Enabling Residuation of Stochastic Min-Plus Servers Using Minimum Arrival Bounds

Vlad-Cristian Constantin, Paul Nikolaus, Jens B. Schmitt
RPTU Kaiserslautern-Landau, Germany
{constantin, nikolaus, jschmitt}@cs.uni-kl.de

***Abstract*—The Stochastic Network Calculus (SNC) is a valuable methodology to derive stochastic performance guarantees in networked and distributed systems. A particular stronghold of SNC is its ability to calculate *per-flow* performance bounds enabled by the *residuation* of service processes. Often, the calculation of such bounds requires the assumption that the service process is work-conserving resp. strict. Yet, some service processes only provide a min-plus linear lower bound (sometimes also called dynamic), thus not being necessarily work-conserving. Building on existing results from the Deterministic Network Calculus (DNC), we extend the SNC to support residuation of min-plus linearly bounded service processes by leveraging minimum arrival bounds through the Laplace transform. Unlike in DNC, this approach requires no additional assumptions, as the Laplace transform always exists.**

I. Introduction

Stringent control of key performance metrics is crucial in modern networked and distributed systems. Over decades, powerful mathematical frameworks have addressed this challenge, including classical queueing theory [1] and large deviations and effective bandwidth theory [2]. Among these, Network Calculus (NC) emerged as a unifying framework for worst-case, non-probabilistic performance analysis based on min-plus algebra [3]. Originating from the seminal work of Cruz in the early 1990s [4], [5], NC introduced deterministic bounds providing tight end-to-end worst-case guarantees, forming what is now known as Deterministic Network Calculus (DNC). Comprehensive treatments are available in [6], [7], [8].

While DNC captures worst-case behavior, it neglects statistical multiplexing gains. This limitation motivated the development of *Stochastic* Network Calculus (SNC), which aims to derive *probable worst-case* guarantees of the form

$$\mathrm{P}(\text{delay} > T) \leq \varepsilon,$$

thus generalizing DNC, which is recovered (a.s.) for $\varepsilon = 0$.

Two main SNC variants exist. The first is a direct stochastic extension of DNC based on tail bounds [9], [10], [11], [12]. The second relies on moment-generating functions (MGFs) [6], [13], exploiting statistical independence to derive probabilistic guarantees in the exponential transform domain. As shown in [14], MGF-based SNC provides significantly tighter bounds under independence than tail-bound methods. It has since developed into a comprehensive framework covering short- and long-range dependent traffic, various scheduling disciplines, and powerful network analyses [15], [16], [17], [18]. In this paper, we adopt the MGF-based variant.

ISBN 978-3-903176-82-9 © 2026 IFIP

A distinct feature of both DNC and SNC is their ability to calculate *per-flow* performance bounds, facilitated by the residuation of service processes (by cross-flows). However, this residuation often requires specific assumptions on the service process, captured in the following definitions:

In SNC, we let service processes $S(s,t)$, which are stochastic processes with $S(t,t) = 0$ for all $t \geq 0$, describe the service offered by servers in the time interval (s,t).

Definition 1 (Min-Plus Service Process)**.** Assume $A(0,t)$ and $D(0,t)$ are the cumulative arrival and departure processes, respectively, of a server. Let $S(s,t)$ for $0 \leq s \leq t$ be a random process. The server is called a *min-plus service process* [6], [19], [20], if for any fixed sample path and all $t \geq 0$

$$D(0,t) \geq A \otimes S(0,t) := \inf_{0 \leq \tau \leq t} \{A(\tau) + S(\tau,t)\}. \quad (1)$$

Definition 2 (Strict Service Process [13])**.** For any $t \geq 0$, let

$$s := \sup\{\tau \in [0,t] : D(0,\tau) = A(0,\tau)\}$$

be the beginning of the last backlogged period before t. Let $\tau \in [s,t]$. A (bivariate) *strict service process* is a positive random process $S(\tau,t)$ such that, for any fixed sample path, the server is non-idling and uses the entire available service $S(\tau,t)$ to serve backlogged data, i.e.,

$$D(0,t) \geq D(0,\tau) + S(\tau,t), \quad \forall s \leq \tau \leq t,$$

in any continuously backlogged period $(\tau,t]$.

Note that a strict service process is also a min-plus service process, but not vice versa. For the sake of consistency between DNC and SNC, we use the terms "service process" and "strict service process" instead of "dynamic" and "work-conserving server", respectively.

Coming back to the residuation issue: in many cases, essentially all schedulers but FIFO, residuation by cross-traffic require a strict (aggregate) service process. Yet, min-plus service processes often result from (previous) residuation and, furthermore, concatenation of strict servers also results in the end-to-end service process being only min-plus instead of strict [8, p. 137]. This creates a fundamental dilemma: residuation is required for per-flow analysis, yet standard transformations destroy the strictness needed to perform it. Addressing this dilemma, there is a recent result in DNC which allows for residuation of min-plus service provided that a minimum arrival curve exists [21]. This facilitates the treatment of negative service, as it may result from the residuation of min-plus service processes.

