On Deriving Stable Backlog Bounds by Stochastic Network Calculus

Yue Wang

Department of Information Management, School of Information, Central University of Finance and Economics, Beijing, China
Email: yue.l.wang@gmail.com

Abstract—**Network calculus is a powerful methodology of characterizing queueing processes and has wide applications. In this work[1], we focus on the fundamental problem of "under what condition can we derive stable backlog bounds using the current state of art of stochastic network calculus". We model an network element (called a "node" here) as a single server with impairment service based on two best-known models in stochastic network calculus (one is first proposed by Cruz and the other is first proposed by Yaron and Sidi). We find that they actually derive equivalent stochastic service curves and backlog bounds. And we prove that stable backlog bounds can be derived by stochastic network calculus as long as the average rate of traffic arrival is less than that of service. This work suggests the effectiveness of stochastic network calculus in theory.**

I. Introduction

Network calculus provides an elegant way to characterize traffic and service processes of network and communication systems. Unlike traditional queueing theory in which one has to make strong assumptions on arrival or service processes (e.g., Poission arrival process, exponential service distribution, etc) so as to derive closed-form solutions in queueing networks [1], network calculus allows general arrival and service processes. Instead of getting exact solutions, one derives network backlog and delay bounds by network calculus. Deterministic network calculus is mature in theory [2] [3] [6] [7]. However, most traffic and service processes are stochastic and deterministic network calculus is often not applicable to them. Therefore, stochastic network calculus was proposed to deal with stochastic arrival and service processes [7]- [20].

In this paper, we focus on the fundamental problem of *"under what condition can we derive stable backlog bounds using the current state of art of stochastic network calculus"*. By stable backlog bounds, we mean that the mean backlog amount calculated by the backlog bounds is a finite number (see Definition 8).

We model an network element (called a "node" here) as an ideal server with an impairment process based on the two best-known models in stochastic network calculus: **model A** (adopted by [12] [18] and etc) and **model B** (adopted by [11] [19] and etc). And we make the following contributions:

- We compare model A and B and prove that they actually derive equivalent stochastic service curves and backlog bounds.
- We prove that stable backlog bounds can be derived by stochastic network calculus as long as the average rate of traffic arrival is less than that of service.

Note that when we prove a statement for ourselves, we call it *Propositions* to differentiate the existing *Theorems* in the literature (see Proposition 1-4).

This paper is organized as follows. In Section II, we give a brief overview of stochastic network calculus. In particular, we present the classic models (namely model A and B), and we also discuss the martingale and independent case analysis techniques. In Section III, we present the two network calculus models. We compare the two models and find that they are equivalent in derivation stochastic service curves and backlog bounds. We also prove the stability condition by the theory of stochastic network calculus in this section. In Section V, we give related works and highlight our contributions. Finally, Section VI concludes the paper.

II. Review of Stochastic Network Calculus

In this section, we first review basic terms of network calculus and then cite some results of the stochastic network calculus theory used in our paper. Jiang classified stochastic arrival curves as the types of *ta (traffic amount centric)*, *vb (virtual backlog centric)* and *mb (max virtual backlog centric)*, and classified stochastic service curve as *ws (weak stochastic)* and *sc (stochastic)*. In this paper, we adopt ta and vb arrival curves as well as the ws service curve, as currently they provides tightest backlog bounds[2]. Note that we just say *"stochastic service curve"* in our paper which means the *ws* one.

A. Basic Terms of Network Calculus

We consider a discrete time system where time is slotted ($t = 0, 1, 2, ...$). A process is a function of time t. By default, we use $A(t)$ to denote the *arrival process* to a network element

[1]This research was supported by National Natural Science Foundation of China (No. 61309030, No. 61100112), Beijing Planning Office of Philosophy and Social Science (No. 12JGA014) and the Discipline Construction Foundation of Central University of Finance and Economics.

[2]As recently known by the network calculus community the bounding probability of a mb arrival curve is either 0 or 1 [22]. The usage of the sc service curve is also restrictive as it is often derived from an impairment process with the mb arrival curve.

978-0-9836283-7-8 © 2013 ITC

with $A(0) = 0$. $A(t)$ is the total amount of traffic arrived to this network element up to time t. We use $A^*(t)$ to denote the *departure process* of the network element with $A^*(0) = 0$. $A^*(t)$ is the total amount of traffic departed from the network element up to time t. Let $\mathcal{F}$ ($\bar{\mathcal{F}}$) represents the set of non-negative wide-sense increasing (decreasing) functions. Clearly, $A(t) \in \mathcal{F}$ and $A^*(t) \in \mathcal{F}$. For any process, say $A(t)$, we define $A(s,t) \equiv A(t) - A(s)$, for $s \leq t$. We define the backlog of the network element at time t by

$$B(t) = A(t) - A^*(t), \tag{1}$$

and the delay of the network element at t by

$$D(t) = \inf\{d : A(t) \leq A^*(t+d)\}. \tag{2}$$

Fig. 1 illustrates an example of $A(t)$ and $A^*(t)$ with $B(t)$ and $D(t)$ at $t = 10$.

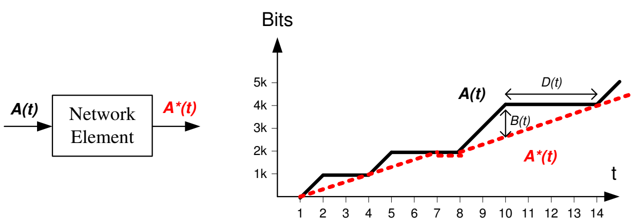

Fig. 1. Illustration of $A(t)$, $A^*(t)$, $B(t)$ and $D(t)$

In deterministic network calculus, $A(t)$ can be upper-bounded by an arrival curve. That is, for all $0 \leq s \leq t$, we have

$$A(s,t) \leq \alpha(t-s),$$

where $\alpha(t)$ is called the *arrival curve* of $A(t)$.

