Dealing with Dependence in Stochastic Network Calculus – Using Independence as a Bound

Paul Nikolaus[1], Jens Schmitt[1], and Florin Ciucu[2]

[1] Distributed Computer Systems (DISCO) Lab,
TU Kaiserslautern,
{nikolaus,jschmitt}@cs.uni-kl.de
[2] Computer Science Department,
University of Warwick,
f.ciucu@warwick.ac.uk

Abstract. Computing probabilistic end-to-end delay bounds is an old, yet still challenging problem. Stochastic network calculus enables closed-form delay bounds for a large class of arrival processes. However, it encounters difficulties in dealing with dependent flows, as standard techniques require to apply Hölder's inequality. In this paper, we present an alternative bounding technique that, under specific conditions, treats them as if flows were independent. We show in two case studies that it often provides better delay bounds while simultaneously significantly improving the computation time.

1 Introduction

Stochastic network calculus (SNC) is a versatile framework to compute stochastic per-flow delay bounds. Developed as a deterministic worst-case analysis in the 1990s by Cruz [6, 7], stochastic extensions of network calculus emerged quickly thereafter. It allows for closed-form solutions for a broad class of arrival and service processes. In [18], it has been shown that the SNC branch using moment-generating functions [4, 11] provides tighter bounds than the approach using envelope functions [5, 8, 12], as it leverages the independence of arrival flows. However, many results limit the end-to-end analysis to pure tandem topologies.

Analyzing more general networks requires to consider also dependent flows at some points in the network, as the sharing of a resource clearly has a mutual impact on the flows' output behavior. Therefore, if we want to obtain the moment-generating function (MGF) of aggregated, yet dependent arrival processes $A_1(s,t)$ and $A_2(s,t)$, we typically invoke Hölder's inequality:

$$\mathrm{E}\left[e^{\theta(A_1(s,t)+A_2(s,t))}\right] \leq \mathrm{E}\left[e^{p\theta A_1(s,t)}\right]^{1/p} \cdot \mathrm{E}\left[e^{q\theta A_2(s,t)}\right]^{1/q},$$

where $0 \leq s \leq t$, $\theta > 0$, and $\frac{1}{p} + \frac{1}{q} = 1$, $p, q \in [1, \infty]$. Hölder's inequality is completely oblivious of the actual dependence structure, thus it often leads to very conservative bounds. Furthermore, it places the burden of an additional, nonlinear parameter for each application to optimize.

Dependence of arrivals does not have to be a negative property per se. Taking advantage of the information about the dependence structure to improve upon the bounds has been attempted before. In [9], the functional dependence is estimated using a copula-based approach. In our work, we investigate a simpler alternative, using the independent scenario as an upper bound. To that end, we rely on a characteristic called negative dependence. We explain the main idea with the help of the following, simplistic example.

Consider a single time slot assuming two arrival processes, A_1 and A_2, that are multiplexed at one server. Both arrivals send one packet, each independently with probability $p \in (0,1)$, and the server serves one packet but strictly prioritizes A_2. Clearly, their two outputs, D_1 and D_2, are strongly dependent, as an arrival of the prioritized flow forces the other one to wait in the queue. Simply put, if one flow get a larger share of the server's capacity, the other is more likely to have less output. For the joint distribution of the output, we have by assumption for the departures both being equal to 1, that $\mathrm{P}(D_1 = 1, D_2 = 1) = 0$. On the other hand, we compute for the product distribution by a simple conditioning, that $\mathrm{P}(D_1 = 1) \cdot \mathrm{P}(D_2 = 1) = (p \cdot (1-p)) \cdot (1-p) > 0$. Hence, if we deliberately forego the knowledge about the dependence structure, we only obtain an upper bound, yet, it allows us to consider just the marginal distributions.

The rest of the paper is structured as follows. Section 2 introduces the necessary network calculus definitions and notations as well as some preliminary results. Section 3 contains the main results obtained in two case studies assuming a conjecture on dependence. The numerical evaluation is presented in Section 4. Section 5 discusses the paper.

2 Definitions and Modeling Assumptions

2.1 SNC Background and Notation

We use the MGF-based SNC in order to bound the probability that the delay exceeds a given value T. The MGF bound on a probability is established by applying Chernoff's bound [16]

$$\mathrm{P}(X > a) \leq e^{-\theta a} \, \mathrm{E}\big[e^{\theta X}\big] , \quad \theta > 0.$$

We define an *arrival flow* by the stochastic process A with discrete time space $\mathbb{N}$ and continuous state space $\mathbb{R}_0^+$ as $A(s,t) \coloneqq \sum_{i=s+1}^{t} a_i$, with a_i as the traffic increment process in time slot i. Network calculus provides an elegant system-theoretic analysis by employing min-plus algebra.

Definition 1 (Convolution and Deconvolution in Min-Plus Algebra [2]). *Let $x(s,t)$ and $y(s,t)$ be real-valued, bivariate functions. The* min-plus convolution *of x and y is defined as*

$$x \otimes y\,(s,t) \coloneqq \inf_{s \leq i \leq t} \{x(s,i) + y(i,t)\} .$$

The min-plus deconvolution *of x and y is defined as*

$$x \oslash y\,(s,t) := \sup_{0 \le i \le s} \{x(i,t) - y(i,s)\}\,.$$

The characteristics of the service process are captured by the notion of a dynamic S-server.

