A Calculus for End-to-end Statistical Service Guarantees *

Technical Report: University of Virginia, CS-2001-19 (2nd revised version)

Almut Burchard † *Jörg Liebeherr* * *Stephen Patek* **

† Department of Mathematics
* Department of Computer Science
** Department of Systems and Information Engineering
University of Virginia
Charlottesville, VA 22904

Abstract

The deterministic network calculus offers an elegant framework for determining delays and backlog in a network with deterministic service guarantees to individual traffic flows. This paper addresses the problem of extending the network calculus to a probabilistic framework with statistical service guarantees. Here, the key difficulty relates to expressing, in a statistical setting, an end-to-end (network) service curve as a concatenation of per-node service curves. The notion of an *effective service curve* is developed as a probabilistic bound on the service received by an individual flow. It is shown that per-node effective service curves can be concatenated to yield a network effective service curve.

Key Words: Quality-of-Service, Service Differentiation, Statistical Service, Network Calculus.

1 Introduction

The deterministic network calculus recently evolved as a fundamental theory for quality of service (QoS) networks, and has provided powerful tools for reasoning about delay and backlog in a network with service guarantees to individual traffic flows. Using the notion of arrival envelopes and service curves [12], several recent works have shown that delay and backlog bounds can be concisely expressed in a min-plus algebra [1, 5, 8].

However, the deterministic view of traffic generally overestimates the actual resource requirements of a flow and results in a low utilization of available network resources. This motivates the search for a statistical network calculus that can exploit statistical multiplexing, while preserving the algebraic aspects of the deterministic calculus. The problem of developing a probabilistic network calculus has been the subject of several studies. Kurose [16] uses the concept of stochastic ordering and obtains bounds on the distribution of delay and buffer occupancy of a flow in a network with FIFO scheduling. Chang [7] presents probabilistic bounds on output burstiness, backlog and delays in a network where the moment generating functions of arrivals

*This work is supported in part by the National Science Foundation through grants ANI-9730103, ECS-9875688 (CAREER), ANI-9903001, DMS-9971493, and ANI-0085955, and by an Alfred P. Sloan research fellowship.

are exponentially bounded. Different bounds for stochastically bounded arrivals are derived by Yaron and Sidi [22] and Starobinski and Sidi [21]. The above results can be used to determine stochastic end-to-end performance bounds. Results on statistical end-to-end delay guarantees in a network have been obtained for specific scheduling algorithms, such as EDF [19, 20], and GPS [15], and a class of coordinated scheduling algorithms [2, 17]. Several researchers have considered probabilistic formulations of service curves. Cruz defines a probabilistic service curve which violates a given deterministic service curve according to a certain distribution [13]. Chang (see [9], Chp. 7) presents exercises which hint at a statistical network calculus for the class of 'dynamic F-servers'. Finally, Knightly and Qiu [18] derive 'statistical service envelopes' as time-invariant lower bounds on the service received by an aggregate of flows.

With exception of ([9], Chp. 7), none of the cited works express statistical end-to-end performance bounds in a min-plus algebra, and it has been an open question whether a statistical network calculus can be developed in this setting. The contribution of this paper is the presentation of a statistical network calculus that uses the min-plus algebra [1, 5, 8]. The advantage of using the min-plus algebra is that end-to-end guarantees can be expressed as a simple concatenation of single node guarantees, which, in turn, can be exploited to achieve simple probabilistic bounds.

We define an *effective service curve*, which is, with high certainty, a probabilistic bound on the service received by a single flow. We will show that the main results of the deterministic network calculus carry over to the statistical framework we present. Our derivations reveal a difficulty that occurs when calculating probabilistic service guarantees for multiple nodes. We show that the problem can be overcome either by adding assumptions on the traffic at nodes or by modifying the definition of the effective service curve. The results in this paper are set in a continuous time model with fluid left-continuous traffic arrival functions, as is common for network delay analysis in the deterministic network calculus. A node represents a router (or switch) in a network. Packetization delays and other effects of discrete-sized packets, such as the non-preemption of packet transmission, are ignored. We refer to [9] for the issues involved in relaxing these assumptions for the analysis of packet networks. When analyzing delays in a network, all processing overhead and propagation delays are ignored. As in the deterministic network calculus, arrivals from a traffic flow to the network satisfy deterministic upper bounds, which are enforced by a deterministic regulator.

The remaining sections of this paper are structured as follows. In Section 2, we review the notation and key results of the deterministic network calculus. In Section 3 we introduce effective service curves and present the results for a statistical network calculus in terms of effective service curves. In Section 4 we provide a discussion that motivates our revised definition of an effective service curve. In Section 5, we present brief conclusions.

2 Network Calculus Preliminaries

The deterministic network calculus, which was created in [10, 11] and fully developed in the last decade, provides concise expressions for upper bounds on the backlog and delay experienced by an individual flow at one or more network nodes. An attractive feature of the network calculus is that end-to-end bounds can often be easily obtained from manipulations of the per-node bounds.

In this section we review some notation and results from the deterministic network calculus. This section is not a comprehensive summary of the network calculus and we refer to [1, 6, 9] for a complete discussion.

2.1 Operators

Much of the formal framework of the network calculus can be elegantly expressed in a min-plus algebra [3], complete with convolution and deconvolution operators for functions. Generally, the functions in this paper are non-negative, non-decreasing, and left-continuous, defined over time intervals $[0, t]$. We assume for a given function f that $f(t) = 0$ if $t \leq 0$.

The *convolution* $f * g$ of two functions f and g, is defined as

$$f * g(t) = \inf_{\tau \in [0,t]} \{ f(t - \tau) + g(\tau) \} \ . \tag{1}$$

The *deconvolution* $f \oslash g$ of two functions f and g is defined as

$$f \oslash g(t) = \sup_{\tau \geq 0} \{ f(t + \tau) - g(\tau) \} \ . \tag{2}$$

For $\tau \geq 0$, the *impulse function* δ_τ is defined as

$$\delta_\tau(t) = \begin{cases} \infty \ , & \text{if } t > \tau \ , \\ 0 \ , & \text{if } t \leq \tau \ . \end{cases} \tag{3}$$

If f is nondecreasing, we have the formulas

$$f(t - \tau) \ = \ f * \delta_\tau(t) \ , \tag{4}$$
$$f(t + \tau) \ = \ f \oslash \delta_\tau(t) \ . \tag{5}$$

We refer to [3, 6, 9] for a detailed discussion of the properties of the min-plus algebra and the properties of the convolution and deconvolution operators.

2.2 Arrival functions and Service Curves

Let us consider the traffic arrivals to a single network node. The arrivals of a flow in the time interval $[0, t)$ are given in terms of a function $A(t)$. The departures of a flow from the node in the time interval $[0, t)$ are denoted by $D(t)$, with $D(t) \leq A(t)$. The backlog of a flow at time t, denoted by $B(t)$, is given by

$$B(t) = A(t) - D(t) \ . \tag{6}$$

The delay at time t, denoted as $W(t)$, is the delay experienced by an arrival which departs at time t, given by

$$W(t) = \inf\{ d \geq 0 \mid A(t - d) \leq D(t) \} \ . \tag{7}$$

We will use $A(x, y)$ and $D(x, y)$ to denote the arrivals and departures in the time interval $[x, y)$, with $A(x, y) = A(y) - A(x)$ and $D(x, y) = D(y) - D(x)$.

We make the following assumptions on the arrival functions.

(A1) *Non-Negativity.* The arrivals in any interval of time are non-negative. That is, for any $x < y$, we have $A(y) - A(x) \ \geq 0$.

(A2) *Upper Bound.* The arrivals A of a flow are bounded by a subadditive[1] function A^*, called the *arrival envelope*,[2] such that $A(t+\tau) - A(t) \leq A^*(\tau)$ for all $t, \tau \geq 0$.

A *minimum service curve* for a flow is a function S which specifies a lower bound on the service given to the flow such that, for all $t \geq 0$,

$$D(t) \geq A * S(t) \ . \tag{8}$$

A *maximum service curve* for a flow is a function $\overline{S}$ which specifies an upper bound on the service given to a flow such that, for all $t \geq 0$,

$$D(t) \leq A * \overline{S}(t) \ . \tag{9}$$

Minimum service curves play a larger role in the network calculus since they provide service guarantees. Therefore, we, as the related literature, often refer to a minimum service curve simply as a service curve. If no maximum service curve is explicitly given, one can use $\overline{S}(t) = Ct$, where C is the link capacity.

The following two theorems summarize some key results of the deterministic network calculus. These results have been derived in [1, 5, 8]. We follow the notation used in [1].