We say, *busy period* is a time period during which the backlog in the network element is always nonzero. For any busy period $(t_0, t]$, suppose we have

$$A^*(t) - A^*(t_0) \geq \beta(t - t_0),$$

if the network element provides a guaranteed service lower-bounded by $\beta(t-t_0)$ during the busy period. We can let t_0 be the beginning of the busy period, that is, the backlog at t_0 is zero or $A^*(t_0) = A(t_0)$. Therefore,

$$A^*(t) - A(t_0) \geq \beta(t - t_0).$$

The above equation infers $A^*(t) \geq \inf_{0 \leq s \leq t} [A(s) + \beta(t-s)]$, which can be written as

$$A^*(t) \geq A \otimes \beta(t), \tag{3}$$

where $\otimes$ is called the operator of *min-plus convolution* and $\beta(t)$ is called the *service curve* of the network element.

B. Stochastic Network Calculus on Bounding Traffic

We consider a server S (i.e. the network element) fed with a flow A. In practice, A's traffic and S's service are often stochastic, which can not be hard bounded by some curves. That is, they can violate the curves but with certain probabilities (we call it *bounding function* here). The theory of stochastic network calculus can get probabilistic bounds for backlogs and delays of the server, suppose we can characterize A by a stochastic arrival curve and S by a stochastic service curve.

In this section, we just consider the derivation of backlog bounds as delay bounds are quite similar to the former. We first give some definitions. Then we cite some results of model A and B. Finally, we make a brief discussion on them.

1) **Definitions***:*

Definition *1 (ta stochastic arrival curve):* A flow is said to have a *ta (traffic-amount-centric) stochastic arrival curve* $\alpha \in \mathcal{F}$ with bounding function $f \in \bar{\mathcal{F}}$, denoted by $A \sim_{ta} < f, \alpha >$, if for all $s, t \geq 0 (s \leq t)$ and all $x \geq 0$, there holds

$$P\{A(s,t) - \alpha(t-s) > x\} \leq f(x). \tag{4}$$

Definition *2 (vb stochastic arrival curve):* A flow is said to have a *vb (virtual-backlog-centric) stochastic arrival curve* $\alpha \in \mathcal{F}$ with bounding function $f \in \bar{\mathcal{F}}$, denoted by $A \sim_{vb} < f, \alpha >$, if for all $t \geq 0$ and all $x \geq 0$, there holds

$$P\{\sup_{0 \leq s \leq t} [A(s,t) - \alpha(t-s)] > x\} \leq f(x). \tag{5}$$

We can see that $A \sim_{vb} < f, \alpha >$ implies $A \sim_{ta} < f, \alpha >$, since $P\{A(s,t) - \alpha(t-s) > x\} \leq P\{\sup_{0 \leq s \leq t}[A(s,t) - \alpha(t-s)] > x\}$.

Definition *3 (Stochastic Service Curve):* A server S is said to provide a *(weak) stochastic service curve* $\beta \in \mathcal{F}$ with bounding function $g \in \bar{\mathcal{F}}$, denoted by $S \sim_{ws} < g, \beta >$ (or just $S \sim < g, \beta >$), if for all $t \geq 0$ and all $x \geq 0$, there holds

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x). \tag{6}$$

Definition *4 (Leftover Service):* Consider a server S provides the ideal service curve $\hat{\beta}(t)$ with the impairment process I to a flow. Then, during any backlogged period $(s,t]$, the output flow $A^*(s,t)$ from the server satisfies

$$A^*(s,t) \geq \hat{\beta}(t-s) - I(s,t). \tag{7}$$

$\hat{\beta}(t) - I(t)$ is the *leftover service* received by the given flow.

The definition of leftover service (also called *stochastic strict server* in [18]) can be applied to many scenarios such as cross traffic and wireless channels.

Definition *5 (θ-MER):* A process A's *minimum envelope rate with respect to θ (θ-MER)*, denoted by $\rho^*(\theta)$, is defined as follows:

$$\rho^*(\theta) = \lim_{t \to \infty} \frac{1}{\theta t} \sup_{s \geq 0} \log \mathrm{E} e^{\theta A(s,s+t)}. \tag{8}$$

We say that A has an *envelope rate with respect to θ* (θ-ER), denoted by $\rho(\theta)$, if $\rho(\theta) \geq \rho^*(\theta)$.

Definition *6 ($(\sigma(\theta), \rho(\theta))$-upper constrained):* A process A is said to be $(\sigma(\theta), \rho(\theta))$*-upper constrained* for some $\theta > 0$, if for all $0 \leq s \leq t$, we have

$$\frac{1}{\theta}\log \mathrm{E}e^{\theta A(s,t)} \leq \rho(\theta)(t-s)+\sigma(\theta). \quad (9)$$

We can derive stochastic arrival and service curves from the $(\sigma(\theta),\rho(\theta))$-upper constrained characterization (Section II-C.