Definition 2 (Dynamic S-Server [4]). *Assume a service element has an arrival flow A as its input and the respective output is denoted by D. Let $S(s,t)$, $0 \le s \le t$, be a stochastic process that is nonnegative and increasing in t. The service element is a* dynamic S-server *iff for all $t \ge 0$ it holds that*

$$D(0,t) \ge A \otimes S\,(0,t) = \inf_{0 \le s \le t} \{A(0,s) + S(s,t)\}\,.$$

The analysis is based on a per-flow perspective. That is, we consider a certain flow, the so-called *flow of interest* (foi). Throughout this paper, for the sake of simplicity, we assume the servers' scheduling to be arbitrary multiplexing [19]. That is, if flow f_2 is prioritized over flow f_1, the leftover service at a dynamic S-server for the corresponding arrival A_1 is $S_{\text{l.o.}}(s,t) = [S(s,t) - A_2(s,t)]^+$. Therefore, we require the server to be work-conserving.

Definition 3 (Work-Conserving Server [4] [11]). *For any $t \ge 0$ let $\tau := \sup\{s \in [0,t] : D(0,s) = A(0,s)\}$ be the beginning of the last backlogged period before t. Assume again the service $S(s,t)$, $0 \le s \le t$, to be a stochastic process that is nonnegative and increasing in t with $S(\tau,\tau) = 0$. A server is said to be* work-conserving *if for any fixed sample path the server is not idle and uses the entire available service, i.e., $D(0,t) = D(0,\tau) + S(\tau,t)$.*

Definition 4 (Virtual Delay). *The* virtual delay *at time $t \ge 0$ is defined as*

$$d(t) := \inf\{\tau \ge 0 : A(0,t) \le D(0,t+\tau)\}\,.$$

It can briefly be described as the time it takes for the cumulated departures to "catch up with" the cumulated arrivals.

Theorem 1 (Output and Delay Bound). *[4] [11] Consider an arrival process $A(s,t)$ with dynamic S-server $S(s,t)$.*

The departure process D is upper bounded for any $0 \le s \le t$ according to

$$D(s,t) \le A \oslash S\,(s,t). \tag{1}$$

The delay at $t \ge 0$ is upper bounded by

$$d(t) \le \inf\{\tau \ge 0 : A \oslash S\,(t+\tau,t) \le 0\}\,.$$

We focus on the analogue of Theorem 1 for moment-generating functions:

Theorem 2 (Output and Delay MGF-Bound [11] [3]). *For the assumptions as in Theorem 1, we obtain:*

The MGF of the departure process D is upper bounded for any $0 \le s \le t$ according to

$$\mathrm{E}\left[e^{\theta D(s,t)}\right] \le \mathrm{E}\left[e^{\theta (A \oslash S\,(s,t))}\right].$$

The violation probability of a given stochastic delay bound $T \ge 0$ at time $t \ge 0$ is bounded by

$$\mathrm{P}(d(t) > T) \le \mathrm{E}\left[e^{\theta (A \oslash S\,(t+T,t))}\right]. \tag{2}$$

In the following definition, we introduce (σ, ρ)-constraints [4] as they enable us to give time-independent, stationary bounds under stability.

Definition 5 ((σ, ρ)-Bound [4]). *An arrival flow is (σ_A, ρ_A)-bounded for some $\theta > 0$, if for all $0 \le s \le t$*

$$\mathrm{E}\left[e^{\theta A(s,t)}\right] \le e^{\theta(\rho_A(\theta)(t-s)+\sigma_A(\theta))}.$$

2.2 Negative Dependence and Acceptable Random Variables

As we discussed in the introduction, we would like to bound the joint distribution of two random variables by their respective product distribution. This concept was captured in the 1960s by Lehmann and his notion of negative dependence.

Definition 6 (Negative Dependence [14]). *A finite family of random variables $\{X_1, \ldots, X_n\}$ is said to be* negatively (orthant) dependent (ND) *if the two following inequalities hold:*

$$\mathrm{P}(X_1 \le x_1, \ldots, X_n \le x_n) \le \prod_{i=1}^{n} \mathrm{P}(X_i \le x_i),$$
$$\mathrm{P}(X_1 > x_1, \ldots, X_n > x_n) \le \prod_{i=1}^{n} \mathrm{P}(X_i > x_i),$$

for all real numbers $x_1, \ldots, x_n$.

The following lemma shows how this characteristic can be used directly in the context of MGFs.

Lemma 1 ([13, 20]). *If $\{X_1, \ldots, X_n\}$ is a set of ND random variables, then for any $\theta > 0$,*

$$\mathrm{E}\left[e^{\theta \sum_{i=1}^{n} X_i}\right] \le \prod_{i=1}^{n} \mathrm{E}\left[e^{\theta X_i}\right]. \tag{3}$$

In other words, treating the aggregate of ND random variables as if they were independent yields an upper bound for the respective MGFs. Random variables that suffice Eqn. (3) are called "acceptable" [1], but are studied in an unrelated context.

