfrac{1}{r_l}.
\end{aligned}$$

Then at time $t = 0$ the rate limit constraint is inactive. Let $W(0)$ denote the workload corresponding to the above mentioned choice of initial number of flows. At any time $t > 0$ as $W(t) \geq W(0)$, it follows that either $\sum_j n_{s_j}(t) \geq \sum_j n_{s_j}(0)$ or $\sum_j n_{l_j}(t) \geq \sum_j n_{l_j}(0)$. From this it follows that the rate constraint still remains inactive. Thus $\forall t \geq 0$ the rate limit is inactive under the choice of initial conditions. As a consequence we have the following lemma.

Lemma 3: If the initial conditions are such that,

$$\begin{aligned}
\textstyle\sum_j n_{s_j}(0) &> \tfrac{1}{r_s}, \\
\textstyle\sum_j n_{l_j}(0) &> \tfrac{1}{r_l},
\end{aligned}$$

then the rate constraint is inactive $\forall t \geq 0$.

Remark 1: If the initial number of flows in the system satisfy the conditions in Lemma 3, then it follows that

$$\min\left(1, (\sum_j n_{s_j} r_s + \sum_j n_{l_j} r_l)\right) = 1,$$

$\forall t \geq 0$. Therefore from Lemma 2 we have $W(t) \equiv W(0)$ $\forall t \geq 0$.

As the number of bits in the system is a constant, this means that the number of flows in the system cannot be arbitrarily large. If Y_i is the number of customers in some phase i, and η_i is the mean file size associated with the phase i, then $W(t) > Y_i \eta_i$. Thus $Y_i(t) < \frac{W(0)}{\eta_i}, \forall t$. This upper bound is crucial in establishing the existence of certain limits as we now show.

Theorem 4: Given any initial condition $(n_{s_i}(0) i \in \mathcal{S}, n_{l_i}(0) i \in \mathcal{L})$, there exists an unique equilibrium point given by (8) where $W(0)$ denotes the work associated with the initial condition $(n_{s_i}(0) i \in \mathcal{S}, n_{l_i}(0) i \in \mathcal{L})$. Furthermore, if $(n_{s_i}(0) i \in \mathcal{S}, n_{l_i}(0) i \in \mathcal{L})$, satisfy the assumptions of Lemma 3 then as $t \to \infty$, $n_{s_i}(t) \to n^*_{s_i}$ $\forall i \in \mathcal{S}$ and $n_{l_i}(t) \to n^*_{l_i}$ $\forall i \in \mathcal{L}$, exponentially.

The proof of this theorem relies on a specific mapping which transforms the system into a simple processor sharing queue where the bandwidth received by a flow does not depend on whether the flow is a short-flow or a long-flow. Then we make use of a result from [5]. To state the result we develop the following preliminaries.

Consider a processor sharing queue operating at the critical load i.e., $\rho = 1$. Let $Q(t)$ denote the number of flows in the system at time t. The flows arrive into the system according to a Poisson process with mean rate λ. Let the file sizes of these flows (denoted by v) be independent and identically distributed according to a distribution function F. The mean file size of these flows is denoted by $\frac{1}{\alpha}$. The file-sizes of the flows that are present in the system at time $t = 0$ are again i.i.d and are distributed according to another distribution function G. The mean file size of these flows is denoted by $\frac{1}{\beta}$. Let $\zeta(t)$ denote the amount of data transferred by a flow by time t assuming that the flow was present in the system at time $t = 0$. In other order words,

$$\zeta(t) = \int_0^t \frac{1}{Q(s)} ds.$$

Let $\xi(z)$ denote the minimum time required to transfer z bits of data. In other words,

$$\xi(z) = inf\{t | \zeta(t) > z\}.$$

It is clear from the above definitions that ξ is the inverse function of $\zeta(t)$. Suppose that a flow has transmitted z bits of data by time t. Then the time taken by the flow to transmit an additional dz bits would be $Q(t)dz$. As the amount of data transmitted by time t is equal to $\zeta(t)$, it follows that

$$\xi'(\zeta(t)) = Q(t).$$

Let F_e (similarly G_e) denote the stationary excess distribution, i.e.,

$$F_e(y) = \alpha \int_0^y (1 - F(u)) du.$$

It has been shown in [5] that the fluid model of the processor sharing queue satisfies the following equation.

$$\xi'(z) = Q(0) G^c(z) + \lambda \int_0^z \xi'(z - u) dF_e(u), \tag{11}$$

where $G^c(z) = 1 - G(z)$ represents the CCDF (Complementary Cumulative Distribution Function). Intuitively, we can explain the equations in the following way. Substituting $z = \zeta(t)$ and $u = \zeta(s)$, in (11) we get,

$$Q(t) = Q(0) G^c(\zeta(t)) + \lambda \int_0^t F^c(\zeta(t) - \zeta(s)) ds. \tag{12}$$

Any flow which initiates transmission at time 0 would have transmitted $\zeta(t)$ bits of data by time t. Therefore, the fraction of flows that are still in the system at time t would be those whose files contain more than $\zeta(t)$ bits of data, i.e., $G^c(\zeta(t))$.

Thus, we get the first term of (12). Now any flow which arrives at time s, would have transmitted $\zeta(t) - \zeta(s)$ bits of data by time t ($t > s$). Thus the number of customers still remaining in the system is $F^c(\zeta(t) - \zeta(s))$, which gives the second term of (12). So the total number of flows in the system at any time t would be that fraction of flows which started transmitting at time 0 and have a file size which is more than $\zeta(t)$ bits and those which arrive at any time $s < t$ and have more than $\zeta(t) - \zeta(s)$ bits to transmit.

Now we state the required result ([5, Proposition 5]).

Lemma 5: When $\rho = 1$,

$$\lim_{t \to \infty} Q(t) = \frac{Q(0)}{\beta} \frac{2\alpha}{1 + c_v^2}, \tag{13}$$

where $c_v^2 = \alpha^2 Var(v)$. Further once the initial conditions are specified, the system converges exponentially fast i.e., $\exists$ constants K, t_0 and σ such that, $\forall\ t > t_0$

$$\left| Q(t) - \frac{Q(0)}{\beta} \frac{2\alpha}{1 + c_v^2} \right| < K e^{-\sigma(t - t_0)}$$

The proof of this theorem relies on the result from [18] which we state below.

Theorem 6: A distribution defined on $[0, \infty)$ is a phase type distribution if and only if

1) it has a point mass at zero, or
2) it has
    a) a strictly positive density on $(0, \infty)$, and
    b) has a rational Laplace Transform such that there exists a pole of maximal real part $-\sigma$ that is real, negative, and such that $-\sigma > \Re(-\gamma)$ where $-\gamma$ is any other pole.

Proof of Theorem 5: A proof of the statement (13) is given in [5], but exponential stability is not characterized there. In this paper, we prove the exponential stability using the final value theorem of Laplace transforms. Exponential stability is required to argue later (Section IV) that the fluid model approximates the actual stochastic system well.