Theorem 1 Deterministic Calculus [1, 5, 8]. *Given a flow with arrival envelope A^* and with minimum service curve S, the following hold:*

1. **Output Envelope.** *The function $D^* = A^* \oslash S$ is an envelope for the departures, in the sense that, for all $t, \tau \geq 0$,*

$$D^*(t) \geq D(t+\tau) - D(\tau) \ . \tag{10}$$

2. **Backlog Bound.** *An upper bound for the backlog, denoted by b_{max}, is given by*

$$b_{max} = A^* \oslash S(0) \ . \tag{11}$$

3. **Delay Bound.** *An upper bound for the delay, denoted by d_{max}, is given by*

$$d_{max} = \inf \left\{ d \geq 0 \mid \forall t \geq 0 : \ A^*(t-d) \leq S(t) \right\} \ . \tag{12}$$

The next theorem states that the service curves of a flow at the nodes on its route can be concatenated to define a network service curve, which expresses service guarantees offered to the flow by the network as a whole.

Theorem 2 Concatenation of Deterministic Network Service Curves [1, 5, 8]. *Suppose a flow passes through H nodes in series, as shown in Figure 1, and suppose the flow is offered minimum and maximum service curves S^h and $\overline{S}^h$, respectively, at each node $h = 1, \ldots, H$. Then, the sequence of nodes provides minimum and maximum service curves S^{net} and $\overline{S}^{net}$, which are given by*

$$S^{net} = S^1 * S^2 * \ldots * S^H \ , \tag{13}$$
$$\overline{S}^{net} = \overline{S}^1 * \overline{S}^2 * \ldots * \overline{S}^H \ . \tag{14}$$

[1] A function f is *subadditive* if $f(x+y) \leq f(x) + f(y)$, for all $x, y \geq 0$, or, equivalently, if $f(t) = f * f(t)$.

[2] A function E is called an *envelope* for a function f if $f(t+\tau) - f(\tau) \leq E(t)$ for all $t, \tau \geq 0$, or, equivalently, if $f(t) \leq E * f(t)$, for all $t \geq 0$.

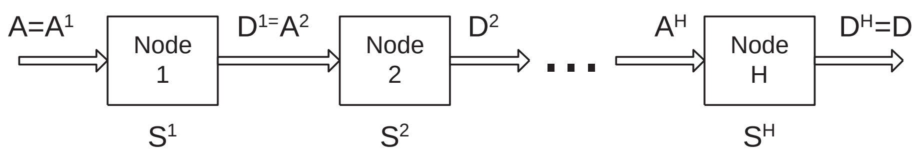

Figure 1: Traffic of a flow through a set of H nodes. Let A^h and D^h denote the arrival and departures at the h-th node, with $A^1 = A$, $A^h = D^{h-1}$ for $h = 2, \dots, H$ and $D^H = D$.

S^{net} and $\overline{S}^{net}$ will be referred to as *network service curves*, and Eqs. (13)–(14) will be called the *concatenation formulas*.

With Theorems 1 and 2 network service curves can be used to determine bounds on delay and backlog for individual flows in a network. There are many additional properties and refinements that have been derived for the deterministic calculus. However, in this paper we will concern ourselves only with the results above.

3 Statistical Network Calculus

We now approach the network calculus in a probabilistic framework. Arrivals and departures from a flow to the network in the time interval $[0, t)$ are described by random processes $A(t)$ and $D(t)$ satisfying assumptions (A1) and (A2). The random processes are defined over an underlying joint probability space that we suppress in our notation. The statistical network calculus makes service guarantees for individual flows, where each flow is allocated a probabilistic service in the form of an 'effective service curve'.

Given a flow with arrival process A, a *(minimum) effective service curve* is a nonnegative function $\mathcal{S}^\varepsilon$ that satisfies for all $t > 0$,

$$Pr\Big\{D(t) \geq A * \mathcal{S}^\varepsilon(t)\Big\} \geq 1 - \varepsilon \,. \tag{15}$$

Note that the effective service curve is a non-random function. We omit the corresponding definition of a *maximum effective service curve*.

The following theorem is a probabilistic counterpart to Theorem 1.

Theorem 3 Statistical Calculus. *Given a flow with arrival process A satisfying assumptions (A1)–(A2), and given an effective service curve $\mathcal{S}^\varepsilon$, the following hold:*

1. **Output Envelope.** *The function $A^* \oslash \mathcal{S}^\varepsilon$ is a probabilistic bound for the departures on $[0, t]$, in the sense that, for all $t, \tau > 0$,*

$$Pr\left\{D(t, t+\tau) \leq A^* \oslash \mathcal{S}^\varepsilon(\tau)\right\} \geq 1 - \varepsilon \,. \tag{16}$$

2. **Backlog Bound.** *A probabilistic bound for the backlog is given by $b_{max} = A^* \oslash \mathcal{S}^\varepsilon(0)$, in the sense that, for all $t > 0$,*

$$Pr\left\{B(t) \leq b_{max}\right\} \geq 1 - \varepsilon \,. \tag{17}$$

3. **Delay Bound.** *A probabilistic bound for the delay is given by,*

$$d_{max} = \inf \{ d \geq 0 \mid \forall t \geq 0 : \; A^*(t-d) \leq \mathcal{S}^{\varepsilon}(t) \} \;, \tag{18}$$

in the sense that, for all $t > 0$*,*

$$Pr \{ W(t) \leq d_{max} \} \geq 1 - \varepsilon \;. \tag{19}$$

By setting $\varepsilon = 0$ in Theorem 3, we can recover the bounds of Theorem 1 with probability one.

Proof. The proof uses on several occasions that the inequality

$$A(t+\tau) - A * g(t) \leq A^* \oslash g(\tau) \tag{20}$$

holds for any nonnegative function g and for all $t, \tau \geq 0$. To see this inequality, we compute

$$\begin{aligned} A(t+\tau) - A * g(t) &= A(t+\tau) - \inf_{x \in [0,t]} \big\{ A(t-x) + g(x) \big\} && (21) \\ &= \sup_{x \in [0,t]} \big\{ A(t-x, t+\tau) - g(x) \big\} && (22) \\ &\leq \sup_{x \geq 0} \big\{ A^*(\tau + x) - g(x) \big\} && (23) \\ &= A^* \oslash g(\tau) \;. && (24) \end{aligned}$$

Eqn. (21) expands the convolution operator. Eqn. (22) takes $A(t+\tau)$ inside the infimum and uses that $A(t+\tau) - A(t-x) = A(t-x, t+\tau)$. Eqn. (23) uses that $A(t-x, t+\tau) \leq A^*(x+\tau)$ for all $x \leq t$ by definition of an arrival envelope, and extends the range of the supremum. Finally, Eqn. (24) uses the definition of the deconvolution operator.

1. Proof of the Output Bound. For any fixed $t, \tau > 0$, we have

$$\begin{aligned} 1 - \varepsilon &\leq Pr \{ D(t) \geq A * \mathcal{S}^{\varepsilon}(t) \} && (25) \\ &= Pr \{ D(t, t+\tau) \leq D(t+\tau) - A * \mathcal{S}^{\varepsilon}(t) \} && (26) \\ &\leq Pr \{ D(t, t+\tau) \leq A(t+\tau) - A * \mathcal{S}^{\varepsilon}(t) \} && (27) \\ &\leq Pr \{ D(t, t+\tau) \leq A^* \oslash \mathcal{S}^{\varepsilon}(\tau) \} \;. && (28) \end{aligned}$$

Eqn. (25) holds by the definition of the effective service curve $\mathcal{S}^{\varepsilon}$. Eqn. (26) uses that $D(t, t+\tau) = D(t+\tau) - D(t)$. Eqn. (27) uses that departures in $[0, t)$ cannot exceed arrivals, that is, $D(t) \leq A(t)$ for all $t \geq 0$. Finally, Eqn. (28) uses that $A(t+\tau) - A * \mathcal{S}^{\varepsilon}(t) \leq A^* \oslash \mathcal{S}^{\varepsilon}(\tau)$ by Eqn. (20).

2. Proof of the Backlog Bound. Since $B(t) = A(t) - D(t)$ and with the definition of the effective service curve, we can write

$$\begin{aligned} 1 - \varepsilon &\leq Pr \{ D(t) \geq A * \mathcal{S}^{\varepsilon}(t) \} && (29) \\ &= Pr \{ B(t) \leq A(t) - A * \mathcal{S}^{\varepsilon}(t) \} && (30) \\ &\leq Pr \{ B(t) \leq A^* \oslash \mathcal{S}^{\varepsilon}(0) \} \;. && (31) \end{aligned}$$

Eqn. (29) holds by definition of the effective service curve. Eqn. (30) uses that $B(t) = A(t) - D(t)$, and Eqn. (31) uses that $A(t) - A * \mathcal{S}^{\varepsilon}(t) \leq A^* \oslash \mathcal{S}^{\varepsilon}(0)$ by Eqn. (20).

