

Does Link Scheduling Matter on Long Paths?

Jörg Liebeherr
Department of ECE
University of Toronto
jorg@comm.utoronto.ca

Yashar Ghiassi-Farrokhfal
Department of ECE
University of Toronto
yashar@comm.utoronto.ca

Almut Burchard
Department of Mathematics
University of Toronto
almut@math.toronto.edu

Abstract—We seek to provide an analytical answer whether the impact of the selection of link scheduling algorithms diminishes on long network paths. The answer is provided through a detailed multi-node delay analysis, which is applicable to a broad class of scheduling algorithms, and which can account for statistical multiplexing. The analysis is enabled by two contributions: (1) We derive a function that can characterize the available bandwidth at a node for various scheduling algorithms. The function has an accuracy that recovers necessary and sufficient conditions for satisfying worst-case delay bounds at a single node; (2) We obtain end-to-end delay bounds by providing an explicit solution to an optimization problem, in which the service received at multiple nodes is subsumed into a single function. By presenting a unified analysis that captures the properties of a broad group of schedulers in a single parameter, we can provide insight how the choice of scheduling algorithms impacts end-to-end delay bounds. An important finding of this paper is that some schedulers show noticeable performance differences which persist in a network setting with long paths.

I. INTRODUCTION

Most carrier-class routers today can support a range of different link scheduling algorithms, at the same time, the view of FIFO (First-In-First-Out) scheduling in the Internet is highly prevalent. As a case in point, measurement methods to estimate the available bandwidth on Internet paths often make explicit or implicit assumptions of FIFO scheduling [23]. This view is justified when the impact of link scheduling algorithms on end-to-end performance metrics diminishes with increased network size. The objective of this paper is to present an analytical method that can shed light on the role of link scheduling algorithms on long paths.

The literature on the impact of link scheduling on end-to-end performance is limited and not conclusive. Most comparative evaluations of scheduling algorithms investigate only a single node scenario, e.g., [26]. Analytical comparisons between Generalized Processor Sharing (GPS) and Earliest Deadline First (EDF) scheduling over multi-hop path have been presented both for worst-case arrival scenarios [10] as well as for statistically multiplexed traffic [22]. These works highlight the differences between end-to-end delays with these schedulers, but do not study how these differences evolve as the number of nodes of a network path is increased. Some studies [2], [9], [25] have found conditions, when, even under the pessimistic assumption of priority scheduling, the output traffic at a network node has similar characteristics as the input, suggesting that the end-to-end delay performance of all (work-conserving) scheduling algorithms is similar. An

analysis of the throughput performance of FIFO scheduling on long paths in an overloaded network with constant bit rate fluid flow traffic in [11] showed that the end-to-end throughput of a flow asymptotically degrades to that of a low-priority flow. Our paper is motivated by the question whether a similar degradation is observable under more complex traffic patterns and in networks that are not permanently overloaded.

The analysis in this paper takes a *network calculus* approach [5], [14], which offers a general method for delay and backlog analysis over a path with multiple nodes. We use as performance metric the end-to-end delay. We assume that the long-term traffic load does not exceed the capacity of any network link. Traffic is characterized in terms of *envelope* functions and service is characterized by *service curve* functions. A key result of the (deterministic) network calculus states that the service curve of an end-to-end network path can be computed from the service curves at the links of the path by applying the convolution operator of the min-plus algebra. More precisely, when S_1, S_2, \dots, S_H are service curves describing the available service for a sequence of nodes, then the service of the entire network can be expressed by a network service curve $S^{net} = S_1 * S_2 * \dots * S_H$. (The min-plus convolution ‘*’ is defined in Section II-B.) In this fashion, a multi-node analysis is reduced to the analysis of a single node representing the entire network path. End-to-end delay bounds computed this way are generally more accurate than adding delay bounds computed for individual nodes. For example, worst-case delay bounds obtained by adding per-node delay bounds scale quadratically with the path length, while delays computed with S^{net} scale linearly [14]. In a probabilistic setting for traffic satisfying exponential bounds, delays computed by adding per-node bounds scale as $O(H^3)$ in the number of traversed nodes H , while delays computed with a network service curve scale with $\Theta(H \log H)$ [4].

In this paper we assess the impact of link scheduling algorithms on probabilistic end-to-end delay bounds using a network calculus analysis. A key difficulty is to find a service curve that can express a probabilistic service guarantee for non-trivial scheduling algorithms. It is possible to express the service available to a traffic flow in terms of the capacity left unused by all other flows with traffic. In a simplified form, such a leftover service description is given by:

$$\text{Leftover service} = \text{Link capacity} - \text{Arrivals from other traffic}$$

This corresponds to an interpretation of service, where the analyzed flow receives a lower priority than all other traffic at the link. For all work-conserving scheduling algorithms (that do not reorder packets from the same flow), such a characterization provides a lower bound for the actually available service. The leftover service interpretation is often referred to as *blind multiplexing*, since bounds on performance metrics hold even if details of the scheduling algorithm are not available. Refinements of a leftover characterization with service curves that account for specific scheduling algorithms have been considered before [8], [16], [20], however, with the exception of FIFO [8], the accuracy of the characterizations has not been accounted for. An application of the blind multiplexing analysis of [6] to FIFO networks in [12] has concluded that the delay bounds for Markov-modulated arrivals can be very pessimistic.

In this paper, we derive an accurate probabilistic leftover service characterization for a broad class of scheduling algorithms, referred to as Δ -schedulers. The class of Δ -schedulers contains a diverse set of algorithms, such as FIFO, priority scheduling, and Earliest-Deadline-First (EDF). For this class of schedulers, we provide a positive answer to the open question of the achievable accuracy of service curve characterizations, by showing that we can recover necessary and sufficient conditions for meeting deterministic delay bounds, since the end-to-end delay bounds derived in this paper extend the analysis of [6] from blind multiplexing to the complete class of Δ -schedulers. It will become evident that the blind multiplexing scenario considered in [6] appears as a simple special case. The end-to-end delay analysis of general Δ -schedulers requires an additional analytical optimization (provided in Section IV). The delay analysis in this paper extends the findings in [3], which studied the impact of the choice of the link scheduling algorithms on probabilistic delay bounds at a single node. Taking advantage of a decade of research on multi-node analysis with the stochastic network calculus [13], we can now study the impact of scheduling algorithms in a network environment. Since our probabilistic analysis contains deterministic delay bounds (where bounds are never violated) as a special case, our work also relates to the literature on worst-case end-to-end analyses [15], [21]. Several methods for computing statistical end-to-end delay bounds have been developed for EDF scheduling [1], [22], [16], but they added assumptions on the operation of the scheduler. For example, Andrews [1] and Li et.al. [16] assume that traffic exceeding an a priori delay constraint at a node is dropped. Sivaraman and Chiussi [22] present an analysis where traffic is re-shaped at each node, resulting in a non-workconserving system. These EDF analyses do not achieve the same $\Theta(H \log H)$ scaling of the delay bounds as the bounds derived in this paper.

We gain new insights and findings on the role of scheduling algorithms in networks. Even though the asymptotic growth of delays in the path length H for exponentially bounded traffic is $\Theta(H \log H)$ for every Δ -scheduler, we find that, in a non-asymptotic regime, the choice of the scheduling algorithm can have a noticeable impact on end-to-end delays. Thus,

accounting for the specifics of a scheduling algorithms can yield probabilistic end-to-end bounds that improve on the conservative bounds provided by blind multiplexing. On the other hand, for long network paths and at low utilizations we find that end-to-end delay bounds of FIFO and other scheduling algorithms often approach those of blind multiplexing.

The paper is structured as follows. In Section II we discuss the probabilistic traffic and service characterization in the network calculus, with traffic envelopes and service curves. In Section III we define and analyze Δ -scheduling algorithms. In Section IV we compute a network service curve for Δ -schedulers and use it for an analytical optimization that result in end-to-end delay bounds. We present numerical examples in Section V and conclude the paper in Section VI.

II. ARRIVAL AND SERVICE MODEL

This paper takes a network calculus [14] modeling and analysis approach where arrivals of a flow and the service given to a flow are expressed in terms of deterministic or probabilistic bounds. We refer to these bounds as *traffic envelopes* and *service curves*, respectively. The concepts presented in this section are developed in [5], [6], [14].