Let $\mathfrak{L}$ denote the Laplace transform operator. Taking Laplace transforms of (11), we get,

$$\chi(s) = \frac{Q(0)\mathcal{G}(s)}{(1 - \mathcal{F}(s))},$$

where, $\chi(s) = \mathfrak{L}(\xi'(x)), \mathcal{G}(s) = \mathfrak{L}(G^c(x))$ and $\mathcal{F}(s) = \mathfrak{L}(\alpha F^c(x))$. To make use of the final value theorem, one should show that the poles of $\chi(s)$ are in the strict L.H.P with the exception of a simple pole at the origin. A complex number s is a pole of $\chi(s)$ if

- s is a pole $\mathcal{G}(s)$
- s is a zero of $(1 - \mathcal{F}(s))$

Let us first consider the zeroes of $1 - \mathcal{F}(s)$. For every s with $\Re(s) > 0$,

$$|\mathcal{F}(s)| = \int_0^\infty |\alpha F^c(x)||e^{-sx}| dx < \int_0^\infty \alpha F^c(x) dx = 1$$

Thus it follows that there cannot exist a zero of $1 - \mathcal{F}(s)$ in the strict R.H.P. Now suppose that for some $\omega \neq 0$, $\mathcal{F}(j\omega) = 1$. Then,

$$\begin{aligned} \mathcal{F}(j\omega) &= \int_0^\infty \alpha F^c(x) \cos(\omega x) dx \\ &\quad + j \int_0^\infty \alpha F^c(x) \sin(\omega x) dx = 1 \end{aligned}$$

This implies that $\int_0^\infty F^c(x) \sin(\omega x) dx = 0$ and $\int_0^\infty \alpha F^c(x) \cos(\omega x) dx = 1$. But the measure of the set of points at which $\cos(\omega x) = 1$ is zero. As Phase type distributions are non-arithmetic, we get

$$\left| \int_0^\infty \alpha F^c(x) \cos(\omega x) dx \right| < \left| \int_0^\infty \alpha F^c(x) dx \right| = 1$$

Thus it follows that for any $\omega \neq 0$, $j\omega$ cannot be a zero of $1 - \mathcal{F}(s)$. Also note that the origin is a zero of $1 - \mathcal{F}(s)$. Thus the origin is a pole of $\chi(s)$. To show that the origin is a simple pole, we can restrict ourselves to the real line. Now, differentiating the denominator we get,

$$\left. \frac{d}{ds} (1 - \mathcal{F}(s)) \right|_{s=0} = \alpha \int_0^\infty x F^c(x) dx > 0$$

It follows that the denominator cannot have multiple zeros at the origin.

Consider the poles of $\mathcal{G}(s)$. For every s with $\Re(s) \geq 0$,

$$|\mathcal{G}(s)| = \left| \int_0^\infty G^c(x) e^{-sx} dx \right| < \int_0^\infty G^c(x) = \frac{1}{\beta}$$

Thus, for all s with $\Re(s) \geq 0$, the value of $\mathcal{G}(s)$ is upper bounded. This implies that all the poles of $\mathcal{G}(s)$ are in the strict L.H.P. Further, as we have restricted ourselves to Phase Type distributions in this paper, by Theorem 6 it is clear that these poles will be bounded away from the $j\omega$ axis. Therefore $\chi(s)$ has all its poles bounded away from the $j\omega$ axis in the L.H.P, with the exception of a simple pole at the origin.

Using final value theorem, we get,

$$\begin{aligned} \lim_{t \to \infty} Q(t) &= \lim_{t \to \infty} \xi'(\zeta(t)) \\ &= \lim_{z \to \infty} \xi'(z) \\ &= \lim_{s \to 0} s\chi(s) = \frac{Q(0)}{\beta} \frac{2\alpha}{1 + c_v^2}, \end{aligned}$$

where the last equality follows from the fact that

$$\frac{1}{\beta} = \int_0^\infty G^c(z) dz$$

and

$$\lim_{s \to 0} \frac{1 - \mathcal{F}(s)}{s} = \int_0^\infty F_e^c(z) dz = \frac{1 + c_v^2}{2\alpha}.$$

Let $Q_\infty = \frac{Q(0)}{\beta} \frac{2\alpha}{1 + c_v^2}$. It is clear that $\xi'(x)$ converges to the equilibrium value exponentially fast. Thus by substituting $x = \zeta(t)$ we get,

$$Q(\xi(x)) = \xi'(x) \to Q_\infty,$$

exponentially. This implies that $\exists$ constants K, x_0 and σ such that $\forall\ x > t_0$,

$$|Q(\xi(x)) - Q_\infty| < K e^{-\sigma(x - x_0)}.$$

Substituting $x = \zeta(t)$ we get,

$$|Q(t) - Q_\infty| < K_1 e^{-\sigma\zeta(t)}.$$

$\forall\ t > \xi(x_0)$ As shown earlier, $Q(t) \to Q_\infty$ as $t \to \infty$, Thus given any $\epsilon > 0$, $\exists\ t_0$ such that $\forall t > t_0\ Q_\infty - \epsilon < Q(t) < Q_\infty + \epsilon$. Therefore $\forall\ t > t_0$

$$\begin{aligned} \zeta(t) &= \int_0^t \frac{1}{Q(s)} ds \\ &> \zeta(t_0) + \int_{t_0}^t \frac{1}{Q(s)} ds \\ &> \zeta(t_0) + \frac{t - t_0}{Q_\infty + \epsilon} \end{aligned}$$

Thus we get,

$$|Q(t) - Q_\infty| < K_2 e^{-\sigma_2 t}$$

for some constants K_2 and σ_2. This implies that $Q(t)$ converges to its equilibrium exponentially fast. ∎

Now we return to the proof of Theorem 4. To show that the convergence is exponential, we need is classical result from perturbation theory. The following lemma and its proof can be found in [13, Theorem 5.8].

Lemma 7: Let

$$\dot{x} = f(x) + g(x)$$

be a non-linear system. Suppose that $x = 0$ be an exponentially stable equilibrium point of the nominal system

$$\dot{x} = f(x)$$

Then if

$$|g(x(t))| \overset{exp}{\underset{}{\xrightarrow{t\to\infty}}} 0,$$

the perturbed system

$$\dot{x} = f(x) + g(x)$$

is also exponentially stable.

Proof of Theorem 4: To begin with, we make the following transformation,

$$\begin{aligned} \tilde{n}_{s_i} &= n_{s_i} w_s, \\ \tilde{n}_{l_i} &= n_{l_i} w_l. \end{aligned}$$

The system of equations is now changed to,

$$\begin{aligned} \dot{\tilde{n}}_{l_i} &= \lambda_l w_l - (w_l \mu_{l_i}) \tilde{n}_{l_i} \tilde{x}_l && i = 1, \\ &= p_{l_{i-1}} (w_l \mu_{l_{i-1}}) \tilde{n}_{l_{i-1}} \tilde{x}_l - w_l \mu_{l_i} \tilde{n}_{l_i} \tilde{x}_l && , i \neq 1, \\ \dot{\tilde{n}}_{s_i} &= \lambda_s w_s - (w_s \mu_{s_i}) \tilde{n}_{s_i} \tilde{x}_s && i = 1, \\ &= p_{s_{i-1}} (w_s \mu_{s_{i-1}}) \tilde{n}_{s_{i-1}} \tilde{x}_s - (w_s \mu_{l_s}) \tilde{n}_{s_i} \tilde{x}_s && i \neq 1. \end{aligned}$$

where

$$\begin{aligned} \tilde{x}_s &= \frac{1}{\sum_{i \in \mathcal{S}} \tilde{n}_{s_i} + \sum_{j \in \mathcal{L}} \tilde{n}_{l_j}}, \\ \tilde{x}_l &= \frac{1}{\sum_{i \in \mathcal{S}} \tilde{n}_{s_i} + \sum_{j \in \mathcal{L}} \tilde{n}_{l_j}}. \end{aligned}$$

The transformation has produced a scaled version of the original fluid model. In the new fluid model, the arrival rates of the long-flows and short-flows have been scaled by factors w_l and w_s respectively. Further, the mean file size of every phase has been scaled appropriately. The load on the new system is the same as before and thus the new fluid model is critically loaded. The main difference between the models is that the bandwidth received by a flow, now no longer depends on whether the flow is a short-flow or a long-flow ($\tilde{x}_s = \tilde{x}_l$), i.e., the resulting fluid model represents the fluid model for simple processor sharing at the router. Now therefore, by Lemma 5, it follows that the number of customers in the system converges to a constant as $t \to \infty$ exponentially fast. Denote

$$\tilde{n}_{tot_e} = \lim_{t \to \infty} \sum_{i \in \mathcal{S}} \tilde{n}_{s_i} + \sum_{i \in \mathcal{L}} \tilde{n}_{l_i}$$

Consider the dynamics of $\tilde{n}_{l_1}$.