3. Proof of the Delay Bound. The delay bound is proven by estimating the probability that the output $D(t)$ exceeds the arrivals $A(t - d_{max})$.

$$\begin{aligned} 1 - \varepsilon &\leq Pr\left\{D(t) \geq A * \mathcal{S}^{\varepsilon}(t)\right\} && (32) \\ &\leq Pr\left\{D(t) \geq A * (A^{*} * \delta_{d_{max}})(t)\right\} && (33) \\ &\leq Pr\left\{D(t) \geq (A * A^{*}) * \delta_{d_{max}})(t)\right\} && (34) \\ &\leq Pr\left\{D(t) \geq A(t - d_{max})\right\} . && (35) \end{aligned}$$

Eqn. (32) uses the definition of the effective service curve $\mathcal{S}^{\varepsilon}$. Eqn. (33) uses the definition of the impulse function in Eqn. (4) and the definition of d_{max} in Eqn. (18). Eqn. (34) follows from the associativity of the convolution, and Eqn. (34) uses the definition of an arrival envelope.

□

A probabilistic counterpart to Theorem 2 can be formulated as follows.

Theorem 4 Concatenation of Effective Service Curves. *Consider a flow that passes through H network nodes in series, as shown in Figure 1. Assume that effective service curves are given by nondecreasing functions $\mathcal{S}^{h,\varepsilon}$ at each node ($h = 1, \ldots, H$). Then, for any $t \geq 0$,*

$$Pr\Big\{D(t) \geq A * \big(\mathcal{S}^{1,\varepsilon} * \ldots * \mathcal{S}^{H,\varepsilon} * \delta_{(H-1)a}\big)(t)\Big\} \geq 1 - \varepsilon \left(1 + (H-1)\frac{t}{a}\right) , \tag{36}$$

where $a > 0$ is an arbitrary parameter.

Again, we can can recover the deterministic result from Theorem 2. By setting $\varepsilon = 0$, the results in Eqn. (36) hold with probability one. Then by letting $a \longrightarrow 0$, we obtain Theorem 2 almost surely.

Proof. We proceed in three steps. In the first step, we modify the effective service curve to give lower bounds on the departures simultaneously for all times in the entire interval $[0, t]$. In the second step, we perform a deterministic calculation. The proof concludes with a simple probabilistic estimate.

Step 1: Uniform probabilistic bound on $[0, t]$. Suppose that S^{ε} is a nondecreasing effective service curve, that is

$$\forall x \in [0, t] : \quad Pr\Big\{D(x) \geq A * \mathcal{S}^{\varepsilon}(x)\Big\} \geq 1 - \varepsilon . \tag{37}$$

We will show that then, for any choice of $a > 0$,

$$Pr\Big\{\forall x \in [0, t] : D(x) \geq A * \mathcal{S}^{\varepsilon}(x - a)\Big\} \geq 1 - \varepsilon\frac{t}{a} . \tag{38}$$

To see this, fix $a > 0$, set $x_j = ja$, and consider the events

$$E_j = \Big\{D(x_j) \geq A * \mathcal{S}^{\varepsilon}(x_j)\Big\} , \quad j = 1, \ldots , n - 1, \tag{39}$$

where $n = \lceil \ell/a \rceil$ is the smallest integer no larger than t/a. Let $x \in [0, t]$ be arbitrary, and let j the largest integer with $x_j \leq x$, so that $x - x_j \leq a$. If E_j occurs, then

$$D(x) \geq D(x_j) \geq A * \mathcal{S}^{\varepsilon}(x_j) \ \geq A * \mathcal{S}^{\varepsilon}(x - a) , \tag{40}$$

where we have used the fact that $\mathcal{S}^\varepsilon$ is nondecreasing in the last step. It follows that

$$\begin{aligned} Pr\big\{\forall x \in [0,t] : D(x) \ge A * \mathcal{S}^\varepsilon(x-a)\big\} &\ge Pr\big\{\forall j = 1, \ldots, n : \ D(x_j) \ge A * \mathcal{S}^\varepsilon(x_j)\big\} \quad (41) \\ &= Pr\Big\{\bigcap_{0<j\le n} E_j\Big\} \quad (42) \\ &\ge 1 - n\varepsilon \ , \quad (43) \end{aligned}$$

which proves Eqn. (38). Thus, the assumptions of the theorem imply that

$$Pr\left\{\begin{array}{lll} \forall x \in [0,t] : & D^h(x) \ge A^h * \mathcal{S}^{h,\varepsilon} * \delta_(x) \ , & h < H \\ & D^H(t) \ge A^h * \mathcal{S}^{H,\varepsilon}(t) \ , & h = H \end{array}\right\} \ge 1 - \varepsilon\left(1 + (H-1)\frac{t}{a}\right) \ . \quad (44)$$

Step 2: A deterministic argument. Suppose that, for a particular sample path,

$$\left\{\begin{array}{lll} \forall x \in [0,t] : & D^h(x) \ge A^h * \mathcal{S}^{h,\varepsilon} * \delta_(x) \ , & h < H \ , \\ & D^H(t) \ge A^h * \mathcal{S}^{H,\varepsilon}(t) \ , & h = H \ . \end{array}\right. \quad (45)$$

Inserting the first line of Eqn. (45) with $h = H - 1$ into the second line yields

$$\begin{aligned} D^H(t) &\ge \inf_{x\in[0,t]}\Big\{\inf_{y\in[0,x]}\big\{A^{H-1}(t-x-y) + \big(\mathcal{S}^{H-1,\varepsilon} * \delta_a\big)(y)\big\} + \mathcal{S}^{H,\varepsilon}(x)\Big\} \quad (46) \\ &= A * \big(\mathcal{S}^{H-1,\varepsilon} * \mathcal{S}^{H,\varepsilon} * \delta_a\big)(t) \ . \quad (47) \end{aligned}$$

An induction over the number of nodes shows that Eqn. (45) implies that $A = A^1$ and $D = D^h$ satisfy

$$D(t) \ge A * \big(\mathcal{S}^{1,\varepsilon} * \ldots * \mathcal{S}^{H,\varepsilon} * \delta_{(H-1)a}\big)(t) \ . \quad (48)$$

Step 3: Conclusion. We estimate

$$\begin{aligned} & Pr\big\{D(t) \ge A * \big(\mathcal{S}^{1,\varepsilon} * \ldots * \mathcal{S}^{H,\varepsilon} * \delta_{(H-1)a}\big)(t) \ \big\} \quad (49) \\ & \ge Pr\{ \text{ Eqn. (45) is satisfied } \} \quad (50) \\ & \ge 1 - \varepsilon\left(1 + (H-1)\varepsilon\frac{t}{a}\right) \ . \quad (51) \end{aligned}$$

The first inequality follows from the fact that Eqn. (45) implies Eqn. (48). The second inequality merely uses Eqn. (44). □

Since the bound in Eqn. (36) deteriorates as t becomes large, Theorem 4 is of limited practical value. To explain why Eqn. (36) deteriorates, consider a network as shown in Figure 1, with $H = 2$ nodes. An effective service curve $\mathcal{S}^{2,\varepsilon}$ in the sense of Eqn. (15) at the second node guarantees that, for any given time t, the departures from this node are with high probability bounded below by

$$D^2(t) \ge A^2 * \mathcal{S}^{2,\varepsilon}(t) = \inf_{\tau\in[0,t]}\big\{A^2(t-\tau) + \mathcal{S}^{2,\varepsilon}(\tau)\big\} \ . \quad (52)$$

Suppose that the infimum in Eqn. (52) is assumed at some value $\hat{\tau} \le t$. Since the departures from the first node are random, even if the arrivals to the first node satisfy the deterministic bound A^*, $\hat{\tau}$ is a random variable. An effective service curve $\mathcal{S}^{1,\varepsilon}$ at the first node guarantees that for any arbitrary but fixed time x, the arrivals $A^2(x) = D^1(x)$ to the second node are with high probability bounded below by

$$D^1(x) \geq A^1 * \mathcal{S}^{1,\varepsilon}(x) \; . \tag{53}$$

Since $\hat{\tau}$ is a random variable, we cannot simply evaluate Eqn. (53) for $x = t - \hat{\tau}$ and use the resulting bound in Eqn. (52). Furthermore, there is, a priori, no time-independent bound on the distribution of $\hat{\tau}$. Note that the above issue does not arise in the deterministic calculus, since deterministic service curves make service guarantees that hold for all values of x.