In this paper, we enable the DNC result from [21] in the MGF-based SNC using bounds on the Laplace transform of the arrivals. Notably, in many cases, Laplace bounds come for free, since the distributional assumptions for MGF bounds already imply them. Assuming an arrival process, a *nontrivial* Laplace transform always exists (unlike the MGF).

In summary, we make the following contributions:

- We enable the residuation of min-plus service processes by exploiting Laplace bounds on arrivals to calculate performance bounds (backlog and delay) in case of partially negative service processes in Sec. III-A.
- We derive a Laplace bound for Markov-modulated arrival processes in Sec. III-B.

II. Preliminaries and Background on SNC

In this section, we provide some preliminaries from probability theory and the necessary background on SNC.

A. Preliminaries on Probability Theory

Here, we state some well-known probability-theoretic results required for our analysis and for SNC in general.

Theorem 3 (Union Bound / Boole's Inequality)**.** *Let $X_1, \dots, X_n$ be random variables and $x \in \mathbb{R}$. Then, the* Union Bound / Boole's Inequality *holds:*

$$\mathrm{P}\left(\max_{i=1,\dots,n} X_i > x\right) \leq \sum_{i=1}^{n} \mathrm{P}(X_i > x)\,. \tag{2}$$

As our approach to SNC relies on MGFs, rooted in effective bandwidth theory [2], [6], we often apply the Chernoff bound.

Theorem 4 (Chernoff Bound)**.** *Let X be a random variable and $\theta > 0$. Then*

$$\mathrm{P}(X \geq a) \leq e^{-\theta a}\,\mathrm{E}\big[e^{\theta X}\big]\,. \tag{3}$$

This highlights the utility of MGFs in representing stochastic processes, making them a natural choice for our analysis.

B. Background on SNC

In this section, we introduce the fundamental concepts and terminology of MGF-based SNC.

We assume a discrete-time space $T \subseteq \mathbb{N}_0 = \{0, 1, 2, \dots\}$ and define the bivariate stochastic arrival processes $A(s,t)$ with $s \leq t$, $s,t \in \mathbb{N}_0$ and state space $\mathbb{R}^+$: $A(s,t) := \sum_{i=s+1}^{t} a_i$, where a_i are the so-called arrival increments at time i. Throughout the paper, $x(t) := x(0,t)$, x being a real-valued bivariate function. We assume a lossless system and also define the output D as the cumulative departure process, such that for all t, we have $D(t) \leq A(t)$, a so-called causal system.

Definition 5 (Min-Plus Operators)**.** Let $x(s,t)$ and $y(s,t)$ be two real-valued bivariate functions. Then, we define the min-plus convolution for $0 \leq s \leq t$ as

$$x \otimes y(s,t) := \inf_{s \leq \tau \leq t} \{x(s,\tau) + y(\tau,t)\}$$

and the min-plus deconvolution as

$$x \oslash y(s,t) := \sup_{0 \leq \tau \leq s} \{x(\tau,t) - y(\tau,s)\},$$

respectively.

We now turn to the performance measures of interest. Specifically, we focus on backlog and delay.

Definition 6 (Performance Measures)**.** The backlog of a server at time t is defined as

$$q(t) := A(t) - D(t) \tag{4}$$

and the virtual delay at time t is defined as

$$d(t) := \inf\{s \geq 0 : A(t) \leq D(t+s)\}. \tag{5}$$

Next, we introduce functions to bound the MGFs of processes that depend solely on the lengths of time intervals.

Definition 7 (f-bounded process)**.** A bivariate stochastic process $U(s,t)$ is $\overline{f}_U(\theta, t-s)$-bounded for $\theta > 0$ if for all $0 \leq s \leq t$

$$\mathrm{E}\Big[e^{\theta U(s,t)}\Big] \leq \overline{f}_U(\theta, t-s) \tag{6}$$

and $\underline{f}_U(\theta, t-s)$-bounded for $\theta > 0$ if for all $0 \leq s \leq t$

$$\mathrm{E}\Big[e^{-\theta U(s,t)}\Big] \leq \underline{f}_U(\theta, t-s). \tag{7}$$

Often, a special class of f-bounded processes is used.

Definition 8 ((σ, ρ)-Bounded Arrivals)**.** An arrival process $A(s,t)$ is (σ_A, ρ_A)-bounded for $\theta > 0$ if for all $0 \leq s \leq t$

$$\mathrm{E}\Big[e^{\theta A(s,t)}\Big] \leq e^{\theta \rho_A(\theta)\cdot(t-s) + \theta \sigma_A(\theta)} \tag{8}$$

and $(\sigma_{\underline{A}}, \rho_{\underline{A}})$-bounded for $\theta > 0$ if for all $0 \leq s \leq t$

$$\mathrm{E}\Big[e^{-\theta A(s,t)}\Big] \leq e^{-\theta \rho_{\underline{A}}(-\theta)\cdot(t-s) + \theta \sigma_{\underline{A}}(-\theta)}. \tag{9}$$

Definition 9 ((σ, ρ)-Bounded Service)**.** A min-plus service process S is (σ_S, ρ_S)-bounded for $\theta > 0$ if for all $0 \leq s \leq t$

$$\mathrm{E}\Big[e^{-\theta S(s,t)}\Big] \leq e^{-\theta \rho_S(-\theta)\cdot(t-s) + \theta \sigma_S(-\theta)}. \tag{10}$$

Having established the foundation for abstracting arrival and service processes within SNC, we continue with deterministic bounds on performance measures at the sample path level.

