A First Course in Stochastic Network Calculus

Dipl.-Math. Michael Alexander Beck

January 9, 2014

Contents

Foreword

This *First Course in Stochastic Network Calculus* is a "First" course in two different perspectives. One can see it as a introductory course to Stochastic Network Calculus. On the other side it was my first course I have given about SNC to a few students in June 2012. It builds on the lecture *Performance Modelling in Distributed Systems [1]* at the University of Kaiserslautern. The concepts of stochastic network calclulus parallels those of deterministic network calculus. This is why I reference on the lecture of 2011 at several points to stress these connections. This document however is thought of a stand-alone course and hence a deep study of [1] is not necessary (but recommended).

This course contains a rather large probability primer to ensure the students can really grasp the expressions, which appear in stochastic network calculus. A student familiar with probability theory might skip this first chapter and delve directly into the stochastic network calculus. For each topic exercises are given, which can (and should) be used to strengthen the understanding of the presented definitions and theory.

This document is still in process and hopefully will evolve at some day into a fully grown course about Stochastic Network Calculus, providing a good overview over this exciting theory. Hence, please provide feedback to `beck@cs.uni-kl.de`.

- Michael Beck

1 Basics of Probability

2 Stochastic Arrivals and Service

3 Stochastic Performance Bounds

Now, based on the models for stochastic arrivals and service guarantees, we ask for performance guarantees in this chapter. We are interested in bounds on the backlog and the delay of a node, as well as in bounds on the departures of a node. Again, the theory here parallels the deterministic approach, with the difference, that the bounds only hold with a high probability. Or, stated more positively: The achieved bounds are only violated with very small probabilities.

If not stated otherwise we assume in this chapter a flow A, which is bounded by $(\sigma_A(\theta), \rho_A(\theta))$ and a stochastically independent node S, which is $(\sigma_S(\theta), \rho_S(\theta))$-bounded for the same $\theta > 0$. Further, for the tail-bounded case, we assume A and S to be tail-bounded by envelopes α_A and α_S with errors η_A and η_S, respectively. The flow A is the input for node S and the corresponding output flow is denoted by D.

3.1 Backlog Bound

Before we can formulate the backlog bound we need another theorem concerning the convolution of MGFs. We will use the following definitions:

Definition 3.1. Let $f, g : \Lambda(\mathbb{N}_0) \to \mathbb{R}$ be two triangular arrays. The *bivariate convolution* $* : \mathcal{D} \times \mathcal{D} \to \mathcal{D}$ of f and g at $(m, n) \in \Lambda(\mathbb{N}_0)$ is defined by:

$$f * g(m,n) = \sum_{k=m}^{n} f(m,k)g(k,n)$$

The *bivariate deconvolution* $\circ : \mathcal{D} \times \mathcal{D} \to \mathcal{D}$ of f and g at $(m, n) \in \Lambda(\mathbb{N}_0)$ is defined by:

$$f \circ g(m,n) = \sum_{k=0}^{m} f(k,n)g(k,m)$$

Theorem 3.2. *Let $X, Y \in \mathcal{S}$ be two stochastically independent random processes. Then it holds that:*

$$\phi_{X \otimes Y(m,n)}(-\theta) \leq \left(\phi_X(-\theta) * \phi_Y(-\theta)\right)(m,n)$$

for all $\theta > 0$ and $(m, n) \in \Lambda(\mathbb{N}_0)$ such that the above MGFs exist.

Further it holds that:

$$\phi_{X \oslash Y(m,n)}(\theta) \leq \left(\phi_X(\theta) \circ \phi_Y(-\theta)\right)(m,n)$$

for all $\theta > 0$ and $(m, n) \in \Lambda(\mathbb{N}_0)$ such that the above MGFs exist.

