An Approach using N-Demisupermartingales for the Stochastic Analysis of Networks

Kishore Angrishi and Ulrich Killat

Abstract—Stochastic network calculus is the probabilistic version of the network calculus, which uses envelopes to perform probabilistic analysis of queueing networks. The accuracy of probabilistic end-to-end delay or backlog bounds computed using the elegant theory of network calculus has been always a concern. In this paper, we propose novel bounds based on Doob’s maximal inequality for N-demisupermartingales to compute end-to-end probabilistic bounds which improve the existing bounds from network calculus. In particular, we show that reasonably accurate bounds are achieved by comparing the new bounds with the exact results from queuing theory for a network of M/M/1 queues.

Index Terms—Network calculus, end-to-end delay and backlog bounds, Doob’s inequality, N-demisupermartingales.

I. INTRODUCTION

Queueing theory is the mathematical study of queues, which generally uses probability mass or density functions to describe arrival traffic and service offered at the network node to compute exact probabilistic delay or backlog measures. However, with few exceptions, stochastic analysis of queueing networks using queueing theory to compute end-to-end probabilistic performance measures is mathematically complex without making simplifying assumptions on arrival traffic or service offered at the network nodes. In most situations, probabilistic bounds on performance measures are as sufficient as the actual values. Stochastic network calculus is an elegant theory, useful for computing probabilistic bounds on end-to-end delay or backlog in queueing networks. Stochastic network calculus is the probabilistic extension of deterministic network calculus, which uses an envelope approach to describe arrival traffic and service offered at the network node. The tightness of the end-to-end probabilistic performance bounds has always been a concern in stochastic network calculus. The concern is mainly due to the use of union bounds for computing the bounds on probabilistic performance measures of the network. Recently, in [1], [2], authors have derived new performance bounds for a GI/GI/1 queue in stochastic network calculus using Doob’s maximal inequality for exponential supermartingales (instead of using union bounds) which are comparable, to the exact results of M/M/1 and M/D/1 queues from queueing theory. A general comparison of results for GI/GI/1 queue from statistical network calculus with the classical queuing theory is made in [3].

In this paper, we compute end-to-end probabilistic performance bounds for a network of GI/GI/1 queues in stochastic network calculus using Doob’s maximal inequality for N-demisupermartingales [4], [5], which yields similar results as in [1], [2] for the single queue case. The rest of the paper is structured as follows: In Section II we introduce the notion and assumptions used in the paper. In Section III we derive the probabilistic end-to-end performance bounds on delay and backlog using the notion of effective bandwidth [6] and effective capacity [7] functions under the stated assumptions. Brief conclusions are presented in Section IV.

II. NOTATION AND ASSUMPTIONS

Our time model is discrete, i.e., $t \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$. We assume that the arrival traffic and the service offered at a node are stationary and have independent increments, i.e., the random process depends only on the length of the interval $(s, t]$ ($\Delta = t - s$) but not on $s$ or $t$ itself and increments over two non-overlapping intervals are independent of each other. In a network of nodes connected in series as shown in Fig. 1, we use non-decreasing, left-continuous processes $A_h$ and $D_h$ to describe the arrivals and the departures at node $h$, respectively. $A_h(s, t)$ and $D_h(s, t)$ represent the cumulative amount of data seen in an interval $(s, t]$ at input and output of the node $h$, respectively, for any $0 \leq s \leq t$. For the arrival and departure processes at node $h$, we assume the initial condition $A_h(0) = 0$ and the causal condition $D_h(t) \leq A_h(t)$, where we denote $A_h(0, t) = A_h(t)$ and $D_h(0, t) = D_h(t)$ for any $t \geq 0$. The backlog $B_h(t)$ and delay $W_h(t)$ at time $t \geq 0$ in a node $h$ is given by $B_h(t) = A_h(t) - D_h(t)$ and $W_h(t) = \inf\{d \geq 0 : A_h(t - d) \leq D_h(t)\}$, respectively.