A. Traffic Envelopes

We consider a continuous time model where arrivals from a traffic flow or flow aggregate in the time interval $[0, t)$ are represented by a random process $A(t)$ whose increments satisfy stationary bounds. Traffic departing from a node in $[0, t)$ is denoted by $D(t)$. Both $A(t)$ and $D(t)$ are non-decreasing in time with $A(t) = D(t) = 0$ for $t \leq 0$, and we have $D(t) \leq A(t)$. For brevity, we use the notation $A(s, t) = A(t) - A(s)$ for any $s \leq t$ to denote arrivals in a time interval. We use a subscript ‘ j ’ to denote traffic arrivals or departures by a flow or flow aggregate j .

We characterize traffic in terms of traffic envelopes that specify the arrivals over given time intervals. In a worst-case analysis, a *deterministic sample path envelope* E provides an upper bound on traffic arrivals satisfying for all $t > 0$ given by

$$\sup_{0 \leq s \leq t} \{A(s, t) - E(t - s)\} \leq 0, \quad (1)$$

with $E(t) = 0$ for $t \leq 0$. By definition, no sample path of A ever violates a deterministic sample path envelope. An example of a deterministic sample path envelope is a leaky bucket with $E(t) = Rt + B$ for a rate parameter R and a burst parameter B . Deterministic envelopes are generally conservative bounds since they do not capture statistical fluctuations of a traffic or statistical multiplexing. This has justified devising a probabilistic analogue to the deterministic envelope as a bound for sample paths of the arrivals that may be violated with a small probability. A *statistical sample path envelope* $\mathcal{G}(t)$ with bounding function $\varepsilon(\sigma)$ satisfies for all $t, \sigma \geq 0$

$$P(\sup_{0 \leq s \leq t} \{A(s, t) - \mathcal{G}(t - s)\} > \sigma) \leq \varepsilon(\sigma), \quad (2)$$

where $0 \leq \varepsilon(\sigma) \leq 1$. \mathcal{G} is non-negative with $\mathcal{G}(t) = 0$ for $t \leq 0$. The function $\varepsilon(\sigma)$ bounds the probability that a sample path

of the arrivals exceeds the envelope \mathcal{G} by more than σ . Over the years, traffic envelopes have been explored for a wide range of arrival processes, including memoryless, Markov-modulated, and even long range dependent processes. We refer to [18] for an informative survey. For $\varepsilon(\sigma) = 0$ if $\sigma > 0$, we recover the deterministic version above.

B. Service Curves

Throughout this paper, we use service curves to describe lower bounds on the service available to a flow. A node offers a (deterministic) service curve S if the input-output relationship of traffic at a node satisfies for all $t \geq 0$

$$D(t) \geq \inf_{0 \leq s \leq t} \{A(s) + S(t-s)\}. \quad (3)$$

The term on the right-hand side is referred to as a min-plus convolution, and denoted by ' $A * S(t)$ '. As examples, a service curve for a constant rate link with capacity C is given by $S(t) = Ct$; the service curve for a delay of d is given by $S(t) = \delta_d$, where

$$\delta_d(t) = \begin{cases} 0, & \text{if } t \leq d \\ \infty, & \text{if } t > d. \end{cases} \quad (4)$$

Using service curves, the service on a path of H nodes, each offering a deterministic service curve of S_1, S_2, \dots, S_H , can be expressed in terms of the convolution $S_1 * S_2 * \dots * S_H$. A probabilistic extension of this concept leads to the formulation of a *statistical service curve* \mathcal{S} with bounding function $\varepsilon(\sigma)$ that satisfies for all $t, \sigma \geq 0$

$$P(D(t) < A * [\mathcal{S} - \sigma]_+(t)) < \varepsilon(\sigma), \quad (5)$$

where $[x]_+ = \max(0, x)$ and $0 \leq \varepsilon(\sigma) \leq 1$, with $\mathcal{S}(t)$ non-negative and non-decreasing, and $\mathcal{S}(t) = 0$ for $t < 0$. Here, $\varepsilon(\sigma)$ bounds the probability that the guarantee of Eq. (3) is violated by more than σ .

When applying the network calculus to practical scheduling methods one finds that some schedulers are more suitable than others for a service curve description. Some scheduling algorithms, such as General Processor Sharing (GPS) [19] and Service Curve Earliest Deadline (SCED) [8], have been specified in terms of service curves. Generally, however, characterizing a scheduler in terms of service curves requires an indirect description in terms of the available bandwidth capacity that is left unused by other flows with traffic at this link. In the next section, we develop such a description and show that we can completely capture the operation of an entire class of scheduling algorithms.

III. Δ -SCHEDULING ALGORITHMS

We consider the arrivals from a set \mathcal{N} of traffic flows (or flow aggregates) to a buffered link with rate C , referred to as a *node*. Schedulers are work-conserving in the sense that they transmit at rate C whenever there is backlogged traffic, and they are locally FIFO in the sense that traffic from the same flow is transmitted in the order of its arrival. In this paper, we ignore that packet transmissions cannot be interrupted. This is

a reasonable assumption when packet sizes are small compared to the transmission rate. The assumption can be relaxed at the cost of additional notation. We do not assume that traffic shaping is performed between nodes.

Let us consider an arrival from flow j at time t , which we refer to as 'tagged arrival'. This arrival will be transmitted when it has higher precedence than all other backlogged traffic. It turns out that for many scheduling algorithms it is feasible to specify constants $\Delta_{j,k}$ ($k \in \mathcal{N}$), so that only arrivals from flow k that occur before $t + \Delta_{j,k}$ have higher precedence than the tagged arrival. Note that any locally FIFO Δ -scheduler satisfies $\Delta_{j,j} = 0$ for all $j \in \mathcal{N}$. We refer to scheduling algorithms whose operations can be completely described by such constants $\Delta_{j,k}$ as Δ -schedulers.

Definition 1: Given a set \mathcal{N} of flows with arrivals to a work-conserving link. A Δ -scheduler is a work-conserving locally FIFO scheduling algorithm if there exist constants $\{\Delta_{j,k}\}_{j,k \in \mathcal{N}}$ such that an arrival at time t from flow j has precedence over all arrivals from flow k that occur after $t + \Delta_{j,k}$.

The class of Δ -schedulers contains a diverse group of scheduling algorithms.

- **FIFO:** If traffic is transmitted in the order of arrival, only arrivals before time t have precedence over an arrival at time t . Consequently, we have $\Delta_{j,k} = 0$ for all $j, k \in \mathcal{N}$.

- **SP, BMUX:** With static priority (SP) scheduling, each flow is assigned a priority level, and the scheduler always transmits backlogged traffic with the highest available priority. FIFO is used to order traffic with the same priority. Since low priority traffic has never higher precedence over higher priority traffic, we obtain for all $j, k \in \mathcal{N}$

$$\Delta_{j,k} = \begin{cases} -\infty, & \text{if } k \text{ has lower priority than } j \\ 0, & \text{if } k \text{ has the same priority as } j \\ \infty, & \text{if } k \text{ has higher priority than } j \end{cases}.$$

A scenario where traffic of one flow j has low priority and all other traffic has high priority is referred to as blind multiplexing (BMUX). BMUX is an important benchmark, since it yields the highest delays for flow j for any work-conserving locally FIFO scheduler.

- **EDF:** Each flow k is associated with an *a priori* delay constraint d_k^* . When traffic from flow k arrives to the scheduler at time t , it is tagged with a deadline $t + d_k^*$. Traffic is transmitted in the order of increasing deadlines. Since, in EDF, traffic has a higher precedence if it has a smaller deadline, we obtain that $\Delta_{j,k} = d_j^* - d_k^*$ for all $j, k \in \mathcal{N}$.

There are numerous schedulers for which we can specify a time limit on the arrivals that have higher precedence than a tagged arrival that are not Δ -schedulers. Consider Generalized Processor Sharing (GPS) [19] as a representative of fair queueing algorithms. Each flow j is assigned a weight ϕ_j and the amount of service given to backlogged flows is proportional to their weights. Here, at a link with fixed capacity C , the service rate of a backlogged flow j at time t is given by $\frac{\phi_j}{\sum_{k \in Q(t)} \phi_k} C$, where $Q(t)$ is the set of flows that

have a backlog at time t . Since the set $Q(t)$ is random, it is not feasible to find constants $\Delta_{j,k}$ that bound the arrival time of traffic from flow k that has higher precedence than a flow j arrival. More generally, any scheduling algorithm for which a time limit $\Delta_{j,k}$ can only be given in terms of a random variable is not a Δ -scheduler.

