On the Calculation of Sample-Path Backlog Bounds in Queueing Systems over Finite Time Horizons

Technical Report

Michael Beck, Jens B. Schmitt

Distributed Computer Systems Lab (DISCO), University of Kaiserslautern, Germany

1 Abstract

The ability to calculate backlog bounds is of key importance for buffer sizing in packet-switched networks. In particular, it is critical to capture the statistical multiplexing gains which, in turn, calls for stochastic backlog bounds. The stochastic network calculus (SNC) is a promising methodology to compute such stochastic backlog bounds. So far in the literature SNC-based backlog bounds apply only to an arbitrary, but fixed single point in time. Yet, from the network engineering perspective, one would rather like to have a sample path backlog bound, i.e., a bound that applies (with a certain fixed violation probability) all of the time. While, in general, such bounds are hard to obtain we investigate in this paper how sample path backlog bounds can be computed over *finite* time horizons. In particular, we show how a simple extension of the known SNC results can lead to sub-optimal bounds by deriving an alternative methodology (based on extreme value theory) for bounding the backlog over finite time horizons. Interestingly, none of the two methods completely dominates the other. For the new method we also discuss how it can be evolved into a corresponding calculus for network analysis analogous to the existing SNC.

2 Introduction

Buffer sizing is a very important task in planning and controlling a packet-switched network. Since the early days of packet-switched networks it has seen much treatment [15], continues to be investigated intensively these days (see, e.g. [1, 13] and very likely will remain an important topic in the future. Thus, it is important to characterize the backlog process $q(n)$ in a queueing system (here, we assume discrete-time). The difficulty in doing this lies in the stochastic nature of arrivals and being able to capture the resulting statistical multiplexing effect, which can be seen as the raison d'être of packet-switched networks. In particular, one is interested in probabilistically bounding the backlog. Ideally, the following sample path bound could be calculated

$$\mathbb{P}(\forall\ n : q(n) > B_\varepsilon) \leq \varepsilon \Leftrightarrow \mathbb{P}(\max_{n\geq 1} q(n) > B_\varepsilon) \leq \varepsilon.$$

Yet, such a sample path bound on the backlog process is under most practical circumstances quasi-deterministic, i.e., ε only takes values 0 or 1. Stochastic network calculus (SNC) is a recent theory which among other performance measures allows to compute bounds on the backlog in a queueing system. In short, it allows to compute the following pointwise backlog bound:

$$\mathbb{P}(q(n) > B_\varepsilon) \leq \varepsilon \quad \forall\, n \geq 1$$

This, however, is often not quite what a network engineer desires as, in the course of time (or, more technically, on the actual sample path of the system), this bound does not give direct information on how often the backlog bound B_ε will be violated. Therefore, in this paper, we in a certain sense aim at the middleground between these two extremes by finding ways to calculate sample path backlog bounds over finite time horizons of the form

$$\mathbb{P}(\forall\, n \leq N : q(n) > B_\varepsilon) \leq \varepsilon \Leftrightarrow \mathbb{P}(\max_{1 \leq n \leq N} q(n) > B_\varepsilon) < \varepsilon.$$

The power of such a finite sample path backlog bound lies in its ability to answer relevant network engineering questions like: What is the probability that my system exceeds a certain backlog of B_ε in the next N time steps? In fact, it may even be a way to work out the (infinite) sample path bound from above if a deterministic bound on the duration of a backlogged period is available (this is the case for example when multiple independendent regulated flows are multiplexed as e.g. in [6, 22, 16, 21]).

As we will see in the course of the paper, it is possible to directly transform the SNC-based pointwise backlog bound into a finite sample path backlog bound (simply using Boole's inequality). Yet, this already "feels" sub-optimal as the violation probability ε grows linearly with the time horizon N, although it is of course bounded by 1. We substantiate this uneasiness of directly applying the SNC results in this way by developing an alternative method to bound backlogs on finite sample paths. The new method naturally lends itself to the calculation of finite sample path backlog bounds and always results in violation probabilities of less than 1. It is based on a simple observation of the system dynamics as well as on extreme value theory (EVT), a tool mainly used in financial and actuarial mathematics to calculate the probability of rare events involving some extremal expression. The new method delivers better bounds than the direct application of existing SNC results, thus exemplifying the problem with simply using Boole's inequality to arrive at finite sample path backlog bounds which was its main purpose in this work. However, motivated by these results we also see the potential for developing an alternative SNC, thus enabling to analyse more complex network scenarios. To that end, we also demonstrate how the corresponding operations, like multiplexing of flows, computation of output bounds, and leftover service computation can be performed.

The rest of the paper is organized as follows: In Section 2 we discuss related work. In Section 3 we briefly review the basics for this work, including our network model and a short introduction into SNC (concretely, we focus on Chang's version of the SNC [5], which is based on moment generating functions (MGF), which is why it is also often simply called MGF-Calculus). In Section 4, we show how to achieve finite sample path backlog bounds using our alternative method and compare it in numerical examples with the direct application of the SNC-based bounds in Section 5. In Section 6, we illustrate how the new method can be applied to more complicated network scenarios using an example. Section 7 concludes the paper and provides an outlook to future work.

3 Related Work

From the domain of classical queuing theory, it is known that exact calculations of the buffer occupancy distribution (in our terms the steady-state backlog distribution) are only possible for some simple source models [17]. However, what has been demonstrated in the literature is that powerful techniques such as large deviations [20], local limit theorems [18], or extreme value theory [10] can provide approximations that work well in the asymptotic domain. As we are, however, interested in non-asymptotic bounds rather than asymptotic approximations for the backlog process, these results, while being interesting and inspiring, do not quite fit our needs. Furthermore, these methods are typically very specifically tailored e.g. to certain assumptions on the arrival processes. In contrast, we follow the framework-oriented approach of SNC where we try to keep the analysis as generic as (far as) possible.

There have been several approaches to develop an SNC: The most prominent branches are the MGF calculus by Chang [4], later refined by Fidler in [11]; the statistical network calculus by Liebeherr et al. [3], extensively and nicely developed in [8]; and the work of Jiang which is well collected in [14]. We do not want to discuss their different merits and drawbacks here (an excellent survey can be found in [12]), but will focus on the MGF calculus in the rest of the paper as it is probably the most popular among those three (admittedly superficially judging based on citation counts from Google Scholar). Anyway, we found none of them or any derivative work to deal with sample-path backlog bounds over finite time horizons as discussed in Section 1. In fact, most SNC papers are very much focussed on delay as performance measure with a few exceptions (e.g. [7]), yet all of these calculate pointwise backlog bounds.

4 Preparations

Throughout the paper, we assume time to be discrete, whereas data can be either discrete or continuous.

4.1 Arrival and Service Processes

We describe the arrival and departure processes at some node as sequences of non-negative real numbers, which can be random. For this denote by $\mathcal{J}$ the space of sequences of non-negative random variables. We denote such a sequence by e.g. $(a_n)_{n\in\mathbb{N}}$ and the cumulative distribution function (cdf) of one element of the sequence by F_a. Further we define $\mathcal{I}$ as the space of sequences of non-negative i.i.d. random variables. Clearly $\mathcal{I} \subset \mathcal{J}$. For the rest of this work capital letters denote the cumulatives of such sequences, for example if $(a_n)_{n\in\mathbb{N}} \in \mathcal{J}$ we have

$$A(n) := \sum_{m=0}^{n} a_m$$

where - as usual - the empty sum is zero, i.e. $A(0) = 0$. For the case, that $a_n = c$ for some c and all $n \in \mathbb{N}$ almost surely, we just write $(a_n)_{n\in\mathbb{N}} = c$. Sometimes it will be convenient to use the zero as index expanding the set $(a_n)_{n\in\mathbb{N}}$ to $(a_n)_{n\in\mathbb{N}_0}$. In this case we always set $a_0 = 0$.