A stochastic service process $S_h$ is said to describe the service offered at node $h$, if the corresponding arrival and departure processes at node $h$ satisfy for any fixed sample path and $t \geq 0$:

$$A_h \otimes S_h(t) \leq D_h(t)$$

(1)

where $\otimes$ is the min-plus convolution of $A_h$ and $S_h$ which is defined as $A_h \otimes S_h(t) = \inf_{0 \leq u \leq t} \{A_h(0, u) + S_h(u, t)\}$. Any random process $S$ satisfying the above relationship is referred to as “dynamic F-server” in [8]. The main advantage of using network calculus to do performance analysis of networks is that the network calculus allows to model a network of nodes as a single virtual node. The stochastic network service process $S_{net}$ characterizing the service offered in a single virtual network node, which represents a network of $H$ nodes connected in series as shown in Fig. 1, can be computed for any fixed sample path using the min-plus convolution of the stochastic service process $S_h$ of constituting
nodes for \( h = 1, \ldots, H \), i.e., \( S_{\text{net}} = S_1 \otimes S_2 \otimes \cdots \otimes S_H \) \cite{8, 9, 10}. Since we assume the arrival traffic \( A_1 \) at the ingress of the network and the service process \( S_0 \) at each hop \( h \) to be stationary with independent increments, we have \( A_1(s, t) = A_1(0, t-s) \) and \( S_h(s, t) = S_h(0, t-s) \) for any \( t \geq s \geq 0 \). We assume that the arrival traffic \( A_1 \) at the ingress of the network and the stochastic service processes \( S_h \), for \( h = 1, \ldots, H \), characterizing the service offered at the nodes of the network are independent of each other.

To compute the probabilistic performance measures using network calculus, we use the notion of effective bandwidth \((\alpha_h)\) \cite{6} and effective capacity \((\beta_h)\) \cite{7} from large deviations theory to describe the stochastic arrival traffic and service offered at a node \( h \), respectively. The effective bandwidth of an arrival traffic \( A_h \) with independent increments from \cite{6}, for any \( \theta, t > 0 \), is given as

\[
\alpha_h(\theta) = \frac{1}{\theta} \log E \left[ e^{\theta A_h(1)} \right] \tag{2}
\]

Similarly, the effective capacity function of a stochastic service process \( S_h \) with independent increments, for any \( \theta, t > 0 \), is defined as

\[
\beta_h(\theta) = -\frac{1}{\theta} \log E \left[ e^{-\theta S_h(1)} \right] \tag{3}
\]

III. PROBABILISTIC BOUNDS ON BACKLOG AND DELAY

In this section, we compute probabilistic bounds on backlog and delay in a network of \( H \) nodes as shown in Fig. 1 using Doob’s maximal inequality for N-semiMartingales. Let \( A_1 = A \) and \( D_H = D \) be the arrival traffic at the ingress of the network and departure traffic from the egress of the network, respectively.

The following theorem provides the probabilistic bounds on end-to-end backlog and delay using the notion of effective bandwidth \((\alpha)\) and effective capacity \((\beta_h)\) of the arrival and service processes at each network node \( h \), respectively.

**Theorem 3.1:** Let the service offered at node \( h \) in a tandem network be modeled by the stochastic service process \( S_h \) with the corresponding effective capacity function \( \beta_h \), for \( h = 1, \ldots, H \). Let \( A \) be the arrival process with effective bandwidth \( \alpha \) and \( D \) be the departure process from the tandem network with \( H \) nodes. Then we have the following bounds.

1) Backlog bound : The probabilistic bound on the backlog in a network, for any \( t \geq 0 \), is given by

\[
P\{B(t) > x\} \leq e^{-\theta x} \sum_{h=0}^{H-1} \frac{(\theta x)^h}{h!} \tag{4}
\]

2) Delay bound : The probabilistic bound on the delay in a network, for any \( t \geq 0 \), is given by

\[
P\{W(t) > d\} \leq e^{-\theta^* (\alpha (\theta^*)) d} \sum_{h=0}^{H-1} \frac{\theta^* (\alpha (\theta^*))^h d^h}{h!} \tag{5}
\]

where \( \theta^* = \sup \{ \theta : \alpha (\theta) \leq \min_{1 \leq h \leq H} \{ \beta_h (\theta) \} \} \)

The proof of the theorem relies on applying Doob’s maximal inequality for N-semiMartingales to compute probabilistic bounds. The key observation is that the random process \( Z(t) = A(t) - S(t) \) and the exponential function \( e^{\theta Z(t)} \) of random process \( Z(t) \) for any \( t \geq 0 \) are N-semiMartingales. This is shown using the following lemma.