A. Service Curves for Δ -Schedulers

We will characterize the service available to a flow $j \in \mathcal{N}$ at a node with fixed capacity C and a Δ -scheduler in terms of a service curve in the sense of Eq. (5). All flows in $\mathcal{N} \setminus \{j\}$ comprise the cross traffic at the node. Before we present the service curve in Theorem 1 we must introduce some notation.

For any locally FIFO scheduler, let $W_j(t)$ denote the delay of an arrival from flow j at time t , defined as

$$W_j(t) = \inf \{s \geq 0 \mid D_j(t+s) \geq A_j(t)\} . \quad (6)$$

It is convenient to view W_j as a function that is defined at all times.

For given constants $\{\Delta_{j,k}\}_{j,k \in \mathcal{N}}$, we define

$$\Delta_{j,k}(y) = \min \{\Delta_{j,k}, y\} \quad (7)$$

to denote arrivals with higher precedence that have already occurred. More specifically, consider an arrival from flow j at time t that still resides in the scheduler at time $t+y$ ($y > 0$), i.e., $W_j(t) > y$. With respect to this arrival, traffic from flow k that is transmitted earlier must arrive no later than $t + \Delta_{j,k}(y)$.

If $\Delta_{j,k} = -\infty$, arrivals from flow k never have precedence over traffic from flow j , and can be excluded from a delay analysis. We define \mathcal{N}_j and \mathcal{N}_{-j} as

$$\mathcal{N}_j = \{k \in \mathcal{N} \mid \Delta_{j,k}(t) > -\infty\} , \quad \mathcal{N}_{-j} = \mathcal{N}_j \setminus \{j\} ,$$

to specify the subset of flows that may have an impact on the delay W_j of traffic from flow j .

Finally, we denote by $I(\cdot)$ the indicator function, where $I(\text{expr}) = 1$ if ‘*expr*’ is true, and $I(\text{expr}) = 0$ otherwise. The following theorem specifies a family of statistical service curves for flow j at a node with Δ -scheduling:

Theorem 1: Given a Δ -scheduler operating at a link with capacity C and a set of flows \mathcal{N} with traffic at the link. For each flow $k \in \mathcal{N}$, let \mathcal{G}_k be a statistical sample path bound with bounding function $\varepsilon_k(\sigma)$ satisfying Eq (2). Then, for each $\theta \geq 0$, the function

$$\mathcal{S}_j(t; \theta) = \left[Ct - \sum_{k \in \mathcal{N}_{-j}} \mathcal{G}_k(t - \theta + \Delta_{j,k}(\theta)) \right]_+ I(t > \theta) \quad (8)$$

is a statistical service curve with bounding function

$$\varepsilon_s(\sigma) = \inf_{\sum \sigma_k = \sigma} \sum_{k \in \mathcal{N}_{-j}} \varepsilon_k(\sigma_k) .$$

The theorem generalizes a service curve characterization that has been known for FIFO [8] to all Δ -schedulers and to a probabilistic setting.

Proof: Fix a time $t \geq 0$. Let $B_{j,k}^t(s)$ denote the backlogged traffic from flow k at time s with higher or equal

precedence than an arrival from flow j at time t . This can be expressed as follows:

$$\begin{aligned} B_{j,k}^t(s) &= A_k(\min\{s, t + \Delta_{j,k}\}) - D_k(s) \\ &= A_k(t + \Delta_{j,k}(s - t)) - D_k(s) . \end{aligned} \quad (9)$$

Let $t - \hat{x}_t$ be the last time before t when the scheduler does not have such a backlog, that is,

$$\hat{x}_t = t - \sup \{s \leq t \mid \sum_{k \in \mathcal{N}_j} B_{j,k}^t(s) \leq 0\} . \quad (10)$$

No traffic that arrives earlier than $t - \hat{x}_t$ contributes to the delay of an arrival from flow j at time t . Note that Eq. (10) implies that $A_j(t - \hat{x}_t) = D_j(t - \hat{x}_t)$. With Eq. (10), we can rewrite Eq. (6) as

$$W_j(t) = \inf \{s \geq 0 \mid D_j(t - \hat{x}_t, t + s) \geq A_j(t - \hat{x}_t, t)\} .$$

Consider first the case where the virtual delay at time t exceeds θ , that is, $W_j(t) > \theta$. Since Δ -schedulers are work-conserving, in the entire interval $[t - \hat{x}_t, t + \theta]$ the link is transmitting traffic from flows in \mathcal{N}_j at rate C . We get:

$$C(\theta + \hat{x}_t) = \sum_{k \in \mathcal{N}_j} D_k(t - \hat{x}_t, t + \theta) . \quad (11)$$

In the interval $[t - \hat{x}_t, t + \theta]$, the link is only transmitting traffic with higher precedence than the tagged arrival. Since arrivals from flow k with higher precedence than the tagged arrival must occur before $t + \Delta_{j,k}(\theta)$, we have for each k that $A_k(t - \hat{x}_t, t + \Delta_{j,k}(\theta)) \geq D_k(t - \hat{x}_t, t + \theta)$. Applying this inequality in Eq. (11) gives a lower bound on the transmitted traffic from class j :

$$D_j(t - \hat{x}_t, t + \theta) \geq [C(\theta + \hat{x}_t) - \sum_{k \in \mathcal{N}_{-j}} A_k(t - \hat{x}_t, t + \Delta_{j,k}(\theta))]_+ .$$

We can view the right hand term as a lower bound on the capacity available to flow j that is left unused by the other flows. Using that $D_j(t - \hat{x}_t) = A_j(t - \hat{x}_t)$ by the choice of \hat{x}_t , we can rewrite the inequality as

$$\begin{aligned} D_j(t + \theta) &\geq A_j(t - \hat{x}_t) + [C(\theta + \hat{x}_t) \\ &\quad - \sum_{k \in \mathcal{N}_{-j}} A_k(t - \hat{x}_t, t + \Delta_{j,k}(\theta))]_+ I(\hat{x}_t > 0) . \end{aligned} \quad (12)$$

Note that the indicator function always evaluates to 1. On the other hand, in the case where $W_j(t) \leq \theta$, we know that

$$D_j(t + \theta) \geq A_j(t) , \quad (13)$$

because the scheduler is locally FIFO.

Eqs. (12) and (13) can be summarized in the statement that

$$\begin{aligned} D_j(t) &\geq \inf_{s \leq t} \left\{ A_j(s) + \left[C(t - s) \right. \right. \\ &\quad \left. \left. - \sum_{k \in \mathcal{N}_{-j}} A_k(s, t - \theta + \Delta_{j,k}(\theta)) \right]_+ I(t - s > \theta) \right\} \end{aligned} \quad (14)$$

for all $t \geq 0$ and all $\theta \geq 0$. To see this, we first replace t by $t - \theta$ in both inequalities, then introduce the new variable $s = t - \theta - \hat{x}_t$ for Eq. (12), and set $s = t - \theta$ for Eq. (13).