A service process at some node is instead given by a doubly indexed stochastic process e.g. $S(m,n)$ with:

$$\begin{aligned} 0 \leq& S(m,n) \qquad \forall\ m,n \in \mathbb{N} \\ S(m,n) \leq& S(m,n') \qquad \forall\ m \in \mathbb{N} \text{ and } n \leq n' \end{aligned}$$

In the special case of $S(m,n) = S(n) - S(m)$ for some S non-decreasing, we can consider the increments $(s_n)_{n\in\mathbb{N}} \in \mathcal{J}$ with $s_n := S(n-1,n)$. We will then just speak of a service $(s_n)_{n\in\mathbb{N}}$. We say a node offers service S if for every arrival $(a_n)_{n\in\mathbb{N}} \in \mathcal{J}$ and its corresponding departures $(d_n)_{n\in\mathbb{N}} \in \mathcal{J}$ holds:

$$D(n) \geq \min_{0\leq k\leq n} \{A(k) + S(k,n)\}$$

with equality if Lindley's equation is fulfilled:

$$q(n+1) = \max\{0,\ q(n) + a(n+1) - s(n+1)\}$$

4.2 Stochastic Network Calculus

In this work, we follow the framework of $(\sigma(\theta), \rho(\theta))$-calculus or simply MGF-Calculus, as presented in [5]. The basic idea is to bound the MGF of the arrivals and service by some exponential. This of course only works, if the corresponding MGF exists. Next, we define how exactly these bounds are calculated and display some results, which allow us to analyse networks and achieve backlog bounds. The proofs for the lemmata in this subsection are omitted and can be found either in [2] or in [5].

Definition 1 (Arrivals and Services) *An arrival $(a_n)_{n\in\mathbb{N}} \in \mathcal{J}$ is $(\sigma(\theta), \rho(\theta))$-bounded iff for some $\theta > 0$:*

$$\sup_{k\geq 0}\{\mathbb{E}(e^{\theta(A(n+k)-A(k))})\} \leq e^{n\theta\rho(\theta)+\theta\sigma(\theta)} \quad \forall\ n \in \mathbb{N}$$

If this is fulfilled we write $(a_n)_{n\in\mathbb{N}} \preceq (\sigma(\theta), \rho(\theta))$.

A service S is $(\sigma(\theta), \rho(\theta))$-bounded iff for some $\theta > 0$:

$$\sup_{k\geq 0}\{\mathbb{E}(e^{-\theta S(k,n+k)})\} \leq e^{n\theta\rho(\theta)+\theta\sigma(\theta)} \quad \forall\, n \in \mathbb{N}$$

If this is fulfilled we write $S \succeq (\sigma(\theta), \rho(\theta))$.

Note that if $S \succeq (\sigma(\theta), \rho(\theta))$ then $\rho(\theta)$ is usually negative.

Now assume a node with service $(s_n)_{n\in\mathbb{N}} \in \mathcal{J}$ serves two arrival processes $(\bar{a}_n)_{n\in\mathbb{N}} \in \mathcal{J}$ and $(\underline{a}_n)_{n\in\mathbb{N}} \in \mathcal{J}$, where $(\bar{a}_n)_{n\in\mathbb{N}}$ has a higher priority than $(\underline{a}_n)_{n\in\mathbb{N}}$. Then the low-priority flow receives only the service, which is left over by the high-priority flow. In expression, if we denote by $(\underline{s}_n)_{n\in\mathbb{N}} \in \mathcal{J}$ the leftover service, we have:

$$\underline{s}_n = \max\{0, s_n - \bar{a}_n - q(n)\}$$

where $q(n)$ denotes the queue of the prioritized flow at time n, i.e. $q(n) = \bar{A}(n-1) - D(n-1)$. This scenario can be generalized to doubly indexed services S and we get the following for the leftover service:

Lemma 1 (Leftover Service) *In the above situation we have*

$$(\underline{S}_n)_{n\in\mathbb{N}} \succeq (\sigma_a(\theta) + \sigma_S(\theta), \rho_a(\theta) + \rho_S(\theta))$$

if S and $(a_n)_{n\in\mathbb{N}} \in \mathcal{J}$ are stochastically independent.
If they are not stochastically independent we still have (using Hölder's inequality)

$$(\underline{S}_n)_{n\in\mathbb{N}} \succeq (\sigma_a(q\theta) + \sigma_S(p\theta), \rho_a(q\theta) + \rho_S(p\theta))$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. We define $\underline{S}(m,n) := \max\{0, S(m,n) - \bar{A}(n) + \bar{A}(m)\}$ as leftover service and show first, that the node indeed offers this service for the second flow. Assume for this $n \geq 0$ arbitrary and choose $m \leq n$ maximal such that $\bar{A}(m) = \bar{D}(m)$ and $\underline{A}(m) = \underline{D}(m)$. The node fulfills then:

$$\bar{D}(n) + \underline{D}(n) \geq \bar{D}(m) + \underline{D}(m) + S(n,m) = \bar{A}(m) + \underline{D}(m) + S(m,n)$$

And since $\bar{D}(n) \leq \bar{A}(n)$ we have:

$$\begin{aligned}\underline{D}(n) &\geq \underline{D}(m) + \bar{A}(m) - \bar{A}(n) + S(m,n)\\ &\geq \min_{0\leq k\leq n}\{\underline{D}(k) + \underline{S}(k,n)\}\end{aligned}$$

Next we show, that the leftover service fulfills the above bound:

$$\begin{aligned}\mathbb{E}(e^{-\theta\underline{S}(m,n)}) &= \mathbb{E}(e^{-\theta\cdot\max\{0,S(m,n)-(A(n)-A(m))\}})\\ &\leq \mathbb{E}(e^{-\theta S(m,n)} \cdot e^{\theta(\bar{A}(n)-\bar{A}(m))})\end{aligned}$$

for the independent case we can continue with

$$\begin{aligned}&\leq e^{\theta\rho_S(\theta)(n-m)+\theta\sigma_S(\theta)}e^{\theta(n-m)\rho_a(\theta)+\theta\sigma_a(\theta)}\\&=e^{(n-m)\theta(\rho_a(\theta)+\rho_S(\theta))}e^{\theta(\sigma_a(\theta)+\sigma_S(\theta))}\end{aligned}$$

For the dependent case we use Hölder's inequality:

$$\begin{aligned}\mathbb{E}(e^{-\theta\underline{S}(m,n)})&\leq \mathbb{E}^{1/p}(e^{-p\theta S(m,n)})\mathbb{E}^{1/q}(e^{q\theta(\bar{A}(n)-\bar{A}(m))})\\&\leq e^{\theta(n-m)\rho_a(p\theta)+\theta\sigma_a(p\theta)}e^{\theta\rho_S(q\theta)(n-m)+\theta\sigma_S(q\theta)}\\&=e^{(n-m)\theta(\rho_a(p\theta)+\rho_S(q\theta))}e^{\theta(\sigma_a(p\theta)+\sigma_S(q\theta)}\end{aligned}$$

Lemma 2 (Multiplexing) *If we have two stochastically independent arrivals $(a_n^{(1)})_{n\in\mathbb{N}}\in\mathcal{J}$ and $(a_n^{(2)})_{n\in\mathbb{N}}\in\mathcal{J}$, which are $(\sigma_{a^{(i)}}(\theta),\rho_{a^{(i)}}(\theta))$-bounded $(i=1,2)$, then it holds for the multiplexed flow that*

$$(a_n^{(1)}+a_n^{(2)})_{n\in\mathbb{N}}\preceq(\sigma_{a^{(1)}}(\theta)+\sigma_{a^{(2)}}(\theta),\rho_{a^{(1)}}(\theta)+\rho_{a^{(2)}}(\theta))$$

For the case that the arrivals are not stochastically independent, we still have

$$(a_n^{(1)}+a_n^{(2)})_{n\in\mathbb{N}}\preceq(\sigma_{a^{(1)}}(q\theta)+\sigma_{a^{(2)}}(p\theta),\rho_{a^{(1)}}(q\theta)+\rho_{a^{(2)}}(p\theta)).$$

Proof. Trivial.

Lemma 3 (Output bound) *Let*

$$(a_n)_{n\in\mathbb{N}}\preceq(\sigma_a(\theta),\rho_a(\theta))$$

and

$$S\succeq(\sigma_S(\theta),\rho_S(\theta)).$$

Denote the output of the node by $(d_n)_{n\in\mathbb{N}}\in\mathcal{J}$, in the case of independence between arrivals and service we get:

$$\begin{aligned}&\mathbb{E}(e^{\theta(D(n)-D(m))})\\\leq&e^{\theta(\sigma_a(\theta)+\sigma_S(\theta))}e^{(n-m)\theta\rho_a(\theta)}\sum_{k=0}^{m}e^{k\theta(\rho_a(\theta)+\rho_S(\theta))}\end{aligned}$$

for all $m,n\in\mathbb{N}$ with $m\leq n$. Also

$$(d_n)_{n\in\mathbb{N}}\preceq(\sigma_a(\theta)+\sigma_S(\theta)+\tilde{\sigma}(\theta),\rho_a(\theta))$$

with:

$$\tilde{\sigma}(\theta)=\frac{1}{\theta}\log(1-e^{\theta(\rho_a(\theta)+\rho_S(\theta))})^{-1}$$

For the dependent case we similarly get

$$(d_n)_{n\in\mathbb{N}}\preceq(\sigma_a(q\theta)+\sigma_S(p\theta)+\tilde{\sigma}(q\theta,p\theta),\rho_a(q\theta))$$

with

$$\tilde{\sigma}(q\theta,p\theta)=(1-e^{\theta(\rho_a(q\theta)+\rho_S(p\theta))})^{-1}$$

and $\frac{1}{p}+\frac{1}{q}=1$.