Definition *7 (Average Rate):* The average rate of a process A, denoted by a_A, is defined as

$$a_A = \lim_{t\to\infty}\sup_{s\geq 0}\frac{\mathrm{E}A(s,s+t)}{t}. \quad (10)$$

Definition *8 (Stable Backlog Bound):* The backlog $B(t)$ is *stable*, if for all t,

$$\mathrm{E}B(t) < \infty, \quad (11)$$

We say that *the backlog bounds are stable* if they can derive stable backlogs.

2) **Model A's Backlog Bounds***:* Model A deals with *vb arrival curves* and stochastic service curves. We have the following theorems for leftover service curves and backlog bounds.

Theorem *1 (Model A's Leftover Stochastic Service Curve):* Suppose a server S providing the ideal service curve $\hat{\beta}(t)$ with the impairment process I. If I has a *vb stochastic arrival curve*, i.e., $I \sim_{vb} < g,\gamma >$, then the server provides the flow the leftover stochastic service curve $S \sim < g,\beta >$ and

$$\beta(t) = \hat{\beta}(t) - \gamma(t). \quad (12)$$

Theorem *2 (Model A's Backlog Bounds):* If the flow A has a *vb stochastic arrival curve* $A \sim_{vb} < f,\alpha >$ and the server S provides a stochastic service curve $S \sim < g,\beta >$ to the flow, then the *backlog* $B(t)$ of the flow in the server at time t satisfies:

$$P\{B(t) > x + \sup_{s\geq 0}[\alpha(s)-\beta(s)]\} \leq f \otimes g(x), \quad (13)$$

for all $t \geq 0$ and all $x \geq 0$.

3) **Model B's Backlog Bounds***:* Model B deals with *ta arrival curves* and stochastic service curves.

In fact, we can derive vb arrival curves from ta arrival curves by introducing the function $\delta(t) = \delta \cdot t$ (δ is an adjustable constant). The following Lemma 1 states this [14].

Lemma *1 (ta to vb Arrival Curves):* Suppose A is a ta stochastic arrival curve, $A \sim_{ta} < f,\alpha >$, then $A \sim_{vb} < \tilde{f}, \alpha_{\delta >}$ with $\alpha_\delta(t) \equiv \alpha(t) + \delta t$ and its bounding function $\tilde{f}(x,\delta) = \sum_{k=0}^{\infty} f(x+k\delta)$ (suppose the sum is finite).

The derivations are as follows.

$$\begin{aligned}
&P\{\sup_{0\leq s\leq t}[A(s,t)-\alpha_\delta(t-s)] > x\} \\
&\leq \sum_{s=0}^{t} P\{A(s,t)-\alpha_\delta(t-s) > x\} \\
&= \sum_{s=0}^{t} P\{A(s,t)-\alpha(t-s) > x + \delta(t-s)\} \\
&\leq \sum_{s=0}^{t} f(x+\delta(t-s)) \leq \sum_{k=0}^{\infty} f(x+k\delta). \quad (14)
\end{aligned}$$

We have the following theorems for the leftover service curves and backlog bounds in model B. Actually, we can derive these results by first converting ta arrival curves to vb ones and then applying model A theorems.

Theorem *3 (Model B's Leftover Stochastic Service Curve):* Suppose a server S providing the ideal service curve $\hat{\beta}(t)$ with the impairment process I. If I has a *ta stochastic arrival curve*, i.e., $I \sim_{ta} < g,\gamma >$, then the server provides the flow the leftover stochastic service curve $S \sim < \tilde{g},\beta >$ and

$$\beta(t) = \hat{\beta}(t) - \gamma_\delta(t), \quad (15)$$

where $\gamma_\delta(t) \equiv \gamma(t) + \delta t$ and $\tilde{g}(x,\delta) \equiv \sum_{k=0}^{\infty} g(x+k\delta)$ by definition.

Theorem *4 (Model B's Backlog Bounds):* If the flow A has a *ta stochastic arrival curve* $A \sim_{ta} < f,\alpha >$ and the server S provides a stochastic service curve $S \sim < g,\beta >$ to the flow, then the *backlog* $B(t)$ of the flow in the server satisfies: for all $t \geq 0$ and all $x \geq 0$,

$$P\{B(t) > x + \sup_{s\geq 0}[\alpha_\delta(s)-\beta(s)]\} \leq \tilde{f} \otimes g(x), \quad (16)$$

where $\alpha_\delta(t) \equiv \alpha(t) + \delta t$ and $\tilde{f}(x,\delta) = \sum_{k=0}^{\infty} f(x+k\delta)$ by definition.

Note that model B can deal with ta arrival curves while model A can not, by introducing a $\delta > 0$ to trade smaller service curve for larger bounding functions.

C. Computation of Stochastic Arrival/Service Curves

We will show in this subsection how to calculate stochastic arrival and service curves from the $(\sigma(\theta),\rho(\theta))$*-upper constrained* characterization [7].

Theorem *5 (Arrival Curves of* $(\sigma(\theta),\rho(\theta))$-upper constrained*):* Suppose $A(t)$ is $(\sigma(\theta),\rho(\theta))$-upper constrained, then it has a ta stochastic arrival curve $A \sim_{ta} < f,\alpha >$, where

$$\begin{aligned}
\alpha(t) &= r \cdot t \\
f(x) &= e^{\theta\sigma(\theta)} \cdot e^{-\theta x}, \quad (17)
\end{aligned}$$

for any $r \geq \rho(\theta)$ and $x \geq 0$. And A has a vb stochastic arrival curve $A \sim_{vb} < f,\alpha >$, where

$$\begin{aligned}
\alpha(t) &= r \cdot t \\
f(x) &= \frac{e^{\theta\sigma(\theta)}}{1-e^{\theta(\rho(\theta)-r)}} \cdot e^{-\theta x}, \quad (18)
\end{aligned}$$

for any $r > \rho(\theta)$ and $x \geq 0$.