Proving that random variables are negatively dependent is a challenging task. Some results exist, e.g., the multinomial and multivariate hypergeometric distribution are ND, or the “Zero-One Lemma” [10], which proves the property for $X_1, \dots, X_n \in \{0,1\}$ such that $\sum_i X_i = 1$. This means that the output processes in the example in Section 1 are indeed ND. Furthermore, it has been shown a related result in [13] that a permutation distribution, and therefore random sampling without replacement, is ND. In our context, this provides a result for a single time slot. In the following, we confine ourselves to conjecture this property for intervals.

Conjecture 1. Let two independent flows with according arrival processes A_1 and A_2 traverse a work-conserving server with finite capacity. Further, both arrival processes have iid increments.

Then, we assume their respective output processes $D_1(s,t)$ and $D_2(s,t)$ to be ND for all $0 \leq s \leq t$.

We do not have a proof but Conjecture 1 held in all our experiments using 10^6 samples to estimate the joint and product (C)CDFs, respectively: For two flows at one server, we tried over 5500 different combinations of intervals, x_1, x_2, (as in the CDF), utilizations (between 0.4 and 0.9), and random packet sizes that were drawn from either exponential, Weibull, Gumbel, or log-normal distribution.

The focus on the same interval for both process is important, as the following, admittedly simplifying, argument suggests: Assume the high priority (HP) flow to send a lot of packets consecutively, i.e., the low priority (LP) flow has no output in this period and queues all its packets. Then, it is more likely for the LP flow to have outputs when the HP flow stops sending, as it is more likely for it to have queued packets.

3 Independence as a Bound

In this section, we investigate two case studies to show in which part of the analysis we exploit the negative dependence.

In the following, we consider the flow f_1 to be the flow of interest (foi) whose delay we stochastically upper bound. All arrival processes A_i are assumed to be discrete time and to have iid increments and all servers S_j are work-conserving and provide a constant rate $c_j \geq 0$. To simplify notation, we denote by $D_i^{(j)}$ the output of flow i at server j.

3.1 Diamond Network

In this case study, we consider the topology in Fig. 1. Assume the foi to have the lowest priority and f_3 to have the highest priority. By SNC literature [5,11], the end-to-end service provided for the flow of interest, also known as the network service curve, can be described by

$$S_{\mathrm{e2e}} = \left[S_1 - \left(\left(\left(A_2 \oslash [S_4 - A_3]^+\right) \oslash S_2\right) + ((A_3 \oslash S_4) \oslash S_3)\right)\right]^+ .$$

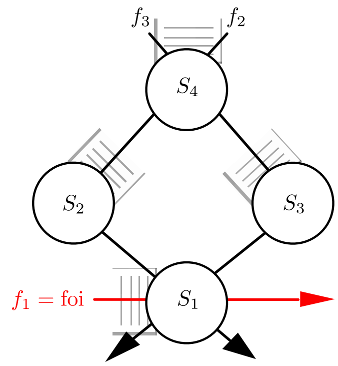

Fig. 1. Diamond network.

Since Conjecture 1 is made on output processes, we postpone the application of the output bound in Eqn. (1) by keeping the exact output at first. That is, we start with

$$S_{\mathrm{e2e}} = \left[S_1 - \left(D_2^{(2)} + D_3^{(3)}\right)\right]^+ , \tag{4}$$

use then the conjecture to bound the MGF of the aggregate by their product (Eqn. (3)), and apply the output bound in a final step.

The probability that the delay process $d(t)$ exceeds a value $T \geq 0$ is upper bounded by

$$\begin{aligned}
&\mathrm{P}(d(t) > T) \\
\overset{(2)}{\leq}\ & \mathrm{E}\left[e^{\theta A_1 \oslash S_{\mathrm{e2e}}(t+T,t)}\right] \\
\leq\ & \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta(A_1(\tau_1,t) - S_{\mathrm{e2e}}(\tau_1,t+T))}\right] \\
\overset{(4)}{=}\ & \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \mathrm{E}\left[e^{-\theta\left[S_1 - \left(D_2^{(2)} + D_3^{(3)}\right)\right]^+(\tau_1,t+T)}\right] \\
\leq\ & \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] e^{-\theta c_1(t+T-\tau_1)}\, \mathrm{E}\left[e^{\theta\left(D_2^{(2)} + D_3^{(3)}\right)(\tau_1,t+T)}\right],
\end{aligned} \tag{5}$$

where we used Theorem 2 in the first inequality and the Union bound in the line below. Since the flows f_2 and f_3 share the server S_4, their according output processes $D_2^{(4)}$ and $D_3^{(4)}$ are dependent and, as a consequence, $D_2^{(2)}$ and $D_3^{(3)}$, as well. However, by the conjecture above, we assume that the resource sharing at S_4 indicates that the dependence on $[\tau_1, t+T]$ is negative which, in turn, is the reason why we upper bound their joint MGF by the product of the marginal MGFs.