Proof. We start by bounding the amount of data which can depart the node in the interval $(m, n]$, which is at most the amount of arriving data plus the buffered data at time m:

$$D(n) - D(m) \leq A(n) - A(m) + q(m)$$

Using that the node offers service S we get:

$$\begin{aligned} D(n) - D(m) &\leq A(n) - A(m) + A(m) - D(m) \leq A(n) - \min_{0\leq k\leq m}\{A(k) - S(k,m)\} \\ &= \max_{0\leq k\leq m}\{A(n) - A(k) - S(k,m)\} \end{aligned}$$

By the monotonicity of the expectation we have in the independent case:

$$\begin{aligned} \mathbb{E}(e^{\theta(D(n)-D(m))}) &\leq \mathbb{E}(e^{\theta \max_{0\leq k\leq m}\{A(n)-A(k)-S(k,m)\}}) \\ &\leq \sum_{k=0}^{m} \mathbb{E}(e^{\theta(A(n)-A(k))})\mathbb{E}(e^{-\theta S(k,m)}) \\ &\leq \sum_{k=0}^{m} e^{\theta(n-k)\rho_a(\theta)+\theta\sigma_a(\theta)} e^{\theta(m-k)\rho_S(\theta)+\theta\sigma_S(\theta)} \\ &= e^{\theta(\sigma_a(\theta)+\sigma_S(\theta))} e^{\theta(n-m)\rho_a(\theta)} \sum_{k=0}^{m} e^{k\theta(\rho_a(\theta)+\rho_S(\theta))} \end{aligned}$$

In the dependent case we use Hölder's inequality:

$$\begin{aligned} \mathbb{E}(e^{\theta(D(n)-D(m))}) &\leq \sum_{l=0}^{m} \mathbb{E}^{1/p}((e^{\theta(A(n)-A(l))})^p)\mathbb{E}^{1/q}((e^{-\theta S(k,m)})^q) \\ &\leq \sum_{l=0}^{m} e^{\theta(n-l)\rho_a(p\theta)+\theta\sigma_a(p\theta)} e^{\theta(m-l)\rho_S(q\theta)+\theta\sigma_S(q\theta)} \\ &= e^{\theta(\sigma_a(p\theta)+\sigma_S(q\theta))} e^{\theta(n-m)\rho_a(p\theta)} \sum_{k=0}^{m} e^{k\theta(\rho_a(p\theta)+\rho_S(q\theta))} \end{aligned}$$

It is left to show that $(d_n)_{n\in\mathbb{N}} \preceq (\sigma_a(\theta) + \sigma_S(\theta) + \tilde{\sigma}(\theta), \rho_a(\theta))$, for this we just have to bound the above sums by its corresponding geometric series.

Lemma 4 (Backlog Bound) *In the same situation as in the previous lemma it holds for all $n \in \mathbb{N}$:*

$$\mathbb{P}(q(n) \leq x) \leq e^{-\theta x} e^{\theta(\sigma_a(\theta)+\sigma_S(\theta))} \sum_{m=0}^{n} e^{m\theta(\rho_a(\theta)+\rho_S(\theta))}$$

if $(a_n)_{n\in\mathbb{N}}$ is stochastically independent of S. If this is not the case we have

$$\mathbb{P}(q(n) \leq x) \leq e^{-\theta x} e^{\theta(\sigma_a(q\theta)+\sigma_S(p\theta))} \sum_{m=0}^{n} e^{m\theta(\rho_a(q\theta)+\rho_S(p\theta))}$$

for all $n \in \mathbb{N}$ and p, q such that $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. We have:

$$q(n) = A(n) - D(n) \leq \max_{0 \leq m \leq n} \{A(n) - A(m) - S(m,n)\}$$

And by the Markov inequality and the monotonicity of the expectation:

$$\begin{aligned}
\mathbb{P}(q(n) \leq x) &\leq e^{-\theta x} \mathbb{E}(e^{\theta \max_{0 \leq m \leq n} \{A(n)-A(m)-S(m,n)\}}) \\
&\leq e^{-\theta x} \sum_{m=0}^{n} \mathbb{E}(e^{\theta(A(n)-A(m)-S(m,n))})
\end{aligned}$$

In the independent case we can continue with:

$$\begin{aligned}
&= e^{-\theta x} \sum_{m=0}^{n} \mathbb{E}(e^{\theta(A(n)-A(m))}) \mathbb{E}(e^{-\theta S(m,n)}) \\
&\leq e^{-\theta x} e^{\theta(\sigma_a(\theta)+\sigma_S(\theta))} \sum_{m=0}^{n} e^{\theta(\rho_a(\theta)+\rho_S(\theta))}
\end{aligned}$$

In the dependent case we have to use Höder's inequality instead:

$$\begin{aligned}
\mathbb{P}(q(n) \leq x) &\leq e^{-\theta x} \sum_{m=0}^{n} \mathbb{E}^{1/p}((e^{\theta(A(n)-A(m))})^p) \mathbb{E}^{1/q}((e^{-\theta S(m,n)})^q) \\
&\leq e^{-\theta x} \sum_{m=0}^{n} (e^{p\theta(\rho_a(p\theta)+\sigma_a(p\theta))})^{1/p} (e^{q\theta(\rho_S(q\theta)+\sigma_S(q\theta))})^{1/q} \\
&= e^{-\theta x} e^{\theta(\sigma_a(p\theta)+\sigma_S(q\theta))} \sum_{m=0}^{n} e^{\theta(\rho_a(p\theta)+\rho_S(q\theta))}
\end{aligned}$$

Here we see, that the violation probability of exceeding a certain backlog is only valid for a single point in time n. To achieve a finite sample path bound we might use the following simple inequality:

$$\mathbb{P}(\max_{1 \leq n \leq N} q(n) < B) = \mathbb{P}\left(\bigcap_{n=1}^{N} q(n) < B\right) \leq \sum_{n=1}^{N} \mathbb{P}(q(n) < B)$$

Here we see, that the violation probability of exceeding a certain backlog is only valid for a single point in time n. To achieve a finite sample path bound we might use the following simple inequality:

$$\mathbb{P}(\max_{1 \leq n \leq N} q(n) < B) = \mathbb{P}\left(\bigcap_{n=1}^{N} q(n) < B\right) \leq \sum_{n=1}^{N} \mathbb{P}(q(n) < B)$$

However by just adding the violation probabilities, we see them (nearly) linearly increasing for growing N. Hence, the violation probabilities grow until they reach the value 1 and are useless henceforth. To achieve a finite sample path bound with violation probability ε, we have to choose the parameter B in such a way that for large intervals of length N the violation probability for the pointwise backlog bound is of order $\frac{\varepsilon}{N}$. Two questions arise at this point. First: how large do we need to choose B, i.e., what is the quality of our bound, for a given violation probability ε and interval length N? Second: Can we do something smarter than just adding the violation probabilities? The next chapter deals with the second question, while the numerical evaluations in chapter 6 give us some insights on the first question.

5 Alternative bound

First, we take a look at a bound, which is valid for finite sample paths "by nature". For this denote by E_N^μ the number of arrivals a_n up to time N exceeding some value μ:

$$E_N^\mu := \sum_{n=1}^{N} \mathbb{1}_{\{a_n > \mu\}} \in \{0, \ldots, N\}$$

The arrivals exceeding μ form a subsequence of $(a_n)_{n\in\mathbb{N}}$, which will be denoted by $(a_{n_i})_{i\in\{0,\ldots,E_N^\mu\}}$.