**Lemma 3.1:** Consider an arrival process \( A \) with independent increments at a node offering a stochastic service characterized by a service process \( S \) with independent increments. If \( Z(t) \) be a random process defined as \( A(t) - S(t) \), then the random process \( Z(t) \) and the exponential function \( e^{\theta Z(t)} \) of random process \( Z(t) \) in an interval \([0, t]\) for \( t \in \mathbb{N}_0 = \{0, 1, 2, \ldots \} \) and any \( \theta \in (0, \theta^* \} \) are N-semiMartingales, where \( \theta^* \) is the supremum of \( \theta \) making \( \alpha (\theta) \leq \beta (\theta) \).

**Proof:** Let \( a(t) \) and \( s(t) \) be the random amount of arrival traffic and data served at a node, respectively, at a time instance \( t \geq 0 \), then \( A(t) = \sum_{u=0}^{t} a(u) \) and \( S(t) = \sum_{u=0}^{t} s(u) \). Let \( z(t) = a(t) - s(t) \) and \( Z(t) = A(t) - S(t) \). Since the arrival process \( A \) and the service process \( S \) are assumed to be independent and have independent increments, the random process \( Z(t) \) will also have independent increments. The stability conditions for the queue at node are \( E[a(1)] \leq E[s(1)] \) and \( \alpha (\theta) \leq \beta (\theta) \) for any finite \( \theta \in (0, \infty) \).

To prove that \( Z(t) \) and \( e^{\theta Z(t)} \) for \( t \in \mathbb{N}_0 = \{0, 1, 2, \ldots \} \) and any \( \theta \in (0, \theta^* \} \) be N-semiMartingales \cite{4, 5}, we need to show that \( E[Z(t+1) - Z(t)] f(Z(1), Z(2), \ldots, Z(t))] \leq 0 \) and \( E[e^{\theta Z(t+1)} - e^{\theta Z(t)}] f(e^{\theta Z(1)}, e^{\theta Z(2)}, \ldots, e^{\theta Z(t)}) \leq 0 \) for \( t \in \mathbb{N}_0 = \{0, 1, 2, \ldots \} \) for every co-ordinatewise non-decreasing, non-negative function \( f \) whenever the expectation is defined.

\[
E[Z(t+1) - Z(t)] f(Z(1), Z(2), \ldots, Z(t))] = E[z(t+1) f(Z(1), Z(2), \ldots, Z(t))] = E[a(t+1) - s(t+1) f(Z(1), Z(2), \ldots, Z(t))] \leq 0
\]

The inequality at the final step is due to the stability condition \( E[a(1)] \leq E[s(1)] \) and \( f \) being a nonnegative function. Similarly,

\[
E[e^{\theta Z(t+1)} - e^{\theta Z(t)}] f(e^{\theta Z(1)}, e^{\theta Z(2)}, \ldots, e^{\theta Z(t)}) \leq 0
\]

The final inequality is due to the stability condition \( \alpha (\theta) \leq \beta (\theta) \) for any \( \theta \in (0, \infty \} \) and \( f(e^{\theta Z(1)}, e^{\theta Z(2)}, \ldots, e^{\theta Z(t)}) = e^{\theta Z(t)} f(e^{\theta Z(1)}, e^{\theta Z(2)}, \ldots, e^{\theta Z(t)}) \) being a nonnegative function. □

Since \( e^{\theta Z(t)}, t \geq 0 \) for any \( \theta \in (0, \theta^* \} \) is a non-negative N-semiMartingale, by Doob’s maximal inequality for N-
demisupermartingale from \cite{4, 5} we have, for any \( \sigma \geq 0 \),
\[
P \left\{ \sup_{\tau \leq u \leq t} Z(u) > \sigma \right\} = P \left\{ \sup_{\tau \leq u \leq t} e^{\theta Z(u)} > e^{\theta \sigma} \right\} \leq e^{\theta \sigma} \mathbb{E} e^{-\theta \sigma} \leq e^{\theta \sigma} \leq e^{-\theta^* \sigma}
\]
(6)