We conclude that, in a probabilistic setting, additional assumptions are required to establish time-independent bounds on the range of the infimum, and, in that way, obtain probabilistic network service curves that do not deteriorate with time. One example of such an assumption is to add the condition that

$$Pr\left\{ D(t) \geq \inf_{x \in [0,T]} \big\{ A(t-x) + \mathcal{S}^{\varepsilon}(x) \big\} \right\} \geq 1 - \varepsilon \; . \tag{54}$$

This condition imposes a limit on the range of the convolution. The condition can be satisfied for a given effective service curve $\mathcal{S}^{\varepsilon}$ and arrival envelope A^* by choosing T such that $A^*(T) \leq \mathcal{S}^{\varepsilon}(T)$, which guarantees that

$$A * \mathcal{S}^{\varepsilon}(t) = \inf_{x \in [0,T]} \big\{ A(t-x) + \mathcal{S}^{\varepsilon}(x) \big\} \; . \tag{55}$$

Theorem 5 *Assume that all hypotheses of Theorem 4 are satisfied, and additionally, that there exists a number $T \geq 0$ such that A^h and $\mathcal{S}^{h,\varepsilon}$ satisfy Eqn. (54) for $h = 1, \ldots, H$. Then, for any choice of $a > 0$,*

$$\mathcal{S}^{net,\varepsilon'} = \mathcal{S}^{1,\varepsilon} * \ldots * \mathcal{S}^{H,\varepsilon} * \delta_{(H-1)a} \tag{56}$$

is an effective network service curve, with violation probability bounded by

$$\varepsilon' \leq \; H\varepsilon \left(1 + (H-1) \frac{T+a}{2a} \right) \; . \tag{57}$$

More precisely, $\mathcal{S}^{net,\varepsilon'}$ satisfies

$$Pr\left\{ D(t) \geq \inf_{x \in [0,H(T+a)]} \big\{ A(t-x) + \mathcal{S}^{net,\varepsilon'}(x) \big\} \right\} \geq 1 - \varepsilon' \; . \tag{58}$$

The bounds of this network service curve deteriorate with the number of nodes H, but, different from Theorem 4, the bounds are not dependent on t. Rather the bounds depend on a time scale T as used in Eqn. (54). A key issue, which is not addressed in this paper, relates to establishing T for an arbitrary node in the network.

Proof. The proof is analogous to the proof of Theorem 4, and proceeds in the same three steps.

Step 1: Uniform probabilistic bounds on intervals of length ℓ. Suppose that $\mathcal{S}^{\varepsilon}$ is a nondecreasing effective service curve satisfying Eqn. (54), and let $\ell > 0$. Fix $a > 0$, set $x_j = t - \ell + ja$, and consider the events

$$E_j = \left\{ D(x_j) \geq \inf_{y \in [0,T]} \big\{ A(x_j - y) + \mathcal{S}^{\varepsilon}(y) \big\} \right\} \; , \quad j = 0, \ldots, n-1, \tag{59}$$

where $n = \lceil \ell/a \rceil$. If $x \in [x_j, x_{j+1})$ and the event E_j occurs, then a computation analogous to Eqn. (40) shows that

$$D(x) \geq \inf_{y \in [0,T+a]} \big\{ A(x-y) + \big(\mathcal{S}^\varepsilon * \delta_a(y)\big) \big\} \, . \tag{60}$$

We conclude as in Eqs. (41)-(43) that

$$\begin{aligned} Pr\left\{ \forall x \in [t-\ell, t] : D(x) \geq \inf_{y \in [0,T+a]} \big\{ A(x-y) + \big(\mathcal{S}^\varepsilon * \delta_a\big)(y) \big\} \right\} \;&\geq\; Pr\Big\{ \bigcap_{0 \leq j < n} E_j \Big\} & (61) \\ &\geq\; 1 - \varepsilon \left\lceil \frac{\ell}{a} \right\rceil \, . & (62) \end{aligned}$$

Step 2: A deterministic argument. A computation analogous to Eqs. (46)-(48) shows that

$$\left\{ \begin{array}{lll} \forall x \in [t-(H-h), t] : & D^h(x) \geq \inf_{y \in [0,T+a]} \big\{ A^h(x-y) + \big(\mathcal{S}^{h,\varepsilon} * \delta_a\big)(y) \big\} & h < H, \\ & D^H(t) \geq \inf_{y \in [0,T+a]} \big\{ A^H(t-y) + \mathcal{S}^{H,\varepsilon}(y) \big\} & h = H \, , \end{array} \right. \tag{63}$$

implies

$$D(t) \geq \inf_{x \in [0, H(T+a)]} \{ A(t-x) + \mathcal{S}^{net,\varepsilon'}(x) \} \, . \tag{64}$$

Step 3: Conclusion. Combining Steps 1 and 2, we obtain

$$\begin{aligned} Pr\big\{ D(t) \geq \inf_{x \in [0, H(T+a)]} \{ A(t-x) + \mathcal{S}^{net,\varepsilon'}(x) \} \big\} \;&\geq\; Pr\big\{ \text{Eqn. (63) is satisfied} \big\} & (65) \\ &\geq\; 1 - \varepsilon \left(1 + \sum_{h=1}^{H-1} \left\lceil \frac{(H-h)T}{a} \right\rceil \right) & (66) \\ &\geq\; 1 - \varepsilon' \, . & (67) \end{aligned}$$

Here, Eqn. (65) follows from Step 2, and Eqn. (66) follows from the assumptions by choosing $\ell_h = (H-h+1)$ for $h = 1, \ldots, H-1$ in Step 1. □

4 Statistical Calculus with Adaptive Service Guarantees

We next define a class of effective service curves where the range of the infimum is bounded independently of time, and then give conditions under which these service curves are also effective service curves in the sense of Eqn. (15). The resulting effective service curves are valid without adding assumptions on a specific arrival distribution or service discipline. Within this context, we obtain an effective network service curve, where the convolution formula has a similarly simple form as in the deterministic network calculus.

4.1 (Deterministic) Adaptive Service Curves

We define a modified convolution operator by setting, for any $t_0 \leq t$,

$$A *_{t_0} g(t) = \min\Big\{ g(t-t_0), B(t_0) + \inf_{\tau \leq t-t_0} \big\{ A(t_0, t-\tau) + g(\tau) \big\} \Big\} \, . \tag{68}$$

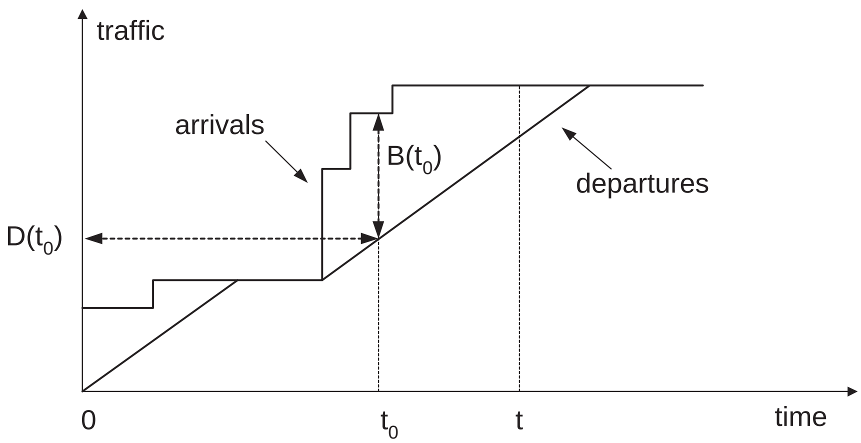

Figure 2: Illustration for the modified convolution operator. The operator $*_{t_0}$ uses the backlog at time t_0 and the arrivals in the interval $[t_0, t]$.

The essential property of this modified operator is that the range over which the infimum is taken is limited to the interval $[t_0, t]$. Note that the function $A *_{t_0} g(t)$ depends on the backlog at time t_0 as well as on the arrivals in the interval $[t_0, t]$. It can be written equivalently as

$$A *_{t_0} g(t) = \min\Big\{ g(t - t_0), \inf_{\tau \le t - t_0} \big\{ A(t - \tau) + g(\tau) \big\} - D(t_0) \Big\} . \tag{69}$$

The usual convolution operator is recovered by setting $t_0 = 0$.