$$\dot{\tilde{n}}_{l_1} = \lambda_l w_l - (w_l \mu_{l_i}) \frac{\tilde{n}_{l_i}}{\tilde{n}_{tot_e}} + g(\tilde{n}(t)) \tag{14}$$

where,

$$\begin{aligned} g(\tilde{\mathbf{n}}(t)) &= (w_l \mu_{l_i}) \left(\frac{\tilde{n}_{l_i}}{\tilde{n}_{tot_e}} - \tilde{n}_{l_i} \tilde{x}_l \right) \\ &= w_l \mu_{l_i} \tilde{n}_{l_i} \frac{\sum_i \tilde{n}_{l_i} + \sum_i \tilde{n}_{s_i} - \tilde{n}_{tot_e}}{\tilde{n}_{tot_e} (\sum_i \tilde{n}_{l_i} + \sum_i \tilde{n}_{s_i})} \end{aligned}$$

It is easy to see that,

$$|g(\tilde{\mathbf{n}}(t))| \overset{exp}{\xrightarrow{t\to\infty}} 0,$$

Therefore by Lemma 7, it follows that (14) is exponentially stable. The same argument can be applied repeatedly, to each of the phases to show that they converge exponentially fast. As the number of phases is finite, it follows that the system converges to its equilibrium point exponentially fast. ∎

III. FLUID MODELS WITH PRIORITIES

In this section, we compare the equilibrium results of the previous section to the equilibrium flow rates of a system where short flows have priority over long flows. We first develop a fluid model for the priority scheme and then perform the comparison between the two schemes.

The short-flows transmit data while the underlying TCP is in the slow-start phase. In the slow-start phase, the window size is slowly increased from 1. The window size doubles each RTT if all the acknowledgments are received. Thus, it is possible that even though enforcing priority ensures that the short-flows are served first, the short-flows may not end up utilizing all the bandwidth since TCP is not designed for data transmissions at such high rates in the slow-start phase. We model this by imposing a limit r_s on the maximum rate at which a short-flow can transmit the data. This would ensure that the maximum rate achieved by the short-flows cannot exceed the inherent limit imposed by TCP. The simulations conducted show that the bandwidth received by the short-flows is very close to this limit.

In this section, we analyze two scenarios. In the first scenario, we assume $r_s < \infty$ and thus model the inherent limit imposed by TCP. In the second scenario, we assume that the underlying transmission protocol has the ability to transmit at arbitrary rates (i.e. $r_s = \infty$). In such a case, the bandwidth of the link can be utilized completely by the short-flows. An interesting result is that, even if $r_s = \infty$, the degradation suffered by long-flows is not significant.

As before, we assume that the arrival processes of the short-flows (long-flows) is Poisson with rate $\lambda_s(\lambda_l)$. The file-sizes of the short-flows are distributed according to a Coxian distribution with mean $\frac{1}{\mu_s}(\frac{1}{\mu_l})$. The bandwidth received by short-flows at any time is given by

$$x_s = \min(r_s, \frac{1}{\sum_{i\in\mathcal{S}} n_{s_i}}). \tag{15}$$

Similarly, the bandwidth received by the long-flows is given by,

$$x_l = \max(\frac{1 - \sum_{i\in\mathcal{S}} n_{s_i} r_s}{\sum_{i\in\mathcal{L}} n_{l_i}}, \frac{I_{\{\sum_{i\in\mathcal{S}} n_{s_i}=0\}}}{\sum_{i\in\mathcal{L}} n_{l_i}}). \tag{16}$$

Again, we resort to the fluid model analysis of the above system under critical load. Using techniques similar to the one described in Section IV, it is easy to show that the stochastic model converges to the fluid model on compact intervals, under appropriate scaling. Then, the evolution of the number of flows in any phase i, is governed by (6) with x_s and x_l given by. (15)-(16).

The stationary point of these differential equations is given by

$$\rho_{s_i} \quad = \quad n^*_{s_i} r_s \quad \rho_{l_i} \quad = \quad n^*_{l_i} x_l. \tag{17}$$

Assuming that the initial conditions satisfy Lemma 3, the equilibrium point can be determined uniquely, if the amount of work left in the system at time $t = 0$ is given.

A. Equilibrium Analysis

At equilibrium, the transmission rates of the short-flows would be equal to r_s. If $x_{s_{wp}}(x_{l_{wp}})$ denotes the bandwidth received by short-flows (long-flows) when priorities are enforced and $x_{s_{wop}}(x_{l_{wop}})$ denotes the bandwidth received by short-flows (long-flows) when priorities are not enforced, then it is easy to see that

$$\frac{x_{s_{wp}}}{x_{s_{wop}}} = \frac{r_s W(0)}{\left(\sum_{i\in\mathcal{S}} \frac{\rho_{s_i}}{\nu_{s_i}} \frac{w_l}{w_s} + \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}}\right)}.$$

The above analysis shows, that if the initial load on the system is large, then when there are no priorities, the bandwidth received by the short-flows is very small. With priorities, the short-flows are guaranteed a rate r_s at equilibrium. Thus one can have arbitrarily large improvement in the performance of the short-flows. We now calculate the degradation seen by the long-flows. From (17) we have,

$$n_{s_i} = \frac{\rho_{s_i}}{r_s}.$$

Define $\mathcal{W}_{se}$ to be the amount of work left in the system due to the presence of short-flows. As the system is critically loaded, it follows that $W(t)$ is invariant with time. Thus,

$$W(0) - \mathcal{W}_{se} = \sum_{i\in\mathcal{L}} n_{l_i} \left(\frac{1}{\mu_{l_i}} + \sum_{k>i:i\in\mathcal{L}} \frac{1}{\mu_{l_k}} \prod_{j=i}^{j=k-1} p_{l_j} \right).$$

Defining $\tilde{\mathcal{W}}_{le} = \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}}$ we get

$$n_{l_i} = \frac{(W(0) - \mathcal{W}_{se})\rho_{l_i}}{\tilde{\mathcal{W}}_{le}}.$$

Therefore, the bandwidth received by the long-flows is given by

$$x_l = \frac{\tilde{\mathcal{W}}_{le}}{(W(0) - \mathcal{W}_{se})}.$$

Thus the degradation faced by the long-flows is given by

$$\frac{x_{l_{wp}}}{x_{l_{wop}}} = \frac{\tilde{\mathcal{W}}_{le} W(0)}{(W(0) - \mathcal{W}_{se})\left(\sum_{i\in\mathcal{S}} \frac{\rho_{s_i}}{\nu_{s_i}} \frac{w_l}{w_s} + \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}}\right)}.$$

To get a bound on this degradation, we now consider the situation where $r_s \to \infty$. Clearly, if there is no transmission rate limit on the short-flows, the long-flows are worse off. Further, if $r_s = \infty$, then we have

$$\dot{\mathcal{W}}_s(t) = \rho_s - \min(\sum_{i\in\mathcal{S}} n_{s_i} r_s, 1) = \rho_s - 1.$$

It follows that $\mathcal{W}_s(t)$ is monotonically decreasing function and $\dot{\mathcal{W}}_s(t) = 0$ if and only if $n_{s_i} = 0 \ \forall i \in \mathcal{S}$. Therefore, at equilibrium, $\mathcal{W}_{se} = 0$. Thus,

$$\frac{x_{l_{wp}}}{x_{l_{wop}}} = \frac{\sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}}}{\left(\sum_{i\in\mathcal{S}} \frac{\rho_{s_i}}{\nu_{s_i}} \frac{w_l}{w_s} + \sum_{i\in\mathcal{L}} \frac{\rho_{l_i}}{\nu_{l_i}}\right)}. \tag{18}$$

Note that for a given $W(0)$ the degradation suffered by the long-flows when r_s is finite is less than the degradation suffered when the r_s is infinite. Thus the degradation suffered by long-flows is bounded and is actually independent of $W(0)$. We now provide some numerical insights into the above expressions. Let us suppose that the short-flows and the long-flows are modeled by a single exponential phase. Approximating the mean behavior of the Internet traffic, we assume $\frac{1}{\mu_l} = \frac{10}{\mu_s}$, $\rho_s = 0.2$, $\rho_l = 0.8$. Further, based on simulation experiments, we take $w_s = 1$ and $w_l = 3$. Using these numbers, we get $\frac{x_{l_{wp}}}{x_{l_{wop}}} = 0.909$. Thus, the long-flows see a reduction in bandwidth of only 10%, while the short-flow performance can improve dramatically, especially if the work in the system is very high. A 10% degradation might still be considerable for some applications, but note that a single exponential phase cannot capture the heavy tailed property of the Internet. In Section V, we show via simulations that if the file-size distribution is heavy tailed, the degradation will be considerably less.