Theorem 10 (Sample Path Bounds: Backlog and Delay [13])**.** *Consider a flow with arrival process $A(s,t)$ traversing a server with min-plus service process $S(s,t) \geq 0$.*

1) *The backlog at time $t \geq 0$ is upper bounded by*

$$q(t) \leq A \oslash S\,(t,t).$$

2) *The virtual delay at time $t \geq 0$ is upper bounded by*

$$d(t) \leq \inf\left\{s \geq 0 : A \hat{\oslash} S(t+s,t) \leq 0\right\}.$$

Note that the deconvolution for the delay extends the domain of the deconvolution in Definition 5 and we define it as

$$A \hat{\oslash} S(t+s,t) := \sup_{0 \leq \tau \leq t} \{A(\tau,t) - S(\tau,t+s)\}\,.$$

Based on the sample path bounds and assuming (σ, ρ)-bounds for arrival and service, we are now able to calculate the desired probable worst-case guarantees with respect to backlog and delay.

Theorem 11 (Violation Probability of Backlog and Delay [13])**.** *Let $\theta > 0$. Suppose we have (σ_A, ρ_A)-bounded arrivals and a server with a (σ_S, ρ_S)-bounded min-plus service process S. Additionally, we require the arrivals and the service to be independent. Further, we assume the* stability condition

$$\rho_A(\theta) < \rho_S(-\theta).$$

1) *Let $B \geq 0$. For the backlog, it holds for all $t \geq 0$ that*

$$\mathrm{P}(q(t) > B) \leq e^{-\theta B} \cdot \frac{e^{\theta(\sigma_A(\theta)+\sigma_S(-\theta))}}{1-e^{\theta(\rho_A(\theta)-\rho_S(-\theta))}}.$$

2) *Let $T \geq 0$. For the virtual delay, it holds for all $t \geq 0$ that*

$$\mathrm{P}(d(t) > T) \leq e^{-\theta \rho_S(-\theta) T} \cdot \frac{e^{\theta(\sigma_A(\theta)+\sigma_S(-\theta))}}{1-e^{\theta(\rho_A(\theta)-\rho_S(-\theta))}}.$$

III. Analytical Results

In this section, we present the main contributions of the paper: we enable the residuation of min-plus service processes and provide a Laplace bound for Markov-modulated arrival processes.

A. Residuation under Min-Plus Service Process

First, we introduce an important result: we generalize the findings in [21] to bivariate functions, enabling us to derive performance bounds despite the server originally only offering a min-plus service process to the aggregate of the flows.

Theorem 12 (Residual Service with Min-Plus Service Process)**.** *Consider two arrival processes A_1, A_2, that share a server with a* min-plus *service process S. Further, we assume A_1 to be the arrivals of the flow of interest (foi) and an* arbitrary multiplexing *between the arrivals. Then, the foi sees a residual min-plus service process*

$$S^{\mathrm{res}}(s,t) = S(s,t) - A_2(s,t) \tag{11}$$

for all $0 \leq s \leq t$.

If we additionally assume the arrival process A_2 and the service process to be independent, $(\sigma_{A_2}, \rho_{A_2})$- and (σ_S, ρ_S)-bounded, respectively, then the residual service process is $(\sigma_{S_{\mathrm{res}}}, \rho_{S_{\mathrm{res}}})$-bounded with

$$\sigma_{S_{\mathrm{res}}}(-\theta) = \sigma_S(-\theta) + \sigma_{A_2}(\theta),$$
$$\rho_{S_{\mathrm{res}}}(-\theta) = \rho_S(-\theta) - \rho_{A_2}(\theta)$$

for some $\theta > 0$.

Proof. The first part is established in [8] for univariate processes and extends directly to bivariate ones, while the second part follows from [13] (see Prop. 1 in [13]). □

Note that $S^{\mathrm{res}}(s,t)$ is not necessarily nonnegative for all $0 \leq s \leq t$. This unusual fact is emphasized from now on by using the notation $\xi(s,t)$ instead of $S^{\mathrm{res}}(s,t)$. Indeed, if we were to assume the service process S to be strict, we could apply the positive part, defined as $[x]^+ := \max\{0, x\}$ (again, see Prop. 1 in [13]).

As the delay bound calculation is based on a nonnegative service process, this requires reconsideration and is the major point addressed in this section. We first show that, under potentially negative service ξ, the sample path backlog bound remains unchanged, but the delay bound needs to be generalized.