Now suppose that we have sample paths so that at time t :

$$\sup_{s \leq t} \left\{ \sum_{k \in \mathcal{N}_{-j}} A_k(s, t - \theta + \Delta_{j,k}(\theta)) - \sum_{k \in \mathcal{N}_{-j}} \mathcal{G}_k(t - s - \theta + \Delta_{j,k}(\theta)) - \sigma \right\} \leq 0. \quad (15)$$

Inserting this into Eq. (14) yields the service curve:

$$\begin{aligned} D_j(t) &\geq \inf_{s \leq t} \left\{ A_j(s) + \left[C(t - s) \right. \right. \\ &\quad \left. \left. - \sum_{k \in \mathcal{N}_{-j}} \mathcal{G}_k(t - s - \theta + \Delta_{j,k}(\theta)) - \sigma \right]_+ I(t - s > \theta) \right\} \\ &\geq A_j * [S_j - \sigma]_+(t; \theta). \end{aligned} \quad (16)$$

Choosing $\{\sigma_k\}_{k \in \mathcal{N}_{-j}}$ such that $\sigma = \sum_{k \in \mathcal{N}_{-j}} \sigma_k$, we estimate the bounding function as follows:

$$\begin{aligned} P(D_j(t) \geq A_j * [S_j - \sigma]_+(t; \theta)) \\ &\geq P(\text{Eq. (15) holds}) \\ &\geq P(\forall k \in \mathcal{N}_{-j} : \sup_{s \leq t} \{ A_k(s, t - \theta + \Delta_{j,k}(\theta)) \\ &\quad - \mathcal{G}_k(t - s - \theta + \Delta_{j,k}(\theta)) \} \leq \sigma_k) \end{aligned} \quad (17)$$

$$\begin{aligned} &\geq 1 - \sum_{k \in \mathcal{N}_{-j}} P\left(\sup_{s \leq t - \theta + \Delta_{j,k}(\theta)} \{ A_k(s, t - \theta + \Delta_{j,k}(\theta)) \right. \\ &\quad \left. - \mathcal{G}_k(t - \theta + \Delta_{j,k}(\theta) - s) \} > \sigma_k \right) \end{aligned} \quad (18)$$

$$\geq 1 - \sum_{k \in \mathcal{N}_{-j}} \varepsilon_k(\sigma_k).$$

In Eq. (17) we restrict the event by requiring that in Eq. (15) each flow $k \in \mathcal{N}_{-j}$ satisfies its envelope \mathcal{G}_k for some allowed choice of σ_k . For Eq. (18) we have first applied the union bound and then used that the supremum cannot be assumed for $s > t - \theta + \Delta_{j,k}(\theta)$ to restrict the range of s . In the last step, we have used that \mathcal{G}_k is a statistical sample path bound. The claim follows by minimizing over $\{\sigma_k\}_{k \in \mathcal{N}_{-j}}$. ■

Given deterministic envelopes E_k satisfying Eq. (1), we immediately obtain from Eq. (14) that for each $\theta \geq 0$, the function

$$S_j(t; \theta) = \left[Ct - \sum_{k \in \mathcal{N}_{-j}} E_k(t - \theta + \Delta_{j,k}(\theta)) \right]_+ I(t > \theta) \quad (19)$$

is a deterministic service curve in the sense of Eq. (3).

B. Tightness of the Service Curve

We make the case that the service curve in Theorem 1 accurately characterizes a scheduler, by showing that we can obtain necessary and sufficient conditions for meeting given delay bounds.

Let S_j be a statistical service curve with bounding function $\varepsilon_s(\sigma)$, as defined in Eq. (5), and let \mathcal{G}_j be a statistical sample path envelope with bounding function $\varepsilon_g(\sigma)$, as defined in Eq. (2). Select $d(\sigma)$ as the smallest value satisfying

$$\mathcal{G}_j(t) + \sigma \leq S_j(t + d(\sigma)), \quad \forall t \geq 0, \quad (20)$$

and set

$$\varepsilon(\sigma) = \inf_{\sigma = \sigma_1 + \sigma_2} \{ \varepsilon_g(\sigma_1) + \varepsilon_s(\sigma_2) \}. \quad (21)$$

From [6] we have that $d(\sigma)$ is a probabilistic delay bound that satisfies

$$P(W_j(t) > d(\sigma)) < \varepsilon(\sigma), \quad \forall t \geq 0, \forall \sigma \geq 0. \quad (22)$$

We apply this probabilistic delay bound to the service curve of the Δ -scheduler from Theorem 1. Setting $\theta = d(\sigma)$ in Eq. (8), and inserting the service curve into Eq. (20) we obtain

$$\sup_{t > 0} \left\{ \mathcal{G}_j(t) + \sigma - \left[C(t + d(\sigma)) - \sum_{k \in \mathcal{N}_{-j}} \mathcal{G}_k(t + \Delta_{j,k}(d(\sigma))) \right]_+ \right\} \leq 0.$$

Since $\Delta_{j,j} = 0$ in all locally FIFO schedulers, we can replace $\mathcal{G}_j(t)$ by $\mathcal{G}_j(t + \Delta_{j,j}(d))$. Collecting terms, we get

$$\sup_{t > 0} \left\{ \sum_{k \in \mathcal{N}_j} \mathcal{G}_k(t + \Delta_{j,k}(d(\sigma))) + \sigma - Ct \right\} \leq Cd(\sigma). \quad (23)$$

For $\sigma = 0$ and $d = d(0)$, this condition has the same structure as the probabilistic schedulability conditions from [3], which were derived without using service curves. This indicates that the service curve from Theorem 1 yields similar delay bounds as those obtained with a ‘direct’ analysis of a specific scheduling algorithm.

In a deterministic setting, Eq. (23) is tight in the sense that it can result in a necessary and sufficient condition for meeting a delay bound. Here, the envelopes \mathcal{G}_k are replaced by deterministic sample-path envelopes E_k (satisfying Eq. (1)), S_j takes the form of Eq. (19). Also, we have $\sigma = 0$, $\varepsilon(\sigma) = 0$. This yields the deterministic condition

$$\sup_{t > 0} \left\{ \sum_{k \in \mathcal{N}_j} E_k(t + \Delta_{j,k}(d)) - Ct \right\} \leq Cd. \quad (24)$$

The following theorem provides the accuracy and tightness of the condition.

Theorem 2: Given a set of flows \mathcal{N} with a deterministic envelope function E_j for each flow j satisfying Eq. (1) and a buffered link with fixed-rate capacity C and a Δ -scheduler. The delay of traffic from flow j does not exceed a delay $d > 0$ if the condition in Eq. (24) is satisfied. The condition is tight if the envelope functions E_k are concave.

This condition recovers the tight (necessary and sufficient) condition for meeting a deterministic delay bound d for FIFO, SP, and EDF from [7], [17]. We point out that deterministic envelope functions that provide a tight description for an arrival flow will generally be concave or well approximated by a concave function. This follows from the fact that the smallest deterministic envelope function for an arrival sample path is always subadditive.

Proof: Sufficiency of Eq. (24) follows directly from the deterministic version of Eq. (20). To prove necessity for

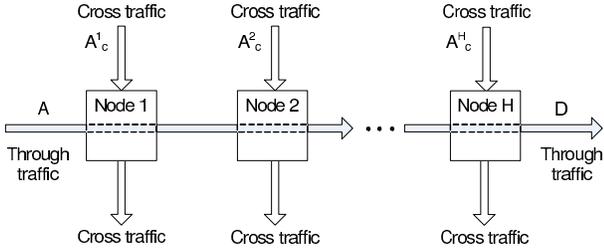


Fig. 1. Multi-node network.

concave envelope functions, we assume that the condition in Eq. (24) is violated for t^* , that is,

$$\sum_{k \in \mathcal{N}_j} E_k(t^* + \Delta_{j,k}(d)) > C(d + t^*). \quad (25)$$

Consider a transmission scenario where the scheduler is empty at time $t = 0$, and immediately after $t = 0$, each flow $k \neq j$ has arrivals to the scheduler from flow k occur such that $A_k(t) = E_k(t)$. Arrivals from flow j also satisfy $A_j(t) = E_j(t)$ with the additional constraint that there is an arrival from flow j at t^* . The backlogged traffic from flow k at time s with higher or equal precedence than an arrival from flow j at time t^* is $B_{j,k}^{t^*}(s)$, as defined in Eq. (9). Defining

$$B_j^{t^*}(s) = \sum_{k \in \mathcal{N}_j} B_{j,k}^{t^*}(s)$$

we obtain that for $s \in [0, t^* + d]$ the following holds:

$$\begin{aligned} B_j^{t^*}(s) &= \sum_{k \in \mathcal{N}_j} (A_k(t^* + \Delta_{j,k}(s - t^*)) - D_k(s)) \\ &= \sum_{k \in \mathcal{N}_j} E_k(t^* + \Delta_{j,k}(s - t^*)) - Cs. \end{aligned} \quad (26)$$

Since, by assumption, all E_k are concave and nondecreasing, and $\Delta_{j,k}(s - t^*)$ is concave, we have that $B_j^{t^*}(t)$ is concave in the interval $[0, t^* + d]$. Since $B_j^{t^*}(0) = 0$ by construction of the arrival pattern, and $B_j^{t^*}(t^* + \Delta_{j,k}(d)) > 0$ due to the assumption in Eq. (25), we have that $B_j^{t^*}(s) > 0$ for $s \in [0, t^* + d]$. Noting that for $t > t^*$, $B_j^{t^*}$ describes the backlogged traffic that is transmitted before the arrival from flow j at time t^* , the arrival cannot be transmitted by time $t^* + d$, resulting in a delay bound violation. ■

IV. END-TO-END DELAY ANALYSIS

We next use the service curve from the previous section to present a stochastic end-to-end delay analysis for Δ -schedulers across a path of multiple nodes. We study the performance of a traffic flow with arrival process A that traverses a sequence of H nodes as shown in Fig. 1, and that experiences cross traffic at each node. We do not assume independence of cross traffic and through traffic. To keep notation minimal we assume that each node has the same fixed capacity link with rate C , that cross traffic at the h -th node is determined by an identically distributed arrival process A_c^h , and that the same scheduling algorithm is used at each node. We denote by $\Delta_{0,c}$ the scheduler dependent constant for the through traffic. Also,

for conciseness of the derivations, we adopt a discrete time system ($t = 0, 1, \dots$).