The final step in the above equation is from the definition of \( \theta^* = \sup \{ \theta : \alpha(\theta) \leq \beta(\theta) \} \). The proof of Theorem 3.1 also relies on Lemma 4.1 from \cite{11}, which states that for any two non-negative independent random variables \( F \) and \( G \) with \( P(F > \sigma) \leq f(\sigma) \) and \( P(G > \sigma) \leq g(\sigma) \) where \( f(\sigma) \) and \( g(\sigma) \) are non-negative, decreasing function for any \( \sigma \geq 0 \), then
\[
P \{ F + G > \sigma \} \leq 1 - \int_0^\sigma \tilde{f}(\sigma - u) d\tilde{g}(u)
\]
(7)

where \( \tilde{f}(\sigma) = 1 - f(\sigma) \) and \( \tilde{g}(\sigma) = 1 - g(\sigma) \).

**Proof of Theorem 3.1** We first prove the probabilistic end-to-end backlog bound. For any fixed sample path and \( t \geq 0 \), we have
\[
P \{ B(t) > x \} = P \{ A(t) - D(t) > x \} \leq P \{ A(t) - A \otimes S_{net}(t) > x \} = P \{ A(t) - A \otimes S_1 \otimes S_2 \cdots \otimes S_H(t) > x \} \leq P \left\{ \sup_{0 \leq k_1 \leq k_2 \leq k_3 \cdots \leq k_H \leq t} \{ A(t) - A(k_1) - S_1(k_1, k_2) - S_2(k_2, k_3) - \cdots - S_H(k_H, t) \} > x \right\}
\]
For any fixed sample path and \( t \geq 0 \), the probabilistic bound on end-to-end delay is given as
\[
P \{ W(t) > d \} = P \{ A(t - d) - D(t) > 0 \} \leq P \{ A(t - d) - A \otimes S_{net}(t) > 0 \} = P \{ A(t - d) - A \otimes S_1 \otimes S_2 \cdots \otimes S_H(t) > 0 \} \leq P \left\{ \sup_{0 \leq k_1 \leq k_2 \leq k_3 \cdots \leq k_H \leq t} \{ A(t) - A(k_1) - S_1(k_1, k_2) - S_2(k_2, k_3) - \cdots - S_H(k_H, t) \} > 0 \right\}
\]
For both parts of the proof, the first and second inequalities are due to the definition of stochastic network service process and the property of supremum operation \cite{11}, respectively. Since \( (0, k_1), (k_1, k_2), \ldots , (k_{H-1}, k_H) \) and \( (d, k_1), (k_1, k_2), \ldots , (k_{H-1}, k_H) \) are non-overlapping intervals, \( \left\{ A(k_1) - S_1(k_1) \right\}, \left\{ A(k_1, k_2) - S_1(k_1, k_2) \right\}, \ldots , \left\{ A(k_H, t) - S_H(k_H, t) \right\} \) are independent of each other. Therefore, by Lemma 5.1, equations 6 and 7, we get the final step.

To analyse the accuracy of the new delay and backlog bounds from Theorem 3.1, we compare them with the accurate results from queueing theory. We apply the network calculus bounds from Theorem 3.1 to analyse a network of M/M/1 queues. In an M/M/1 queuing system with one server, both the arrival and the service processes are of Poisson type. The customers arrive at rate \( \lambda \) and the server works at rate \( \mu \). The M/M/1 queue is a birth and death process. We denote the utilization factor by \( \rho = \lambda / \mu \) and assume for stability that \( \rho < 1 \). The effective bandwidth and effective capacity of the arrival and service processes (Poisson process) are \( \lambda e^{\frac{\rho}{1-\rho}} \) and \( \mu e^{\frac{\rho}{1-\rho}} \), respectively, with \( \theta^* = -\log \rho \). We consider a special case of the network from Fig. 1 with \( H \) M/M/1 queues connected in series to analyse the accuracy of end-to-end network calculus delay bound. A Poisson flow with rate \( \lambda \) traverses through the entire network. The arrival process at the downstream queue is the departure process of the upstream queue. Let each queue in the network be served by a similar service process \( S \) with effective capacity \( \beta(\theta) \) and the service processes at each hop of the network are independent of each other. The exact distribution of steady state end-to-end delay \( W_{\text{net}} \) in a M/M/1 queueing network from queueing theory is given by \cite{12}
\[
P \{ W_{\text{net}} > d \} = \sum_{h=0}^{H-1} \frac{(\mu(1-\rho)) h^d}{h!} e^{-\mu(1-\rho) d}
\]
(8)