IV. CONVERGENCE TO THE FLUID MODEL

The validity of the fluid model is shown in two steps. Over compact intervals, the validity follows from a celebrated result by Kurtz [15]. We then make use of the fact that the fluid model is stable to show convergence over larger intervals.

A. Convergence on Compact Intervals

As described earlier, the validity of the fluid model is established from a theorem by Kurtz [15]. To state the result we need the following preliminaries. For each $N \in \mathcal{N}$, let $X^N(t)$ denote a continuous time Markov process on the countable state space $\{\frac{k}{N} : k \in \mathcal{Z}_+^m\}$ with transition rates $q_{\frac{k}{N},\frac{k+p}{N}} \equiv N f_p^N(\frac{k}{N})$. The functions f_p^N are assumed to satisfy the following properties.

$$\begin{aligned} |f_p^N(\mathbf{x})| &< \epsilon_p(1+\|x\|) \\ |f_p^N(\mathbf{x}) - f_p(\mathbf{x})| &< \frac{\Gamma\epsilon_p}{N}(1+\|x\|), \end{aligned} \tag{19}$$

for some Γ, ϵ such that $0 < \Gamma, \epsilon < \infty$. Then it can be shown that $X_N(t)$ can be obtained as a solution of the following stochastic equation.

$$X_N(t) = X_N(0) + \sum_p \frac{1}{N}\mathbf{p}Y_p(N\int_0^t f_p^N(X_N(s))ds), \tag{20}$$

where $\{Y_p\}$ are Poisson processes with unit mean (see [15] for more details). Let $X(t)$ denote the corresponding fluid model.

$$X(t) = X(0) + \int_0^t \sum_p \mathbf{p} f_p(X(s))ds \tag{21}$$

We want to show that in some sense, for large N, the stochastic process will have a sample path that would be very close to the solution of the fluid model. The following theorem by Kurtz shows that this is indeed true under some assumptions.

Theorem 8: Let

$$F(x) = \sum_p \mathbf{p} f_p(x).$$

If

$$|F(x) - F(y)| \leq M\|x-y\| \quad \textstyle\sum_p |\mathbf{p}|\epsilon_p < \infty,$$

and if

$$\lim_{N\to\infty} X_N(0) = X(0),$$

then assuming (19) holds,

$$\lim_{N\to\infty} \sup_{t\leq T} |X_N(t) - X(t)| = 0 \ a.s,$$

on every compact interval $[0, T]$.

To make use of Kurtz'z lemma, one should define appropriate functions f_p^N and show that the limiting function f_p is Lipshitz continuous. As discussed before, We assume that the short-flows and the long-flows can transmit at maximum rates r_s and r_l. Further we assume, that the rate limits satisfy the following property.

$$\frac{r_l}{r_s} = \frac{w_l}{w_s}.$$

Under these assumptions, the rates received by the short-flows and long-flows are as follows:

$$\begin{aligned} x_s &= r_s \text{ if } \textstyle\sum_j n_{s_i} r_s + \sum_j n_{l_i} r_l \leq 1 \\ &= \frac{w_s}{\sum_j n_{s_i} w_s + \sum_j n_{l_i} w_l} \text{ otherwise} \\ x_l &= r_l \text{ if } \textstyle\sum_j n_{s_i} r_s + \sum_j n_{l_i} r_l \leq 1 \\ &= \frac{w_l}{\sum_j n_{s_i} w_s + \sum_j n_{l_i} w_l} \text{ otherwise} \end{aligned} \tag{22}$$

We now describe the Markov process that dictates the evolution of the queue under DPS. As the time spent by a file in each phase is exponential, the number of files in each phase i, would denote the Markov process. i.e. $X^1(t) = (\{n_{s_i}\}, \{n_{l_i}\})$ (Here we use the convention that $X^1(t)$ is a vector of dimension $m_s + m_l$ and the number of files in short-flow phases are in positions $1, 2 ... m_s$ and the number of files in long-flow phases are in positions $m_s+1, ..., m_s+m_l$). The functions $f_p^1(\mathbf{n})$ are given as follows :

$$\begin{aligned} f_p^1(\mathbf{n}) &= \lambda_s & \mathbf{p} &= e_1 \\ &= \lambda_l & \mathbf{p} &= e_{m_s+1} \\ &= \mu_{l_i} n_{l_i} x_l & \mathbf{p} &= e_{m_s+i+1} - e_{m_s+i} \\ &= \mu_{s_i} n_{s_i} x_s & \mathbf{p} &= e_{i+1} - e_i \\ &= \mu_{l_{m_l}} n_{l_{m_l}} x_l & \mathbf{p} &= -e_{m_s+m_l} \\ &= \mu_{l_{m_s}} n_{l_{m_s}} x_s & \mathbf{p} &= -e_{m_s} \\ &= 0 & &\text{otherwise} \end{aligned} \tag{23}$$

The Markov process $X^1(t)$ can be expressed as a solution of the stochastic differential equation (20). The corresponding fluid limit is given by the following set of differential equations.

$$\begin{aligned} \dot{n_{l_i}} &= \begin{cases} \lambda_l - \mu_{l_i} n_{l_i} x_l & i = 1, \\ p_{l_{i-1}} \mu_{l_{i-1}} n_{l_{i-1}} x_l - \mu_{l_i} n_{l_i} x_l & i \in \mathcal{L}, i \neq 1, \end{cases} \\ \dot{n_{s_i}} &= \begin{cases} \lambda_s - \mu_{s_i} n_{s_i} x_s & i = 1, \\ p_{s_{i-1}} \mu_{s_{i-1}} n_{s_{i-1}} x_s - \mu_{l_s} n_{l_s} x_s & i \in \mathcal{S}, i \neq 1. \end{cases} \end{aligned} \tag{24}$$

To prove the fluid limit, we need the following lemma and its corollary.

Lemma 9: Let $F_i(x)$ be a function on p-dimensional vector $\mathbf{x}$ defined as follows.

$$F_i(\mathbf{x}) = \begin{cases} x_i r & \text{If } r\sum_j x_j \leq 1 \\ \frac{x_i}{\sum_j x_j} & \text{otherwise} \end{cases}$$

Then $F_i(x)$ is Lipshitz continuous i.e. $\exists\, M < \infty$ such that,

$$|F_i(\mathbf{x}) - F_i(\mathbf{y})| < M\|x-y\|$$

where $\|\cdot\|$ denotes the $\mathcal{L}^2$ norm.