We compute end-to-end delay bounds for exponentially bounded arrival processes, where the arrivals of the through traffic A satisfy for all $s \leq t$ and for $\sigma \geq 0$

$$P(A(s, t) > \rho(t - s) + \sigma) \leq M e^{-\alpha \sigma}, \quad (27)$$

where $M \geq 1$ and $\rho, \alpha > 0$ are constants. Such traffic is said to have *Exponentially Bounded Burstiness (EBB)* [24]. We will write $A \sim (M, \rho, \alpha)$ to denote an EBB arrival process with the given constants. The EBB traffic model is expressive enough to capture Markov-Modulated processes.

Remark: For EBB arrivals and a network as in Fig. 1, the asymptotic growth of end-to-end delays in H is $\Theta(H \log H)$ for all Δ -schedulers. This follows from the scaling analysis in [4] for a packetized arrival model. The upper bound of $O(H \log H)$ for delays in BMUX from [4] holds for all Δ -schedulers since delays with BMUX are larger than those of any Δ -scheduler. Additionally, the lower bound of $\Omega(H \log H)$ in [4] was derived for a network without any cross traffic. Since all Δ -schedulers are locally FIFO, the lower bound holds for any Δ -scheduler.

A statistical sample path envelope for an EBB arrival process can be computed with the union bound as

$$\mathcal{G}(t) = (\rho + \gamma)t, \quad \varepsilon(\sigma) = \frac{M e^{-\alpha \sigma}}{1 - e^{-\alpha \gamma}},$$

for any choice of $\gamma > 0$. We assume $A_c^h \sim (M, \rho_c, \alpha)$ for the cross traffic, and denote by \mathcal{G}_c and $\varepsilon_c(\sigma)$ its statistical sample path envelope and bounding function.

With these statistical sample path envelopes, Theorem 1 provides a service curve of the through traffic at the h -th node for any choice of $\theta^h, \gamma > 0$ by

$$\mathcal{S}^h(t; \theta^h) = \left[Ct - (\rho_c + \gamma)[t - \theta^h + \Delta_{0,c}(\theta^h)] \right]_+ I(t > \theta^h) \quad (28)$$

$$\varepsilon^h(\sigma) = \frac{M e^{-\alpha \sigma}}{1 - e^{-\alpha \gamma}}. \quad (29)$$

Since Eq. (29) satisfies the condition $\int_0^\infty \varepsilon^h(x) dx < \infty$, we can apply a result in [6] (Theorem 1) which provides a statistical network service curve satisfying Eq. (5). In the discrete time version of the network service curve, we obtain that, for any choice of γ , the through flow receives a service curve for the entire network of H nodes given by

$$\mathcal{S}^{net}(t; \underline{\theta}) = \mathcal{S}^1 * \mathcal{S}_\gamma^2 * \dots * \mathcal{S}_{(H-1)\gamma}^H(t; \underline{\theta}), \quad (30)$$

where $\underline{\theta} = (\theta^1, \dots, \theta^H)$ and $\mathcal{S}_{h\gamma}^{h+1}(t; \underline{\theta}) = \mathcal{S}^{h+1}(t; \underline{\theta}) - h\gamma t$. Here, the convolution of statistical service curves incurs at each node an additional rate degradation of γ . The corresponding bounding function is given by

$$\varepsilon^{net}(\sigma) = \inf_{\sum_{h=1}^H \sigma^h = \sigma} \left\{ \varepsilon^H(\sigma^H) + \sum_{h=1}^{H-1} \sum_{j=0}^{\infty} \varepsilon^h(\sigma^h + j\gamma) \right\}. \quad (31)$$

The desired probabilistic end-to-end delay bound is obtained by using Eq. (30) in the single-node delay bound from Eq. (20).

We first choose $\gamma > 0$ to satisfy

$$(H+1)\gamma < C - \rho_c - \rho, \quad (32)$$

and compute the bounding function in Eq. (31) with ε^h given in Eq. (29). The computation becomes straightforward when we take advantage of the following property derived in [6]: For any set of constants (M_j, ρ_j, α_j) with $j = 1, 2, \dots, N$,

$$\inf_{\sum_{j=1}^N \sigma_j = \sigma} \sum_{j=1}^N M_j e^{\alpha_j \sigma_j} = \prod_{j=1}^N (M_j \alpha_j w) e^{\sigma/w}, \quad (33)$$

where $w = 1/\alpha_1 + 1/\alpha_2 + \dots + 1/\alpha_N$. Then ε^{net} can be computed as

$$\varepsilon^{net}(\sigma) = \frac{MH}{(1 - e^{-\alpha\gamma})^{\frac{2H-1}{H}}} e^{-\frac{\sigma}{H}}.$$

We next apply the single node delay bound from Eq. (21) to obtain a probabilistic delay bound $d(\sigma)$. With $\varepsilon_s = \varepsilon^{net}$ and $\varepsilon_g = \frac{Me^{-\alpha\sigma}}{1 - e^{-\alpha\gamma}}$, we evaluate Eq. (20) with another application of Eq. (33) as

$$\Pr\{W(t) \geq d(\sigma)\} = \frac{M(H+1)}{(1 - e^{-\alpha\gamma})^{\frac{2H+1}{H}}} e^{-\frac{\alpha}{H+1}\sigma}. \quad (34)$$

We next compute \mathcal{S}^{net} (as in Eq. (30)), and then apply Eq. (20). By Eq. (32), and since $\theta^h - \Delta_{0,c}(\theta^h)$ is non-decreasing in θ^h , the term in the outer brackets of Eq. (28) is non-negative, and we can write $\tilde{\mathcal{S}}_{(h-1)\gamma}^h$ as

$$\tilde{\mathcal{S}}_{(h-1)\gamma}^h(t; \theta^h) = \tilde{S}_{(h-1)\gamma}^h * \delta_{\theta^h}(t; \theta^h),$$

with δ as defined by Eq. (4), and

$$\tilde{S}_{(h-1)\gamma}^h(t; \theta^h) = (C - (h-1)\gamma)(t + \theta^h) - (\rho_c + \gamma)[t + \Delta_{0,c}(\theta^h)]_+. \quad (35)$$

Exploiting basic properties of the min-plus algebra, we derive

$$\begin{aligned} \mathcal{S}^{net}(t; \underline{\theta}) &= (\tilde{\mathcal{S}}^1 * \delta_{\theta^1}) * \dots * (\tilde{\mathcal{S}}_{(H-1)\gamma}^H * \delta_{\theta^H})(t; \underline{\theta}) \\ &= (\tilde{\mathcal{S}}^1 * \dots * \tilde{\mathcal{S}}_{(H-1)\gamma}^H) * (\delta_{\theta^1} * \dots * \delta_{\theta^H})(t; \underline{\theta}) \\ &= \min_{h=1, \dots, H} \{\tilde{\mathcal{S}}_{(h-1)\gamma}^h\} * \delta_{\sum_{h=1}^H \theta^h}(t; \underline{\theta}) \\ &= \min_{h=1, \dots, H} \{\tilde{\mathcal{S}}_{(h-1)\gamma}^h(t - \sum_{h=1}^H \theta^h; \theta^h)\} I(t > \sum_{h=1}^H \theta^h). \end{aligned}$$

The second line above applies the associativity and commutativity of the min-plus convolution. The third line uses that the convolution of concave functions is their pointwise minimum, and that $\delta_a * \delta_b = \delta_{a+b}$.