We now reconsider the definition of a service curve in a deterministic regime. We introduce a revised definition of a (deterministic) service curve, which is presented in [1, 14], and is referred to as *adaptive service curve* in [6]. A *(minimum) adaptive service curve* is defined as a function S which specifies a lower bound on the service given to a flow such that, for all $t, t_0 \ge 0$, with $t_0 \le t$,

$$D(t_0, t) \ge A *_{t_0} S(t) . \tag{70}$$

A maximum adaptive service curve can be defined accordingly.[3] Eqn. (70) is equivalent to requiring that S satisfies Eqn. (8) for the time-shifted arrivals and departures

$$\tilde{A}(x) = B(t_0) + A(t_0, t_0 + x) , \quad \tilde{D}(x) = D(t_0, t_0 + x) . \tag{71}$$

Figure 2 illustrates the time-shifted arrivals. Many service curves with applications in packet networks, such as shapers, schedulers with delay guarantees, and rate-controlled schedulers such as GPS, can be expressed in terms of adaptive service curves. By setting $t_0 = 0$, one can see that each adaptive service curve is a service curve. However, the converse does not hold [6].

We next define a *(minimum) ℓ-adaptive service curve*, denoted by S^ℓ, as a function for which Eqn. (70) is satisfied whenever $t - t_0 \le \ell$. If $\ell = \infty$, we obtain an *adaptive service curve*, and drop the superscript in the notation. The difference between a service curve according to Eqn. (8) and an ℓ-adaptive service curve is that the former involves arrivals over the entire interval $[0, t]$, while the latter uses information

[3] We note that the adaptive service curve in [6] is more general and is defined using $D(t_0, t) \ge \min\Big\{ f(t - t_0), B(t_0) + \inf_{\tau \le t - t_0} \big\{ A(t_0, t - \tau) + g(\tau) \big\} \Big\}$. In our context we set $f = g$.

about arrivals and departures in intervals $[t_0, t]$ whose length does not depend on t. Performing a time shift as in Eqn. (71) and applying Theorem 2 shows that the convolution of ℓ-adaptive service curves yields an ℓ-adaptive network service curve.

The following lemma shows that for ℓ sufficiently large, but finite, an ℓ-adaptive service curve is a service curve in the sense of Eqn. (8). In particular, the conclusions of Theorems 1 and 2 hold for such service curves.

Lemma 1 *Suppose that the arrival function A of a flow has arrival envelope A^*. Let S^ℓ be an ℓ-adaptive service curve. If*

$$\exists t \in [0, \ell] : \quad A^*(t) \leq S^\ell(t) \; , \tag{72}$$

then $S = S^\ell$ is an adaptive service curve for intervals of arbitrary length. In particular, S satisfies Eqn. (8) for all $t \geq 0$.

The proof of the lemma is given in the appendix.

4.2 Effective Adaptive Service Curves

Next we introduce a probabilistic version of the ℓ-adaptive service curve. We define an *effective ℓ-adaptive service curve* to be a nonnegative function $\mathcal{S}^{\ell,\varepsilon}$ such that

$$Pr\big\{D(t_0, t) \geq A *_{t_0} \mathcal{S}^{\ell,\varepsilon}(t)\big\} \;\geq 1 - \varepsilon \tag{73}$$

for all $t_0, t \geq 0$ with $t - t_0 \leq \ell$. If $\ell = \infty$, we call the resulting function an *effective adaptive service curve*, and drop the superscript. Note that the infimum in the convolution on the right hand side of Eqn. (73) ranges over an interval of length at most ℓ. With this bound on the range of the infimum, we derive the following effective network service curve. Technically, ℓ plays a similar role as the bound T on the range of the convolution in Eqn. (54).

Theorem 6 Concatenation of Effective ℓ-Adaptive Service Curves. *Consider a flow passing through nodes numbered $h = 1, \ldots H$, and assume that, at each node, an effective ℓ-adaptive service curve is given by a nondecreasing function $\mathcal{S}^{h,\ell,\varepsilon_h}$. Then the function*

$$\mathcal{S}^{net,\ell,\varepsilon'} = \mathcal{S}^{1,\ell,\varepsilon} * \ldots * \mathcal{S}^{H,\ell,\varepsilon} * \delta_{(H-1)a}(t) \tag{74}$$

is an effective ℓ-adaptive network service curve for any choice of $a > 0$, with violation probability bounded by

$$\varepsilon' \leq \varepsilon \left(1 + (H-1) \left\lceil \frac{\ell}{a} \right\rceil \right) \; . \tag{75}$$

Proof. We need to show that, for any t_0, t with $t - t_0 \leq \ell$ and any choice of the parameter a, we have

$$Pr\big\{D(t_0, t) \geq A *_{t_0} \mathcal{S}^{net,\ell,H\varepsilon\ell/a}(t)\big\} \geq 1 - \varepsilon' \; . \tag{76}$$

Performing a time shift as in Eqn. (71), we may assume without loss of generality that $t_0 = 0$ and $t \in [0, \ell]$. The claim now follows immediately from Theorem 4.

$\square$

Even though the concatenation formula in Theorem 6 results in a significant improvement over Theorem 4, a drawback of Theorem 6 is that the construction of the network service curve results in a degradation of the violation probability ε' and introduces a time shift $\delta_{(H-1)a}$, which grow significantly when ℓ and H become large. To avoid this successive degradation of the service guarantees, we further strengthen the effective service curve. We define a *strong effective adaptive service curve* for intervals of length ℓ to be a function $\mathcal{T}^{\ell,\varepsilon}$ which satisfies for any interval I_ℓ of length ℓ,

$$Pr\big\{\forall[t_0,t] \subseteq I_\ell :\ D(t_0,t) \geq A *_{t_0} \mathcal{T}^{\ell,\varepsilon}(t)\big\} \geq 1-\varepsilon\ . \tag{77}$$

This definition differs from the definition of an effective service curve in Eqn. (15) in two ways: it uses the modified convolution operator, and it provides lower bounds on the departures simultaneously in all subintervals of an interval I_ℓ.

With the strong effective adaptive network service curve, we obtain a probabilistic version of a network service curve, with a similar concatenation formula as in the deterministic calculus. This is the content of the following theorem.

Theorem 7 Concatenation of Strong Effective Adaptive Service Curves. *Consider a flow that passes through H network nodes in series. Assume that the functions $\mathcal{T}^{h,\ell,\varepsilon}$ define strong effective adaptive service curves for intervals of length ℓ at each node ($h = 1,\ldots,H$). Then*

$$\mathcal{T}^{net,\ell,H\varepsilon}(t) = \mathcal{T}^{1,\ell,\varepsilon} * \ldots * \mathcal{T}^{H,\ell,\varepsilon}(t) \tag{78}$$

is a strong effective adaptive service curve for intervals of length ℓ.

Note the similarity of the convolution formula in Eqn. (78) with the corresponding expression in the deterministic calculus. Thus, in the statistical calculus, obtaining a statistical end-to-end service curve via a simple convolution operation comes at the price of significant modifications to the definition of a service curve.

Proof. We need to show that $T^{net,\ell,H\varepsilon}$ satisfies for any interval I_ℓ of length ℓ

$$Pr\big\{\forall[t_0,t] \subseteq I_\ell :\ D(t_0,t) \geq A *_{t_0} \mathcal{T}^{net,\ell,H\varepsilon}(t)\big\} \geq 1-H\varepsilon\ . \tag{79}$$

The argument closely follows Steps 2 and 3 from the proof of Theorem 4. If, for a particular sample path,

$$\forall[t_0,x] \subseteq I_\ell,\ \forall h = 1,\ldots,H: \quad D^h(t_0,x) \geq A^h *_{t_0} \mathcal{T}^{h,\ell,\varepsilon}(x)\ , \tag{80}$$

then, for any fixed $[t_0,t] \subseteq I_\ell$, the time-shifted arrivals and departures defined by Eqn. (71) satisfy

$$\forall x \leq t-t_0,\ \forall h = 1,\ldots,H: \quad \tilde{D}^h(x) \geq \tilde{A}^h * \mathcal{T}^{h,\ell,\varepsilon}(x)\ . \tag{81}$$

By Step 2 of proof of Theorem 4, this implies

$$\tilde{D}(t) \geq \tilde{A} * \big(\mathcal{T}^{1,\ell,\varepsilon} * \ldots * \mathcal{T}^{H,\ell,\varepsilon}\big)(t)\ . \tag{82}$$

Reversing the time shift and using that $[t_0, t] \subseteq I_\ell$ was arbitrary, we arrive at

$$\forall [t_0, t] \subseteq I_\ell : \; D(t) \geq A *_{t_0} \big(\mathcal{T}^{1,\ell,\varepsilon} * \ldots * \mathcal{T}^{H,\ell,\varepsilon} \big)(t) \; . \tag{83}$$

We conclude that

$$Pr\big\{ \forall [t_0, t] \subseteq I_\ell : \; D(t_0, t) \geq A *_{t_0} \mathcal{T}^{net,\ell,H\varepsilon}(t) \big\} \tag{84}$$
$$\geq Pr\big\{ \forall [t_0, x] \subseteq I_\ell, \, \forall h = 1, \ldots H : \; D^h(t_0, x) \geq A^h *_{t_0} \mathcal{T}^{h,\ell,\varepsilon}(x) \big\} \tag{85}$$
$$\geq 1 - H\varepsilon \; . \tag{86}$$

The first inequality follows from the definition of $\mathcal{T}^{net,\ell,H\varepsilon}$ and the fact that Eqn. (80) implies Eqn. (83), and the second inequality uses the defining property of strong effective ℓ-adaptive service curves. □

A comparison of the definition of the strong effective adaptive service curve in Eqn. (77) with Eqn. (73) shows that a strong effective adaptive service curve is an effective ℓ-adaptive service curve which provides service guarantees simultaneously on all subintervals of an interval of length ℓ. A comparison of Theorem 7 with Theorems 4, 5, and 6 shows that the more stringent strong effective adaptive service curve expresses the statistical calculus more concisely. Therefore, unless additional assumptions are made on the arrival processes and the service curves, the network calculus with strong effective adaptive service curves offers the preferred framework.