Note that we have $r \geq \rho(\theta)$ in ta and $r > \rho(\theta)$ in vb. And Eq.(18) applies Boole's inequality to the bounding functions $f(x)$ which are loose in general.

How to derive stochastic service curves? If we can model the server S with the ideal service curve $\hat{\beta}$ with the impairment process $I(t)$, we can first characterize $I(t)$ by vb (ta) arrival curves, and then we use Theorem 1 (Theorem 3) to get its stochastic service curves.

The following theorem states the relation between θ-ER and $(\sigma(\theta), \rho(\theta))$-upper constrained. We will use it in proving the stability condition of backlog bounds in Section IV-B.

Theorem *6 (θ-ER vs $(\sigma(\theta), \rho(\theta))$-upper constrained):* If the process $A(t)$ has a θ-envelop rate (θ-ER) $\rho(\theta) < \infty$, then for every $\epsilon > 0$ there exists $\sigma_\epsilon(\theta) < \infty$ so that A is $(\sigma_\epsilon(\theta), \rho(\theta) + \epsilon)$-upper constrained.

D. Bounds Improvement

There are two ways of improving bounds in current literature. One way is to apply independent case analysis. The other way is to improve the bounding functions of stochastic service curves for time-independent arrivals.

The first way says that suppose the impairment process of the server S is independent from the traffic arrival process, we can derive tighter backlog bounds using independent probability analysis.

Theorem *7 (Backlog Bounds under Independent Cases):* Suppose the server S provides the flow (satisfying $A \sim_{vb}< f, \alpha >$) the ideal service curve $\hat{\beta}(t)$ with the impairment process $I \sim_{vb}< g, \gamma >$ (thus $S \sim< g, \beta >$ where $\beta(t) = \hat{\beta(t)} - \gamma(t)$). Suppose *$A$ and I are independent*, we have

$$\begin{aligned} & P\{B(t) > \sup_{s \geq 0}[\alpha(s) - \beta(s)] + x\} \\ & \leq \sum_{k=0}^{x} (\bar{g}(k) - \bar{g}(k-1)) \bar{f}(x-k) \end{aligned} \tag{19}$$

where $\bar{f}(x) = 1 - f(x)$, $\bar{g}(x) = 1 - g(x)$, and we set $\bar{g}(-1) = 0$.

This theorem of independent case analysis can be applied to model B. However, we first need to convert the ta arrival curves to the vb ones with new bounding functions $\hat{f}(x, \delta)$ and $\hat{g}(x, \delta)$ by Lemma 1. Then we apply the above theorem by plugging in $\hat{f}$ and $\hat{g}$.

Another way of tightening backlog bounds is to derive tighter bounding functions of stochastic arrival and service curves. Ciucu first proposed to use martingale to tighten the bounds for $M/M/1$ and $M/D/1$ queues [20]. In the following proposition, we provide a more general result following his idea. The proof is given in Appendix-A.

Proposition *1:* **[vb Arrival Curves of Time-Independent Process]** Suppose $A(t)$ is $(\sigma(\theta), \rho(\theta))$-upper constrained. On condition that $a(t) \equiv A(t) - A(t-1)$ is independent of each t, it has a vb stochastic arrival curve $A \sim_{vb}< f, \alpha >$, where

$$\begin{aligned} \alpha(t) &= r \cdot t \\ f(x) &= e^{-\theta x}, \end{aligned} \tag{20}$$

for any $r \geq \rho(\theta) + \sigma(\theta)$ and $x \geq 0$.

E. Discussion on Model A and B

The key difference between model A and B is: A uses vb traffic arrival curves while B uses ta ones variant stochastic service curves. Which one can derive tighter backlog bounds?

In general, ta arrival curves provide tighter bounding functions than vb. Actually, $A \sim_{vb}< f, \alpha >$ implies $A \sim_{ta}< f, \alpha >$ and the inverse is not true generally. In particular, the bounding function of ta is tighter than that of vb (especially when r is close to $\rho(\theta)$) in Theorem 5. But one can not conclude that model B is always better than model A, as it has looser bounding functions for leftover service curves and backlog bounds (see Theorem 3 and Theorem 4). The situation becomes even more uncertain when consider time-independent processes and independent A and I. Interestingly, we find that two models can derive equivalent stochastic service curves and backlog bounds in our studied case (Section IV-A).

III. A Node's Stochastic Network Calculus Model

In this section, we model a node by stochastic network calculus. In general, we can define one time slot ($t = 1$) to be any small duration and measure traffic amount in any unit (e.g. bits, bytes or packets).

Let $A(t)$ denote the traffic arrived at the node from the application layer. We assume A is $(\sigma_A(\theta_1), \rho_A(\theta_1))$-upper constrained, which is a right assumption for many cases.

We model the service of a node as an ideal server curve with an impairment process.Let the capacity of the node be c traffic units per slot. The departure process $A^*(s, t) = \hat{\beta}(s, t) - I(s, t)$ during any backlogged period $[s, t]$, where $\hat{\beta}(t) = c \cdot t$ is the ideal service curve and I is the impairment process. Since $I(s, t) \leq c \cdot (t - s)$, there exist $\sigma_I(\theta_2)$ and $\rho_I(\theta_2)$ so that I is $(\sigma_I(\theta_2), \rho_I(\theta_2))$-upper constrained. In here, θ_1 and θ_2 are adjustable parameters.