Per Eq. (20),

$$\forall t \geq 0 : \mathcal{S}^{net}(t + d(\sigma); \underline{\theta}) \geq (\rho + \gamma)t + \sigma \quad (36)$$

is a sufficient condition for $d(\sigma)$ to be a probabilistic delay bound. For $t = 0$, we obtain the necessary condition that

$$d(\sigma) \geq \sum_{h=1}^H \theta^h. \quad (37)$$

Under this condition, the left hand side in Eq. (36) is concave in t . Our choice of γ guarantees that Eq. (36) is automatically satisfied for t sufficiently large, since $\mathcal{S}^{net}(t)$ grows linearly with rate $C - \rho_c - H\gamma > \rho + \gamma$. Since the left hand side of Eq. (36) is concave in t , it suffices to verify the condition in Eq. (36) for $t = 0$. Thus, we must find the smallest d such that

$$\forall h = 1, \dots, H : \tilde{\mathcal{S}}_{(h-1)\gamma}^h(d - \sum_{h=1}^H \theta^h; \theta) \geq \sigma,$$

subject to Eq. (37). The θ^h 's should be selected to maximize the left hand side of the inequality, yielding the following optimization problem:

$$\begin{aligned} d(\sigma) = \min \left\{ d \geq \sum_{h=1}^H \theta^h \mid \right. \\ \left. \max_{\theta^1, \dots, \theta^H} \min_{h=1, \dots, H} \left\{ (C - (h-1)\gamma)(d + \theta^h - \sum_{j=1}^H \theta^j) \right. \right. \\ \left. \left. - (\rho_c + \gamma)[d + \Delta_{0,c}(\theta^h) - \sum_{j=1}^H \theta^j]_+ \right\} \geq \sigma \right\}. \end{aligned}$$

For $H = 1$, since $\theta^1 - \Delta_{0,c}(\theta^1)$ is non-decreasing in θ^1 , the optimal choice is $\theta^1 = d$ for all Δ -schedulers. This will give the same results as the single node analysis in Section III-B. For $H > 1$, the optimal choice of θ^h depends on the value of $\Delta_{0,c}$.

With a change of variables $X = d(\sigma) - \sum_{h=1}^H \theta^h$, the optimizing problem takes the following form:

$$\text{Minimize } d(\sigma) = X + \sum_{h=1}^H \theta^h \quad (38)$$

subject to

$$(C - (h-1)\gamma)(X + \theta^h) - (\rho_c + \gamma)[X + \Delta_{0,c}(\theta^h)]_+ \geq \sigma, \forall h \\ \theta^1, \dots, \theta^H, X \geq 0.$$

Note that this optimization problem is not generally convex when $\Delta_{0,c} > 0$, since $\Delta_{0,c}(\theta^h)$ is concave in θ^h . However, since the objective function in Eq. (38) is linear, it assumes its minimum at a point where $H+1$ constraints hold with equality. In such a point, we must have $\theta^1 = \dots = \theta^K = 0$, and $0 < \theta^{K+1} \leq \dots \leq \theta^H$ for some K . We will now describe how to identify this point.

We reduce the problem to a minimization problem in the single variable X . For $X \geq 0$, let $\theta^h(X)$ be the smallest nonnegative solution of

$$(C - (h-1)\gamma)(X + \theta^h) - (\rho_c + \gamma)[X + \Delta_{0,c}(\theta^h)]_+ \geq \sigma.$$

This choice ensures that the h -th constraint in Eq. (38) is satisfied with equality whenever $\theta_h(X) > 0$, and we obtain

$$d(\sigma) = \min_{X \geq 0} \left\{ X + \sum_{h=1}^H \theta_h(X) \right\}. \quad (39)$$

Clearly, $\theta^h(X)$ is a strictly decreasing function of X in the range where $\theta^h(x) > 0$. Once we have identified the index K described above, we can drop the first K summands and consider only $\theta^h(x)$ for $h > K$. The minimum is assumed at a point where the derivative $\frac{d}{dX} d(\sigma)$ changes from negative to

positive. We distinguish two cases. For $\Delta_{0,c} \geq 0$, we compute its derivative as

$$\frac{d}{dX}\theta^h(X) = \begin{cases} -\frac{C-\rho_c-h\gamma}{C-(h-1)\gamma}, & \theta^h > \Delta, \\ -1, & 0 < \theta^h < \Delta_{0,c}. \end{cases}$$

To summarize our analysis, the optimal K satisfies

$$\sum_{h>K} \frac{C-\rho_c-h\gamma}{C-(h-1)\gamma} < 1. \quad (40)$$

Given $K \geq 1$, we set

$$X = \frac{\sigma}{C-\rho_c-K\gamma}. \quad (41)$$

For $K = 0$ we set $X = 0$. The optimal K must also satisfy $\theta^h(X) > \Delta_{0,c}$ for all $h > K$. We take K to be the smallest index with these properties. Finally, we determine $d(\sigma)$ from Eq. (39).

For $\Delta_{0,c} \leq 0$, the function $\theta^h(X)$ is convex. Then

$$\frac{d}{dX}\theta^h(X) = \begin{cases} -1, & X < -\Delta_{0,c}, \\ -\frac{C-\rho_c-h\gamma}{C-(h-1)\gamma}, & X > -\Delta_{0,c}. \end{cases}$$

As above, K should be chosen so that Eq. (40) holds. If $K \geq 1$, set

$$X = \max \left\{ \frac{\sigma}{C-(K-1)\gamma}, \frac{\sigma + (\rho_c + \gamma)\Delta_{0,c}}{C-\rho_c-K\gamma} \right\}. \quad (42)$$

If $K = 0$ set $X = -\Delta_{0,c}$. We take K to be the smallest index satisfying Eq. (40).

We do not claim that these choices are optimal. However, in practice, K is usually close to H , resulting in a near-optimal choice.

To compute a probabilistic delay bound, we first set the right hand side of Eq. (34) equal to the desired violation probability and solve for σ , and then find $d(\sigma)$ by solving the optimization problem in Eq. (38) according to the procedure outlined above. Since there is no explicit term for γ , we optimize numerically over γ .

There are several special cases where we can find explicit solutions to the optimization problem in Eq. (38):

- $\Delta_{0,c} = \infty$: This is the case of blind multiplexing. Here, we obtain that the optimal solution is $\theta^1 = \dots = \theta^H = 0$, and we get

$$d(\sigma) = X = \frac{\sigma}{C-\rho_c-H\gamma}, \quad (43)$$

which is the same delay bound found in [6].

- $\Delta_{0,c} = 0$: This is the case of FIFO. The optimization problem greatly simplifies, because the constraints are linear. The second condition in Eq. (40) is always satisfied, leaving only the first condition which can be directly computed. We choose X according to Eq. (41) and compute

$$\theta^h = \frac{(h-K)\gamma X}{C-(h-1)\gamma}, \quad \text{for } h > K,$$

and obtain for the delay bound

$$d(\sigma) = \frac{\sigma}{C-\rho_c-K\gamma} \left(1 + \sum_{h>K} \frac{(h-K)\gamma}{C-(h-1)\gamma} \right). \quad (44)$$

- $\gamma = 0$: This cases arises in a deterministic scenario, in which bounds are never violated. Note that by setting $M = e^{B\alpha}$ and letting $\alpha \rightarrow \infty$ in the EBB model in Eq. (27), we obtain a leaky bucket with $E(t-s) = R(t-s) + B$. From Eq. (38), we see that necessarily $\theta^h = \theta$ for all h . Then the solution to the optimization gives either $\theta = 0$ or $X = 0$. For different values of $\Delta_{0,c}$, we obtain end-to-end delay bounds that apply to the deterministic network calculus. For FIFO scheduling, these bounds will be weaker than those obtained in [15]. While this may appear as a principal limitation of applying the stochastic network calculus to a deterministic network, a modification to the network service curve from Eq. (30) can strengthen the result.