Our next result shows how to construct a strong effective adaptive service curve from an effective adaptive service curve. The lemma indicates that the choice of working with a strong effective adaptive service curve rather than an effective adaptive service curve is purely a matter of technical convenience.

Lemma 2 *If $\mathcal{S}^{\ell,\varepsilon}$ is a nondecreasing function which defines an effective ℓ-adaptive service curve for a flow, then, for any choice of $a > 0$, the function*

$$\mathcal{T}^{\ell,\varepsilon'} = \mathcal{S}^{\ell,\varepsilon} * \delta_a \tag{87}$$

is a strong effective service curve for intervals of length ℓ, with violation probability given by

$$\varepsilon' = \lceil 2\ell/a \rceil^2 \varepsilon/2 \; . \tag{88}$$

Proof. We will show that for any interval I_ℓ of length ℓ,

$$\forall [t_0, t] \subseteq I_\ell : \quad Pr\big\{ D(t_0, t) \geq A *_{t_0} \mathcal{S}^{\ell,\varepsilon}(t) \big\} \geq 1 - \varepsilon \tag{89}$$

implies

$$Pr\big\{ \forall [t_0, t] \subseteq I_\ell : \; D(t_0, t) \geq A *_{t_0} \mathcal{T}^{\ell,\varepsilon'}(t) \big\} \geq 1 - \varepsilon' \; , \tag{90}$$

where ε' and $\mathcal{T}^{\ell,\varepsilon'}$ are as given in the statement of the lemma. By performing a suitable time shift as in Eqn. (71), we may assume without loss of generality that $I_\ell = [0, \ell]$.

The strategy is similar to the construction of strong effective envelopes from effective envelopes in [4], and uses the same techniques as the first step in the proof of Theorem 4. We first use the fact that the departures satisfy the positivity assumption (A1) to translate service guarantees given on a subinterval into a service guarantee on a longer interval. In the second step, we establish probabilistic bounds for the departures simultaneously in a finite number of subintervals I_{ij} of I_ℓ, and then bound the departures in general subintervals of I_ℓ from below in terms of the departures in the I_{ij}.

Step 1: A property of the modified convolution. Let g be a nondecreasing function, let $[t_1, t_2] \subseteq [t_0, t]$, and $a \geq (t - t_0) - (t_2 - t_1)$. Then

$$D(t_1, t_2) \geq A *_{t_1} g(t_2) \tag{91}$$

implies

$$D(t_0, t) \geq A *_{t_0} \big(g * \delta_a\big)(t) \; . \tag{92}$$

To see this, note that Eqn. (91) implies that either

$$D(t_1, t_2) \geq g(t_2 - t_1) \; , \tag{93}$$

or

$$D(t_1, t_2) \geq B(t_1) + \inf_{\tau \in [t_1, t_2]} \{A(t_1, \tau) + g(t_2 - \tau)\} \; . \tag{94}$$

If Eqn. (93) holds, then

$$D(t_0, t) \geq D(t_1, t_2) \geq g(t_2 - t_1) \geq g * \delta_a(t - t_0) \geq A *_{t_0} \big(g * \delta_a\big)(t) \; , \tag{95}$$

proving Eqn. (92) in this case. We have used that g is nondecreasing in the last inequality. If Eqn (94) holds, then

$$\begin{aligned} D(t_0, t) &\geq D(t_0, t_1) + B(t_1) + \inf_{\tau \in [t_1, t_2]} \{A(t_1, \tau) + g(t_2 - \tau)\} && (96)\\ &= B(t_0) + \inf_{\tau \in [t_1, t_2]} \{A(t_0, \tau) + g(t_2 - \tau)\} && (97)\\ &\geq B(t_0) + \inf_{\tau \in [t_0, t]} \{A(t_0, \tau) + \big(g * \delta_a\big)(t_2 - \tau)\} && (98)\\ &\geq A *_{t_0} \big(g * \delta_a\big)(t) \; . && (99) \end{aligned}$$

which proves Eqn. (92) in the second case. In Eqn. (96) we have used Eqn. (94). In Eqn. (97), we have used that $D(t_0, t_1) + B(t_1) = B(t_0) + A(t_0, t_1)$ and taken $A(t_0, t_1)$ under the infimum. Eqn. (98) uses the monotonicity of g and extends the range of the infimum.

Step 2: Uniform probabilistic bounds on I_ℓ. Fix $a > 0$, set $x_i = i\, a/2$, and consider the intervals

$$I_{ij} = [x_i, x_j] \; , \quad 0 \leq i < j < n \; , \tag{100}$$

where $n = \lceil 2\ell/a \rceil$ is the smallest integer no less than $2\ell/a$. Consider the events

$$E_{ij} := \big\{D(x_i, x_j) \geq A *_{x_i} \mathcal{S}^{\ell,\varepsilon}(x_j)\big\} \; . \tag{101}$$

Let $[t_0, t] \subseteq [0, \ell]$ be arbitrary, and choose $I_{ij} \subseteq [t_0, t]$ be as large as possible. If E_{ij} occurs, we apply Step 1 with $t_1 = x_i$ and $t_2 = x_j$, and use that $(x_i - t_0) + (t - x_j) \leq a$ to see that

$$D(t) \geq A *_{t_0} \big(\mathcal{S}^{\ell,\varepsilon} * \delta_a\big)(t) \; . \tag{102}$$

It follows that

$$
\begin{aligned}
& Pr\big\{\forall [t_0,t] \subseteq [0,\ell]:\ D(t_0,t) \geq A *_{t_0} \big(\mathcal{S}^{\ell,\varepsilon} * \delta_a\big)(t)\big\} \\
& \quad \geq Pr\Big\{\bigcap_{0\leq i<j<n} E_{ij}\Big\} \qquad (103) \\
& \quad \geq 1 - n^2\varepsilon/2 \ , \qquad (104)
\end{aligned}
$$

as claimed. Here, Eqn. (103) uses Step 2, and Eqn. (104) uses the definition of $\mathcal{S}^{\ell,\varepsilon}$. □

4.3 Recovering an effective service curve from effective adaptive service curves

We next show that the adaptive versions of the effective service curve can yield effective service service curves in their original definition. This, however, requires us to add appropriate assumptions on the traffic at a node. The following lemma gives a sufficient conditions for an effective ℓ-adaptive curve to be an effective service curve in the sense of Eqn. (15). Combining Theorem 6 with Lemma 3 yields an effective network service curve, which by Theorem 3 guarantees probabilistic bounds on output, backlog, and delay.

Lemma 3 *Let $\mathcal{S}^{\ell,\varepsilon}$ be a nondecreasing function which defines an effective ℓ-adaptive service curve for a flow with arrival process A, and an*

1. *If*

$$Pr\big\{\exists t_0 \in [t-\ell,t]:\ B(t_0)=0\big\} \geq 1-\varepsilon_1 \qquad (105)$$

for all $t>0$, then, for any choice of $a>0$, $\mathcal{S}^{\varepsilon\ell/a+\varepsilon_1} = \mathcal{S}^{\ell,\varepsilon} * \delta_a$ is an effective adaptive service curve for intervals of arbitrary length, with violation probability $\varepsilon\ell/a+\varepsilon_1$. In particular,

$$Pr\big\{D(t) \geq A * \mathcal{S}^{\ell,\varepsilon} * \delta_a(t)\big\} \geq 1-(\varepsilon\ell/a+\varepsilon_1) \qquad (106)$$

for all $t>0$.