Theorem 1. *Assume a node with service S and an incoming flow described by $(a_n)_{n\in\mathbb{N}} \in \mathcal{J}$. Then the following finite sample-path backlog bound holds for all $\mu \in [0,\infty)$:*

$$\begin{aligned}
&\mathbb{P}(\max_{1\le n\le N} q(n) > B)\\
&\le 1 - \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m)\\
&\quad \cdot \mathbb{P}\Big(\Big\{\bigcap_{\substack{1\le n\le N\\ 0\le k\le n}} S(k,n) \ge (n-k)\mu\Big\} \cap \Big\{\max_{1\le i\le m} a_{n_i} \le \mu + \frac{B}{m}\Big\}\Big| E_N^\mu = m\Big)
\end{aligned}$$

And if S is stochastically independent of $(a_n)_{n\in\mathbb{N}}\mathcal{J}$ we have:

$$\begin{aligned}
&\mathbb{P}(\max_{1\le n\le N} q(n) > B)\\
&\le 1 - \mathbb{P}\Big(\bigcap_{\substack{1\le n\le N\\ 0\le k\le n}} S(k,n) \ge (n-k)\mu\Big)\\
&\quad \cdot \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m)\mathbb{P}\left(\max_{1\le i\le m} a_{n_i} \le \mu + \frac{B}{m}\middle| E_N^\mu = m\right)
\end{aligned}$$

Proof. Assume for a while that $E_N^\mu = m$ and

$$\max_{1\le i\le m} a_{n_i} \le \mu + \frac{B}{m}$$

and

$$S(k,n) \geq (n-k)\mu \qquad \forall\ 0 \leq k \leq n \leq N$$

holds. Then we can imply for every $n \in \{1,\dots,N\}$:

$$\begin{aligned}
q(n) = A(n) - D(n) &\leq \max_{0\leq k\leq n} \{A(n) - A(k) - S(k,n)\} \\
&= \max_{0\leq k'\leq n} \{A(n) - A(n-k') - S(n-k',n)\} \\
&= \max_{0\leq k'\leq m} \{A(n) - A(n-k') - S(n-k',n)\} \\
&\quad \vee \max_{m+1\leq k'\leq n} \{A(n) - A(n-k') - S(n-k',n)\} \\
&\leq \max_{0\leq k'\leq m} \left\{k'\left(\mu + \frac{B}{m}\right) - k'\mu\right\} \\
&\quad \vee \max_{m+1\leq k'\leq n} \left\{m\left(\mu + \frac{B}{m}\right) + (k'-m)\mu - k'\mu\right\} \\
&= B
\end{aligned}$$

Hence we get by the law of total probability:

$$\begin{aligned}
&\mathbb{P}(\max_{1\leq n\leq N} q(n) > B) \\
&= 1 - \mathbb{P}(\max_{1\leq n\leq N} q(n) \leq B) \\
&= 1 - \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m)\mathbb{P}(\max_{1\leq n\leq N} q(n) \leq B | E_N^\mu = m) \\
&\leq 1 - \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m) \\
&\quad \cdot \mathbb{P}\Bigg(\Bigg\{\bigcap_{\substack{1\leq n\leq N\\0\leq k\leq n}} S(k,n) \geq (n-k)\mu\Bigg\} \cap \left\{\max_{1\leq i\leq m} a_{n_i} \leq \mu + \frac{B}{m}\right\}\Bigg| E_N^\mu = m\Bigg)
\end{aligned}$$

For the case of independence we continue by applying $\mathbb{P}(A \cap B|C) = \mathbb{P}(B|C)\mathbb{P}(A|B \cap C)$

$$\begin{aligned}
&= 1 - \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m)\mathbb{P}(\bigcap_{\substack{1\leq n\leq N\\0\leq k\leq n}} S(k,n) \geq (n-k)\mu | E_N^\mu = m) \\
&\quad \cdot \mathbb{P}\left(\max_{1\leq i\leq m} a_{n_i} \leq \mu + \frac{B}{m}\middle| E_N^\mu = m\right) \\
&= 1 - \mathbb{P}(\bigcap_{\substack{1\leq n\leq N\\0\leq k\leq n}} S(k,n) \geq (n-k)\mu) \\
&\quad \cdot \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m)\mathbb{P}\left(\max_{1\leq i\leq m} a_{n_i} \leq \mu + \frac{B}{m}\middle| E_N^\mu = m\right)
\end{aligned}$$

In this bound the parameter μ is left open as subject to optimization. Note that there are no assumptions about the service or the arrivals having corresponding MGFs or being i.i.d. sequences. Further, we see that the above bound is always smaller 1, as we expect it of a violation probability. For the special case of $S(k,n) = S(n) - S(k)$ the probabilities simplify to:

$$\begin{aligned}&\mathbb{P}(\max_{1\le n\le N} q(n) > B)\\ &\le 1 - \mathbb{P}(\min_{1\le n\le N} s_n \ge \mu)\sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m)\mathbb{P}\left(\max_{1\le i\le m} a_{n_i} \le \mu + \frac{B}{m}\middle| E_N^\mu = m\right)\end{aligned}$$

The above bound relies only on the analysis of an expression of the form:

$$\mathbb{P}(\max_{1\le n\le N} \boldsymbol{x}_n \le \boldsymbol{y})$$

where $(\boldsymbol{x}_n)_{n\in\mathbb{N}} \in \mathcal{J}^d$ is a sequence of d-dimensional random vectors and $y \in \mathbb{R}^d$. Describing this probability is one of the main goals of Extreme Value Theory. The above probability is well studied under different assumptions on $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ (see for example [19, 10, 9]). The following very small selection of results from EVT assumes $d = 1$ and $(x_n)_{n\in\mathbb{N}} \in \mathcal{I}$.

If we denote by F_x the distribution of x_n we have:

$$\mathbb{P}(\max_{1\le n\le N} x_n \le y) = F_x^N(y)$$

For simple distributions F_x we can directly use the result in the previous theorem to compute finite sample-path backlog bounds. However taking the N-th power of F might be computationally very unstable and the question arises if this expression cannot be approximated by some other expression which is easier to calculate. It is clear that in this case, without some kind of scaling, the above probability just converges to either zero or one. Hence we ask for sequences α_N, β_N such that:

$$\mathbb{P}(\max_{1\le n\le N} x_n \le \alpha_N y + \beta_N) \xrightarrow{N\to\infty} G \tag{1}$$

for some non-degenerate distribution G. We present here some results, as they can be found in [19], to address this question.

5.1 A Brief Introduction to EVT

Denote the right endpoint of some distribution F by $x_0 := \sup\{y : F(y) < 1\}$.

Definition 2 (von Mises Function) *We call a distribution F a von Mises function if there exists a $z_0 < x_0$ such that for all $z_0 < x < x_0$ and some $c > 0$ holds*

$$1 - F(x) = c\exp\left(-\int_{z_0}^{x} \frac{1}{f(u)}du\right)$$

where $f(u) > 0$ for all $z_0 < u < x_0$ and absolutely continuous on (z_0, x_0) and $\lim_{u\uparrow x_0} f'(u) = 0$. We call f an auxiliary function.

The notion of von Mises functions is very important, since one can show that every von Mises function, as defined above, converges to the Gumbel distribution in the sense of (1). Another important equivalent definition (under the assumption that F is twice differentiable) is the von Mises condition. For this define the function ϕ by

$$\phi := -\log(-\log(F)).$$

Definition 3 (von Mises Condition) *We say a distribution F fulfills the von Mises condition if:*

$$\begin{aligned} h(x) :&= \left(\frac{1}{\phi'(x)}\right)' \\ &= -\log F(x) + \frac{F(x)F''(x)\log F(x)}{(F'(x))^2} \xrightarrow{x \to x_0} 1 \end{aligned}$$

If some distribution F fulfills the von Mises Condition define $g(x) := \sup_{y \geq x} |h(x)|$ and $f(x) := \frac{1}{\phi'(x)}$.

One can show that the von Mises condition is fulfilled iff F is a twice differentiable von Mises function. We use the above condition, since it is not only sufficient for the convergence of F to the Gumbel distribution in the sense of (1), but also allows us to derive the speed of convergence.