Proof: Note that F_i is a scalar function. The gradient of F_i is

$$\nabla F = \begin{cases} \hat{e}_i & \text{if } \sum_j x_j \leq 1 \\ \hat{e}_i \dfrac{\sum_{j \neq i} x_j}{(\sum_j x_j)^2} - \sum_j \hat{e}_j \dfrac{x_i}{(\sum_j x_j)^2} & \text{otherwise} \end{cases}$$

where $\hat{e}_x, \hat{e}_y$ denote the unit vectors in the directions x and y. It follows that the gradient $|\nabla F|$ is bounded $\forall$ $\mathbf{x} \in \Re^{m_s+m_l}$. (Note: Where the function is not differentiable, gradient is defined as that vector in which the directional derivative (sub gradient) achieves the maximum norm. It is easy to show that the directional derivatives are bounded.) Further, as F is a scalar potential function, it follows that the line integral of ∇F is independent of the path chosen. Therefore, for a linear path γ such that $\gamma(0) = x$ and $\gamma(1) = y$

$$F(\mathbf{y}) - F(\mathbf{x}) = \int_\gamma \nabla F \cdot dl$$

From this it follows that,

$$|F(\mathbf{x}) - F(\mathbf{y})| \leq \int_\gamma |\nabla F||dl| \leq M \big\| x - y \big\|$$

■

Corollary 10: Suppose function $F_i(\mathbf{x})$ is defined as follows:

$$F_i(\mathbf{x}) = \begin{cases} x_i r_i & \text{If } \sum_j x_j r_j \leq 1 \\ \dfrac{x_i w_i}{\sum_j x_j w_j} & \text{otherwise} \end{cases}$$

where, $\dfrac{r_i}{w_i} = \dfrac{r_j}{w_j}$ $\forall i, j$. Then $F_i(\mathbf{x})$ is Lipshitz continuous.

Proof: Substituting $N_i = x_i w_i$, it is easy to see that the function $F_i(\cdot)$ satisfies the assumptions of Lemma 9. Thus we have,

$$\begin{aligned} |F(\mathbf{x}) - F(\mathbf{y})| &\leq M \left(\sum_j (x_j w_j - y_j w_j)^2 \right)^{\frac{1}{2}} \\ &\leq M \max_j |w_j| \big\| x - y \big\| \end{aligned}$$

■

We can now state the main result of this section in the form of the following theorem.

Theorem 11: Consider the DPS queue described by (20) with the functions $f_p(\mathbf{n}), f_p^N(\mathbf{n})$ given by (23). (In our model, N can be thought of as the parameter that scales the arrival rates of the flows and the link capacity) The corresponding fluid model is (21) whose differential form is given by (24). Then

$$\lim_{N \to \infty} \sup_{[0,T]} \big\| X_N(t) - X(t) \big\| \xrightarrow{N \to \infty} 0 \text{ a.s.},$$

over every compact interval $[0, T]$.

Proof: To describe a DPS queue, the functions $f_p^N(\mathbf{n}) \equiv f_p^1(\mathbf{n}) \equiv f_p(\mathbf{n})$ $\forall p$. Thus (19) hold trivially. Further, as proved in Lemma 9, $f_p(\mathbf{n})$ is Lipshitz continuous. Therefore, the conditions of Theorem 8 are satisfied. Thus, the stochastic model converges to the fluid limit (24) uniformly on compact intervals. ■

B. Convergence of the Fluid Model on Larger Intervals

So far, we showed that the scaled stochastic system converges to its fluid model on finite interval. In this section, we show that the scaled stochastic system converges to the fluid model over large intervals of time (of order N^δ). Our work is along the lines of [20], [8] in principle, though the proof techniques are different.

The following lemma might be hidden in some textbook. But we are unaware of this precise result. As this is needed to prove Theorem 14, we present the lemma and its proof below.

Lemma 12: Consider the following non-linear system

$$\dot{x} = f(x), \tag{25}$$

where $x \in \Re^n$, $f : D \to \Re^n$ is a continuously differentiable map from a domain $D \subset \Re^n$ into $\Re^n$. Further suppose that the origin is exponentially stable. Then given any $\epsilon > 0$ $\exists$ $\delta, \delta_1 > 0$ such that if $|g(t)| < \delta$ and $|x(0)| < \delta_1$ then the trajectories of the perturbed system

$$\dot{x} = f(x + g(t)) \tag{26}$$

satisfy $\big\| x(t) \big\| \leq \epsilon$, $\forall$ $t > 0$.

Proof: As (25) is exponentially stable, by the converse Lyapunov theorem (see Theorem 3.12 in [13]) $\exists$ a Lyapunov function $V(x)$ that satisfies,

$$\begin{aligned} c_1 \big\| x \big\|^2 \quad \leq \quad V(x) &\leq c_2 \big\| x \big\|^2 \\ \frac{\partial V}{\partial x} f(x) &\leq -c_3 \big\| x \big\|^2 \\ \Big\| \frac{\partial V}{\partial x} \Big\| &\leq c_4 \big\| x \big\| \end{aligned}$$

for some positive constants c_1, c_2, c_3, c_4.

Fix $\delta_1 < \infty$ such that $\forall$ x s.t $\big\| x \big\| < \delta_1$, $x \in D$. As the system is exponentially stable, it follows that

$$\Big\| \frac{\partial f}{\partial x} \Big\|_{x=0} = c < \infty$$

As f is continuously differentiable in D, it follows that

$$\sup_{\| x \| \leq \delta_1} \big\| \frac{\partial f}{\partial x} \big\| = c_6 < \infty.$$

Consider the perturbed system (26). By the Mean Value Theorem, $\exists$ $z_i \in [x, x + g(t)]$ such that

$$f_i(x + g(t)) = f(x) + \frac{\partial f_i}{\partial x}\Big|_{x=z_i} g(t)$$

Therefore, the perturbed system can be written as follows

$$\dot{x} = f(x) + \frac{\partial f}{\partial x}\Big|_{x=x} g(t) + h(x, t),$$

where

$$h_i(x,t) = \left(\frac{\partial f_i}{\partial x}\bigg|_{x=z_i} - \frac{\partial f}{\partial x}\bigg|_{x=x} \right) g(t).$$

As f is assumed to be differentiable, it follows that as $\|g(t)\| \to 0$,

$$\frac{|h_i(x,t)|}{\|g(t)\|} \to 0.$$

This implies that for as $\|g(t)\| \to 0$,

$$\frac{\|h(x,t)\|}{\|g(t)\|} \leq \sum_{i=1}^{n} \frac{|h_i(x,t)|}{\|g(t)\|} \to 0.$$

Consider the Lyapunov function $V(x)$. Differentiating w.r.t t we get,

$$\dot{V} = \frac{\partial V}{\partial x} f(x+g(t)) = \frac{\partial V}{\partial x} f(x) + \frac{\partial V}{\partial x}\frac{\partial f}{\partial x} g(t) + \frac{\partial V}{\partial x} h(x,t)$$

Assuming $\|g(t)\| \leq \delta$, we get,

$$\dot{V} \leq -c_3 \|x\|^2 + c_4 c_6 \|x\| \delta + c_4 \|x\| \|h(x,t)\| \delta$$

Therefore $\forall$ δ sufficiently small and $\forall$ x such that $\|x\| > \left(\frac{c_4(c_6+1)}{c_3}\right)\delta$

$$\dot{V} < 0$$

Therefore, if $\|x(0)\| < \left(\frac{c_4(c_6+1)}{c_3}\right)\delta$, it is clear that $\forall t$ $\|x(t)\| < \left(\frac{c_4(c_6+1)}{c_3}\right)\delta$. By making δ arbitrarily small, we have the desired result. ■

Lemma 13: Let Y_p denote a Poisson process with rate λ. Then

$$\sup_{[0,N^\delta T]} \left| \frac{Y_p(Nt)}{N} - \lambda t \right| \overset{N\to\infty}{\longrightarrow} 0 \text{ a.s}$$

for every $\delta \in [0,1)$ and for every $T \in [0,\infty)$.