A. Model A's Backlog Bounds

Because A is $(\sigma_A(\theta_1), \rho_A(\theta_1))$-upper constrained, by Theorem 5, $A \sim_{vb}< f, \alpha >$ where

$$\begin{aligned} \alpha(t) &= r_A \cdot t \\ f(x) &= \frac{e^{\theta_1 \sigma_A(\theta_1)}}{1 - e^{\theta_1(\rho_A(\theta_1) - r_A)}} \cdot e^{-\theta_1 x}, \end{aligned} \tag{21}$$

for any $r_A > \rho_A(\theta_1)$.

In the same way, $I \sim_{vb}< g, \gamma >$ where

$$\begin{aligned} \gamma(t) &= r_I \cdot t \\ g(x) &= \frac{e^{\theta_2 \sigma_I(\theta_2)}}{1 - e^{\theta_2(\rho_I(\theta_2) - r_I)}} \cdot e^{-\theta_2 x}, \end{aligned} \tag{22}$$

for any $r_I > \rho_I(\theta_2)$.

By Theorem 1, the node provides a stochastic service curve $S \sim< g, \beta >$, where

$$\beta(t) \quad = \quad (c - r_I) \cdot t, \tag{23}$$

for any $c > r_I$.

Finally, by Theorem 2, we must let $\alpha(t) \leq \beta(t)$, i.e., $r_A \leq c - r_I$, in order to get meaningful backlog bounds which are $P\{B(t) > x\} \leq f \otimes g(x)$.

We note that $f(x)$ ($g(x)$) is the decreasing function of r_A (r_I). Considering the above conditions, we get the following optimal backlog bounds,

$$\begin{aligned}
&P\{B(t) > x\} \leq \min_{\theta_1,\theta_2,r_A,r_I} [f \otimes g(x)] \\
&\text{subject to} \\
&r_A > \rho_A(\theta_1), r_I > \rho_I(\theta_2) \\
&r_A + r_I = c \\
&\theta_1, \theta_2 > 0.
\end{aligned} \tag{24}$$

In here, $\rho_A(\theta_1)$ ($\rho_I(\theta_2)$) is the function of θ_1 (θ_2).

B. Model B's Backlog Bounds

Because A is $(\sigma_A(\theta_1), \rho_A(\theta_1))$-upper constrained, by Theorem 5, $A \sim_{ta}< f, \alpha >$, where

$$\begin{aligned}
\alpha(t) &= r_A \cdot t \\
f(x) &= e^{\theta_1 \sigma_A(\theta_1)} \cdot e^{-\theta_1 x},
\end{aligned} \tag{25}$$

for any $r_A \geq \rho_A(\theta_1)$.

In the same way, $I \sim_{ta}< g, \gamma >$, where

$$\begin{aligned}
\gamma(t) &= r_I \cdot t \\
g(x) &= e^{\theta_2 \sigma_I(\theta_2)} \cdot e^{-\theta_2 x},
\end{aligned} \tag{26}$$

for any $r_I \geq \rho_I(\theta_2)$.

By Lemma 1, we have $A \sim_{vb}< \tilde{f}, \alpha_{\delta_1} >$ where

$$\begin{aligned}
\alpha_{\delta_1}(t) &= (r_A + \delta_1) \cdot t \\
\hat{f}(x, \delta_1) &= \frac{e^{\theta_1 \sigma_A(\theta_1)}}{1 - e^{-\theta_1 \delta_1}} \cdot e^{-\theta_1 x},
\end{aligned} \tag{27}$$

for any $r_A \geq \rho_A(\theta_1)$ and $\delta_1 > 0$. In here, we can get the close form of $\tilde{f}(x, \delta) = \sum_{k=0}^{\infty} f(x + k\delta)$ for the particular $f(x)$ in Eq.(25).

In the same way, $I \sim_{vb}< \tilde{g}, \gamma_{\delta_2} >$ where

$$\begin{aligned}
\gamma_{\delta_2}(t) &= (r_I + \delta_2) \cdot t \\
\hat{g}(x, \delta_2) &= \frac{e^{\theta_2 \sigma_I(\theta_2)}}{1 - e^{-\theta_2 \delta_2}} \cdot e^{-\theta_2 x},
\end{aligned} \tag{28}$$

for any $r_I \geq \rho_I(\theta_2)$ and $\delta_2 > 0$.

By Theorem 3, the node provides a stochastic service curve $S \sim< \tilde{g}, \beta >$, where

$$\beta_{-\delta_2}(t) \quad = \quad (c - r_I - \delta_2) \cdot t, \tag{29}$$

for any $c > r_I + \delta_2$.

Finally, by Theorem 4, we must have $\alpha_{\delta_1}(t) \leq \beta_{-\delta_2}(t)$, i.e., $r_A + \delta_1 \leq c - r_I - \delta_2$ in order to get meaningful backlog bounds which are $P\{B(t) > x\} \leq \tilde{f} \otimes \tilde{g}(x)$. We note that $f(x)$ ($g(x)$) is the decreasing function of δ_1 (δ_2). Considering the above conditions, we get the following optimal backlog bounds,

$$\begin{aligned}
&P\{B(t) > x\} \leq \min_{\theta_1,\theta_2,\delta_1,\delta_2,r_A,r_I} [\tilde{f} \otimes \tilde{g}(x)] \\
&\text{subject to} \\
&r_A > \rho_A(\theta_1), r_I > \rho_I(\theta_2) \\
&r_A + r_I + \delta_1 + \delta_2 = c \\
&\theta_1, \theta_2 > 0, \delta_1, \delta_2 > 0.
\end{aligned} \tag{30}$$

In here, $\rho_A(\theta_1)$ ($\rho_I(\theta_2)$) is the function of θ_1 (θ_2).

