On the Output Rate of Overloaded Link Schedulers

Jörg Liebeherr, *Fellow, IEEE*, Yashar Ghiassi-Farrokhfal, *Member, IEEE*

***Abstract*—We derive and compare properties of the output rates at overloaded links for three types of link scheduling algorithms: First-in-First-Out (FIFO), Static-Priority (SP), and Earliest-Deadline-First (EDF). Under most general assumptions, i.e., each traffic flow has a long-term average rate, we show that the output rates of flows at overloaded FIFO and EDF links are proportional to their input rates. As a consequence, the service rate guarantee offered by EDF and FIFO cannot exceed the guarantee given to a low-priority flow under SP scheduling.**

I. Introduction

THE service experienced by a traffic flow in a packet network is influenced by the properties of the scheduling algorithms that determine the transmission order of backlogged traffic in packet switches. The performance of scheduling algorithms has been extensively studied in the underloaded regime, where the aggregate arrival rate to a buffered link is smaller than the service rate of the link. For overloaded links, fair queuing algorithms, which include Round Robin and Generalized Processor Sharing are known to provide isolation between traffic flows [1]. However, the properties of other scheduling algorithms in overload, in particular, FIFO, have been largely ignored.

The need to understand the properties of classical scheduling algorithms in overload became apparent when a study on bandwidth estimation for the Internet [2] resorted to empirical measurements to determine the output rate of probe traffic at an overloaded FIFO link with cross traffic. The study reported that the output rate of probe traffic is proportional to the share of the total offered load. The result was used in [3] to conjecture that the service guarantee to a traffic flow at an overloaded FIFO scheduler is as bad as the service given to a flow with lowest priority at a priority scheduler. These observations and conjectures were proved analytically in [4] for constant-bit-rate (CBR) traffic, and extended to random traffic with exponentially bounded burstiness and burstiness bounded by a power-law [5]. This has raised the question if the same properties hold for more general traffic scenarios.

In this letter, we show that rate proportional sharing at overloaded FIFO links is valid for any deterministic or random traffic scenario, as long as arrivals have a long term average traffic rate. We also show that the same result extends to links with EDF scheduling. As a consequence, the service guarantee of EDF and FIFO, when expressed in terms of the rate of a lower service curve, cannot exceed that of a low-priority flow with SP scheduling.

The remainder is structured as follows. In Sec. II we present the system model. In Sec. III we derive properties of general work-conserving links. In Sec. IV we analyze EDF scheduling in overload, where we treat FIFO as a special case. In Sec. V, we present the corresponding results for SP. In Sec. VI, we derive the long-term rate of the service guarantee offered by FIFO and EDF. We present brief conclusions in Sec. VII.

J. Liebeherr (jorg@comm.utoronto.ca) is with the Dept. ECE, University of Toronto, Toronto, ON, Canada M5S 3G4 and Y. Ghiassi-Farrokhfal (yashar.ghiassi@uwaterloo.ca) is with the CS Dept., University of Waterloo, Waterloo, ON, Canada N2L 3G1.

II. System Model

We model a buffered link as a work-conserving queueing system with an infinite buffer and a fixed service rate C. Arrivals from a set of flows can be *discrete*, representing instantaneous arrivals of packets with a given size, *fluid flow*, representing arrivals that occur at a constant rate, or a mix of both. Fluid flow arrivals occurring to an empty buffer at a total rate C or less depart immediately without delay. In all other cases, arrivals are added to the buffer and served at rate C. The departure time of a discrete-sized arrival is the departure of the last piece of the arrival (e.g., the last bit of a packet). Arrivals can be deterministic or described by a random process. Random service times of tasks are expressed in terms of the arrival process. For example, an M/M/1 queue is represented using discrete-size arrivals of exponentially distributed size with exponential interarrival times.