Lemma 5 *Let $(a_n)_{n\in\mathbb{N}} \in \mathcal{I}$ and the corresponding distribution F_a fulfills the von Mises Condition. Then holds for all $N \in \mathbb{N}$ and $x \geq 0$:*

$$\mathbb{P}(\max_{1\leq n\leq N} a_n \leq x\beta_N + \alpha_N) \leq \Lambda(x) - e^{-1}g(\alpha_N)$$

Here $\Lambda(x) = \exp(-e^{-(x)})$ is the Gumbel distribution, $\phi(\alpha_n) := \log n$ and $\beta_n := \frac{F(\alpha_n)}{nF'(\alpha_n)}$.

There exist similar conditions and results for the convergence to the Fréchet or the Weibull distribution (again in the sense of (1)). As example we give here the parallel results for a convergence against the Fréchet distribution.

Definition 4 (von Mises Condition II) *We say a differentiable distribution F fulfills the von Mises condition for some $\alpha > 0$ if:*

$$h(x) := x\phi'(x) - \alpha = \frac{xF'(x)}{F(x)(-\log F(x))} - \alpha \to 0$$

Under the assumption that F is differentiable one can show that this condition is equivalent to

$$\lim_{x\to\infty} \frac{xF'(x)}{1 - F(x))} = \alpha$$

for some $\alpha > 0$. These conditions imply, that the distribution F converges to the Fréchet distribution in the sense of (1).

Lemma 6 *Keeping the notations of Definition 4 let* $(a_n)_{n\in\mathbb{N}} \in \mathcal{I}$ *and* F_a *be its corresponding distribution fulfilling the second von Mises condition. Then holds for all* $N \in \mathbb{N}$ *and* $x \geq 0$*:*

$$\mathbb{P}(\max_{1\leq n\leq N} a_n \leq x\beta_N) \leq \Phi_\alpha(x) + 0.2701 \cdot (\alpha - g(\beta_n))^{-1} g(\beta_n))$$

where $\Phi_\alpha(x) = \exp(-x^{-\alpha})$ *is the Fréchet distribution,* $g(x) = \sup_{y\geq x} |h(y)|$ *and* β_n *is given by* $-\log F(\beta_n) = n^{-1}$*.*

In the following we only need the case of convergence to the Gumbel distribution. However, we wanted to point out that for some distributions one needs to check another von Mises condition and gets a different convergence speed.

The von Mises condition takes a similar role, as the existence of the moment generating function for the MGF-calculus in the previous section. Yet, there exist a lot of distributions, which fulfill the von Mises conditition without having a MGF. Some heavy-tailed examples are the Cauchy distribution, the Fréchet distribution itself and the Pareto distribution which all converge to the Fréchet distribution in the sense of (1). Another difference is that achieving backlog bounds in the way of theorem 1 is not tied to the von Mises condition, but instead to the analysis of:

$$\mathbb{P}\left(\left\{\bigcap_{\substack{1\leq n\leq N\\0\leq k\leq n}} S(k,n) \geq (n-k)\mu\right\} \cap \left\{\max_{1\leq i\leq m} a_{n_i} \leq \mu + \frac{B}{m}\right\} \middle| E_N^\mu = m\right)$$

When analysing whole networks the above service and arrivals can be the result of network operations, as for example when arrivals $(a_n)_{n\in\mathbb{N}}$ at some node are actually the output of another node, with its own service and other arrivals. So, in general it is hard to use theorem 1 directly. To solve this we compare the service or the arrival distribution to other distributions, which we know more about. If, for example, the arrivals $(a_n)_{n\in\mathbb{N}}$ are the output of another node, we reformulate them in terms of the service and the arrivals of this preceding node. This allows us to investigate more complex network scenarios.

5.2 Network Operations

We prove now a series of results which follow this idea and are in their structure similar to the results in subsection 4.2.

Lemma 7 (Output Bound) *Let* S *be the service of some node, serving the arrivals* $(a_n)_{n\in\mathbb{N}} \in \mathcal{J}$ *and denote by* $(d_n)_{n\in\mathbb{N}} \in \mathcal{J}$ *the departures of that node. Then for all* $x \in [0,\infty)$ *and* $\mu \in [0,x]$ *holds:*

$$\begin{aligned}&\mathbb{P}(\max_{1\leq n\leq N} d_n \leq x)\\ \geq\ &\mathbb{P}\left(\left\{\max_{1\leq n\leq N} a_n \leq \frac{x}{N} + \frac{N+1}{N}\mu\right\} \cap \left\{\bigcap_{\substack{1\leq n\leq N\\0\leq k\leq n-1}} S(k,n-1) \geq (n-1-k)\mu\right\}\right)\end{aligned}$$

Proof. By the definition of service we know:

$$\begin{aligned} d_n &= D(n) - D(n-1) \\ &\le D(n) - \min_{0\le k\le n-1}\{A(k) + S(k,n-1)\} \\ &\le \max_{0\le k\le n-1}\{A(n) - A(k) - S(k,n-1)\} \\ &= \max_{0\le k\le n-1}\left\{\sum_{l=k+1}^{n} a_l - S(k,n-1)\right\} \\ &= \max_{0\le k\le n-1}\left\{a_n - S(k,n-1) + \sum_{l=k+1}^{n-1} a_l\right\} \end{aligned}$$

Assume now for a while that

$$a_k \le \frac{x}{N} + \frac{N-1}{N}\mu \quad \forall\ k = 1,\dots,N$$

and

$$S(k,n-1) \ge (n-1-k)\mu \qquad \forall\ 0 \le k < n \le N$$

holds, for some $\mu \in [0,x]$. Then we would have:

$$\begin{aligned} &\max_{\substack{1\le n\le N\\ 0\le k\le n-1}}\left\{a_n - S(k,n-1) + \sum_{l=k+1}^{n-1} a_l\right\} \\ \le &\max_{\substack{1\le n\le N\\ 0\le k\le n-1}}\left\{\frac{x}{N} + \frac{N-1}{N}\mu - (n-1-k)\mu + (n-1-k)\Big(\frac{x}{N} + \frac{N-1}{N}\mu\Big)\right\} \\ = &\max_{\substack{1\le n\le N\\ 0\le k\le n-1}}\left\{\frac{x}{N} + \frac{N-1}{N}\mu + (n-k-1)\Big(\frac{x}{N} + \frac{N-1}{N}\mu - \mu\Big)\right\} \\ = &\frac{x}{N} + \frac{N-1}{N}\mu + (N-1)\Big(\frac{x}{N} + \frac{N-1}{N}\mu - \mu\Big) = x \end{aligned}$$

Hence, we get for all $\mu \in [0,x]$:

$$\begin{aligned} &\mathbb{P}(\max_{1\le n\le N} d_n \le x) \\ \ge &\,\mathbb{P}\left(\bigcap_{n=1}^{N} \max_{0\le m\le n-1}\left\{a_n - S(k,n-1) + \sum_{l=k+1}^{n-1} a_l\right\} \le x\right) \\ \ge &\,\mathbb{P}\left(\left\{\max_{1\le n\le N} a_n \le \frac{x}{N} + \frac{N-1}{N}\mu\right\} \cap \left\{\bigcap_{\substack{1\le n\le N\\ 0\le k\le n-1}} S(k,n-1) \ge (n-1-k)\mu\right\}\right). \end{aligned}$$

The parameter $\mu \in [0,x]$ is subject to optimization and it is easy to check, that there is no gain in letting μ being larger than x. Note that in the special case

$(s_n)_{n\in\mathbb{N}} = c$ we can choose μ optimally by $\mu = x$ and get the (somewhat trivial) bound:

$$\mathbb{P}(\max_{1\le n\le N} d_n \le x) \ge \mathbb{P}(\max_{1\le n\le N} a_n \le x)$$

Lemma 8 (Leftover Service) *Assume again the scenario as presented before lemma 1. It holds for all $x \in [0,\infty)$ and $\mu \in [0,\infty)$:*

$$\begin{aligned} &\mathbb{P}\Big(\bigcap_{\substack{1\le n\le N\\ 0\le k\le n}} \underline{S}(k,n) \ge (n-k)x\Big) \\ &\ge \mathbb{P}\Big(\{\max_{1\le n\le N} \bar{a}_n \le \mu\} \cap \Big\{\bigcap_{\substack{1\le n\le N\\ 0\le k\le n}} S(k,n) \ge (x+\mu)(n-k)\Big\}\Big) \end{aligned}$$

Proof. Let $x \in [0,\infty)$. Assume for a while that

$$\max_{1\le n\le N} \bar{a}_n \le \mu$$

and

$$S(k,n) \ge (x+\mu)(n-k) \qquad \forall\ 0 \le k \le n \le N$$

holds. Then we have for all $0 \le k \le n \le N$:

$$\begin{aligned} \underline{S}(k,n) &= \max\{0, S(k,n) - A(n) + A(k)\} \\ &= \max\{0, S(k,n) - \sum_{l=k+1}^{n} a_l\} \\ &\ge \max\{0, (x+\mu)(n-k) - (n-k)\mu\} = x(n-k) \end{aligned}$$

Hence: The assertion follows then, as in the previous proof.