Proof. We show the first inequality, since the second can be proven similarly:

$$\begin{aligned}\phi_{X\otimes Y}(-\theta) &= \mathbb{E}(e^{-\theta \min_{m\le k\le n}\{X(m,k)+Y(k,n)\}}) = \mathbb{E}(e^{\max_{m\le k\le n}\{-\theta X(m,k)-\theta Y(k,n)\}}) \\ &= \mathbb{E}(\max_{m\le k\le n}\{e^{-\theta X(m,k)}\cdot e^{-\theta Y(k,n)}\}) \le \sum_{k=m}^{n}\mathbb{E}(e^{-\theta X(m,k)}\cdot e^{-\theta Y(k,n)}\}) \\ &= \sum_{k=m}^{n}\phi_{X(m,k)}(-\theta)\phi_{Y(k,n)}(-\theta) = (\phi_X(-\theta)*\phi_Y(-\theta))\,(m,n)\end{aligned}$$

For all $\theta > 0$ such that the above MGFs exist. □

We present the backlog bound for the MGF-case first and briefly discuss it. Thereafter, we present the corresponding bound for the tail-bound-case.

Theorem 3.3. *For the backlog $q(n) := A(n) - D(n)$ at time n and all $x \in \mathbb{R}_0^+$ it holds that:*

$$\mathbb{P}(q(n) > x) \le e^{-\theta x+\theta(\sigma_A(\theta)+\sigma_S(\theta))}\sum_{k=0}^{n} e^{\theta k(\rho_A(\theta)+\rho_S(\theta))}.$$

Proof. We know by the previous chapter:

$$q(n) \le A(n) - A\otimes S(0,n) = \max_{0\le k\le n}\{A(k,n) - S(k,n)\} = A \oslash S(n,n)$$

Hence from $q(n) > x$ follows $A \oslash S(n,n) > x$ and using Chernoff's inequality yields:

$$\begin{aligned}\mathbb{P}(q(n) > x) &\le \mathbb{P}(A\oslash S(n,n) > x) \le e^{-\theta x}\mathbb{E}(e^{\theta A\oslash S(n,n)}) \\ &\le e^{-\theta x}\mathbb{E}(e^{\theta A})\circ\mathbb{E}(e^{-\theta S})(n,n) \\ &= e^{-\theta x}\sum_{k=0}^{n}\mathbb{E}(e^{\theta A(k,n)})\mathbb{E}(e^{-\theta S(k,n)}) \\ &\le e^{-\theta x}\sum_{k=0}^{n} e^{\theta(\sigma_A(\theta)+(n-k)\rho_A(\theta))}e^{\theta(\sigma_S(\theta)+(n-k)\rho_S(\theta))} \\ &\le e^{-\theta x+\theta(\sigma_A(\theta)+\sigma_S(\theta))}\sum_{k'=0}^{n} e^{\theta k'(\rho_A(\theta)+\rho_S(\theta))}\end{aligned}$$

□

There is another way to derive the above bound. Instead of using first Chernoff's inequality and the inqeuality $a + b \ge a \vee b$ thereafter, we can interchange their order:

$$\begin{aligned}\mathbb{P}(\max_{0\le k\le n}\{A(k,n) - S(k,n) > x\} &= \mathbb{P}\left(\bigcup_{k=0}^{n} A(k,n) - S(k,n) > x\right) \\ &\le \sum_{k=0}^{n}\mathbb{P}(A(k,n) - S(k,n) > x) \\ &\le \sum_{k=0}^{n} e^{-\theta x}\mathbb{E}(e^{\theta(A(k,n)-S(k,n)})\end{aligned}$$

and then proceed as above. The result is the same, even if A and S are not stochastically independent (see the exercises). However one should keep these two different ways in mind, since both of them may contain possible improvements on the bound.

Before we continue, we analyse the structure of the above proof. It consists of the following steps:

1. We expressed the quantity $q(n)$ in terms of A and S.
2. We translate the achieved inequality into an inequality of probabilities.
3. We use Chernoff to bound the new probability and moment generating functions arise.
4. We use the MGF-bounds of A and S.

We will encounter this structure again in the next two sections.

As a last note we discuss θ. A little assumption was made, by assuming A and S can be bounded with the same parameter θ. This assumption is not very restrictive, since we normally can bound A by $(\sigma_A(\theta'), \rho_A(\theta'))$ for all $0 < \theta' < \theta$ if it is already bounded by $(\sigma_A(\theta), \rho_A(\theta))$. The same holds for S. Hence if the θ in the two bounds differ, we can decrease one of them to the minimal θ.