A scheduling algorithm selects backlogged arrivals for service, where the backlog consists of arrivals that have not yet departed. We consider three scheduling algorithms: *First-in-first-out (FIFO)*, *Earliest-deadline-first (EDF)*, and *Static Priority (SP)*. FIFO services the backlog in the order of arrivals. EDF services in the order of deadlines that are determined as follows. Each flow j is associated with a constant $0 \leq d_j < \infty$. An arrival from flow j at time t is tagged with $t + d_j$ as its deadline. We assume that EDF services the backlog in the order of deadlines, even when deadlines have been missed. (An alternative realization of EDF discards backlog with expired deadlines.) Note that FIFO can be viewed as a special case of EDF with $d_j = 0$ for all flows j. SP assigns a priority level to each flow, and services backlog with the highest priority first. Within the same priority, backlog is served in the order of arrivals. The analysis in this paper is done for preemptive scheduling, which assumes that the transmission of a packet can be interrupted, however, all results in this paper are easily extended to a non-preemptive service model with some additional notation.

Denote by $A_j(t)$ and $D_j(t)$ the arrivals and departures, respectively, from flow j in the time interval $[0, t)$. These functions are non-negative and non-decreasing with $A_j(t) = D_j(t) = 0$ for $t \leq 0$, and satisfy the causality condition $A_j(t) \geq D_j(t)$. We assume that arrivals of flow j have a long-term average rate ρ_j in the sense that

$$\rho_j = \lim_{t \to \infty} \frac{A_j(t)}{t} > 0 \,. \qquad (1)$$

We denote the total arrivals and departures, respectively, from all flows by $A(t) = \sum_j A_j(t)$ and $D(t) = \sum_j D_j(t)$. Also, we write $\rho = \sum_j \rho_j$. When convenient, we use the notation $A(s,t) = A(t) - A(s)$ or $D(s,t) = D(t) - D(s)$ for $s \le t$.

The backlog at time t is $B(t) = A(t) - D(t)$. The (virtual) delay at time t is $W(t) = \inf\{y \ge 0 \mid D(t) \ge A(t-y)\}$. A busy period is a time period with positive backlog. We define $\underline{t} \le t$ as the start time of the busy period containing time t. In other words, $\underline{t}$ is the last time before or at time t where the buffer is empty. We can write this as

$$\underline{t} = \sup\{\upsilon \le t \mid A(\upsilon) = D(\upsilon)\} .$$

The backlog, delay, and start of a busy period of an individual flow j, B_j, W_j, and $\underline{t}_j$, are defined analogously by replacing arrivals and departures by A_j and D_j, respectively.

III. Work-conserving Schedulers in Underload and Overload

For the above system model, we first derive results that hold for all work-conserving scheduling algorithms. We refer to a system as overloaded when $\rho > C$, i.e., the aggregate average arrival rate exceeds the service rate. In this case, some arrivals experience infinite delays. We refer to a system with $\rho < C$ as underloaded. When $\rho = C$, the state of the system depends on the arrival scenario. With deterministic arrivals, finite delays may be achievable for $\rho = C$. However, with random arrivals delays generally become unbounded for $\rho \to C$.

Lemma 1: A work-conserving link with $\rho < C$ satisfies *(i)* $\lim_{t\to\infty} \underline{t} = \infty$, and *(ii)* $\lim_{t\to\infty} \frac{\underline{t}}{t} = 1$.

The first claim implies that an underloaded scheduler has infinitely many busy periods of finite length. The second claim is stronger, stating that the beginning of the most recent busy period has the same scaling as t.

Proof: For every time t, we have

$$A(t) \ge D(t) = A(\underline{t}) + C(t - \underline{t}) . \tag{2}$$

Suppose to the contrary that $\lim_{t\to\infty} \underline{t}$ is finite, that is, there exists a final busy period of infinite length. Then, we obtain that the link cannot be underloaded:

$$\rho = \lim_{t\to\infty} \frac{A(t)}{t} \ge \lim_{t\to\infty} \left(\frac{A(\underline{t})}{t} + C \frac{t - \underline{t}}{t} \right) = C ,$$