Theorem 13 (Generalized Sample Path Bounds: Backlog and Delay [21])**.** *Consider a flow with arrival process $A(s,t)$ traversing a server with min-plus service process $\xi(s,t)$.*

1) *The backlog at time $t \geq 0$ is upper bounded by*

$$q(t) \leq A \oslash \xi\,(t,t).$$

2) *The virtual delay at time $t \geq 0$ is upper bounded by*

$$d(t) \leq \inf\big\{s \geq 0 : A \,\hat{\oslash}\, \xi(t+s,t) \leq 0 \textit{ AND } A \hat{\otimes} \xi(t,t+s) \geq 0\big\}, \tag{12}$$

$$x \hat{\otimes} y(t,t+u) := \inf_{t+1 \leq \tau \leq t+u}\{x(t,\tau) + y(\tau,t+u)\}.$$

Proof. 1) By assumption of a service process, we can conclude for any fixed sample path that

$$\begin{aligned} q(t) &\overset{(4)}{=} A(t) - D(t) \\ &\overset{(1)}{\leq} A(t) - \inf_{0 \leq \tau \leq t}\{A(\tau) + \xi(\tau,t)\} \\ &= \sup_{0 \leq \tau \leq t}\{A(\tau,t) - \xi(\tau,t)\}. \end{aligned}$$

2) From the definition of the delay, for any fixed sample path and all $t \geq 0$

$$\begin{aligned} &d(t) \\ &\overset{(5)}{=} \inf\{s \geq 0 \mid A(t) \leq D(t+s)\} \\ &\overset{(1)}{\leq} \inf\left\{s \geq 0 \mid A(t) \leq \inf_{0 \leq \tau \leq t+s}\{A(\tau) + \xi(\tau,t+s)\}\right\} \\ &= \inf\left\{s \geq 0 \mid \sup_{0 \leq \tau \leq t+s}\{A(\tau,t) - \xi(\tau,t+s)\} \leq 0\right\} \\ &= \inf\left\{s \geq 0 \mid \sup_{0 \leq \tau \leq t}\{A(\tau,t) - \xi(\tau,t+s)\} \vee \right. \\ &\qquad \left. \sup_{t+1 \leq \tau \leq t+s}\{A(\tau,t) - \xi(\tau,t+s)\} \leq 0\right\} \\ &= \inf\big\{s \geq 0 \mid A \hat{\oslash} \xi(t+s,t) \leq 0 \text{ AND } \\ &\qquad A \hat{\otimes} \xi(t,t+s) \geq 0\big\}. \end{aligned}$$

□

As mentioned in the introduction, this theorem recovers a recent DNC result, but for the bivariate instead of the univariate case [21].

Next, we prove a generalized stochastic delay bound without requiring the service process to be nonnegative, so we can apply it to Eq. (11) in Th. 12, i.e., when the aggregate server has a min-plus service process. This is based on Th. 13 and is enabled by a Laplace bound on the arrivals, effectively representing a minimal arrival guarantee.

Theorem 14 (Generalized Stochastic Delay Bound)**.** *Let $\theta > 0$. Let an arrival process be $\overline{f}_A$- and $\underline{f}_A$-bounded and a server with min-plus service process ξ be $\underline{f}_\xi$-bounded. We also assume arrivals and service to be independent*[1]*. For $T \geq 0$, it holds for all $t \geq 0$ that*

$$\begin{aligned} \mathrm{P}(d(t) > T) \leq & \sum_{s=0}^{t} \overline{f}_A(\theta, t-s) \cdot \underline{f}_\xi(\theta, t+T-s) \\ & + \sum_{k=1}^{T} \underline{f}_A(\theta, k) \cdot \underline{f}_\xi(\theta, T-k). \end{aligned} \tag{13}$$

Proof. By (12), we know that if there exists a $T \geq 0$ such that $A \hat{\oslash} \xi(t+T, t) \leq 0$ AND $A \hat{\otimes} \xi(t, t+T) \geq 0$, then $d(t) \leq T$. Vice versa, for $T \geq 0$:

$$d(t) > T \Rightarrow A \hat{\oslash} \xi(t+T, t) > 0 \text{ OR } A \hat{\otimes} \xi(t, t+T) < 0.$$

Therefore,

$$\begin{aligned} & \mathrm{P}(d(t) > T) \\ \leq & \mathrm{P}\big(A \hat{\oslash} \xi(t+T, t) > 0 \text{ OR } A \hat{\otimes} \xi(t, t+T) < 0\big) \\ \overset{(2)}{\leq} & \mathrm{P}\big(A \hat{\oslash} \xi(t+T, t) > 0) + \mathrm{P}\big(A \hat{\otimes} \xi(t, t+T) < 0\big) \\ = & \mathrm{P}\left(\sup_{0 \leq \tau \leq t} \{A(\tau, t) - \xi(\tau, t+T)\} > 0\right) \\ & + \mathrm{P}\left(\inf_{t+1 \leq \tau \leq t+T} \{A(t, \tau) + \xi(\tau, t+T)\} < 0\right) \\ \overset{(2)}{\leq} & \sum_{\tau=0}^{t} \mathrm{P}(A(\tau, t) - \xi(\tau, t+T) > 0) \\ & + \sum_{\tau=t+1}^{t+T} \mathrm{P}(A(t, \tau) + \xi(\tau, t+T) < 0) \\ \overset{(3),\text{(ind.)}}{\leq} & \sum_{\tau=0}^{t} \mathrm{E}\left[e^{\theta A(\tau, t)}\right] \mathrm{E}\left[e^{-\theta \xi(\tau, t+T)}\right] \\ & + \sum_{\tau=t+1}^{t+T} \mathrm{E}\left[e^{-\theta A(t, \tau)}\right] \mathrm{E}\left[e^{-\theta \xi(\tau, t+T)}\right] \\ \overset{(6),(7)}{\leq} & \sum_{s=0}^{t} \overline{f}_A(\theta, t-s) \cdot \underline{f}_\xi(\theta, t+T-s) \\ & + \sum_{k=1}^{T} \underline{f}_A(\theta, k) \cdot \underline{f}_\xi(\theta, T-k). \end{aligned}$$