Proof: Let Z denote the inverse process of $Y_p(t)$ i.e.,

$$Z(n) = \min\{t | Y_p(t) \geq n\}$$

Thus, we have

$$Z(n) = \sum_{i=1}^{n} \nu_i,$$

where ν_i are independent and exponentially distributed with mean $\frac{1}{\lambda}$. if n is not an integer then an appropriate flooring function will be used. Applying Chernoff bounds we get,

$$Prob\left\{\frac{Z(Nt)}{N} - \frac{t}{\lambda} > \epsilon\right\} = Prob\left\{ \frac{\sum_{i=1}^{\lfloor Nt \rfloor} \nu_i - \frac{Nt}{\lambda}}{Nt} > \frac{\epsilon}{t} \right\}$$
$$\leq Ke^{-Ntl(\frac{\epsilon}{t})}$$

where the error exponent $l(a)$ satisfies,

$$l(a) = \frac{(a\lambda)^2}{a\lambda+1} + \left(\frac{a\lambda}{a\lambda+1} - \log(1+a\lambda) \right).$$

Assuming a is small and expanding $\log(1+a\lambda)$ around 1 we get,

$$\begin{aligned} l(a) &\geq \frac{(a\lambda)^2}{a\lambda+1} + \left(\frac{a\lambda}{a\lambda+1} - a\lambda + \frac{(a\lambda)^2}{2} - \frac{(a\lambda)^3}{3} \right) \\ &= \frac{(a\lambda)^2}{2} - \frac{(a\lambda)^3}{3} \end{aligned}$$

Therefore by applying union bound we get,

$$\begin{aligned} & Prob\left\{ \sup_{[0,N^\delta T]} \frac{Z(Nt)}{N} - \frac{t}{\lambda} > \epsilon \right\} \\ &\leq \sum_{i=1}^{\lfloor N^\delta T \rfloor} Prob\left\{\frac{Z(Ni)}{N} - \frac{t}{\lambda} > \epsilon\right\} \\ &\leq N^\delta T K e^{-(N^{1-\delta}T) l_1(\epsilon,T)} \end{aligned}$$

where,

$$l_1(\epsilon,T) = \frac{\epsilon^2\lambda^2}{2T^2} - \frac{\epsilon^3\lambda^3}{3N^\delta T^3} > 0$$

Therefore

$$Prob\left\{ \sup_{[0,N^\delta T]} \frac{Z(Nt)}{N} - \frac{t}{\lambda} > \epsilon \right\} \overset{N\to\infty}{\longrightarrow} 0,$$

exponentially fast. Similarly it is easy to show that,

$$Prob\left\{ \sup_{[0,N^\delta T]} \frac{Z(Nt)}{N} - \frac{t}{\lambda} < -\epsilon \right\} \overset{\overset{exp}{N\to\infty}}{\longrightarrow} 0.$$

Therefore,

$$Prob\left\{ \sup_{[0,N^\delta T]} \left| \frac{Z(Nt)}{N} - \frac{t}{\lambda} \right| > \epsilon \right\} \overset{\overset{exp}{N\to\infty}}{\longrightarrow} 0.$$

By Borel-Cantelli Lemma, it follows that

$$\sup_{[0,N^\delta T]} \left| \frac{Z(Nt)}{N} - \frac{t}{\lambda} \right| \overset{N\to\infty}{\longrightarrow} 0 \text{ a.s}$$

Now consider the Poisson process $Y_p(t)$. By definition,

$$\begin{aligned} & Prob\left\{ \sup_{t\in[0,N^\delta T]} \left| \frac{Y_p(Nt) - \lambda Nt}{N} \right| > \epsilon \right\} \\ &\leq Prob\left\{ \sup_{[0,N^\delta \lambda T+\epsilon]} \left| \frac{Z(Nt)}{N} - \frac{t}{\lambda} \right| > \epsilon \right\} \overset{\overset{exp}{N\to\infty}}{\longrightarrow} 0. \end{aligned}$$

Applying Borel Cantelli Lemma again, we get,

$$\sup_{[0,N^\delta T]} \left| \frac{Y_p(Nt)}{N} - \lambda t \right| \overset{N\to\infty}{\longrightarrow} 0 \text{ a.s}$$

for every $\delta \in [0,1)$ and for every $T \in [0,\infty)$. ■

Remark 2: If $f(x)$ is any continuous bounded non-negative function, then

$$\sup_{[0,N^\delta T]} \left| \frac{Y_p(\int_0^t f(s)ds)}{N} - \lambda \int_0^t f(s)ds \right| \overset{N\to\infty}{\longrightarrow} 0 \text{ a.s}$$

The proof is very similar to the proof of Lemma 13.

Theorem 14: Let $\{X_N(t)\}$ denote a sequence of scaled stochastic systems representing the modified DPS queue. The modified DPS queue is assumed to be critically loaded. Further, suppose that the initial conditions of the stochastic system and the fluid limit satisfy Lemma 3. Then,

$$\sup_{[0,N^{\delta}T]} |X_N(t) - X(t)| \overset{N\to\infty}{\longrightarrow} 0 \text{ a.s}$$

for every $T > 0$ and for every $\delta \in [0.1)$.

Proof: The N^{th} scaled system is given by

$$X_N(t) = X_N(0) + \frac{1}{N}\sum_{\mathbf{p}} \mathbf{p}Y_p(N\int_0^t f_l(X_N(s))ds)$$

Adding and subtracting $\sum_p \int_0^t f_p(X_N(s))ds$ we get,

$$X_N(t) = X_N(0) + \sum_{\mathbf{p}} \int_0^t f_p(X_N(s))ds + g_N(t),$$

where

$$\begin{aligned} g_N(t) = \quad & \frac{1}{N}\sum_{\mathbf{p}} \mathbf{p}Y_p(N\int_0^t f_l(X_N(s))ds) \\ & -\sum_{\mathbf{p}} \int_0^t f_p(X_N(s))ds \end{aligned}$$

The function $g_N(t)$ is a perturbation about the nominal fluid model. As the functions $\{f_p(x)\}$ are bounded uniformly, by Lemma 13, it follows that

$$\sup_{[0,N^{\delta}T]} \left|g_N(t)\right| \overset{N\to\infty}{\longrightarrow} 0 \text{ a.s}$$

The fluid model is given by,

$$X(t) = X(0) + \sum_{\mathbf{p}} \int_0^t f_p(X_N(s))ds$$

Defining $Z_N(t) = X_N(t) - g_N(t)$, and differentiating we get

$$\dot{Z}_N = f(Z_N + g_N)$$

We have shown before that the fluid model *converges exponentially* to its equilibrium determined by the initial conditions. Given $W(0)$ and $\{n_{s_1}, n_{s_2}...n_{s_{m_s}}, n_{l_1}, ., n_{l_{i-1}}, n_{l_{i+1}}, ...n_{l_{m_l}}\}$, it is possible to determine n_{l_i} exactly. Therefore, using W(0), one can eliminate n_{l_i} from the equations. Therefore, we get a modified fluid model whose state-space has only $m_s + m_l - 1$ dimensions. As $W(0)$ is fixed, this implies that this modified dynamics will be *exponentially stable.* Also, since the modified dynamics represent the same system as the fluid limit, the trajectories of both the systems will be exactly same.

Consider the modified fluid model and the modified stochastic model with $m_s + m_l - 1$ dimensions. The modified dynamics would be denoted by $X^{m_i}(t)$, $X_N^{m_i}(t)$ and $Z_N^{m_i}(t)$ respectively. Similarly, the "perturbation" function in this smaller dimensional state space would be denoted by $g_N^{m_i}(t)$. To make the notation simple, we will assume that the origin is the equilibrium point.

By Lemma 12, it follows that if the $|Z_N^{m_i}(0)| < \delta_1$ and $\left\|g_N(t)\right\| < \delta_2$ then $\forall t$ $\left\|Z_N^{m_i}(t)\right\| \leq \epsilon$ for any given $\epsilon > 0$. Thus, as $X^{m_i}(t)$ is exponentially stable, it follows that $\exists$ δ_3 such that if $\left\|X^{m_i}(0)\right\| < \delta_3$, $\left\|X^{m_i}(t)\right\| < \epsilon$ $\forall$ $t \geq 0$.