where we used (2) and $\lim_{t\to\infty} \frac{A(\underline{t})}{t} = 0$. This contradicts the assumption $\rho < C$, and the first claim follows. For the second claim, we consider again (2), and write

$$\begin{aligned} \frac{A(t)}{t} &\ge \frac{\underline{t}}{t} \frac{A(\underline{t})}{\underline{t}} + \frac{t - \underline{t}}{t} \rho + \frac{t - \underline{t}}{t} (C - \rho) \\ &\ge \min\left\{ \frac{A(\underline{t})}{\underline{t}} , \rho \right\} + \frac{t - \underline{t}}{t} (C - \rho) , \end{aligned}$$

where we use that a convex combination is at least as large as the smaller of its terms. Taking the limit $t \to \infty$, we get

$$\rho \ge \lim_{t\to\infty} \left(\min\left\{ \frac{A(\underline{t})}{\underline{t}} , \rho \right\} \right) + \limsup_{t\to\infty} \frac{t - \underline{t}}{t} (C - \rho) .$$

Since $\lim_{t\to\infty} \frac{A(\underline{t})}{\underline{t}} = \rho$ according to the first claim, the second term must be equal to zero. Thus, it must hold that $\lim_{t\to\infty} \frac{t - \underline{t}}{t} = 0$, or, equivalently, $\lim_{t\to\infty} \frac{\underline{t}}{t} = 1$. ■

This leads to the following result for underloaded links.

Theorem 1: A work-conserving link with $\rho < C$ satisfies for all flows j that

$$\lim_{t\to\infty} \frac{D_j(t)}{t} = \rho_j .$$

The theorem implies that the departure rate in the underloaded regime is not sensitive to the scheduling algorithm.

Proof: As $\lim_{t\to\infty} \frac{A_j(t)}{t} = \rho_j$, and, by Lemma 1, we have

$$\lim_{t\to\infty} \frac{A_j(\underline{t})}{t} = \lim_{t\to\infty} \frac{A_j(\underline{t})}{\underline{t}} \lim_{t\to\infty} \frac{\underline{t}}{t} = \rho_j .$$

Since the departures of flow j at time t are bounded by $A_j(t) \ge D_j(t) \ge A_j(\underline{t})$, the claim follows by the squeeze lemma. ■

An overloaded scheduler eventually becomes permanently backlogged. As shown in the next lemma, such a system has a last idle time, which implies that there is a final busy period of infinite length.

Lemma 2: A work-conserving scheduler with $\rho > C$ satisfies $\lim_{t\to\infty} \underline{t} < \infty$.

Proof: Suppose there is no last idle time. Then, there exists an increasing sequence of time instants $\{I_n\}_{n\in\mathbb{N}}$ which grows to ∞ where the link is idle. Clearly,

$$\rho = \lim_{t\to\infty} \frac{A(t)}{t} = \lim_{n\to\infty} \frac{A(I_n)}{I_n} .$$

Since I_n is an idle time, the arrivals $A(I_n)$ are less than or equal to CI_n, the maximum output until time I_n. This yields $A(I_n) \le CI_n$ for all n, and the limit

$$\lim_{n\to\infty} \frac{A(I_n)}{I_n} \le C .$$

Thus, the link is not overloaded. Therefore, there is a last time when the scheduler is idle. ■

IV. Output Rates of EDF and FIFO

We next look at the output rate of a link with EDF scheduling. Clearly, in an overloaded state, the EDF scheduler does not meet the deadlines of arrivals.

Lemma 3: At a link with EDF scheduling where $\rho > C$, each flow j satisfies $\lim_{t\to\infty} \underline{t}_j < \infty$.