This can be interpreted as if we analyzed a new system, where the server S_4 would be split into two servers. That is, one provides the same service as the

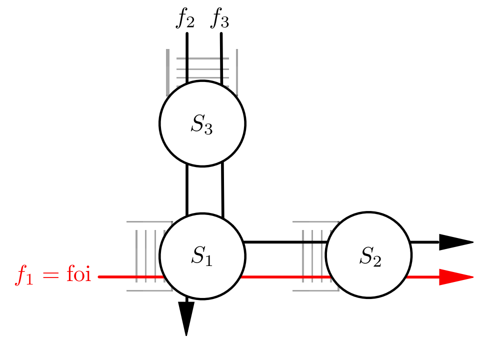

Fig. 2. The $\mathbb{L}$.

original (for the high priority flow f_3), and the other provides the leftover service $[S_4' - A_3']^+$, where S_4' has the same service rate as S_4 and A_3' is a new arrival process, but with the same distribution as A_3.

Hence, the second factor is upper bounded by

$$\begin{aligned}
&\mathrm{E}\left[e^{\theta\left(D_2^{(2)}+D_3^{(3)}\right)(\tau_1,t+T)}\right] \\
&\le \mathrm{E}\left[e^{\theta D_2^{(2)}(\tau_1,t+T)}\right]\mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)}\right] \\
&\le \mathrm{E}\left[e^{\theta\left(\left(A_2 \oslash [S_4-A_3]^+\right)\oslash S_2\right)(\tau_1,t+T)}\right]\mathrm{E}\left[e^{\theta((A_3\oslash S_4)\oslash S_3)(\tau_1,t+T)}\right].
\end{aligned}$$

Further assuming all A_i to be (σ_A, ρ_A)-bounded yields a closed-form for the delay bound under stability:

$$\begin{aligned}
\mathrm{P}(d(t) > T) \le & \frac{e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_1\right)T+\sigma_1(\theta)+\sigma_{A_2}(\theta)+2\sigma_{A_3}(\theta)\right)}}{1-e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_1\right)}} \\
& \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)-c_2\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_3}(\theta)-c_3\right)}} \\
& \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_4\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_3}(\theta)-c_4\right)}}.
\end{aligned}$$

For detailed calculations we refer to Appendix A.1.

In contrast, standard techniques proceed at Eqn. (5) by applying the output Bound Eqn. (1) immediately and continue with Hölder's inequality to deal with the dependence.

3.2 The $\mathbb{L}$

In this case study, we analyze the topology in Fig. 2. The foi has the lowest priority and f_2 the highest. Similarly to Subsection 3.1, we assume the outputs processes of f_2 and f_3 to be ND, based on Conjecture 1. Here, the end-to-end service is

$$S_{\mathrm{e2e}} = \left[\left(\left[S_1 - (A_2 \oslash S_3)\right]^+ \otimes S_2\right) - \left(A_3 \oslash [S_3 - A_2]^+\right)\right]^+ .$$

Again, we postpone the output bounding and start with

$$S_{\mathrm{e2e}} = \left[\left(\left[S_1 - D_2^{(3)}\right]^+ \otimes S_2\right) - D_3^{(3)}\right]^+ . \tag{6}$$

The crucial difference is that, in order to obtain the delay bound for the foi, the so-called min-plus convolution has to be applied to the service processes of S_1 and S_2 forcing us to analyze the output processes at different intervals:

$$\begin{aligned}
&\mathrm{P}(d(t) > T) \\
&\overset{(2)}{\le} \mathrm{E}\left[e^{\theta A_1 \oslash S_{\mathrm{e2e}}(t+T,t)}\right] \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta(A_1(\tau_1,t) - S_{\mathrm{e2e}}(\tau_1,t+T))}\right] \\
&\overset{(6)}{=} \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \mathrm{E}\left[e^{-\theta\left[\left(\left[S_1 - D_2^{(3)}\right]^+ \otimes S_2\right) - D_3^{(3)}\right]^+(\tau_1,t+T)}\right] \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)} e^{-\theta\left(\left[S_1 - D_2^{(3)}\right]^+ \otimes S_2\right)(\tau_1,t+T)}\right] \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} \mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)} e^{-\theta\left[S_1 - D_2^{(3)}\right]^+(\tau_1,\tau_2)} e^{-\theta S_2(\tau_2,t+T)}\right] \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1 \cdot(\tau_2-\tau_1)} e^{-\theta c_2 \cdot (t+T-\tau_2)} \mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)} e^{\theta D_2^{(3)}(\tau_1,\tau_2)}\right],
\end{aligned}$$

where we used the Union bound for each application of the convolution / deconvolution. This scenario is not covered by Conjecture 1 (see also the discussion at the end of Subsection 2.2). Our work-around is to leverage the monotonicity of $D_2^{(3)}$:

$$\mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)} e^{\theta D_2^{(3)}(\tau_1,\tau_2)}\right] \le \mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)} e^{\theta D_2^{(3)}(\tau_1,t+T)}\right].$$