In the last equation, we substitute $k := \tau - t$. □

With (σ, ρ)-bounded arrivals and service, we can achieve time-independent bounds.

Corollary 15 (Generalized Delay Bound (Linear Case))**.** *Let $\theta > 0$. Suppose we have an arrival process A that is (σ_A, ρ_A)-bounded as well as $(\sigma_{\underline{A}}, \rho_{\underline{A}})$-lower-bounded and a min-plus service process ξ that is (σ_ξ, ρ_ξ)-bounded. Additionally, we require the arrivals and the service to be independent. Further, we assume the* stability condition

$$\rho_A(\theta) < \rho_\xi(-\theta). \tag{14}$$

Let $T \geq 0$. For the virtual delay, it holds for all $t \geq 0$ that

$$\begin{aligned} & \mathrm{P}(d(t) > T) \\ \leq & e^{-\theta \rho_\xi(-\theta) T} e^{\theta \sigma_\xi(-\theta)} \cdot \left(e^{\theta \sigma_A(\theta)} \frac{1}{1 - e^{\theta(\rho_A(\theta) - \rho_\xi(-\theta))}} \right. \\ & + e^{\theta \sigma_{\underline{A}}(-\theta)} e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right)} \\ & \left. \cdot \frac{1 - e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right) T}}{1 - e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right)}} \right). \end{aligned}$$

Proof. Starting from Eq. (13), we obtain

$$\begin{aligned} & \mathrm{P}(d(t) > T) \\ \leq & \sum_{\tau=0}^{t} \overline{f}_A(\theta, t-\tau) \cdot \underline{f}_\xi(\theta, t+T-\tau) \\ & + \sum_{\tau=1}^{T} \underline{f}_A(\theta, \tau) \cdot \underline{f}_\xi(\theta, T-\tau) \\ \overset{(8),(9),(10)}{=} & \sum_{\tau=0}^{t} e^{\theta(\rho_A(\theta)(t-\tau) + \sigma_A(\theta))} e^{-\theta(\rho_\xi(-\theta)(t+T-\tau) - \sigma_\xi(-\theta))} \\ & + \sum_{\tau=1}^{T} e^{-\theta\left(\rho_{\underline{A}}(-\theta)\tau - \sigma_{\underline{A}}(-\theta)\right)} e^{-\theta(\rho_\xi(-\theta)(T-\tau) - \sigma_\xi(-\theta))} \\ = & e^{-\theta \rho_\xi(-\theta) T} e^{\theta \sigma_\xi(-\theta)} \cdot \left(e^{\theta \sigma_A(\theta)} \sum_{j=0}^{t} e^{\theta(\rho_A(\theta) - \rho_\xi(-\theta)) j} \right. \\ & \left. + e^{\theta \sigma_{\underline{A}}(-\theta)} \sum_{\tau=1}^{T} e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right)\tau} \right). \end{aligned}$$

By substituting $j := t - \tau$ in the previous expression, we upper bound it using a geometric series, exploiting the stability condition $\left|e^{\theta(\rho_A(\theta) - \rho_\xi(-\theta))}\right| < 1$

$$\begin{aligned} & \mathrm{P}(d(t) > T) \\ \leq & e^{-\theta \rho_\xi(-\theta) T} e^{\theta \sigma_\xi(-\theta)} \cdot \left(e^{\theta \sigma_A(\theta)} \sum_{j=0}^{\infty} e^{\theta(\rho_A(\theta) - \rho_\xi(-\theta)) j} \right. \\ & \left. + e^{\theta \sigma_{\underline{A}}(-\theta)} \sum_{k=1}^{T} e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right) k} \right) \\ \overset{(14)}{=} & e^{-\theta \rho_\xi(-\theta) T} e^{\theta \sigma_\xi(-\theta)} \cdot \left(\frac{e^{\theta \sigma_A(\theta)}}{1 - e^{\theta(\rho_A(\theta) - \rho_\xi(-\theta))}} \right. \\ & \left. + e^{\theta \sigma_{\underline{A}}(-\theta)} e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right)} \frac{1 - e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right) T}}{1 - e^{-\theta\left(\rho_{\underline{A}}(-\theta) - \rho_\xi(-\theta)\right)}} \right). \end{aligned}$$

□

We now turn to the backlog bound. Since the sample-path expression remains unchanged, compare parts (1) of Th. 13 and Th. 10, the violation probability is unaffected. This means

[1]If arrivals and service were not independent, we could resort to using Hölder's inequality.

for the backlog that the existing results can be recovered, i.e., here the partial negativity of the service process does not play a role. For the sake of completeness, we state the results below without proofs (since they proceed along the same lines as for the delay).