Nevertheless the parameter θ can and should be optimized in the above bound:

Corollary 3.4. *Assume A is $(\sigma_A(\theta), \rho_A(\theta))$-bounded for all $\theta \in [0, b]$ and S is $(\sigma_S(\theta), \rho_S(\theta))$-bounded for all $\theta \in [0, b]$ and some $b \in \mathbb{R}^+ \cup \{\infty\}$. Then*[1]*:*

$$\mathbb{P}(q(n) > x) \leq \inf_{\theta \in [0,b]} e^{-\theta x + \theta(\sigma_A(\theta) + \sigma_S(\theta))} \sum_{k'=0}^{n} e^{\theta k'(\rho_A(\theta) + \rho_S(\theta))}$$

We proceed with the tail-bounded version of a backlog bound.

Theorem 3.5. *For all $n \in \mathbb{N}_0$ and $\varepsilon > 0$ it holds that:*

$$\mathbb{P}(q(n) > (\alpha_A(\varepsilon) + \alpha_S(\varepsilon)) \oslash S(n, n)) \leq \eta_S(n, \varepsilon) + \sum_{m=0}^{n} \eta_A(m, \varepsilon)$$

Proof. Let $n \in \mathbb{N}_0$ and $\varepsilon > 0$ be arbitrary. Assume for a while that

$$A(m, n) \leq \alpha_A(n - m, \varepsilon) \qquad \forall\, m \leq n \tag{3.1}$$

and

$$D(0, n) \geq A \otimes (S - \alpha_S(\varepsilon))(0, n) \tag{3.2}$$

would hold. We would have then:

$$\begin{aligned} q(n) = A(0, n) - D(0, n) &\leq \max_{0 \leq k \leq n} \{A(n) - A(k) - S(k, n) + \alpha_S(n - k, \varepsilon)\} \\ &\leq \max_{0 \leq k \leq n} \{\alpha_S(n - k, \varepsilon) + \alpha_A(n - k, \varepsilon) - S(k, n)\} \\ &= (\alpha_A(\varepsilon) + \alpha_S(\varepsilon)) \oslash S(n, n) \end{aligned}$$

[1] We investigate this parameter θ, which we have already called *acuity*, in the exercises.

Since we have derived the above inequality from (3.1) and (3.2), we have:

$$\begin{aligned}
&\mathbb{P}(q(n) > (\alpha_A(\varepsilon) + \alpha_S(\varepsilon)) \oslash S(n,n)) \\
\leq &\mathbb{P}\left(D(0,n) \geq A \otimes (S - \alpha_S(\varepsilon))(0,n) \cup \bigcup_{m=0}^{n} A(m,n) > \alpha_A(n-m,\varepsilon)\right) \\
\leq &\mathbb{P}(D(0,n) \geq A \otimes (S - \alpha_S(\varepsilon))(0,n)) + \sum_{m=0}^{n} \mathbb{P}(A(m,n) > \alpha_A(n-m,\varepsilon)) \\
\leq &\eta_S(n,\varepsilon) + \sum_{m=0}^{n} \eta_A(m,\varepsilon)
\end{aligned}$$

□

Let us have a closer look at the above bound: the crucial point here is that - compared to the MGF-backlog-bound - we do not have a bound for the event $\{q(n) > x\}$ with some x chosen by ourselves. Instead $q(n)$ is compared with the rather involved expression $\max_{0\leq k\leq n}\{\alpha_S(n-k,\varepsilon) + \alpha_A(n-k,\varepsilon) - S(k,n)\}$. This, in fact, raises some challenges, since we need to solve:

$$\max_{0\leq k\leq n}\{\alpha_S(n-k,\varepsilon) + \alpha_A(n-k,\varepsilon) - S(k,n)\} = x$$

In general this equation might or might not have a unique solution for ε. However, this alters drastically, if we state the more general

Corollary 3.6. *For all $n \in \mathbb{N}_0$ and $\varepsilon, \varepsilon' > 0$ holds:*

$$\mathbb{P}(q(n) > (\alpha_A(\varepsilon) + \alpha_S(\varepsilon')) \oslash S(n,n)) \leq \eta_S(n,\varepsilon') + \sum_{m=0}^{n} \eta_A(m,\varepsilon)$$

Now the solution to

$$\max_{0\leq k\leq n}\{\alpha_S(n-k,\varepsilon') + \alpha_A(n-k,\varepsilon) - S(k,n)\} = x$$

does not neet to be unique and we are confronted with an optimization problem: How to choose $\varepsilon, \varepsilon'$ such that above equality holds and $\eta_S(n,\varepsilon') + \sum_{k=0}^{n} \eta_A(k,\varepsilon)$ becomes minimal?