The rest of the analysis employs similar techniques as for the diamond network. See also Appendix A.2. Under the assumption of (σ_A, ρ_A)-bounded arrivals, we again obtain a closed form for a bound on the delay's violation probability under stability:

$$\begin{aligned}
\mathrm{P}(d(t) > T) \le &\frac{e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-\min\{c_1,c_2\}\right)\cdot T + \sigma_{A_1}(\theta)+2\sigma_{A_2}(\theta)+\sigma_{A_3}(\theta)\right)}}{1 - e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-\min\{c_1,c_2\}\right)}} \\
&\cdot \frac{1}{1 - e^{\theta\left(\rho_{A_2}(\theta)-c_3\right)}} \cdot \frac{1}{1 - e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_3\right)}} \cdot \frac{1}{1 - e^{-\theta|c_1-c_2|}}.
\end{aligned}$$

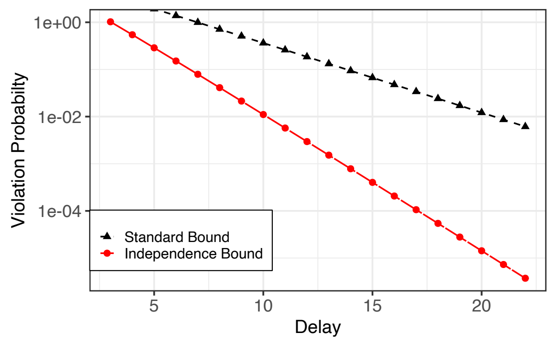

Fig. 3. Delay bound diamond network.

4 Numerical Evaluation

We present the results of a numerical evaluation for both case studies. We ran 10^4 Monte-Carlo simulations to sample the parameters for different server rates and packet sizes, the latter sampled from an exponential distribution. The scenarios are then filtered to ensure a utilization $\in [0.5, 1)$.

4.1 Quality of the Bounds

Diamond Network: This topology, after above the mentioned filtering, yields 485 remaining scenarios, of which 371 are improved. The fact that not all are improved despite the avoidance of Hölder's inequality can be explained as follows: In the analysis, the Union bound is applied after Hölder's inequality. The exponentiation before the summing followed by a square root can have a "mitigating" effect. A similar observation has been exploited in SNC literature before [17].

We also measured the extent of the improvement by computing the ratio of the delay violation probability of the standard approach over the "independence bound". Clearly, values above 1 are desirable. Here, we obtain a median improvement of 6.04. In Fig. 3, we depict the delay bounds for specific parameters.

The $\mathbb{L}$: For this topology, we expect a weaker performance, as our approach using independence as a bound requires the additional step of extending the interval of one output process. The numerical results confirm this expectation: Out of the 729 scenarios, only half of them (384) yield a performance gain. The median of the improvement ratio confirms this, being relatively close to 1 (1.27). Again, we show the delay bounds for fixed parameters (Fig. 4).

4.2 Computation Run Time

Our proposed approach does not only often substantially improve the bounds but it also has a much lower computation complexity than the standard approach. The reason is that the latter relies on an additional Hölder parameter. The

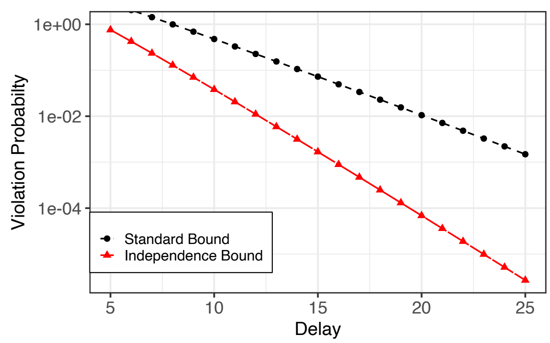

Fig. 4. Delay bound in the $\mathbb{L}$.

optimizations are conducted using a grid search followed by a downhill simplex algorithm. The improvements ratios are in the median 337.5 (1.62 sec compared to 0.0048 sec) for the diamond scenario and 458.1 for the $\mathbb{L}$ (1.42 sec compared to 0.0031 sec). These improvements due to the reduction of the optimization parameters indicates a significant potential for an analysis of larger networks, as the optimization step in the MGF-based SNC can severely limit its scalability.

5 Discussion

In this paper, we found interesting results indicating that by using independence as a bound, one can often times improve the delay bound while also speeding up the run time significantly. Obviously, the crucial next step is to find scenarios in which the conjecture can be proved rigorously. One potential technique might be to use the coupling method [15,21], as it is can be applied to derive relations between tail probabilities. Furthermore, more scenarios can be analyzed in which the negative dependence can be exploited. In particular, this includes large-scale experiments that require many invocations of Hölder's inequality.