Again we can consider the special case $(s_n)_{n\in\mathbb{N}} = c$. Then the optimal μ is given by $c - x$ if $x \in [0,c]$, resulting in:

$$\mathbb{P}(\min_{1\le n\le N} \underline{s}_n \ge x) \ge \mathbb{P}(\max_{1\le n\le N} a_n \le c - x)$$

Lemma 9 (Multiplexing) *Let $(a_n^{(i)})_{n\in\mathbb{N}} \in \mathcal{J}$ be two arrivals ($i = 1,2$). Define $a_n := a_n^{(1)} + a_n^{(2)}$ for all $n \in \mathbb{N}$. Then for all $x \in [0,\infty)$ and $\mu \in [0,x]$ holds:*

$$\mathbb{P}(\max_{1\le n\le N} a(n) \le x) \ge \mathbb{P}(\{\max_{1\le n\le N} a_n^{(1)} \le x - \mu)\} \cap \{\max_{1\le n\le N} a_n^{(2)} \le \mu\})$$

The proof is very similar to the arguments in the previous proofs and hence omitted.

We can use these operations to compute backlog bounds at nodes which lie in the middle or at the end of a network path. In the next section, we show how the presented results of EVT and network operations work together, to achieve a finite sample-path backlog bound, which is competitive to the corresponding MGF-calculus bound.

6 Numerical Evaluation

To compare the two methods we investigate the following scenario: We have a constant rate node, which serves a high and a low priority flow, denoted by $(\bar{a}_n)_{n\in\mathbb{N}}$ and $(\underline{a}_n)_{n\in\mathbb{N}}$, respectively. We are interested in the finite sample-path backlog bound for the low priority flow. For the sake of simplicity, we consider the high and low priority flows to be i.i.d. exponentially distributed with parameter λ, i.e.

$$F_{\bar{a}}(x) = F_{\underline{a}}(x) = 1 - e^{-\lambda x} \quad \forall\, x \in [0, \infty)$$

and equal to zero for all $x \in (-\infty, 0)$. The service rate of the node is given by c.

6.1 MGF-Calculus Bound

Denote the leftover service at the node by $(\underline{s}_n)_{n\in\mathbb{N}} \in \mathcal{J}$. First we derive the $(\sigma(\theta), \rho(\theta))$-bound for the arrivals:

$$\sup_{k\geq 0} \mathbb{E}(e^{\theta(\bar{A}(n+k)-\bar{A}(k))}) = \prod_{m=1}^{n} \mathbb{E}(e^{\theta \bar{a}_m}) = \left(\frac{\lambda}{\lambda-\theta}\right)^n = e^{\theta n \rho(\theta)}$$

with $\rho(\theta) := \frac{1}{\theta}\log(\frac{\lambda}{\lambda-\theta})$ and $\theta \in (0, \lambda)$. Hence the high and low priority flows are $(0, \rho(\theta))$-bounded. Using Lemma 1 and the fact that a constant rate node is $(0, c)$-bounded we have for the leftover service

$$(\underline{s}_n)_{n\in\mathbb{N}} \succeq (0, \rho(\theta) + c).$$

Hence we can use lemma 4 to calculate the following finite sample path backlog bound:

$$\mathbb{P}(\max_{1\leq n\leq N} q(n) \geq B) \leq \min_{0\leq\theta<\lambda} \sum_{n=0}^{N} e^{-\theta B} \frac{1 - e^{\theta(n+1)(2\rho(\theta)+c)}}{1 - e^{\theta(2\rho(\theta)+c)}}$$

$$= \min_{0\leq\theta<\lambda} \sum_{n=0}^{N} e^{-\theta B} \frac{1 - \left(\frac{\lambda}{\lambda-\theta}\right)^{2(n+1)} e^{-\theta(n+1)c}}{1 - \left(\frac{\lambda}{\lambda-\theta}\right)^{2} e^{-\theta c}}$$

To compute a competitive backlog bound we optimize the parameter θ numerically.

6.2 Alternative Bound

We keep the previous notations and begin as in the proof of theorem 1. Now denote by E_N^{μ} the number of low priority arrivals, which exceed the value μ and denote these arrivals by the subsequence $(\underline{a}_{n_i})_{i\in\{0,\ldots,E_N^{\mu}\}}$. We then have:

$$\mathbb{P}(\max_{1\leq n\leq N} q(n) > B)$$

$$\leq 1 - \mathbb{P}(\min_{1\leq n\leq N} \underline{s}_n \geq \mu) \sum_{m=0}^{N} \mathbb{P}(E_N^{\mu} = m)\mathbb{P}\left(\max_{1\leq n\leq m} \underline{a}_j \leq \mu + \frac{B}{m} \middle| E_N^{\mu} = m\right)$$

and with Lemma 8

$$
\begin{aligned}
\leq & 1 - \mathbb{P}(\max_{1\leq n\leq N} \bar{a}_n \leq c-\mu) \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m) \mathbb{P}\left(\max_{1\leq n\leq m} \underline{a}_j \leq \mu + \frac{B}{m} \middle| E_N^\mu = m \right) \\
\leq & 1 - \mathbb{P}(\max_{1\leq n\leq N} \bar{a}_n \leq c-\mu) \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m) \mathbb{P}\left(\max_{1\leq n\leq m} \underline{a}_j \leq \frac{B}{m} \right)
\end{aligned}
$$

In the last step we have used the memoryless-property of the exponential distribution and that the arrivals are i.i.d.

Due to the simple nature of the arrivals we have the choice to use the EVT-approximation or directly compute the above expression by using $\mathbb{P}(\max_{1\leq n\leq N} a_n \leq x) = F_a^N(x)$. We perform both in order to test the quality of the EVT approximation. To use the von Mises condition one can easily verify that the exponential distribution with parameter λ fulfills the conditions of Lemma 5 with the norming sequences

$$
\alpha_n = -\frac{\log(1-e^{-1/n})}{\lambda}
$$

and

$$
\beta_n = \frac{1}{n\lambda(e^{1/n}-1)}
$$

and g given by:

$$
g(x) = -\frac{\log(1-e^{-\lambda x})}{e^{-\lambda x}} - 1
$$

Inserting this into our backlog bound yields:

$$
\begin{aligned}
& \mathbb{P}(\max_{1\leq n\leq N} q(n) > B) \\
\leq & \; 1 - \left(\exp(-e^{-\gamma_N(\lambda(c-\mu)+\log(1-e^{-1/N}))}) - \tilde{g}(N) \right) \\
& \cdot \sum_{m=0}^{N} \mathbb{P}(E_N^\mu = m) \left(\exp(-e^{-\gamma_m(\lambda\frac{B}{m}+\log(1-e^{-1/m}))}) - \tilde{g}(m) \right)
\end{aligned}
$$

where

$$
\tilde{g}(n) := \frac{1}{e \cdot n(1-e^{-1/n})}
$$

and

$$
\gamma_n := n(e^{1/n}-1).
$$

Similar to the MGF-bound we optimize a parameter - in this case $\mu \in [0, c]$ - numerically to achieve a competitive backlog bound.