Proof: Consider an arrival from flow j at time t that occurs in the final busy period. Let $B^j(t)$ denote the backlog in the buffer at time t with deadlines equal to or less than $t + d_j$. We can provide a lower bound by

$$\begin{aligned} B^j(t) &\ge \sum_k A_k(\underline{t}, t - [d_k - d_j]^+) - C(t - \underline{t}) \\ &\ge A(\underline{t}, t - \max_k d_k) - C(t - \underline{t}) , \end{aligned}$$

where we use $[x]^+ = \max\{x, 0\}$. Note that equality need not hold in the first line since, in $[\underline{t}, t)$, the link may serve backlog with a deadline greater than $t + d_j$. Dividing by t and taking the limit, we have

$$\begin{aligned} \liminf_{t\to\infty} \frac{B^j(t)}{t} &\ge \sum_k \lim_{t\to\infty} \frac{A_k(t - \max_k d_k)}{t} \\ &\qquad - \lim_{t\to\infty} \frac{A(\underline{t})}{t} - C \lim_{t\to\infty} \frac{t - \underline{t}}{t} \\ &= \rho - C , \end{aligned}$$

which holds due to Lemma 2, and since

$$\lim_{t\to\infty} \frac{A_k(t-\max_k d_k)}{t-\max_k d_k} \cdot \lim_{t\to\infty} \frac{t-\max_k d_k}{t} = \rho_k \,,$$

where we used that the limit of a sum (product) is equal to the sum (product) of the limits. This yields

$$\liminf_{t\to\infty} B^j(t) = \liminf_{t\to\infty} \frac{B^j(t)}{t} \cdot \lim_{t\to\infty} t \geq (\rho - C) \lim_{t\to\infty} t \,,$$

Hence, the backlog with a deadline at or before $t + d_j$ grows at least linearly in t, i.e., $\liminf_{t\to\infty} B^j(t) = \infty$.

The delay of a flow-j arrival at time t, $W_j(t)$, must account for the service backlog with earlier deadlines, that is, $W_j(t) \geq \frac{B^j(t)}{C}$, which gives

$$\liminf_{t\to\infty} \frac{W_j(t)}{t} \geq \frac{\rho - C}{C} > 0 \,,$$

meaning that $W_j(t)$ also grows at least linearly in t. Then, the claim follows immediately, since an infinitely growing waiting time implies that there exists a time after which flow j is permanently backlogged. ■

Theorem 2: The output of a flow j at a link with EDF scheduling satisfies

$$\lim_{t\to\infty} \frac{D_j(t)}{t} = \rho_j \min\Big\{1, \frac{C}{\rho}\Big\} \,. \tag{3}$$

Thus, when the link is overloaded, EDF achieves a bandwidth allocation that is proportional to the average input rate.

Proof: For $\rho < C$, the claim follows from Theorem 1. For $\rho > C$, we consider a time t with a departure from flow j. Let u_t denote the arrival time of the departure, that is, $D_j(t) = A_j(u_t)$. By Lemma 3, we can select t sufficiently large so that all flows are permanently backlogged. Hence, we can express the departures from each flow $k \neq j$ until time t as

$$D_k(t) = A_k(u_t - [d_k - d_j]^+) \,.$$

Summing over all flows, dividing by t, and taking the limit yields

$$\begin{aligned} \lim_{t\to\infty} \frac{D(t)}{t} &= \sum_k \Bigg(\lim_{t\to\infty} \frac{A_k(u_t - [d_k - d_j]^+)}{u_t - [d_k - d_j]^+} \\ &\qquad \cdot \lim_{t\to\infty} \frac{u_t - [d_k - d_j]^+}{t} \Bigg) \\ &= \rho \lim_{t\to\infty} \frac{u_t}{t} \,. \end{aligned} \tag{4}$$

Another way to express the long-term output rate at an overloaded scheduler is

$$\begin{aligned} \lim_{t\to\infty} \frac{D(t)}{t} &= \lim_{t\to\infty} \frac{A(\underline{t}) + D(\underline{t}, t)}{t} \\ &= \lim_{t\to\infty} \frac{A(\underline{t})}{t} + \lim_{t\to\infty} \frac{C(t - \underline{t})}{t} = C \,, \end{aligned} \tag{5}$$

where we used that, according to Lemma 2, $\underline{t}$ is bounded. Combining (4) and (5), we obtain for $\rho < C$ that

$$\lim_{t\to\infty} \frac{u_t}{t} = \frac{C}{\rho} \,. \tag{6}$$

Using (6), we now express the long-term output rate of the departures of flow j as

$$\lim_{t\to\infty} \frac{D_j(t)}{t} = \lim_{t\to\infty} \frac{A_j(u_t)}{t} = \frac{\rho_j}{\rho} \, C \,.$$