2. *If the arrival process A has arrival envelope A^* and*

$$Pr\big\{B(t) \leq \mathcal{S}^{\ell,\varepsilon}(\ell) - A^*(\ell)\big\} \geq 1-\varepsilon_1 \ , \qquad (107)$$

for all $t \geq 0$, then $\mathcal{S}^{\varepsilon+\varepsilon_1} = \mathcal{S}^{\ell,\varepsilon}$ is an effective adaptive service curve for intervals of arbitrary length, with violation probability $\varepsilon+\varepsilon_1$. In particular,

$$Pr\big\{D(t) \geq A * \mathcal{S}^{\ell,\varepsilon}(t)\big\} \geq 1-(\varepsilon+\varepsilon_1) \ , \qquad (108)$$

for all $t \geq 0$.

The lemma should be compared with Lemma 1, as both provide sufficient conditions under which service guarantees on intervals of a given finite length imply service guarantees on intervals of arbitrary length. While the condition on ℓ in Eqn. (72) involves only the deterministic arrival envelope and the service curve, Eqs. (105) and (107) represent additional assumptions on the backlog process. This points out a fundamental difference between the deterministic and the statistical network calculus.

Proof. The proof consists of three steps. If Eqn. (107) holds, the first step can be omitted. In the first step, we modify a given effective ℓ-adaptive service curve to give uniform probabilistic lower bounds on the departure of all intervals of the form $[t_0, t]$, where t is fixed and $t_0 \in [t-\ell, t]$. This is analogous to the first step in the proof of Theorem 4. The second step contains a deterministic argument. We conclude with a probabilistic estimate.

Step 1: Uniform probabilistic bounds. Suppose that $S^{\ell,\varepsilon}$ is a nondecreasing effective ℓ-adaptive service curve, that is, for any $t \geq 0$,

$$\forall t_0 \in [t-\ell, t]: \quad Pr\big\{D(t_0, t) \geq A *_{t_0} \mathcal{S}^{\ell,\varepsilon}(t)\big\} \geq 1 - \varepsilon \,. \tag{109}$$

We will show that then, for any choice of $a > 0$,

$$Pr\big\{\forall t_0 \in [t-\ell, t] : D(t_0, t) \geq A *_{t_0} \mathcal{S}^{\ell,\varepsilon}(t-a)\big\} \geq 1 - \varepsilon\ell/a \,. \tag{110}$$

To see this, assume without loss of generality that $t = \ell$, and consider the events

$$E_j = \big\{D(x_j, t) \geq A *_{x_j} \mathcal{S}^{\ell,\varepsilon}(t)\big\} \,, \quad 0 \leq i < n \,. \tag{111}$$

If E_j occurs, we have for $x \in [x_{j-1}, x_j)$ by the first step of the proof of Lemma 2 (with $t_0 = x$, $t_1 = x_j$, $t_2 = t$), that

$$D(x, t) \geq A *_{x_j} \big(\mathcal{S}^{\ell,\varepsilon} * \delta_a\big)(t) \,. \tag{112}$$

It follows that

$$\begin{aligned} Pr\big\{&\forall t_0 \in [0, \ell] : D(t) \geq A *_{t_0} \mathcal{S}^{\ell,\varepsilon}(t-a)\big\} \\ &= Pr\Big\{\bigcap_{0 \leq i < n} E_j\Big\} && (113) \\ &\geq 1 - n\varepsilon \,, && (114) \end{aligned}$$

which proves Eqn. (110) in the case $t = \ell$.

Step 2: Deterministic argument. Fix $t \geq 0$, and suppose that for a particular sample path, we have

$$\forall x \in [t-\ell, t] : \; D(x, t) \geq A *_x \mathcal{T}^{\ell,\varepsilon}(t) \,, \tag{115}$$

and either

$$1. \qquad \exists t_0 \in [t-\ell, t] : \; B(t_0) = 0 \text{ , or} \tag{116}$$

$$2. \qquad B(t-\ell) \leq \mathcal{S}^{\ell,\varepsilon}(\ell) - A^*(\ell) \,, \text{ where } A^* \text{ is an arrival envelope.} \tag{117}$$

In the first case, we can set $x = t_0$ in Eqn. (115) to obtain

$$D(t) \geq \inf_{\tau \leq t - t_0} \{A(t-\tau) + \mathcal{S}^{\ell,\varepsilon}(\tau)\}. \tag{118}$$

In the second case, we note that from Eqn. (117) it follows that

$$B(t-\ell) + \inf_{\tau \leq \ell} \{A(t-\ell, t-\tau) + \mathcal{S}^{\ell,\varepsilon}(\tau)\} \leq \mathcal{S}^{\ell,\varepsilon}(\ell) - A^*(\ell) + A(t-\ell, t) \leq \mathcal{S}^{\ell,\varepsilon}(\ell) \,, \tag{119}$$

which implies that

$$A *_{t-\ell}(t) = B(t-\ell) + \inf_{\tau \le \ell}\{A(t-\ell,\tau) + \mathcal{S}^{\ell\varepsilon}(\tau)\} \ . \tag{120}$$

Inserting this into Eqn. (115) with $x = t - \ell$ yields again Eqn. (118).

Step 3: Probabilistic estimate. If Eqn. (105) holds, we use Step 2 to see that

$$\begin{aligned} & Pr\big\{D(t) \ge \inf_{\tau \le \ell}\{A(t-\tau) + D(\tau)\}\big\} \\ & \quad \ge Pr\Big\{ \ \forall x \in [t-\ell,t]: \ D(x,t) \ge A *_x \mathcal{T}^{\ell,\varepsilon}(t), \ \text{and } \exists t_0 \in [t-\ell,t]: \ B(t_0) = 0 \ \Big\} & (121) \\ & \quad \ge 1 - (\varepsilon\ell/a + \varepsilon_1) \ , & (122) \end{aligned}$$

where we have used the result of Step 2 in the second line. If in Eqn. (107) holds, we have by Step 2

$$\begin{aligned} & Pr\big\{D(t) \ge \inf_{\tau \le \ell}\{A(t-\tau) + D(\tau)\}\big\} \\ & \quad \ge Pr\Big\{ \ D(t-\ell,t) \ge A *_{t-\ell} \mathcal{T}^{\ell,\varepsilon}(t), \ \text{and } B(t-\ell) \le \mathcal{S}^{\ell,\varepsilon}(\ell) - A^*(\ell) \ \Big\} & (123) \\ & \quad \ge 1 - (\varepsilon + \varepsilon_1) \ , & (124) \end{aligned}$$

where we have used the definition of $\mathcal{S}^{\ell,\varepsilon}$ in the second line. □

Note that Eqs. (122) and (124) supply time-independent bounds on the range of the convolution, of the form given in Eqn. (54). A similar, but simpler result holds for strong effective ℓ-adaptive service curves:

Lemma 4 *Given a flow with arrival process A, and a strong effective adaptive service curve $\mathcal{T}^{\ell,\varepsilon}$ on intervals of length ℓ. Assume that for every $t \ge 0$, either Eqn. (105) or or Eqn. (107) is satisfied. Then, for any $t \ge 0$ and any $t_1 \le t$,*

$$Pr\big\{D(t_1,t) \ge A *_{t_1} \mathcal{T}^{\ell,\varepsilon}(t)\big\} \ge 1 - (\varepsilon + \varepsilon_1) \ . \tag{125}$$

In particular, for $t_1 = 0$, $\mathcal{S}^{\varepsilon+\varepsilon_1} = \mathcal{T}^{\ell,\varepsilon}$ is an effective service curve in the sense of Eqn. (15).

Proof. We need to show that under the assumptions of the lemma, we have for any $t > 0$ and any $t_1 \le t$

$$Pr\big\{D(t_1,t) \ge A *_{t_1} \mathcal{T}^{\ell,\varepsilon}(t)\big\} \ge 1 - (\varepsilon + \varepsilon_1) \ . \tag{126}$$

By considering time-shifted arrivals and departures as in Eqn. (71), we may assume without loss of generality that $t_1 = 0$. The first step in the proof of Lemma 3 shows that

$$\begin{aligned} & Pr\big\{D(t) \ge \inf_{\tau \le \ell}\{A(t-\tau) + D(\tau)\}\big\} \\ & \quad \ge Pr\left\{ \begin{array}{l} \forall [x,y] \subset [t-\ell,t]: \ D(x,y) \ge A *_x \mathcal{T}^{\ell,\varepsilon}(x), \\ \text{and} \big\{ \text{ either } \ \exists t_0 \in [t-\ell,t]: \ B(t_0) = 0 \\ \qquad \text{or } B(t-\ell) \le \mathcal{T}^{\ell,\varepsilon}(\ell) - A^*(\ell) \big\} \end{array} \right\} & (127) \\ & \quad \ge 1 - (\varepsilon + \varepsilon_1) \ , & (128) \end{aligned}$$

as claimed. □

5 Conclusions

We have presented a network calculus with probabilistic service guarantees where arrivals to the network satisfy a deterministic arrival bound. We have introduced the notion of *effective service curves* as a probabilistic bound on the service received by individual flows in a network. We have shown that some key results from the deterministic network calculus can be carried over to the statistical framework by inserting appropriate probabilistic arguments.