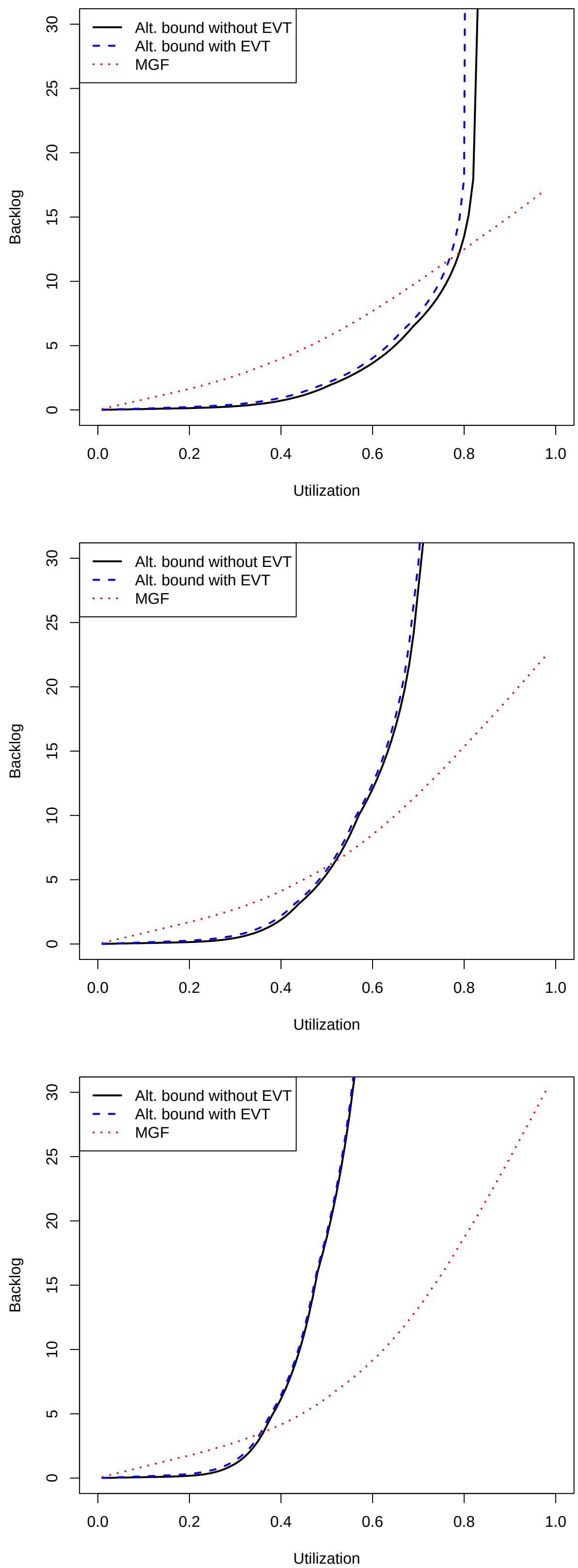

Fig. 1. From top $N = 10$, $N = 20$, $N = 40$. $\varepsilon = 10^{-6}$.

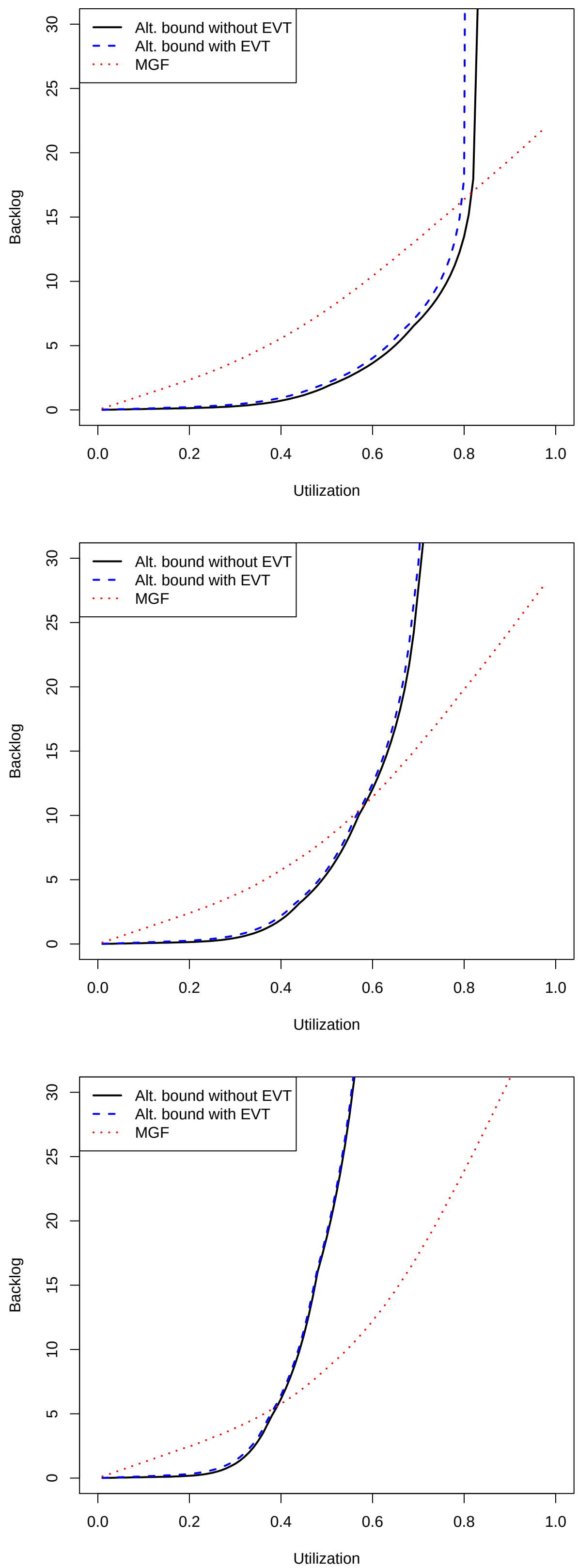

Fig. 2. From top $N = 10$, $N = 20$, $N = 40$. $\varepsilon = 10^{-9}$.

Results To present the results we choose $c = 1$ and investigate different utilizations of the node. The utilization of the node is given by $u = \frac{2}{\lambda}$. In our experiments we ask for the smallest B we can choose, such that we do not exceed a certain violation probability. This violation probability is set to 10^{-6} and 10^{-9} in the experiments. Of course the results are dependent on the considered sample path length N. To find reasonable values of N we simulated the queuing system and observed the duration of the backlogged periods. The startpoint of a backlogged period is defined as the timestep, in which the node starts to accumulate backlog and the endpoint is defined as the next timestep thereafter, in which no backlog occurs any more at the node. In the simulations, we observed 100,000 backlogged periods under different utilizations. For example, for a utilization of 80%. we obtained an average period-length of 3.2 and 99.9% of the periods had a length smaller than 37. For this reason we considered for our scenario sample path lengths of $N = 10$, $N = 20$ and $N = 40$.

The results for B under different utilizations are displayed in Figures 1 and 2. We can make different observations from the graphs. First we can compare the alternative bound with and without EVT-approximation, which are shown in the graph by the blue dashed line and the solid black line, respectively. We see that the approximation is very close and we do not loose much by it. This gives us hope that the approximation is also a good choice for more complex arrivals, in which we cannot use a direct computation. We also see that the alternative method outperforms the MGF-method, given by the dotted red line, in the region of lower utilizations. However the alternative method has for large N some tipping point after which, only by an immense increment of B the wished violation probabilities can be achieved. Comparing the three methods under increasing N the MGF-method loses the least. All three methods are quite robust against the transition from a violation probability of 10^{-6} to 10^{-9}, however the MGF-method seems to loose a bit more here.

7 Two Node Scenario

In this example we show how the results of chapter 4.2 and chapter 5 work together to achieve a backlog bound in more complex networks. The considered example is similar to the just analysed one, but instead of traversing a single node, we now have to cross two nodes. Both nodes are constant rate servers and the priorities of the flows are preserved in the transition from the first to the second node. For this scenario we denote the intermediate flows by $(\bar{i}_n)_{n\in\mathbb{N}} \in \mathcal{J}$ and $(\underline{i}_n)_{n\in\mathbb{N}} \in \mathcal{J}$.

7.1 MGF Bound

We start by computing bounds for the intermediate flows. Using our previous results we obtain

$$(\bar{i}_n)_{n\in\mathbb{N}} \preceq (\bar{\sigma}(\theta), \rho(\theta))$$

with $\bar{\sigma}(\theta) = \frac{1}{\theta}\log(1 - e^{\theta(\rho(\theta)+c_1)})^{-1}$ and

$$(\underline{i}_n)_{n\in\mathbb{N}} \preceq (\underline{\sigma}(\theta), \rho(\theta))$$

with $\underline{\sigma}(\theta) = \frac{1}{\theta}\log(1-e^{\theta(2\rho(\theta)+c_1)})^{-1}$. This leads to a leftover service at the second node $(\underline{s}_n)_{n\in\mathbb{N}} \in \mathcal{J}$ with

$$(\underline{s}_n)_{n\in\mathbb{N}} \succeq (\bar{\sigma}(\theta), \rho(\theta) + c_2)$$

We can now compute the backlog bound at the second node, but have to watch out for a stochastic dependency between the leftover service at the second node and the intermediate low priority arrivals. This dependency results from the fact that after the first node the two intermediate arrivals are dependent, which in turn makes the leftover service (which is a function of the high priority intermediate arrivals) stochastically dependent:

$$\begin{aligned}
&\mathbb{P}(\max_{1\leq n\leq N} q(n) > B) \\
&\leq \sum_{n=1}^{N} e^{-\theta B} e^{\theta(\underline{\sigma}(q\theta)+\bar{\sigma}(p\theta))} \sum_{m=0}^{n} e^{m\theta(\rho(q\theta)+\rho(p\theta))}
\end{aligned}$$

By the dependence of the two intermediate flows, we now have a second parameter p, next to θ, which we need to optimize. In more complex scenarios a large set of these parameters can occur. In practice this means that often the parameters need to be set to certain values, to keep the formulas tractable (in our example a convenient choice of p would be 2). This leads to looser bounds.