Theorem 16 (Violation Probability of Backlog)**.** *Let $\theta > 0$. Suppose we have $\overline{f}_A$-bounded arrivals and a $\underline{f}_\xi$-bounded min-plus service ξ. Additionally, we require the arrivals and the service to be independent. Let $B \geq 0$. For the backlog, it holds for all $t \geq 0$ that*

$$\mathrm{P}(q(t) > B) \leq e^{-\theta B} \cdot \sum_{\tau=0}^{t} \overline{f}_A(\theta, t-\tau) \cdot \underline{f}_\xi(\theta, t-\tau). \quad (15)$$

Proof. The proof proceeds along the same lines as Th. 14 and is therefore omitted. □

Corollary 17 (Violation Probability of Backlog (Linear Case))**.** *Let $\theta > 0$. Suppose we have (σ_A, ρ_A)-bounded arrivals and a (σ_ξ, ρ_ξ)-bounded min-plus service ξ. Additionally, we require the arrivals and the service to be independent. Further, we assume the* stability condition

$$\rho_A(\theta) < \rho_\xi(-\theta). \quad (16)$$

Let $B \geq 0$. For the backlog, it holds for all $t \geq 0$ that

$$\mathrm{P}(q(t) > B) \leq e^{-\theta B} \cdot \frac{e^{\theta(\sigma_A(\theta)+\sigma_\xi(-\theta))}}{1 - e^{\theta(\rho_A(\theta)-\rho_\xi(-\theta))}}. \quad (17)$$

Proof. The proof proceeds along the same lines as Cor. 15 and is therefore omitted. □

B. Markov-Modulated Arrival Processes

Markov-modulated arrival processes constitute a natural application scenario for our framework [6]. To that end, we provide upper bounds to their corresponding Laplace transform.

Assume a Markov-modulated arrival process A with the finite state space S, i.e., the distribution of the increments of A only depend on the current state of an underlying Markov chain Y. The Markov chain is described by its state space S and transition matrix $T = [t_{ij}]$ such that $t_{ij} > 0$ for all $i, j \in S$, where we assume that T is irreducible and aperiodic. We define the increments of the arrival process $a(t) = X_{Y(t)}(t)$, where $X_i(t), i \in S$, is an i.i.d. process with existing MGF, and denote by $E \in \mathrm{Diag}(S)$ the matrix with entries $E_i := E_{ii} := \mathrm{E}\left[e^{-\theta a(t)} \mid Y_t = i\right]$ for all states $i \in S$. Similarly, let $E' \in \mathrm{Diag}(S)$ with entries $E'_i := E'_{ii} := \mathrm{E}\left[e^{\theta a(t)} \mid Y_t = i\right]$ for all states $i \in S$.

We first require the following lemma.

Lemma 18. *Let $B = [b_{ij}]$ be a nonnegative matrix with index set S and $B^t = [b_{ij}^{(t)}]$ be its t-th power. If B has a positive eigenvector $\bar{x} = [\bar{x}_i]$, then for all $t \in \mathbb{N}$ and all $i \in S$ the inequality*

$$\sum_{j \in S} b_{ij}^t \leq \frac{\max_{k \in S} \bar{x}_k}{\min_{k \in S} \bar{x}_k} \cdot \mathrm{sp}(B)^t$$

holds, where $\mathrm{sp}(\cdot)$ denotes the spectral radius *of a matrix.*

Proof. See (8.1.33) in [22]. □

We can build on existing MGF bounds for Markov-modulated arrival processes from the literature.

Lemma 19 (MGF Bound for Markov-Modulated Arrival Processes [6], [23])**.** *For the above holds that the MGF of A is (σ, ρ)-bounded with*

$$\sigma(\theta) = \frac{1}{\theta} \log\left(\left(\max_{i \in S} E'_i\right) \cdot \frac{\max_{k \in S} \bar{x}_k}{\min_{k \in S} \bar{x}_k} \cdot \frac{1}{\mathrm{sp}(E'T)}\right),$$
$$\rho(\theta) = \frac{1}{\theta} \log\left(\mathrm{sp}(E'T)\right),$$

where $\bar{x}$ is a positive eigenvector of $E'T$.

Now, we just need to derive the required Laplace bound to apply our analytical results.

Lemma 20 (Laplace Bound for Markov-Modulated Arrival Processes)**.** *For the above holds that the Laplace transform of A is (σ, ρ)-bounded with*

$$\sigma(-\theta) = \frac{1}{\theta} \log\left(\left(\max_{i \in S} E_i\right) \cdot \frac{\max_{k \in S} \bar{x}_k}{\min_{k \in S} \bar{x}_k} \cdot \frac{1}{\mathrm{sp}(ET)}\right),$$
$$\rho(-\theta) = -\frac{1}{\theta} \log\left(\mathrm{sp}(ET)\right),$$

where $\bar{x}$ is a positive eigenvector of ET.