Note that in the terminology of queueing theory, the delay \( W_{\text{net}} \) includes the time spent in the server, hence \( W_{\text{net}} = W + S_{\text{net}} \), whereas in network calculus the term “delay” is used for the waiting time \( W_{\text{net}} \) neglecting the service time \( S_{\text{net}} \). We observe that the delay in a queueing system \( W_{\text{net}} \) is always an upper bound to the waiting time in the queue \( W_{\text{net}} \), i.e., \( W_{\text{net}} \leq W_{\text{net}} \) and by Little’s law \( E[W_{\text{net}}] = E[W_{\text{net}}] + 1/\mu \). For \( \theta^* = -\log \rho \), we have steady state end-to-end waiting time \( W_{\text{net}} \) in a virtual network queue from equation 5.
\[
P \{ W_{\text{net}} > d \} \leq \sum_{h=0}^{H-1} \frac{(\mu(1-\rho)) h^d}{h!} e^{-\mu(1-\rho) d}
\]
(9)

It is apparent from equations 8 and 9 that the probabilistic end-to-end network waiting time \( W_{\text{net}} \) computed using Theorem 3.1 provides reasonably good bounds.

**IV. CONCLUSIONS**

In this paper we used Doob’s maximal inequality for N-Demisupermartingales to compute end-to-end probabilistic delay and backlog bounds within the framework of network
calculus. The tightness of the computed end-to-end probabilistic performance bounds is explored by comparing new bounds with the exact results from queueing theory for a network of M/M/1 queues.

REFERENCES

[1] F. Ciucu, “Network calculus delay bounds in queueing networks with exact solutions,” in Proceedings of International Teletraffic Congress (ITC-20), 2007.

[2] ——, “Exponential supermartingales for evaluating end-to-end backlog bounds,” in Proceedings of Ninth Workshop on Mathematical Performance Modeling and Analysis (MAMA), 2007.

[3] Y. Jiang, “Network calculus and queueing theory: two sides of one coin,” in Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools, October 20-22, 2009.

[4] T. Christofides, “Maximal inequalities for n-demimartingales,” Archives of Inequalities and Applications, vol. 1, pp. 397–408, 2003.

[5] B. P. Rao, “On some maximal inequalities for demisubmartingales and n-demisupermartingales,” Journal of Inequalities in Pure and Applied Mathematics, vol. 8, no. 4, 2007.

[6] F. P. Kelly, “Notes on effective bandwidths,” Stochastic Networks: Theory and Applications, vol. Oxford, Royal Statistical Society Lecture Notes Series., pp. 141–168, 1996.

[7] K. Angrishi and U. Killat, “Analysis of a real-time network using statistical network calculus with effective bandwidth and effective capacity,” in Proceedings of 14. GI/ITG Konferenz Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen (MMB 2008), 2008.

[8] C.-S. Chang, Performance Guarantees in Communication Networks. Springer-Verlag, 2000.

[9] M. Fidler, “An end-to-end probabilistic network calculus with moment generating functions,” in Proceedings of IWQoS, 2006.

[10] K. Angrishi and U. Killat, “On the linear scaling of stochastic performance bounds in network calculus,” 2011, submitted for publication.

[11] Y. Jiang, “A basic stochastic network calculus,” in Proceedings of ACM SIGCOMM, 2006, pp. 123–134.

[12] L. Kleinrock, Queueing Systems. John Wiley and Sons, 1975.