References

1. Antonini, R.G., Kozachenko, Y., Volodin, A.: Convergence of series of dependent φ-subgaussian random variables. J. Math. Anal. Appl. (2008)
2. Baccelli, F., Cohen, G., Olsder, G.J., Quadrat, J.P.: Synchronization and linearity: an algebra for discrete event systems. John Wiley & Sons Ltd (1992)
3. Beck, M.A.: Advances in Theory and Applicability of Stochastic Network Calculus. Ph.D. thesis, TU Kaiserslautern (2016)
4. Chang, C.S.: Performance guarantees in communication networks. Springer, London (2000)
5. Ciucu, F., Burchard, A., Liebeherr, J.: Scaling properties of statistical end-to-end bounds in the network calculus. IEEE/ACM ToN (2006)

6. Cruz, R.L.: A calculus for network delay, part I: Network elements in isolation. IEEE Transactions on information theory **37**(1), 114–131 (1991)
7. Cruz, R.L.: A calculus for network delay, part II: Network analysis. IEEE Transactions on information theory **37**(1), 132–141 (1991)
8. Cruz, R.L.: Quality of service management in integrated services networks. In: Proc. Semi-Annual Research Review, CWC. vol. 1, pp. 4–5 (1996)
9. Dong, F., Wu, K., Srinivasan, V.: Copula analysis for statistical network calculus. In: Proc. IEEE INFOCOM'15. pp. 1535–1543 (2015)
10. Dubhashi, D., Ranjan, D.: Balls and bins: A study in negative dependence. Random Structures & Algorithms **13**(2), 99–124 (1998)
11. Fidler, M.: An end-to-end probabilistic network calculus with moment generating functions. In: Proc. IEEE IWQoS'06. pp. 261–270 (Jun 2006)
12. Jiang, Y., Liu, Y.: Stochastic network calculus, vol. 1. Springer (2008)
13. Joag-Dev, K., Proschan, F.: Negative association of random variables with applications. The Annals of Statistics **11**(1), 286–295 (1983)
14. Lehmann, E.L.: Some concepts of dependence. The Annals of Mathematical Statistics pp. 1137–1153 (1966)
15. Lindvall, T.: Lectures on the coupling method. Courier Corporation (2002)
16. Nelson, R.: Probability, stochastic processes, and queueing theory: the mathematics of computer performance modeling. Springer (1995)
17. Nikolaus, P., Schmitt, J., Schütze, M.: h-Mitigators: Improving your stochastic network calculus output bounds. Computer Communications (May 2019)
18. Rizk, A., Fidler, M.: Leveraging statistical multiplexing gains in single-and multi-hop networks. In: Proc. IEEE IWQoS '11. pp. 1–9 (2011)
19. Schmitt, J., Zdarsky, F.A., Fidler, M.: Delay bounds under arbitrary multiplexing: When network calculus leaves you in the lurch ... In: Proc. IEEE International Conference on Computer Communications (INFOCOM'08). Phoenix, AZ, USA (Apr 2008)
20. Sung, S.H.: On the exponential inequalities for negatively dependent random variables. J. Math. Anal. Appl. **381**(2), 538–545 (2011)
21. Thorisson, H.: Coupling, Stationarity and Regeneration. Springer (2000)

A Appendix

A.1 Diamond Network

By using the conjecture, we have obtained so far that

$$\begin{aligned}
&\mathrm{P}(d(t) > T) \\
\leq &\sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] e^{-\theta c_1(t+T-\tau_1)} \,\mathrm{E}\left[e^{\theta\left(D_2^{(2)}+D_3^{(3)}\right)(\tau_1,t+T)}\right] \\
\leq &\sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] e^{-\theta c_1(t+T-\tau_1)} \,\mathrm{E}\left[e^{\theta\left(\left(A_2 \oslash [S_4-A_3]^+\right)\oslash S_2\right)(\tau_1,t+T)}\right] \mathrm{E}\left[e^{\theta((A_3\oslash S_4)\oslash S_3)(\tau_1,t+T)}\right].
\end{aligned}$$

This leads to

$$
\begin{aligned}
&\mathrm{P}(d(t) > T) \\
&\leq \sum_{\tau_1=0}^{t} \mathrm{E}\Big[e^{\theta A_1(\tau_1,t)}\Big] e^{-\theta c_1(t+T-\tau_1)} \left\{ \sum_{\tau_2=0}^{\tau_1} \mathrm{E}\Big[e^{\theta\left(A_2 \oslash [S_4 - A_3]^+\right)(\tau_2,t+T)}\Big] \mathrm{E}\Big[e^{-\theta S_2(\tau_2,\tau_1)}\Big] \right\} \\
&\quad \cdot \left\{ \sum_{\tau_2=0}^{\tau_1} \mathrm{E}\Big[e^{\theta(A_3 \oslash S_4)(\tau_2,t+T)}\Big] \mathrm{E}\Big[e^{-\theta S_3(\tau_2,\tau_1)}\Big] \right\} \\
&\leq \sum_{\tau_1=0}^{t} \mathrm{E}\Big[e^{\theta A_1(\tau_1,t)}\Big] e^{-\theta c_1(t+T-\tau_1)} \\
&\quad \cdot \left\{ \sum_{\tau_2=0}^{\tau_1} \left\{ \sum_{\tau_3=0}^{\tau_2} \mathrm{E}\Big[e^{\theta A_2(\tau_3,t+T)}\Big] \mathrm{E}\Big[e^{\theta A_3(\tau_3,\tau_2)}\Big] \mathrm{E}\Big[e^{-\theta S_4(\tau_3,\tau_2)}\Big] \right\} \mathrm{E}\Big[e^{-\theta S_2(\tau_2,\tau_1)}\Big] \right\} \\
&\quad \cdot \left\{ \sum_{\tau_2=0}^{\tau_1} \left\{ \sum_{\tau_3=0}^{\tau_2} \mathrm{E}\Big[e^{\theta A_3(\tau_3,t+T)}\Big] \mathrm{E}\Big[e^{-\theta S_4(\tau_3,,\tau_2)}\Big] \right\} \mathrm{E}\Big[e^{-\theta S_3(\tau_2,\tau_1)}\Big] \right\},
\end{aligned}
$$