Studying the optimization problem, we discover that the delay bounds of FIFO approach that of blind multiplexing when the utilization of the cross traffic ρ_c is small or H is large. The reason is that the first condition in Eq. (40) forces K to be close to H , which in turn forces Eq. (44) (with $\Delta_{0,c} = 0$) to converge to the blind multiplexing delay bound Eq. (43). In the numerical examples in the next section, we will see that FIFO approaches the delay bounds of BMUX even for modest path lengths and link utilizations.

Let us add some remarks on the case of non-homogeneous networks, where the link rates C^h , the scheduling parameters $\Delta_{0,h}$, and the cross traffic rates ρ_c^h , and $\varepsilon^h(\sigma)$ may take different values for each node $h = 1, \dots, H$. Then the computation of the delay bound reduces to the optimization problem in Eq. (39), where $\theta^h(X)$ is the smallest nonnegative solution of

$$(C^h - (h-1)\gamma)(X + \theta^h) - (\rho_c^h + \gamma) [X + \Delta_{0,h}(\theta^h)]_+ \geq \sigma.$$

As above, this is reduced to a minimization problem in a single variable. The minimum is achieved at one of at most $H + 1$ explicitly computable points.

V. NUMERICAL EXAMPLES

We present numerical examples of the end-to-end delay bounds derived in this paper, where we assume that the time unit is $T = 1$ ms. As traffic, we use discrete-time on-off Markov-Modulated Process with two states (OFF= 1, ON= 2), where the probability of the transitions ON \rightarrow OFF and OFF \rightarrow ON, are denoted by p_{12} and p_{21} . In one time unit in the ON state, the process transmits a fixed amount of data denoted by P . We assume that $p_{12} + p_{21} \leq 1$. The effective bandwidth $eb(s, t) = \frac{1}{st} \log E[e^{sA(t)}]$ of such a process is bounded by [5]

$$eb(s, t) \leq \frac{1}{s} \log \frac{1}{2} (p_{11} + p_{22}e^{sP} + \sqrt{(p_{11} + p_{22}e^{sP})^2 - 4(p_{11} + p_{22} - 1)e^{sP}}),$$

where $p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$. If A denotes the arrivals of an aggregate of N independent such flows, the traffic complies to the EBB model in Eq. (27) with $A \sim (1, N \cdot eb(s, t), s)$. We set the parameters to $P = 1.5$ kilobits, $p_{11} = 0.989$ and $p_{22} = 0.9$, resulting in a peak rate of 1.5 Mbps and an average rate of 0.15 Mbps.

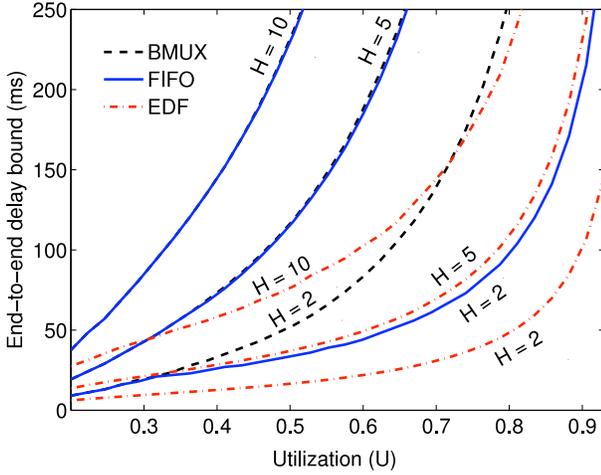


Fig. 2. Example 1: End-to-end delay bounds of through traffic for EDF ($d_0^* < d_c^*$), BMUX, and FIFO as a function of the total utilization U ($H = 2, 5, 10$, $U_0 = 15\%$ (constant), $\varepsilon = 10^{-9}$).

We consider a network as in Fig. 1, where all nodes have the same capacity $C = 100$ Mbps. The number of through flows is N_0 and the number of cross flows at each node is N_c . All flows have the same characteristics as discussed above. The load on a link is denoted by the utilization U , defined as a percentage of utilized capacity $U = (N_0 + N_c) * 0.15/100$. We use U_0 and U_c , to denote the utilization due to through and cross traffic, respectively, with $U = U_0 + U_c$. In all examples, we compute end-to-end delay bounds for the through flows with a violation probability set to $\varepsilon = 10^{-9}$.

A. Example 1

We evaluate the end-to-end delay bounds computed in Section IV as a function of the total utilization U . We consider the scheduling algorithms BMUX (for reference), FIFO, and EDF. For EDF we set the a priori delay constraints at each node to $d_0^* = \frac{d^{e2e}}{H}$ and $d_c^* = \frac{10d^{e2e}}{H}$ for the through and the cross traffic, respectively, where d^{e2e} is the computed end-to-end delay bound of the through traffic with EDF. In the example, the number of through flows is kept constant at $N_0 = 100$, corresponding to a utilization of $U_c = 15\%$, while the number of through flows is increased, so that the total utilization $U = U_0 + U_c$ covers a range of $20\% \leq U \leq 95\%$. We include results for path lengths of $H = 2, 5, 10$.

In Fig. 2 we see that the delay bounds of FIFO are indistinguishable from those of BMUX as early as $H = 5$. At the same time, the delay bounds of BMUX and FIFO are noticeably larger than those of EDF, and the difference seems to increase with the network size. Note that, since the BMUX bounds from [6] represent the state-of-the-art for end-to-end delay bound analysis, the graphs for EDF illustrate the benefits of applying the scheduler-aware analysis from this paper.

B. Example 2

This example examines in more detail how scheduling algorithms influence delays in a network. We use the same setup as in Example 1, however, we keep the link utilization constant at $U = 50\%$ and instead vary the traffic mix U_c/U .

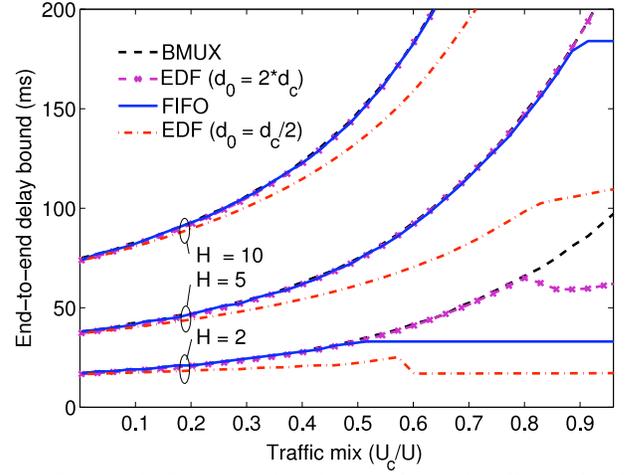


Fig. 3. Example 2: End-to-end delay bounds of through traffic as a function of the traffic mix U_c/U ($H = 2, 5, 10$, $U = 50\%$ (constant), $\varepsilon = 10^{-9}$).

For EDF we consider two cases: shorter delay constraints for through traffic ($d_0^* = d_c^*/2$) and longer delay constraints for through traffic ($d_0^* = 2d_c^*$). The results are shown in Fig. 2. Even though the total utilization at the node is constant, we observe different delay bounds for the evaluated schedulers. We can observe several counteracting effects. Interpreting FIFO as an EDF scheduler with identical a priori delay constraints, i.e., $d_0^* = d_c^*$, and BMUX as an EDF scheduler with $d_0^* = \infty$ and $d_c^* < \infty$, we see, for $H = 2$, that a larger ratio d_0^*/d_c^* causes end-to-end delay bounds to become more sensitive to increased cross traffic. In particular for EDF ($d_0^* = d_c^*/2$), the end-to-end delay bounds are almost insensitive to the traffic mix at $H = 2$, and even decrease when the fraction of cross traffic is increased. This is due to the fact that with the choice of delay bounds, cross traffic mostly has lower precedence, while through traffic maintains its locally FIFO property. Thus, increasing cross traffic creates more opportunities to ‘jump’ ahead in the buffer, leading to shorter delays. When the path length H is increased, these effects are diminished as the scaling law for the delays of Δ -schedulers makes all schedulers have a similar performance as BMUX, where increased cross traffic results in a steep increase of delay bounds.