For $\rho = C$, the terms in the minimum in (3) are identical, and the claim holds since, for each t, $D_j(t)$ grows monotonically if $A_j(t)$ is increased (while holding other arrivals fixed). ■

The same rate proportional allocation exists in a FIFO system. Thus, in overload, the service of FIFO and EDF becomes indistinguishable when observed over longer time periods.

Corollary 1: The output of a flow j at a link with FIFO scheduling satisfies (3).
This follows immediately since EDF scheduling with $d_k = 0$ for each flow k is equal to FIFO.

V. Output Rate of Static Priority

In SP, each flow has a priority level p. We use the convention that a larger priority level indicates a higher priority. For simplicity, we assume one flow per priority level, which allows us to use the priority level as flow index.

Consider an arbitrary time t, and define $\underline{t}_{\geq p}$ as the last time when the buffer did not contain any arrivals from priority p or higher, i.e.,

$$\underline{t}_{\geq p} = \sup\{ \upsilon \leq t \mid A_q(\upsilon) = D_q(\upsilon), q \geq p \} \,.$$

In a time period $[\underline{t}_{\geq p}, t)$, a link with SP scheduling only services arrivals from priority p or higher.

Lemma 4: A link with SP scheduling satisfies:

(i) If $\sum_{q\geq p} \rho_q < C$, then $\lim_{t\to\infty} \underline{t}_{\geq p} = \infty$ and $\lim_{t\to\infty} \frac{\underline{t}_{\geq p}}{t} = 1$.

(ii) If $\sum_{q\geq p} \rho_q > C$, then $\lim_{t\to\infty} \underline{t}_{\geq p} < \infty$.

As a consequence, in SP, the condition that a link is overloaded depends on the priority level.

Proof: For every time t, a link with SP scheduling satisfies

$$\sum_{q\geq p} A_q(t) \geq \sum_{q\geq p} D_q(t) = \sum_{q\geq p} A_q(\underline{t}_{\geq p}) + C(t - \underline{t}_{\geq p}) \,. \tag{7}$$

Comparing (7) with (2), the proofs of the claims in *(i)* proceed as in Lemma 1, with the following substitutions:

$$\begin{array}{lcllcl} \rho & \rightarrow & \sum_{q\geq p} \rho_q \,, & A(t) & \rightarrow & \sum_{q\geq p} A_q(t) \,, \\ \underline{t} & \rightarrow & \underline{t}_{\geq p} \,, & D(t) & \rightarrow & \sum_{q\geq p} D_q(t) \,. \end{array}$$

The proof of claim *(ii)* is analogous to Lemma 2, by substituting $\sum_{q\geq p} \rho_q$ for ρ and $\underline{t}_{\geq p}$ for $\underline{t}$, and instead of arguing that the scheduler is idle, we argue that the scheduler does not have a backlog from priority levels $q \geq p$. ■

Theorem 3: Given a link with SP scheduling as described above. The output of flow p satisfies

$$\lim_{t\to\infty} \frac{D_p(t)}{t} = \min\Big\{ \rho_p, \big[C - \sum_{q>p} \rho_q\big]^+ \Big\} \,. \tag{8}$$

Proof: Using (7), dividing by t, and taking the limit we obtain

$$\limsup_{t\to\infty} \frac{\sum_{q\geq p} D_p(t)}{t} = \lim_{t\to\infty} \frac{\sum_{q\geq p} A_q(\underline{t}_{\geq p})}{t} + \lim_{t\to\infty} \frac{C(t - \underline{t}_{\geq p})}{t} \,.$$