after applying the Union bound for each usage of the deconvolution. Further assuming all A_i to be (σ_A, ρ_A)-bounded yields a closed-form for the delay bound under the stability condition

$$
\begin{aligned}
\rho_{A_1}(\theta) + \rho_{A_2}(\theta) + \rho_{A_3}(\theta) &< c_1, \\
\rho_{A_2}(\theta) &< c_2, \\
\rho_{A_3}(\theta) &< c_3, \\
\rho_{A_2}(\theta) + \rho_{A_3}(\theta) &< c_4 :
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{P}(d(t) > T) \\
&\overset{\text{(Def. 5)}}{\leq} \sum_{\tau_1=0}^{t} e^{\theta\left(\rho_{A_1}(\theta)(t-\tau_1)+\sigma_1(\theta)\right)} e^{-\theta c_1(t+T-\tau_1)} \\
&\quad \cdot \left\{ \sum_{\tau_2=0}^{\tau_1} \left\{ \sum_{\tau_3=0}^{\tau_2} e^{\theta\left(\rho_{A_2}(\theta)(t+T-\tau_3)+\sigma_{A_2}(\theta)\right)} e^{\theta\left(\rho_{A_3}(\theta)(\tau_2-\tau_3)+\sigma_{A_3}(\theta)\right)} e^{-\theta c_4(\tau_2-\tau_3)} \right\} e^{-\theta c_2(\tau_1-\tau_2)} \right\} \\
&\quad \cdot \left\{ \sum_{\tau_2=0}^{\tau_1} \left\{ \sum_{\tau_3=0}^{\tau_2} e^{\theta\left(\rho_{A_3}(\theta)(t+T-\tau_3)+\sigma_{A_3}(\theta)\right)} e^{-\theta c_4(\tau_2-\tau_3)} \right\} e^{-\theta c_3(\tau_1-\tau_2)} \right\} \\
&\leq e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_1\right)T+\sigma_1(\theta)+\sigma_{A_2}(\theta)+2\sigma_{A_3}(\theta)\right)} \\
&\quad \cdot \sum_{\tau_1=0}^{t} e^{\theta\left(\rho_{A_1}(\theta)-c_1\right)(t-\tau_1)} \left\{ \sum_{\tau_2=0}^{\tau_1} \frac{e^{\theta\rho_{A_2}(\theta)(t-\tau_2)}}{1-e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_4\right)}} e^{-\theta c_2(\tau_1-\tau_2)} \right\} \\
&\quad \cdot \left\{ \sum_{\tau_2=0}^{\tau_1} \frac{e^{\theta\rho_{A_3}(\theta)(t-\tau_2)}}{1-e^{\theta\left(\rho_{A_3}(\theta)-c_4\right)}} e^{-\theta c_3(\tau_1-\tau_2)} \right\}
\end{aligned}
$$

$$
\begin{aligned}
&\le e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_1\right)T+\sigma_1(\theta)+\sigma_{A_2}(\theta)+2\sigma_{A_3}(\theta)\right)} \\
&\quad \cdot \sum_{\tau_1=0}^{t} \frac{e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_1\right)(t-\tau_1)}}{1-e^{\theta\left(\rho_{A_2}(\theta)-c_2\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_3}(\theta)-c_3\right)}} \\
&\quad \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_4\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_3}(\theta)-c_4\right)}} \\
&\le \frac{e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_1\right)T+\sigma_1(\theta)+\sigma_{A_2}(\theta)+2\sigma_{A_3}(\theta)\right)}}{1-e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_1\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)-c_2\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_3}(\theta)-c_3\right)}} \\
&\quad \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_4\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_3}(\theta)-c_4\right)}},
\end{aligned}
$$

where we used the convergence of the geometric series.

A.2 The $\mathbb{L}$

We have that

$$
\begin{aligned}
&\mathrm{P}(d(t)>T) \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)}\, \mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)} e^{\theta D_2^{(3)}(\tau_1,\tau_2)}\right] \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)}\, \mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)} e^{\theta D_2^{(3)}(\tau_1,t+T)}\right].
\end{aligned}
$$