Let $\delta_4 = \frac{1}{2}\min(\delta_1, \delta_2, \delta_3, \epsilon)$. Fix $T > 0$ and $0 < \delta < 1$. By Lemma 13, $\exists$ N_1^* large enough such that $\forall$ $N > N_1^*$,

$$\left\|g_N(t)\right\| \leq \delta_4 \ \forall \ t \in [0, N^{\delta}T] \text{ a.s .}$$

Also, as the fluid model is exponentially stable, it follows that $\exists$ t_0 such that $\forall$ $t > t_0$,

$$\left\|X^{m_i}(t)\right\| \leq \delta_3.$$

By Corollary 11, we know that the stochastic model converges to the fluid model u.o.c. Therefore, $\exists$ N_2^* large enough such that such that $\forall$ $N > N_2*$,

$$\left\|X(t) - X_N(t)\right\| \leq \delta_4 \text{ a.s}$$

$\forall$ $t \in [0, t_0]$. Therefore it follows that

$$\left\|Z_N^{m_i}(t_0)\right\| \leq 2\delta_4 \text{ a.s}$$

By Lemma 12, it therefore follows that $\forall$ $N > \max(N_1^*, N_2^*)$ and $\forall$ $t \in [t_0, N^{\delta}T]$,

$$\left\|Z_N^{m_i}(t)\right\| \leq \epsilon \text{ a.s}$$

Therefore, $\forall t \in [0, N^{\delta}T]$, we have,

$$\left\|X^{m_i}(t) - X_N^{m_i}(t)\right\| \leq 2\epsilon \text{ a.s}$$

So far we have shown that the trajectories in the $m_s + m_l - 1$ dimensions, stay "very close" to the fluid limit over large intervals of time. Since the dimension that was eliminated was arbitrary and due to the fact that N_1^* and N_2^* do not depend on the dimension eliminated, it follows that $\forall$ $N > \max(N_1^*, N_2^*)$

$$\begin{aligned} \left\|X(t) - X_N(t)\right\| & \leq \left\|X^{m_i}(t) - X_N^{m_i}(t)\right\| \\ & \quad + \left\|X^{m_{i-1}}(t) - X_N^{m_{i-1}}(t)\right\| \\ & \leq 4\epsilon \text{ a.s} \end{aligned}$$

$\forall$ $t \in [0, N^{\delta}T]$.

■

V. Simulation results

In this section, we study the effect of priorities on the system via NS-2 simulations. We consider a 100 Mb bottleneck link, which is accessed by many flows (Fig. 1). All the flows are TCP file transfers. The round trip time (RTT) of the flows are uniformly distributed between 40ms to 60ms. Also, we fix the packet size of all flows to be 1000 bytes.

Parameter	Short-flows	Long-flows
c	10kB	100kB
d	100kB	100MB
α	1.1	1.1

TABLE I
PARAMETERS OF THE BOUNDED PARETO DISTRIBUTION

It has been suggested that the traffic in the Internet is approximately heavy tailed [7]. A heavy tailed distribution is of the form

$$P(X > x) \sim x^{-\alpha} \; 1 < \alpha < 2.$$

In the case of the Internet, it is reasonable to assume that the file-sizes have minimum and maximum values. Therefore, we model the file-sizes to be i.i.d according to a distribution that has heavy tail form, but has finite upper and lower bounds. This truncated distribution is called a bounded Pareto (b.p) distribution [4]. A random variable X distributed according to a b.p distribution takes values in the interval $[c \; d]$ with the following CDF :

$$Prob(X < x) = \frac{c^{-\alpha} - x^{-\alpha}}{c^{-\alpha} - d^{-\alpha}}, \; 1 < \alpha < 2.$$

In the current Internet, the short-flows are a few tens of kilobytes long while the long-flows range from several hundred kilobytes to a few megabytes. To capture this behavior, we assume that short-flow file-sizes are distributed according to a b.p. distribution and long-flow file-sizes are distributed according to another independent b.p. distribution. The chosen parameter values are given in Table I. As the b.p. distribution is not a phase-type distribution, we cannot directly compare the simulation results with the results in the preceding sections. So we approximate the b.p distribution by a 10-phase Coxian distribution.

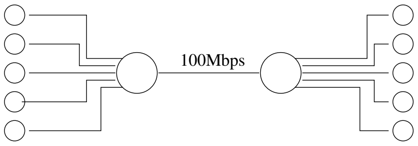

Fig. 1. Simulated Topology

Fitting Coxian distributions to various distributions has been well studied in the literature (For example, see [3], [2], [11] and the references there in). In this section, we use the EM (Expectation-Maximization) algorithm developed by Asmussen et al [3]. The EMpht program [2], which is based on EM algorithm, was used to fit phase type distributions. The results of the fit are shown in Figures 2 and 3.

The files arrive into the system according to a Poisson process. Measurements of the Internet have indicated that around $70 - 90\%$ of the flows in the Internet are short-flows. Hence, in the simulations, we model each arrival to be a short-flow with a probability p where $p \in [0.7, 0.9]$.

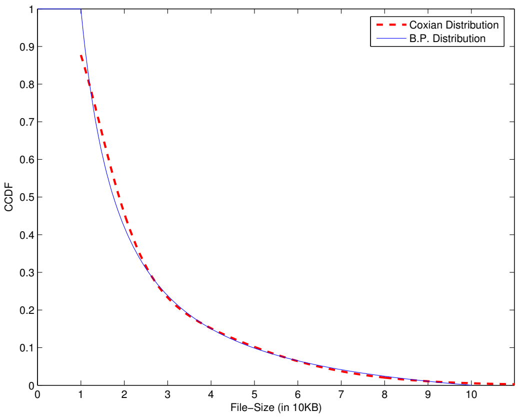

Fig. 2. Coxian fit for the Short Flow distribution

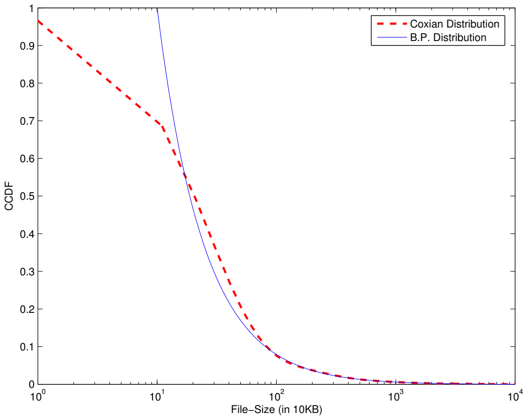

Fig. 3. Coxian fit for the Long Flow distribution

The mean file-size of the short-flows is 25kB. The mean file-size of the long-flows is 540kB. When the priorities are not enforced, the buffer size is chosen to be 1000 packets. When the priorities are enforced, 2 queues are formed at the router, one containing the packets belonging to the short-flows and the other containing the packets belonging to the long-flows. The packets of the long-flows are served only when the there are no packets belonging to short-flows in the queue. The buffer size for these queues are chosen to be proportional to the offered loads. In other words, if b_s and b_l denote the buffer sizes of short-flow queue and long-flow queue then $\frac{b_s}{b_l} = \frac{\rho_s}{\rho_l}$. To make a fair comparison, the total buffer size is unchanged, i.e., $b_s + b_l = 1000$ packets.

The load on the system is varied by varying the average file arrival rate. The load offered by short-flows is varied by varying the probability p. For $p = 0.7$, the load offered by the short-flows is about 10% of the overall load. Similarly for $p = 0.9$, the load offered by the short-flows is about 30% of the overall load.

The file-sizes of long-flows (similarly short-flows) are randomly generated according to the phase type distribution fit to its b.p. distribution. After the transmission is complete, the file leaves the system. By measuring the time between the file's arrival and departure, one can estimate the mean bandwidth received by the file. Averaging this over all long-flows (similarly short-flows) would give us the average bandwidth received by the long-flows (similarly short-flows). The results of these NS-2 simulations are given in Figures 4, 5 and 6. In these simulations, the short-flows offer a total load of 0.3.