IV. Stability Condition of Backlog Bounds

We show the derivation flow in Fig. 2.

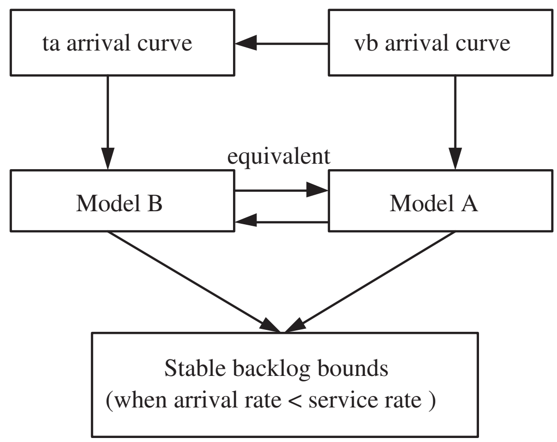

Fig. 2. The derivation flow of the stability condition of backlog bounds

A. Equivalency Condition of Model A and B

We find that the two models actually can derive the same stochastic service curves and backlog bounds. To understand this, we note that the key difference is the traffic model. The following proposition shows that we can derive the same vb arrival curves from the two models[3].

Proposition *2 (***Equivalency of ta and vb arrival curves):** If A is a $(\sigma(\theta), \rho(\theta))$-upper constrained process, its vb arrival curve immediately generated by applying Theorem 5's Eq.(18) and the one generated by applying Theorem 5's Eq.(17) and then Lemma 1 are equivalent.

Proof: Following the discussions above, the vb arrival curves generated immediately by applying Theorem 5's Eq.(18) are $A \sim_{vb}< f, \alpha >$ where

$$\begin{aligned}
\alpha(t) &= r \cdot t \\
f(x) &= \frac{e^{\theta \sigma(\theta)}}{1 - e^{\theta(\rho(\theta) - r)}} \cdot e^{-\theta x},
\end{aligned} \tag{31}$$

for any $r > \rho(\theta)$.

[3]Ciucu and Hohlfeld got the average backlog bounds in relationship to our proposition [21].

The vb arrival curves by applying Theorem 5's Eq.(17) and then Lemma 1 (converted by the ta arrival curves) are $A \sim_{vb} < \tilde{f}, \alpha_\delta >$ where

$$\begin{aligned} \alpha_\delta(t) &= (r+\delta)\cdot t \\ \hat{f}(x,\delta) &= \frac{e^{\theta\sigma(\theta)}}{1-e^{-\theta\delta}}\cdot e^{-\theta x}, \end{aligned} \tag{32}$$

for any $r \geq \rho(\theta)$ and $\delta > 0$.

For the same value of $(r+\delta)$ in Eq.(32), we should maximize δ to get tighter $\hat{f}(x,\delta)$; in other words, we should minimize r and let it to be $\rho(\theta)$. In this optimized case we find that Eq.(31) and Eq.(32) are in the same form. This establishes the equivalence between them. ∎

Proposition *3 (***Backlog bounds equivalency of two models***):* Consider a single server S with an ideal service curve $\hat{\beta}$ and an impairment process I. Suppose the traffic arrival process A and the impairment process I are $(\sigma(\theta), \rho(\theta))$-upper constrained for some θ respectively, then the vb arrival curves, the stochastic service curve and backlog bounds derived by Model A and B are equivalent.

Proof: Proposition 2 can be applied to the vb arrival curves of impairment processes of the server. Since model B can be derived from model A (see Section II-B3), the two models can derive exactly the same backlog bounds. ∎

Note: By equivalence, we do not mean it is general for all situations but derivation stable backlog bounds. Actually, model B can be extended to multiple concatenated nodes while model A can not, and they are different models. Even for the single-node case, we only prove the equivalence property for *linear* arrival curves. And it is still an open problem for more general cases.

B. Stability Condition

One fundamental question is under what condition we can derive *stable* backlog bounds (i.e., $\mathrm{E}B(t) < \infty$) by stochastic network calculus. The following proposition shows the stability condition. We adopt model A in our proof since the two models are equivalent in derivation backlog bounds.

Proposition *4 (Stability Condition):* Suppose there exist θ-MERs (θ-Minimum Envelop Rates) for the traffic arrival process A and the impairment process I of the node for $0 < \theta < \hat{\theta}$ where $\hat{\theta}$ is some constant value, then stochastic network calculus can derive stable backlogs if

$$a_A < c - a_I, \tag{33}$$

where c is the transmission rate of the ideal channel, a_A and a_I are the average rate of A and I defined in Definition 7, respectively.

Proof:

The proof consists of two phases. First, we show that $a_A < c - a_I$ can lead to $r_A \leq c - r_I$. Next, we show that if $r_A \leq c - r_I$ then stochastic network calculus can derive $\mathrm{E}B(t)$ which is less than a finite value.

We adopt model A in Section III-A where A is the traffic arrival process and I is the impairment process of the server S since the two models are equivalent(see Proposition 3). We have shown that $P\{B(t) > x\} \leq f \otimes g(x)$ if $r_A \leq c - r_I$ holds.