With the conjecture, we compute

$$
\begin{aligned}
&\mathrm{P}(d(t)>T) \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)}\, \mathrm{E}\left[e^{\theta D_3^{(3)}(\tau_1,t+T)}\right] \mathrm{E}\left[e^{\theta D_2^{(3)}(\tau_1,t+T)}\right] \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)}\, \mathrm{E}\left[e^{\theta (A_2\oslash S_3)(\tau_1,t+T)}\right] \mathrm{E}\left[e^{\theta\left(A_3\oslash[S_3-A_2]^+\right)(\tau_1,t+T)}\right] \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)} \left\{\sum_{\tau_3=0}^{\tau_1} \mathrm{E}\left[e^{\theta A_2(\tau_3,t+T)}\right] e^{-\theta c_3(\tau_1-\tau_3)}\right\} \\
&\quad \cdot \left\{\sum_{\tau_3=0}^{\tau_1} \mathrm{E}\left[e^{\theta A_3(\tau_3,t+T)}\right] \mathrm{E}\left[e^{-\theta[S_3-A_2]^+(\tau_3,\tau_1)}\right]\right\} \\
&\le \sum_{\tau_1=0}^{t} \mathrm{E}\left[e^{\theta A_1(\tau_1,t)}\right] \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)} \left\{\sum_{\tau_3=0}^{\tau_1} \mathrm{E}\left[e^{\theta A_2(\tau_3,t+T)}\right] e^{-\theta c_3(\tau_1-\tau_3)}\right\} \\
&\quad \cdot \left\{\sum_{\tau_3=0}^{\tau_1} \mathrm{E}\left[e^{\theta A_3(\tau_3,t+T)}\right] \mathrm{E}\left[e^{\theta A_2(\tau_3,\tau_1)}\right] e^{-\theta c_3(\tau_1-\tau_3)}\right\}.
\end{aligned}
$$

If we again assume all A_i to be (σ_A, ρ_A)-bounded, we obtain for

$$\begin{aligned} \rho_{A_1}(\theta) + \rho_{A_2}(\theta) + \rho_{A_3}(\theta) &< \min\{c_1, c_2\}, \\ \rho_{A_2}(\theta) + \rho_{A_3}(\theta) &< c_3, \end{aligned}$$

and $c_1 \neq c_2$:

$$\begin{aligned}
&\mathrm{P}(d(t) > T) \\
\overset{\text{(Def. 5)}}{\leq} &\sum_{\tau_1=0}^{t} e^{\theta\left(\rho_{A_1}(\theta)(t-\tau_1)+\sigma_{A_1}(\theta)\right)} \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)} \\
&\cdot \left\{ \sum_{\tau_3=0}^{\tau_1} e^{\theta\left(\rho_{A_2}(\theta)(t+T-\tau_3)+\sigma_{A_2}(\theta)\right)} e^{-\theta c_3(\tau_1-\tau_3)} \right\} \\
&\cdot \left\{ \sum_{\tau_3=0}^{\tau_1} e^{\theta\left(\rho_{A_2}(\theta)(\tau_1-\tau_3)+\sigma_{A_2}(\theta)\right)} e^{\theta\left(\rho_{A_3}(\theta)(t+T-\tau_3)+\sigma_{A_3}(\theta)\right)} e^{-\theta c_3(\tau_1-\tau_3)} \right\} \\
\leq & e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)\right)\cdot T+\sigma_{A_1}(\theta)+2\sigma_{A_2}(\theta)+\sigma_{A_3}(\theta)\right)} \\
&\cdot \sum_{\tau_1=0}^{t} e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)\right)(t-\tau_1)} \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)} \\
&\cdot \left\{ \sum_{\tau_3=0}^{\tau_1} e^{\theta\left(\rho_{A_2}(\theta)-c_3\right)(\tau_1-\tau_3)} \right\} \left\{ \sum_{\tau_3=0}^{\tau_1} e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_3\right)(\tau_1-\tau_3)} \right\} \\
\leq & e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)\right)\cdot T+\sigma_{A_1}(\theta)+2\sigma_{A_2}(\theta)+\sigma_{A_3}(\theta)\right)} \\
&\cdot \sum_{\tau_1=0}^{t} \frac{e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)\right)(t-\tau_1)}}{1-e^{\theta\left(\rho_{A_2}(\theta)-c_3\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_3\right)}} \\
&\cdot \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)} \\
\leq & e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)\right)\cdot T+\sigma_{A_1}(\theta)+2\sigma_{A_2}(\theta)+\sigma_{A_3}(\theta)\right)} \\
&\cdot \sum_{\tau_1=0}^{t} \frac{e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)\right)(t-\tau_1)}}{1-e^{\theta\left(\rho_{A_2}(\theta)-c_3\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_3\right)}} \\
&\cdot \sum_{\tau_2=\tau_1}^{t+T} e^{-\theta c_1\cdot(\tau_2-\tau_1)} e^{-\theta c_2\cdot(t+T-\tau_2)} \\
\leq & \frac{e^{\theta\left(\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-\min\{c_1,c_2\}\right)\cdot T+\sigma_{A_1}(\theta)+2\sigma_{A_2}(\theta)+\sigma_{A_3}(\theta)\right)}}{1-e^{\theta\left(\rho_{A_1}(\theta)+\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-\min\{c_1,c_2\}\right)}} \\
&\cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)-c_3\right)}} \cdot \frac{1}{1-e^{\theta\left(\rho_{A_2}(\theta)+\rho_{A_3}(\theta)-c_3\right)}} \cdot \frac{1}{1-e^{-\theta|c_1-c_2|}},
\end{aligned}$$

where we used again the convergence of the geometric series.