Proof. Fix $\theta > 0$, for every $i \in S$ holds the backward equation

$$\begin{aligned}
E_i(t) :=& \mathrm{E}\left[e^{-\theta A(t)} \mid Y_1 = i\right] \\
=& \mathrm{E}\left[e^{-\theta a(1)} \mid Y_1 = i\right] \cdot \mathrm{E}\left[e^{-\theta(A(t)-a(1))} \mid Y_1 = i\right] \\
=& E_i \sum_{j \in S} \mathrm{E}\left[e^{-\theta(A(t)-a(1))} \mid Y_1 = i, Y_2 = j\right] \\
&\cdot \mathrm{P}\left(Y_2 = j \mid Y_1 = i\right) \\
=& E_i \sum_{j \in S} \mathrm{E}\left[e^{-\theta A(t-1)} \mid Y_1 = j\right] \cdot t_{ij} \\
=& E_i \sum_{j \in S} E_j(t-1) \cdot t_{ij}.
\end{aligned}$$

Hence, the vector $E(t)$ with entries $E_i(t)$ has the form $E(t) = ET \cdot E(t-1)$. Applying this recursion results in $E(t) = (ET)^{t-1} E \cdot \mathbf{1}$, where $\mathbf{1}$ is the unit column vector on S. Assume now the beginning state of the chain is not given but follows a distribution $(\pi_i)_{i \in S}$. Then, an application of the law of total probability yields

$$\begin{aligned}
E_\pi(t) := \mathrm{E}\left[e^{-\theta A(t)}\right] &= \sum_{j \in S} \mathrm{P}(Y_1 = j)\, E_j(t) = \sum_{j \in S} \pi_j E_j(t) \\
&= \sum_{j \in S} \pi_j ((ET)^{t-1} E \cdot \mathbf{1})_j = \pi \cdot (ET)^{t-1} E \cdot \mathbf{1}.
\end{aligned}$$

Since T is strictly positive and E has only strictly positive entries on the diagonal, the matrix ET is also strictly positive (i.e., every entry is > 0). This allows us to apply the Perron-Frobenius theorem [24], [25], which guarantees an eigenvector with strictly positive entries. Denote this eigenvector by $\bar{x} \in$

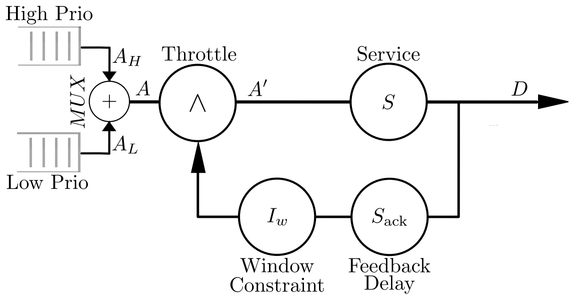

Fig. 1: Window flow control scheme with priorities.

$\mathbb{R}^S$; it fulfills the conditions of Lem. 18. Thus

$$\begin{aligned} E_\pi(t) &= \sum_{i\in S} \pi_i \sum_{j\in S} (ET)_{ij}^{t-1} (E \cdot \mathbf{1})_j \\ &\leq \sum_{i\in S} \pi_i \left(\max_{k\in S} E_k\right) \sum_{j\in S} (ET)_{ij}^{t-1} \\ &\leq \left(\max_{k\in S} E_k\right) \sum_{i\in S} \pi_i \frac{\max_{k\in S} \bar{x}_k}{\min_{k\in S} \bar{x}_k} \cdot \mathrm{sp}(ET)^{t-1} \\ &= \left(\max_{k\in S} E_k\right) \cdot \frac{\max_{k\in S} \bar{x}_k}{\min_{k\in S} \bar{x}_k} \cdot \mathrm{sp}(ET)^{t-1} \end{aligned}$$

for every starting distribution π and all $t \in \mathbb{N}$. Then,

$$\begin{aligned} E_{Y_s}(t-s) &= \mathrm{E}\left[e^{-\theta(A(t)-A(s))} \mid Y_s\right] \\ &\leq e^{-\theta\left(-\frac{1}{\theta}\log\left((\max_{k\in S} E_k)\cdot \frac{\max_{k\in S}\bar{x}_k}{\min_{k\in S}\bar{x}_k}\cdot \mathrm{sp}(ET)^{t-s-1}\right)\right)}. \end{aligned}$$

□

IV. Possible Application

As a possible application, we identify window flow control (WFC) systems, since transforming the closed-loop system into an open-loop one via sub-additive closure results in a service process that is no longer strict [8, p. 137]. Equipped with our new results presented in Sec. III, we can now analyse *multiple* flow scenarios in WFC systems (see Fig. 1), as we are able to perform the residuation step *after* the transformation. For instance, building on the model of [26], one can incorporate Markov-modulated arrivals and derive delay bounds not only at the aggregate level but also per flow, especially for lower priority traffic, as well as calculate the required source buffer sizes to avoid excessive packet loss at the foi's source buffer, thereby enabling true per-flow analysis.