From Eq.(21) and (22), we let $\epsilon_1 = r_A - \rho_A(\theta_1)$ and $\epsilon_2 = r_I - \rho_I(\theta_2)$ for $\theta_1, \theta_2 > 0$ and $\epsilon_1, \epsilon_2 > 0$. To simplify the arguments, we let $\theta_1 = \theta_2 = \theta$ and $\epsilon_1 = \epsilon_2 = \epsilon$.

Thus, $r_A \leq c - r_I$ holds if

$$\rho_A(\theta) \leq c - \rho_I(\theta) - 2\epsilon. \tag{34}$$

From Theorem 6, we can construct the $(\sigma(\theta), \rho(\theta))$-upper constrained characterization by letting $\rho_A(\theta) = \rho_A^*(\theta) + \epsilon$ and $\rho_I(\theta) = \rho_I^*(\theta) + \epsilon$ for any $\epsilon > 0$, where $\rho_A^*(\theta_1)$ and $\rho_I^*(\theta_2)$ are θ-MERs of A and I, respectively. And Eq.(34) holds if

$$\rho_A^*(\theta) \leq c - \rho_I^*(\theta) - 4\epsilon. \tag{35}$$

Because $\rho_A^*(\theta)$ exists, applying Taylor's expansion,

$$\begin{aligned} &\rho_A^*(\theta) = \lim_{t\to\infty} \frac{1}{\theta t} \sup_{s\geq 0} \log \mathrm{E} e^{\theta A(s,s+t)} \\ &= \lim_{t\to\infty}\sup \frac{1}{\theta t} \sup_{s\geq 0} \log \mathrm{E}(1 + \theta A(s,s+t) + O(\theta^2 A(s,s+t)^2)) \\ &= \lim_{t\to\infty}\sup \frac{1}{\theta t} \sup_{s\geq 0} \log\left(1 + \theta \mathrm{E}A(s,s+t) + O(\theta^2 A(s,s+t)^2)\right) \\ &= \lim_{t\to\infty}\sup \frac{1}{\theta t} \sup_{s\geq 0} [\theta \mathrm{E}A(s,s+t) + O(\theta^2 A(s,s+t)^2)]. \end{aligned}$$

Let θ go to 0,

$$\lim_{\theta\to 0} \rho_A^*(\theta) = \lim_{t\to\infty} \sup_{s\geq 0} \frac{\mathrm{E}A(s,s+t)}{t} = a_A. \tag{36}$$

Similarly,

$$\lim_{\theta\to 0} \rho_I^*(\theta) = a_I. \tag{37}$$

Therefore, there exists some $\theta < \hat{\theta}$ so that $\rho_A^*(\theta) \leq a_A + \epsilon$ and $\rho_I^*(\theta) \leq a_I + \epsilon$. So Eq. (35) holds if

$$a_A \leq c - a_I - 6\epsilon. \tag{38}$$

Since ϵ can be arbitrarily small, Eq.(38) holds if

$$a_A < c - a_I. \tag{39}$$

Following the above derivations backwards, we prove that $a_A < c - a_I$ leads to $r_A < c - r_I$.

Next, we prove that stochastic network calculus can derive $\mathrm{E}B(t)$ which is less than a finite value if $r_A < c - r_I$.

Since $f(x)$ and $g(x)$ are exponentially decreasing functions according to Eq. (21) and Eq. (22), we can show that $\mathrm{E}B(t)$ is upper-bounded by some finite constant value as follows. Note that $B(t)$ is a discrete value in practice (e.g., in bits or packets).

$$
\begin{aligned}
\mathrm{E}B(t) &= \sum_{k=0}^{\infty} P\{B(t)=k+1\}\cdot(k+1) \\
&< \sum_{k=0}^{\infty} P\{B(t)>k\}\cdot(k+1) \\
&\le \sum_{k=0}^{\infty} f\otimes g(k)\cdot(k+1) \\
&\le \sum_{k=0}^{\infty} (f(\lfloor \frac{k}{2}\rfloor)+g(\lceil \frac{k}{2}\rceil))\cdot(k+1) < \infty. \qquad (40)
\end{aligned}
$$

$\blacksquare$

Remarks: Since the proof is based on the theory of stochastic network calculus, it indicates that we can get stable backlog bounds by stochastic network calculus on the condition that the average arrival rate is less than the average service rate. As this condition is very general, we think stochastic network calculus is effective in theory.

V. Related Work

The increasing demand on transmitting multimedia and other real time applications over the Internet has motivated the study of quality of service guarantees. Towards it, deterministic and stochastic network calculus has been recognized by researchers as a promising step.

Essentially, the network calculus is the theory of queueing systems that comes from the seminal work by Cruz on the (σ, ρ) traffic characterization [2] [3] and work on the service curve characterization of Generalized Processor Sharing (GPS) schedulers [4] [5]. The theory has been developed by many researchers since then. The elegance of network calculus is due to the fundamental convolution formulas (under the min-plus algebra) that determine the departure process of a system from its arrivals and its service curve. The notable strength of the min-plus convolution is the ability to concatenate tandem nodes along a network path, and therefore network calculus has the ability to characterize the whole network as a single server, which is generally intactable by traditional queueing theory [1]. Le Boudec's book covers deterministic network calculus and its applications in the Internet [6]. Chang's book substantially presented the first approaches to stochastic network calculus besides deterministic network calculus [7]. Jiang summarized different types of stochastic arrival and service curves in a unified framework and proposed a new stochastic network calculus model stemmed from mb (maximal backlog centric) arrival curves, although its application conditions have some unsolved controversy. Jiang also wrote a book on the theory of stochastic network calculus [8]. Ciucu proposed an effective stochastic service curve that can be applied to concatenated systems and calculating end-to-end delay and backlog bounds, which exhibits a good scaling property of $O(H \log H)$ where H is the number of nodes traversed by a flow [19]. Ciucu also showed that his model can derive quite accurate delay bounds in M/M/1 and M/D/1 queueing systems by using the martingale technique [20]. More recently, Fidler proposed a novel solution of the queue system using expectations instead of probabilities [29], and he also made a comprehensive survey on the recent progress of stochastic network calculus [9]. Besides, Jiang wrote an overview on this topic from the queueing principle perspective and he presented a nice outlook by discussing many open challenges [10]. And Ciucu demonstrates that the network calculus can capture actual system behavior tightly when applied carefully, but the unification of deterministic network calculus and stochastic network calculus remains an open [22].