We showed that the deterministic bounds on output, delay, and backlog from Theorem 1 have corresponding formulations in the statistical calculus (Theorem 3). We have extended the concatenation formula of Theorem 2 for network service curves to a statistical setting (Theorems 4, 5, 6, and 7). We showed that a modified effective service curve, called *strong effective adaptive service curve* yields the simplest concatenation formula. In order to connect the different notions of effective service curves, we have made an additional assumption on the backlog in Lemmas 3 and 4. The results in this paper showed that a multi-node version of the statistical network calculus requires us to make assumptions that limit the range of the convolution operation when concatenating effective service curves. Such limits on a 'maximum relevant time scale', can follow from assumptions on the traffic load (as in Theorems 5, Lemma 3 and Lemma 4), or from appropriately modified service curves. While the question is open whether one can dispense with these additional assumptions, we have made an attempt to justify the need for them.

Acknowledgments

We thank Rene Cruz for pointing out problems in earlier versions of the effective network service curve. The authors gratefully acknowledge the valuable comments from Chengzhi Li.

References

[1] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan. Performance bounds for flow control protocols. *IEEE/ACM Transactions on Networking*, 7(3):310–323, June 1999.

[2] M. Andrews. Probabilistic end-to-end delay bounds for earliest deadline first scheduling. In *Proceedings of IEEE Infocom 2000*, pages 603–612, Tel Aviv, March 2000.

[3] F. L. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity: An Algebra for Discrete Event Systems*. John Wiley and Sons, 1992.

[4] R. Boorstyn, A. Burchard, J. Liebeherr, and C. Oottamakorn. Statistical service assurances for traffic scheduling algorithms. *IEEE Journal on Selected Areas in Communications. Special Issue on Internet QoS*, 18(12):2651–2664, December 2000.

[5] J. Y. Le Boudec. Application of network calculus to guaranteed service networks. *IEEE/ACM Transactions on Information Theory*, 44(3):1087–1097, May 1998.

[6] J. Y. Le Boudec and P. Thiran. *Network Calculus*. Springer Verlag, Lecture Notes in Computer Science, LNCS 2050, 2001.

[7] C. S. Chang. Stability, queue length, and delay of deterministic and stochastic queueing networks. *IEEE Transactions on Automatic Control*, 39(5):913–931, May 1994.

[8] C. S. Chang. On deterministic traffic regulation and service guarantees: a systematic approach by filtering. *IEEE/ACM Transactions on Information Theory*, 44(3):1097–1110, May 1998.

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

[10] R. L. Cruz. A calculus for network delay, Part I : Network elements in isolation. *IEEE Transaction of Information Theory*, 37(1):114–121, 1991.

[11] R. L. Cruz. A calculus for network delay, Part II : Network analysis. *IEEE Transactions on Information Theory*, 37(1):132–141, January 1991.

[12] R. L. Cruz. Quality of service guarantees in virtual circuit switched networks. *IEEE Journal on Selected Areas in Communications*, 13(6):1048–1056, August 1995.

[13] R. L. Cruz. Quality of service management in integrated services networks. In *Proceedings of the 1st Semi-Annual Research Review, CWC*, UCSD, June 1996.

[14] R. L. Cruz and C. Okino. Service gurantees for flow control protocols. In *Proceedings of the 34th Allerton Conference on Communications, Control and Computating*, October 1996.

[15] A. Elwalid and D. Mitra. Design of generalized processor sharing schedulers which statistically multiplex heterogeneous QoS classes. In *Proceedings of IEEE INFOCOM'99*, pages 1220–1230, New York, March 1999.

[16] J. Kurose. On computing per-session performance bounds in high-speed multi-hop computer networks. In *ACM Sigmetrics'92*, pages 128–139, 1992.

[17] C. Li and E. Knightly. Coordinated network scheduling: A framework for end-to-end services. In *Proceedings of IEEE ICNP 2000*, Osaka, November 2000.

[18] J. Qiu and E. Knightly. Inter-class resource sharing using statistical service envelopes. In *Proceedings of IEEE Infocom '99*, pages 36–42, March 1999.

[19] V. Sivaraman and F. M. Chiussi. Statistical analysis of delay bound violations at an earliest deadline first scheduler. *Performance Evaluation*, 36(1):457–470, 1999.

[20] V. Sivaraman and F. M. Chiussi. Providing end-to-end statistical delay guarantees with earliest deadline first scheduling and per-hop traffic shaping. In *Proceedings of IEEE Infocom 2000*, pages 603–612, Tel Aviv, March 2000.

[21] D. Starobinski and M. Sidi. Stochastically bounded burstiness for communication networks. *IEEE Transaction of Information Theory*, 46(1):206–212, 2000.

[22] 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.

APPENDIX

A Proof of Lemma 1

Let S^ℓ be an ℓ-adaptive service curve. We need to show that

$$D(t_1,t) \geq A *_{t_1} S^\ell(t) \tag{129}$$

holds for all $t, t_1 \geq 0$ with $t_1 \leq t$. By considering the time-shifted arrivals and departures as in Eqn. (71), we may assume without loss of generality that $t_1 = 0$.

Consider intervals $I_\ell^k = [k\ell, (k+1)\ell]$, where $k \geq 0$ is an integer. We will show by induction, that for any integer $k \geq 0$,

$$\forall t \in I_\ell^k : \quad D(t) \geq A * S^\ell(t) \ . \tag{130}$$

Applying the definition of an ℓ-adaptive network service curve with $I_\ell = [0,\ell]$, we see that Eqn. (130) clearly holds for $k = 0$.

For the inductive step, suppose that Eqn. (130) holds for some integer $k \geq 0$. Fix $t \in I_\ell^{k+1}$, and let $t_0 = t - \ell \in I_\ell^k$. By the inductive assumption, $D(t_0) \geq A * S^\ell(t_0)$. Eqn. (70) says that either

$$D(t_0,t) \geq S^\ell(t-t_0) \tag{131}$$

or

$$D(t_0,t) \geq B(t_0) + \inf_{\tau \leq t-t_0} \left\{ A(t-t_0, t-\tau) + S^\ell(\tau) \right\} \ . \tag{132}$$

If Eqn. (131) holds, then

$$\begin{aligned} D(t) &= D(t_0,t) + D(t_0) & (133)\\ &\geq S^\ell(t-t_0) + \inf_{\tau \leq t_0} \left\{ A(t_0-\tau) + S^\ell(\tau) \right\} & (134)\\ &\geq S^\ell(t-t_0) - A^*(t-t_0) + \inf_{\tau \leq t_0} \left\{ A(t-\tau) + S^\ell(\tau) \right\} & (135)\\ &\geq A * S^\ell(t) \ . & (136) \end{aligned}$$

In Eqn. (134), we have used Eqn. (131) and the inductive assumption. In Eqn. (135), we have used that $A(t_0-\tau, t-\tau) \leq A^*(t-t_0)$ and pulled $A^*(t-t_0)$ out of the infimum. In Eqn. (136), we have inserted $t - t_0 = \ell$, used the assumption that $A^*(\ell) \leq S^\ell(\ell)$, and extended the range of the infimum.

If Eqn. (132) holds, then

$$\begin{aligned} D(t) &= D(t_0) + D(t-t_0) & (137)\\ &\geq A(t_0) + \inf_{\tau \leq t-t_0} \left\{ A(t_0, t-\tau) + S^\ell(\tau) \right\} & (138)\\ &= \inf_{\tau \leq t-t_0} \left\{ A(t-\tau) + S^\ell(\tau) \right\} & (139)\\ &\geq A * S^\ell(t) \ . & (140) \end{aligned}$$

In Eqn. (138), we have used Eqn. (132), and the fact that $D(t_0) + B(t_0) = A(t_0)$. In Eqn. (139), we have taken $A(t_0)$ under the infimum and used that $A(t_0) + A(t_0, t-\tau) = A(t-\tau)$. In Eqn. (140), we have extended the range of the infimum and used the definition of the convolution.

Since $t \in I_\ell^{k+1}$ was arbitrary, this proves the inductive step, and the lemma.