By Lemma 4, the limits appearing on the right-hand side exist. If $\sum_{q\geq p}\rho_q < C$ the limits are

$$\lim_{t\to\infty}\frac{\sum_{q\geq p}A_q(t_{\geq p})}{t} = \sum_{q\geq p}\rho_q \quad \text{and} \quad \lim_{t\to\infty}\frac{t - t_{\geq p}}{t} = 0\,.$$

If $\sum_{q\geq p}\rho_q > C$, the limits are

$$\lim_{t\to\infty}\frac{\sum_{q\geq p}A_q(t_{\geq p})}{t} = 0 \quad \text{and} \quad \lim_{t\to\infty}\frac{t - t_{\geq p}}{t} = 1\,.$$

The case $\sum_{q\geq p}\rho_q = C$ follows by a monotonicity argument analogously to the case $\rho = C$ in the proof of Theorem 2. Considering all cases, we have

$$\lim_{t\to\infty}\frac{\sum_{q\geq p}D_p(t)}{t} = \min\Big\{\sum_{q\geq p}\rho_q, C\Big\}\,. \tag{9}$$

Using (9), since $D_p(t) = \sum_{q\geq p}D_p(t) - \sum_{q>p}D_p(t)$, we obtain

$$\begin{aligned}\lim_{t\to\infty}\frac{D_p(t)}{t} &= \min\Big\{\sum_{q\geq p}\rho_q, C\Big\} - \min\Big\{\sum_{q>p}\rho_q, C\Big\}\\ &= \begin{cases}\rho_p\,, & \text{if } \sum_{q\geq p}\rho_q \leq C\,,\\ C - \sum_{q\geq p}\,, & \text{if } \sum_{q>p}\rho_q \leq C < \sum_{q\geq p}\rho_q\,,\\ 0\,, & \text{if } C < \sum_{q>p}\rho_q\,,\end{cases}\end{aligned}$$

which is equal to (8). ■

VI. Service Guarantees

In the network calculus [6], a work-conserving link with cross traffic is represented by a service element that provides a service guarantee to a flow. The minimum service guarantee of a flow j is expressed by a (lower) service curve S_j, a non-decreasing non-negative function, which satisfies $D_j(t) \geq \inf_{0\leq s\leq t}\{A_j(s) + S_j(t-s)\}$ for all $t \geq 0$ for any arrival and departure function A_j and D_j. The following corollary to Theorem 2 (and Corollary 1) states that the long-term rate of a service curve of a flow at a link with EDF or FIFO scheduling cannot exceed the available rate left unused by the other flows, and is in fact identical to the rate obtained by a flow with lowest priority at a link with SP scheduling.

Corollary 2: Consider a FIFO or EDF link where all arrivals satisfy (1). When the service of a flow j is described by a service curve S_j for flow j with a longterm rate (i.e., $\lim_{t\to\infty}\frac{S_j(t)}{t}$ exists), it holds that

$$\lim_{t\to\infty}\frac{S_j(t)}{t} \leq \Big[C - \sum_{k\neq j}\rho_k\Big]^+.$$

Proof: Suppose the rate of the service curve of flow j is $\lim_{t\to\infty}\frac{S_j(t)}{t} = \mu_j > 0$. Then, for every $\varepsilon > 0$, there exists a t_o such that for all $t > t_o$ we have

$$\Big|\rho_j - \frac{A_j(t)}{t}\Big| < \varepsilon \quad \text{and} \quad \Big|\mu_j - \frac{S_j(t)}{t}\Big| < \varepsilon\,.$$

Select $t > 2t_o$. Since S_j is a service curve it holds that

$$\frac{D_j(t)}{t} \geq \inf_{0\leq s\leq t}\Big\{\underbrace{\frac{A_j(s)}{t} + \frac{S_j(t-s)}{t}}_{=:\mathrm{F(s,t)}}\Big\}\,.$$