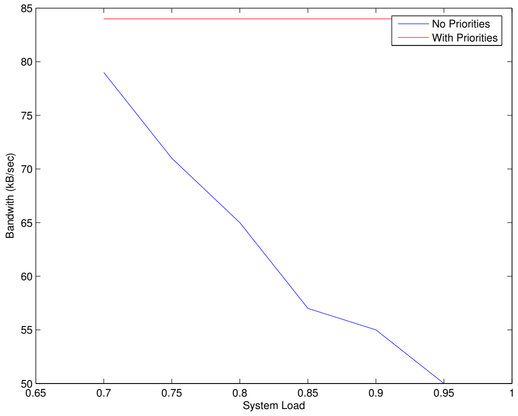

Fig. 4. Bandwidth received by short-flows with and without priorities

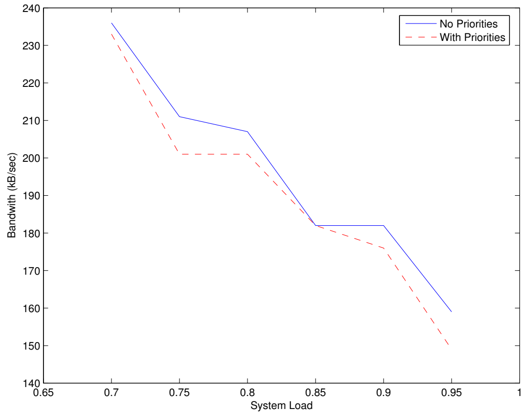

Fig. 5. Bandwidth received by long-flows with and without priorities

These results clearly indicate that by giving priorities one can significantly improve the performance of short-flows while the loss seen by long-flows is minimal. When no priorities are enforced, at low system loads, the effect of priorities is not significant due to inherent limitations of TCP. However, as the load increases, the effects of congestion dominates and short-flows receive inferior throughputs. With priority, the short-flows do not see any congestion and therefore, their bandwidths are significantly improved. As Figure 6 demonstrates, the short-flow bandwidths can be improved by a factor of 70% while the degradation seen by long-flows is almost non-existent. Numerous experiments have been done at various other loads and similar behavior is seen.

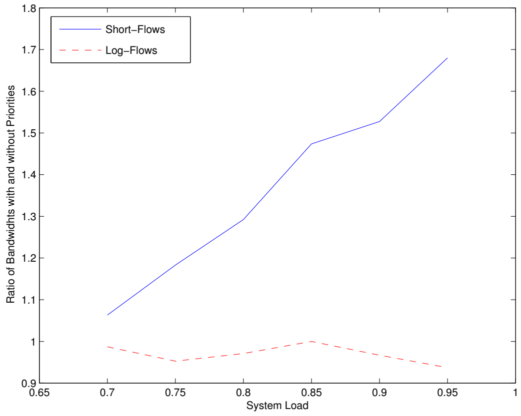

Fig. 6. The ratio of bandwidths received by short-flows and long-flows with and without Priorities

In the above results, we find that when priorities are enforced the short-flows get about 84kB/sec. This result seems to be independent of the system load and the load offered by the short-flows. We conjecture that this behavior is due to the inherent limitation of the TCP. To justify our conjecture we do the following. We calculate a bound on the average rate at which the short-flows can transmit when priorities are enforced. To this end, we assume that since the short-flows contribute to a very small fraction of the total load, they do not see any losses when priorities are enforced. The short-flows are in the slow-start phase of TCP and thus the window size is exponentially increased starting from a window size of 1. Thus, if a particular flow has a file-size y kB, then the time taken by that flow to transmit the entire file is,

$$T(y,t) = (\lceil \log_2(y+1) - 1 \rceil + 2) \cdot t,$$

where t denotes the RTT of the flow and $\lceil \cdot \rceil$ denotes the ceiling function. The additional 2 RTTs appear in the equation do to the fact that TCP uses one RTT to exchange the SYN messages and another RTT to exchange FIN messages. The bandwidth received by the flow is

$$B_s(y,t) = \frac{y}{T(y,t)}$$

Assuming y is distributed according to a distribution F and RTT is uniformly distributed between a and b, we have

$$r_s = \frac{1}{b-a} \int_a^b \int_0^\infty \frac{y}{T(y,t)} dF(y) dt$$

Assuming the file-sizes are distributed according to a heavy tailed distribution with parameter values given in Table I, the above expression can be numerically evaluated. From this we get $r_s = 89$kB/sec. From the simulation results we see that the short-flows on average receive about 84 kB/sec which is very close to the limiting rate r_s.

VI. CONCLUSIONS

In this paper, we have introduced a simple fluid model to study the performance of long and short-flows in the Internet. We have shown that when the router is severely congested, giving priority to short-flows will significantly enhance their throughput without significantly degrading the data rates received by the long-flows. The improvement seen by short flows is more pronounced if the initial load on the system is large. Further, irrespective of the average number of bits left in the system, the long-flows suffer a only a very small degradation due to the nature of the traffic distribution in the Internet.

REFERENCES

[1] C. Psounis A. Ghosh and B. Prabhakar. SIFT: a simple algorithm for identifying large flows, 2004. Presented at the Stochastic Networks Conference, Montreal, Canada.

[2] S. Asmussen, O. Nerman, and M. Olson. EMpht-program for fitting Phase type distributions via EM algorihtm, 1996. Available at http://www.maths.lth.se/matstat/ staff/asmus/pspapers.html.

[3] S. Asmussen, O. Nerman, and M. Olson. Fitting Phase type distributions via EM algorithm. *Scandinavian Journal of Statistics*, pages 419–441, 1996.

[4] N. Bansal and M. Harchol-Balter. Analysis of SRPT scheduling: Investigating unfairness. In *Proceedings of ACM Sigmetrics*, 2001.

[5] H. Chen, O. Kella, and G. Weiss. Fluid approximations for a processor-sharing queue. *Queueing Systems*, pages 99–125, 1997.

[6] X. Chen and J Heidemann. Preferential treatment for short flows to reduce web latency. *Computer Networks*, 2003.

[7] M.E. Crovella and A. Bestavros. Self-similarity in the World Wide Web traffic: Evidence and possible causes. *IEEE/ACM transactions on Networking*, pages 835–846, 1997.

[8] S. Deb, S. Shakkottai, and R. Srikant. Asymptotic behavior of internet congestion controllers in a many-flows regime. *Mathematics of Operations Research*.

[9] C. Estan and G. Varghese. New directions in traffic measurement and accounting: Focussing on the Elephants, ignoring the Mice. *ACM Transactions on Computer Systems*, August 2003.

[10] G. Fayolle, I. Mitrani, and R. Iasnogorodski. Sharing a processor among many classes. *Journal of the ACM*, pages 519–532, 1980.

[11] A Feldmann and W Whitt. Fitting mixtures of exponentials to long tail distributions to analyze network performance analysis. In *Proceedings of INFOCOM*, 1997.

[12] N Bansal M. Harchol-Balter. Analysis of SRPT scheduling: Investigating unfairness. In *Proceedings of the 2001 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, 2001.

[13] H. Khalil. *Nonlinear Systems*. 2nd edition, Prentice Hall, Upper Saddle River, NJ, 1996.

[14] L. Kleinrock. Time shared systems : A theoretical treatment. *Journal of ACM*, pages 242–261, 1967.

[15] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and Applications*, pages 223–240, 1978.

[16] A. Lakshmikantha, R. Srikant, and C. L. Beck. Processor sharing versus priority sharing schemes for tcp flows in internet routers. Submitted to Conference on Decision and Control, 2005.

[17] A. Lakshmikantha, R. Srikant, and C. L. Beck. Performance analysis of priority queueing schemes in Internet routers. In *Conference on Information Sciences and Systems*, 2005.

[18] O'Cinneide. A characterization of phase-type distributions. *communications in Statistics Stochastic Models*, pages 1–57, 1990.

[19] T.M. O'Donovan. Direct solutions of M/G/1 processor sharing models. *Operations Research*, pages 1232–1235, 1974.

[20] S. Shakkottai and R. Srikant. Mean FDE models for Internet congestion control. Technical report. A shorter version to appear in the *Proceedings of IEEE INFOCOM*, 2002.