C. Example 3

We illustrate the scaling of end-to-end delay bounds as a function of the path length H . Using the same MMPP traffic parameters as before, we set $N_0 = N_c$, and compute the delay bounds for different values of $U = 10, 50, 100\%$. For EDF, we set delay constraints at each node to $d_0^* = \frac{d^{e2e}}{H}$, $d_c^* = \frac{10d^{e2e}}{H}$. For BMUX, FIFO, and EDF we compute delay bounds as discussed in Section IV. For BMUX, we additionally present a delay analysis that is based on adding up per-node delay bounds (as sketched in the first paragraph of this paper), using a discrete-time version of such an analysis in [6]. In Fig. 4 we see that the method of adding per-node delay bounds yields loose bounds. In fact, the growth of delays can be shown to be $O(H^3 \log H)$ with discrete time.

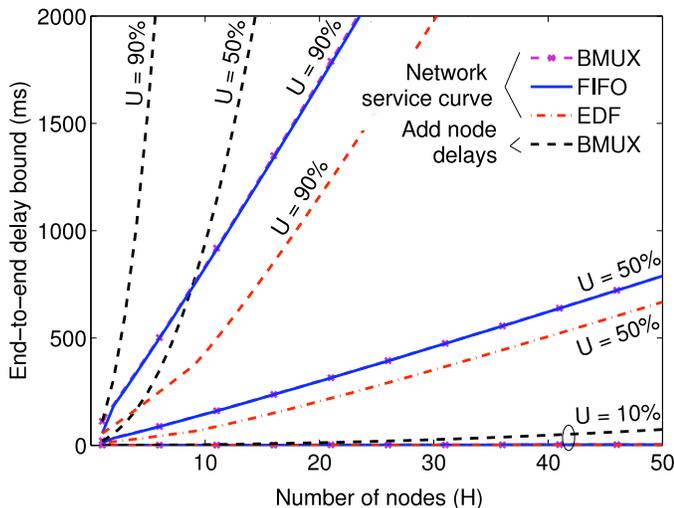


Fig. 4. Example 3: End-to-end delay bounds of through traffic vs. path length ($U_0 = U_c$, $U = 10, 50, 90\%$, $\varepsilon = 10^{-9}$).

The delay bounds using our network service curve increase essentially linearly for all schedulers, following the predicted asymptotic growth of $\Theta(H \log H)$. These findings generalize a comparison of a network delay analysis with a node-by-node analysis for BMUX in [6] for blind multiplexing. Note that, for the depicted range of H , the delay bounds for FIFO and BMUX appear identical, while delays for EDF (with the selected a priori delay constraints) are noticeably lower at higher utilizations.

VI. CONCLUSION

We presented an end-to-end analysis of probabilistic delay bounds for Δ -schedulers, a class of algorithms that can be defined in terms of constants which specify time intervals when arrivals have precedence over buffered traffic. We derived a statistical service curve that can characterize the operations of any given Δ -scheduler to a degree that it recovers tight delay bound conditions. We performed an end-to-end delay bound analysis by providing explicit solutions to an optimization problem. We presented numerical examples that illustrated the impact of the choice of the scheduling algorithm on end-to-end delays. We observed that delay bounds of FIFO frequently approach those of blind multiplexing when the number of traversed nodes is large. At the same time, the delays of EDF scheduling were generally markedly different, and maintained the difference on long paths. Thus, our study does not confirm findings in other works that longer routes in a network reduce the effectiveness of link scheduling for providing delay differentiation to flows.

ACKNOWLEDGMENTS

The research in this paper is supported in part by the National Science Foundation under grants CNS-0435061, and by the Natural Sciences and Engineering Research Council of Canada under two Discovery grants and a Strategic project grant.

REFERENCES

- [1] M. Andrews. Probabilistic end-to-end delay bounds for earliest deadline first scheduling. In *Proc. IEEE Infocom*, pages 603–612, March 2000.
- [2] T. Bonald, A. Proutiere, and J. W. Roberts. Statistical guarantees for streaming flows using expedited forwarding. In *Proc. IEEE Infocom*, pages 1104–1112, 2001.
- [3] R. Boorstyn, A. Burchard, J. Liebeherr, and C. Oottamakorn. Statistical service assurances for traffic scheduling algorithms. *IEEE Journal on Selected Areas in Communications*, 18(12):2651–2664, December 2000.
- [4] A. Burchard, J. Liebeherr, and F. Ciucu. On $\Theta(H \log H)$ scaling of network delays. In *Proc. of IEEE Infocom*, pages 1866–1874, May 2007.
- [5] C.-S. Chang. *Performance Guarantees in Communication Networks*. Springer Verlag, 2000.
- [6] F. Ciucu, A. Burchard, and J. Liebeherr. Scaling properties of statistical end-to-end bounds in the network calculus. *IEEE Transactions on Information Theory*, 52(6):2300–2312, June 2006.
- [7] R. Cruz. A calculus for network delay, parts I and II. *IEEE Transactions on Information Theory*, 37(1):114–141, January 1991.
- [8] R. L. Cruz. SCED+: Efficient management of quality of service guarantees. In *Proc. IEEE Infocom*, pages 625–634, March 1998.
- [9] D. Y. Eun and N. B. Shroff. Network decomposition: Theory and practice. *to appear in IEEE/ACM Trans. on Networking*, 13(3):526–539, June 2005.
- [10] L. Georgiadis, R. Guerin, V. Peris, and K. Sivarajan. Efficient network QoS provisioning based on per node traffic shaping. *IEEE/ACM Transactions on Networking*, 4(4):482–501, Aug. 1996.
- [11] Y. Ghiassi-Farrokhhfal and J. Liebeherr. Output characterization of constant bit rate traffic in fifo networks. *IEEE Communications Letters*, 13(8), August 2009.
- [12] P. Giacomazzi. Closed-form analysis of end-to-end network delay with markov-modulated poisson and fluid traffic. *Computer Communications*, 32(4):640–648, March 2009.
- [13] Y. Jiang and Y. Liu. *Stochastic Network Calculus*. Springer, 2008.
- [14] J. Y. Le Boudec and P. Thiran. *Network Calculus*. Springer Verlag, Lecture Notes in Computer Science, LNCS 2050, 2001.
- [15] L. Lenzini, E. Mingozzi, and G. Stea. Delay bounds for fifo aggregates: A case study. *Computer Communications*, 28(3), 2005.
- [16] C. Li, A. Burchard, and J. Liebeherr. A network calculus with effective bandwidth. *IEEE/ACM Trans. on Networking*, 15(6):1442–1453, December 2007.
- [17] J. Liebeherr, D. Wrege, and D. Ferrari. Exact admission control for networks with bounded delay services. *IEEE/ACM Transactions on Networking*, 4(6):885–901, December 1996.
- [18] S. Mao and S. S. Panwar. A survey of envelope processes and their applications in quality of service provisioning. *IEEE Communications Surveys & Tutorials*, 8(3):2–20, 3rd Quarter 2006.
- [19] A. K. Parekh and R. G. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: the Single-Node Case. *IEEE/ACM Transactions on Networking*, 1(3):344–357, June 1993.
- [20] J. Qiu and E. W. Knightly. Inter-class resource sharing using statistical service envelopes. In *Proc. IEEE Infocom*, pages 1404–1411, March 1999.
- [21] J. B. Schmitt, F. A. Zdarsky, and M. Fidler. Delay bounds under arbitrary multiplexing: When network calculus leaves you in the lurch. In *Proc. IEEE Infocom*, pages 1669–1677, 2008.
- [22] V. Sivaraman and F. M. Chiussi. Providing end-to-end statistical delay guarantees with earliest deadline first scheduling and per-hop traffic shaping. In *Proc. IEEE Infocom*, pages 603–612, March 2000.
- [23] J. Strauss, D. Katabi, and F. Kaashoek. A measurement study of available bandwidth estimation tools. In *Proc. 3rd ACM Sigcomm Conference on Internet Measurement*, pages 39–44, October 2003.
- [24] O. Yaron and M. Sidi. Performance and stability of communication networks via robust exponential bounds. *IEEE/ACM Transactions on Networking*, 1(3):372–385, June 1993.
- [25] Y. Ying, R. Mazumdar, C. Rosenberg, and F. Guillemin. The burstiness behavior of regulated flows in networks. In *Proc. 4th IFIP Networking 2005*, pages 918–929, May 2005.
- [26] H. Zhang. Service disciplines for guaranteed performance service in packet-switching networks. *Proceedings of the IEEE*, 83(10):1374–1396, Oct. 1995.