For the evaluation of the infimum, we break its range into three subintervals $0 \leq s \leq t_o$, $t_o \leq s \leq t - t_o$, and $t - t_o \leq s \leq t$, and take the minimum of the results. For the first subinterval, since $A_j(s)$ is non-decreasing, we obtain that

$$\inf_{0\leq s\leq t_o}\mathrm{F(s,t)} \geq (\mu_\mathrm{j} - \varepsilon)\cdot\frac{\mathrm{t}-\mathrm{t_o}}{\mathrm{t}}\,.$$

For the second subinterval, we get

$$\inf_{t_o\leq s\leq t-t_o}\mathrm{F(s,t)} \geq (\rho_\mathrm{j}-\varepsilon)\cdot\frac{\mathrm{s}}{\mathrm{t}} + (\mu_\mathrm{j}-\varepsilon)\cdot\frac{\mathrm{t}-\mathrm{s}}{\mathrm{t}} \geq \min\{\rho_\mathrm{j},\mu_\mathrm{j}\} - \varepsilon\,,$$

since both $s \geq t_o$ and $t - s \geq t_o$. For the third subinterval, since $S_j(s)$ is non-decreasing, we get

$$\inf_{t-t_o\leq s\leq t}\mathrm{F(s,t)} \geq (\rho_\mathrm{j} - \varepsilon)\cdot\frac{\mathrm{t}-\mathrm{t_o}}{\mathrm{t}}\,.$$

Collecting terms, taking $t \to \infty$ (and using that t_o is finite), and then selecting ε arbitrarily small, we obtain

$$\lim_{t\to\infty}\frac{D_j(t)}{t} \geq \lim_{t\to\infty}\inf_{0\leq s\leq t}\mathrm{F(s,t)} \geq \min\{\rho_\mathrm{j},\mu_\mathrm{j}\}\,. \tag{10}$$

Now consider an overloaded link with $\rho > C$ and $\rho_j = [C - \sum_{k\neq j}\rho_k]^+ + \varepsilon$, where $\varepsilon > 0$. In this case, (3) and (10) provide that $\lim_{t\to\infty}\frac{D_j(t)}{t} = \rho_j\frac{C}{\rho} \geq \mu$. Inserting the value for ρ_j results in

$$\Big([C - \sum_{k\neq j}\rho_k]^+ + \varepsilon\Big)\frac{C}{\rho} \geq \mu\,.$$

Letting $\varepsilon \to 0$ gives the claim. ■

VII. Conclusion

We showed that the output rates of flows at an overloaded link with EDF and FIFO scheduling are proportional to their arrival rates. Since we merely assumed that arrivals have a long-term average rate, the results holds for any deterministic or random traffic scenario. An implication of this is that the service of rate-proportional fair scheduling algorithms can be realized with lower overhead by simply resorting to a FIFO system. We also showed that the service guarantees to a flow at a link with FIFO or EDF are identical to that of a low-priority flow in a static priority system. Since this is vastly more pessimistic than the service achievable in an underloaded regime, our work quantifies the consequences of not controlling the total load admitted to a network.

References

[1] S. Keshav, "An Engineering Approach to Computer Networking: ATM Networks, the Internet, and the Telephone Network," Addison-Wesley, 1997.

[2] B. Melander, M. Bjorkman, and P. Gunningberg, "First-come-first-served packet dispersion and implications for TCP," In *Proc. IEEE Globecom*, vol. 3, pp. 2170–2174, Nov. 2002.

[3] J. Liebeherr, M. Fidler, and S. Valaee, "A min-plus system interpretation of bandwidth estimation," in *Proc. IEEE Infocom*, May 2007, pp. 1127–1135.

[4] Y. Ghiassi-Farrokhfal and J. Liebeherr. "Output characterization of constant bit rate traffic in FIFO networks." *IEEE Communication Letters*, 13(8):618–620, August 2009.

[5] F. Ciucu and O. Hohlfeld, and L. Y. Chen. "On the convergence to fairness in overloaded FIFO systems." In *Proc. IEEE Infocom*, pp. 1988–1996. April 2011.

[6] C.-S. Chang. *Performance Guarantees in Communication Networks*. Springer Verlag, 2000.