V. Conclusion

We extended the SNC by accommodating a broader class of service processes, potentially negative ones. This enables us to perform residuation in the case of only min-plus service processes instead of strict ones. To that end, the delay bound was generalized leveraging minimum arrival bounds, represented through the Laplace transform in MGF-based SNC. Notably, Laplace bounds "come for free", as nontrivial Laplace transforms always exist, given arrival distribution assumptions. In particular, we derived bounds on the Laplace transform of Markov-modulated processes.

References

[1] L. Kleinrock, *Queueing systems*. wiley New York, 1975, vol. 66.

[2] F. P. Kelly, “Notes on Effective Bandwidths,” in *Stochastic Networks: Theory and Applications*, 1996.

[3] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and linearity*. Wiley New York, 1992, vol. 1.

[4] R. L. Cruz, “A calculus for network delay. i. network elements in isolation,” *IEEE Transactions on information theory*, vol. 37, no. 1, pp. 114–131, 1991.

[5] ——, “A calculus for network delay. ii. network analysis,” *IEEE Transactions on information theory*, vol. 37, no. 1, pp. 132–141, 1991.

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

[7] J.-Y. Le Boudec and P. Thiran, *Network calculus: a theory of deterministic queuing systems for the internet*. Springer, 2002.

[8] A. Bouillard, M. Boyer, and E. Le Corronc, *Deterministic Network Calculus: From Theory to Practical Implementation*. John Wiley & Sons, 2018.

[9] R. L. Cruz, “Quality of service management in integrated services networks,” in *Proceedings of the 1st Semi-Annual Research Review, CWC*, vol. 1, no. 7. Citeseer, 1996, pp. 4–5.

[10] R. R. Boorstyn, A. Burchard, J. Liebeherr, and C. Oottamakorn, “Statistical service assurances for traffic scheduling algorithms,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, pp. 2651–2664, 2002.

[11] F. Ciucu, A. Burchard, and J. Liebeherr, “Scaling properties of statistical end-to-end bounds in the network calculus,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2300–2312, 2006.

[12] Y. Jiang, Y. Liu *et al.*, *Stochastic network calculus*. Springer, 2008, vol. 1.

[13] M. Fidler, “An end-to-end probabilistic network calculus with moment generating functions,” in *14th IEEE International Workshop on Quality of Service*. IEEE, 2006, pp. 261–270.

[14] A. Rizk and M. Fidler, “Leveraging statistical multiplexing gains in single-and multi-hop networks,” in *2011 IEEE Nineteenth IEEE International Workshop on Quality of Service*. IEEE, 2011, pp. 1–9.

[15] F. Ciucu, A. Burchard, and J. Liebeherr, “Scaling properties of statistical end-to-end bounds in the network calculus,” *IEEE/ACM Transactions on Networking (ToN)*, vol. 14, no. SI, pp. 2300–2312, 2006.

[16] J. Liebeherr, A. Burchard, and F. Ciucu, “Delay bounds in communication networks with heavy-tailed and self-similar traffic,” *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1010–1024, 2012.

[17] A. Rizk and M. Fidler, “Non-asymptotic end-to-end performance bounds for networks with long range dependent fbm cross traffic,” *Computer Networks*, vol. 56, no. 1, pp. 127–141, 2012.

[18] A. Bouillard, P. Nikolaus, and J. Schmitt, “Unleashing the power of paying multiplexing only once in stochastic network calculus,” *Proceedings of the ACM on Measurement and Analysis of Computing Systems (POMACS)*, vol. 6, no. 2, p. 34, 2022.

[19] C.-S. Chang and R. L. Cruz, “A time varying filtering theory for constrained traffic regulation and dynamic service guarantees,” in *Proc. IEEE International Conference on Computer Communications (INFOCOM'99)*, vol. 1, 1999, pp. 63–70.

[20] C.-S. Chang, R. L. Cruz, J.-Y. Le Boudec, and P. Thiran, “A min,+ system theory for constrained traffic regulation and dynamic service guarantees,” *IEEE/ACM Transactions on Networking*, vol. 10, no. 6, pp. 805–817, 2002.

[21] A. Hamscher, V.-C. Constantin, and J. B. Schmitt, “Extending network calculus to deal with min-plus service curves in multiple flow scenarios,” in *2024 IEEE 30th Real-Time and Embedded Technology and Applications Symposium (RTAS)*. IEEE, 2024, pp. 95–107.

[22] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[23] M. Beck, “Advances in theory and applicability of stochastic network calculus,” Ph.D. dissertation, Technische Universität Kaiserslautern, 2016.

[24] F. G. Frobenius, “Über Matrizen aus nicht negativen Elementen,” 1912.

[25] O. Perron, “Zur Theorie der Matrices,” *Mathematische Annalen*, vol. 64, no. 2, pp. 248–263, 1907.

[26] A. Shekaramiz, J. Liebeherr, and A. Burchard, “Window flow control systems with random service,” *arXiv preprint arXiv:1507.04631*, 2015.