7.2 Alternative Bound

For the EVT-bound we also have to consider the dependencies, but there is a way to get rid of them. However, we have to pay this way by a much worse bound. Denote by $(t_n)_{n\in\mathbb{N}} \in \mathcal{J}$ the service at the second node and by $(\underline{t}_n)_{n\in\mathbb{N}} \in \mathcal{J}$ the leftover service at the second node. We start similar as in the case of one node, but we cannot use the law of total probability.

$$
\begin{aligned}
&\mathbb{P}(\max_{1\le n\le N} q(n) > B) \\
&\le 1-\mathbb{P}\left(\left\{\min_{1\le n\le N} \underline{t}_n \ge \mu\right\}\cap\left\{\max_{1\le n\le N} \underline{i}_n \le \mu+\frac{B}{N}\right\}\right) \\
&\le 1-\mathbb{P}\left(\left\{\max_{1\le n\le N} \bar{i}_n \ge c_2-\mu\right\}\cap\left\{\max_{1\le n\le N} \underline{i}_n \le \mu+\frac{B}{N}\right\}\right) \\
&\le 1-\mathbb{P}\left(\left\{\max_{1\le n\le N} \bar{a}_n \le \frac{c_2-\mu}{N}+\frac{N-1}{N}\mu'\right\}\right. \\
&\quad\cap\left\{\max_{1\le n\le N} \underline{a}_n \le \frac{\mu+\frac{B}{N}}{N}+\frac{N-1}{N}\mu''\right\} \\
&\quad\left.\cap\left\{\min_{1\le n\le N} s_n \ge \mu'\right\}\cap\left\{\min_{1\le n\le N} \underline{s}_n \ge \mu''\right\}\right) \\
&\le 1-\mathbb{P}\left(\left\{\max_{1\le n\le N} \bar{a}_n \ge \frac{c_2-\mu}{N}+\frac{N-1}{N}c_1\right\}\right. \\
&\quad\left.\cap\left\{\max_{1\le n\le N} \underline{a}_n \le \frac{\mu+\frac{B}{N}}{N}+\frac{N-1}{N}\mu''\right\}\cap\left\{\max_{1\le n\le N} \underline{a}_n \ge c_1-\mu''\right\}\right) \\
&\le 1-\mathbb{P}\left(\left\{\max_{1\le n\le N} \bar{a}_n \ge \frac{c_2-\mu}{N}+\frac{N-1}{N}c_1\right\}\right. \\
&\quad\left.\cap\left\{\max_{1\le n\le N} \underline{a}_n \le \frac{\mu+\frac{B}{N}}{N}+\frac{N-1}{N}\mu''\wedge c_1-\mu''\right\}\right)
\end{aligned}
$$

with $\mu \in [0, c_2]$ and $\mu'' \in [0, \frac{B}{N}+\mu]$. The optimal μ'' can be found by setting

$$\frac{\frac{B}{N}+\mu}{N}+\frac{N-1}{N}\mu'' = c_1-\mu''$$

Using the independence of $(\underline{a}_n)_{n\in\mathbb{N}}$ and $(\bar{a}_n)_{n\in\mathbb{N}}$, we eventually get the backlog bound:

$$
\begin{aligned}
\mathbb{P}(\max_{1\le n\le N} q(n) > B) \le \mathbb{P}&\left(\max_{1\le n\le N} \bar{a}_n \le \frac{c_2-\mu}{N}+\frac{N-1}{N}c_1\right) \\
\cdot\,\mathbb{P}&\left(\max_{1\le n\le N} \underline{a}_n \le \frac{(N-1)c_1+\mu+\frac{B}{N}}{2N-1}\right)
\end{aligned}
$$

8 Conclusion and Outlook

In this paper, we have dealt with the practically important issue of sample path backlog bounds and have compared two methods to achieve such backlog bounds. The first is derived directly from the MGF-calculus, which cannot be optimal, since the violation probabilities are simply added, leading to a linear growth, which eventually exceeds 1. The second is a new method, which asks directly for finite sample path backlog bounds and is based on extreme value theory results. We have shown how to extend this new bound to an alternative SNC, which can be applied to more complex networks. Comparing the two methods in a simple example shows no clear winner: while the EVT-bound has trouble with high utilizatione it outperforms the MGF-bound for smaller utilizations. Nevertheless, we see by this that the new method provides an alternative, which needs to be considered, to achieve low sample path backlog bounds. Besides this the new bound has same interesting conceptual properties. First it does not rely on the existence of an MGF. Hence by this method we can tackle also heavy-tailed distributions and to some extent solve dependent cases. Fully exploring and exploiting these conceptual strengths will be subject to future work. In general, we also believe that our new method is supported by a versatile tool: EVT. With its help computationally problemtic expressions can be approximated. For future work the results of EVT can be mined to include a broader class of sequences, such as non-i.i.d. arrivals or stochastically dependent arrival flows. Further directions to which this theory can be extended include concatenation results and sample-path delay bounds.

Bibliography

[1] G. Appenzeller, I. Keslassy, and N. McKeown. Sizing router buffers. In *ACM SIGCOMM*, 2004.

[2] M. A. Beck and J. B. Schmitt. On the calculation of sample-path backlog bounds in queueing systems over finite time horizons. Technical report, University of Kaiserslautern, Distributed Computer Systems Lab (DISCO), May 2012.

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

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

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

[6] C.-S. Chang. On the performance of multiplexing independent regulated inputs. In *ACM SIGMETRICS*, 2001.

[7] F. Ciucu. Exponential supermartingales for evaluating end-to-end backlog bounds. *SIGMETRICS Perform. Eval. Rev.*, 35(2):21–23, 2007.

[8] F. Ciucu. *Scaling Properties in the Stochastic Network Calculus*. PhD thesis, University of Virginia, 2007.

[9] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

[10] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.

[11] M. Fidler. An end-to-end probabilistic network calculus with moment generating functions. In *Quality of Service, 2006. IWQoS 2006. 14th IEEE International Workshop on*, pages 261 –270, 2006.

[12] M. Fidler. Survey of deterministic and stochastic service curve models in the network calculus. *Communications Surveys Tutorials, IEEE*, 12(1):59–86, 2010.

[13] J. Gettys and K. Nichols. Bufferbloat: Dark buffers in the internet. *Communications of the ACM*, 55(1):57–65, 2012.

[14] Y. Jiang and Y. Liu. *Stochastic Network Calculus*. Springer, 2008.

[15] L. Kleinrock. *Queueing Systems*. Wiley, 1975.

[16] L. Massoulié and A. Busson. Stochastic majorization of aggregates of leaky-bucket constrained traffic streams. Preprint, Microsoft Research, Cambridge, 2000.

[17] O. Ozturk, R. R. Mazumdar, and N. Likhanov. Large buffer asymptotics for fluid queues with heterogeneous $m/g/\infty$ weibullian inputs. *QUESTA*, 45:333–356, 2003.

[18] V. V. Petrov. *Sums of Independent Random Variables*. Springer, 1975.

[19] S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, 1987.

[20] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis: Queues, Communications and Computing*. Chapman and Hall, 1995.

[21] M. Vojnovic and J.-Y. Le Boudec. Bounds for independent regulated inputs multiplexed in a service curve network element. *IEEE Trans. on Communications*, 51(5):735–740, 2003.

[22] Y. Ying, F. Guillemin, R. Mazumdar, and C. Rosenberg. Buffer overflow asymptotics for multiplexed regulated traffic. *Performance Evaluation*, 65(8):555–572, 2008.