Another important note on the tail-bound is, that it works without the assumption of stochastic independence between A and S. This is a great advantage of the tail-bounds.

Exercises

Exercise 3.7. Prove the second part of 3.2!

Exercise 3.8. Use Hölder's inequality to derive a backlog bound for the case that the flow A and the service S are not stochastically independent.

Exercise 3.9. In this exercise we derive a backlog bound for an arrival flow, which is heavy-tailed, i.e. $\phi_{A(m,n)}(\theta)$ does not exist for any choice of $m < n \in \mathbb{N}_0$ and $\theta \in \mathbb{R}^+$. To achieve this we need a determinstic lower bound for the dynamic server S:

$$S(m,n) \geq f(m,n)$$

Use the alternative proof of 3.3 and Markov's inequality instead of Chernoff's inequality, to show:

$$\mathbb{P}(q(n) > x) \leq \sum_{k=0}^{n} \left(\frac{1}{x + f(k,n)} \right)^a \mathbb{E}(A(k,n)^a) \qquad \forall\, a \in \mathbb{N}$$

Use this bound to compute the probability that a node with service $S(m,n) = (n-m)c$ serving an arrival with Pareto distributed increments ($x_{\text{min}} = 0.5$, $\alpha = 3$) has a backlog higher 1 at time n. To achieve the best possible bound optimize the parameter $a \in \{1,2,3\}$.

Exercise 3.10. We want to investigate how a system with increasing number of flows scales. Assume the following scenario: We have one node with a constant service rate $c = 1$ (i.e. $S(m,n) = c(n-m)$ for all $m \leq n \in \mathbb{N}_0$) and N stochastically independent flows arrive at this node. All of the flows share the same distribution for their increments a_i ($i = 1, \ldots, N$). All increments are i.i.d., we consider three different cases of what this distribution looks like:

- $a_i(m) = 1$ with probability $\frac{1}{N}$ and $a_i(m) = 0$ with probability $1 - \frac{1}{N}$
- $a_i(m) = \frac{2}{N}$ with probability $\frac{1}{2}$ and $a_i(m) = 0$ with probability $a_i(m) = \frac{1}{2}$
- $a_i(m) = N$ with probability $\frac{1}{N^2}$ and $a_i(m) = 0$ with probability $1 - \frac{1}{N^2}$

Show that the expected number of arrivals in one time step is in each case equal to 1 (Hence the node has an utilization of 100%).

Give for each case the corresponding $f(\theta, n)$-bound (compare example **??**) and calculate the backlog bound at time $n = 1$ for a fixed N. How do these bounds evolve for large N? What is

$$\lim_{N \to \infty} \mathbb{P}(q(1) > 2)$$

for each case? Why do they behave differently? Analyse the variance of the increments!

(Hints: For the first case you need that $(1 + \frac{t}{N})^N \xrightarrow{N \to \infty} e^t$, for the second case that $(\frac{1}{2}(e^{\frac{2\theta}{N}} + 1))^N \xrightarrow{N \to \infty} e^{\theta}$. The variance of a binomial distribution with parameters N and p is given by $\text{Var}(B) = Np - Np^2$)

Exercise 3.11. Remember exercise **??** in which we have investigated the parameter θ and called it *acuity*. We now check what happens for the backlog bound, when the acuity is altered. For this consider a node, which offers a constant rate service c and an arrival with i.i.d. exponentially distributed increments with parameter $\lambda = 10$. Compute the backlog bound for x at time $n = 1$ and give it in the form

$$\mathbb{P}(q(1) > x) \leq h(x,\theta)(1 + f(\lambda,\theta)g(c,\theta))$$

Here the three functions f, g and h are only dependent on θ and one further parameter. We can identify f and g as the parts of the bound, which result from our arrival bound and the service bound, respectively. Plot for varying $\theta \in [0, 10)$ the functions f, g and the above backlog bound for $x = 1$.

Bibliography

[1] *J. B. Schmitt. Lecture notes: Performance modeling of distributed systems. http://disco.informatik.uni-kl.de/content/PDS_ WS1112, October 2011.*