Many works have applied network calculus, for example, in measurement-based admission control schemes [23], in conformance testing, [24], in wireless sensor networks [25], in Aloha systems [30], in speeding up network simulations [26] [27], in bandwidth estimation [28] and even in manufacturing blocking systems in management science [31].

Compared with the existing theories of stochastic network calculus, we study the stability of backlog bounds. We focus on two best known models: model A and B. Interestingly, we find that the two models can derive equivalent stochastic service curves and backlog bounds, and stable backlog bounds can be derived as long as the average rate of traffic arrival is less than that of service.

VI. Conclusion

In this paper, we model an network element as an ideal server with an impairment process based on two best-known models in stochastic network calculus, and we find that they actually derive equivalent stochastic service curves and backlog bounds. Finally, we prove that stable backlog bounds can be derived by stochastic network calculus as long as the average rate of traffic arrival is less than that of service, suggesting the effectiveness of stochastic network calculus in theory.

VII. Acknowledgement

We thank Prof. Yuming Jiang with NTNU, Prof. Florin Ciucu with TU-Berlin, and Dr. Kai Wang with Tsinghua and Caltech for their nice discussions on the theories of stochastic network calculus. We thank Prof. Jianming Zhu and Prof. Haiqi Feng in our school for holding related seminars.

References

[1] L. Kleinrock, "Queueing Systems. Volume 1: Theory," *Wiley Interscience*, 1975.

[2] R. L. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation," *IEEE Trans. Information Theory*, vol. 37(1), pp. 114-131, 1991.

[3] R. L. Cruz, "A Calculus for Network Delay, Part II: Network Analysis," *IEEE Trans. Information Theory*, vol. 37(1), pp. 132C141, 1991.

[4] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-node Case," *IEEE/ACM Trans. Networking*, vol. 1(3), pp. 344-357, 1993.

[5] A. K. Parekh and R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple-node Case," *IEEE/ACM Trans. Networking*, vol. 2(2), pp. 137-150, 1994.

[6] J.-Y. Le Boudec and P. Thiran, "Network Calculus: A Theory of Deterministic Queueing Systems for the Internet," *Springer-Verlag*, 2001.

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

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

[9] M. Fidler, "A Survey of Deterministic and Stochastic Service Curve Models in the Network Calculus," *IEEE Communications Surveys & Tutorials*, vol. 12(1), 2010.

[10] Y. Jiang, "Stochastic Network Calculus for Performance Analysis of Internet Networks C An Overview and Outlook," Proc. International Conference on Computing, Networking and Communications (ICNC), 2012.

[11] O. Yaron and M. Sidi, "Performance and stability of communication networks via robust exponential bounds," Proc. *IEEE/ACM Transactions on Networking*, vol. 1(3), pp.372-385, 1993.

[12] R. L. Cruz, "Quality of Service Guarantees in Virtual Circuit Switched Networks," IEEE J. Selected Areas Communication, vol. 13(6), pp. 1048-1056, 1995.

[13] C.-S. Chang, "On Deterministic Traffic Regulation and Service Guarantees: A systematic Approach by Filtering," *IEEE Trans. Information Theory*, vol. 44(3), pp. 1097-1110, 1998.

[14] Q. Yin, Y. Jiang, S. Jiang, and P. Y. Kong, "Analysis on generalized stochastically bounded bursty traffic for communication networks," Proc. IEEE Local Computer Networks (LCN), pp. 141-149, 2002.

[15] C. Li, A. Burchard, and J. Liebeherr, "A Network Calculus with Effective Bandwidth," Technical Report, CS-2003-20, University of Virginia, 2003; *IEEE/ACM Trans. Networking*, vol. 15(6), pp. 1442-1453, 2007.

[16] Y. Jiang and P. J. Emstad, "Analysis of Stochastic Service Guarantees in Communication Networks: A Traffic Model," Proc. *International Teletraffic Congress*, 2005.

[17] Y. Jiang and P. J. Emstad, "Analysis of Stochastic Service Guarantees in Communication Networks: A Server Model," Proc. *IEEE IWQoS*, 2005.

[18] Y. Jiang, "A Basic Stochastic Network Calculus," Proc. *ACM SIGCOMM*, pp. 123-134, 2006.

[19] F. Ciucu, A. Burchard, and J. Liebeherr, "A Network Service Curve Approach for the Stochastic Analysis of Networks," Proc. *ACM SIGMETRICS*, vol. 33(1), 2005.

[20] F. Ciucu, "Network Calculus Delay Bounds in Queueing Networks with Exact Solutions," Proc. *International Teletraffic Congress*, 2007.

[21] F. Ciucu and O. Hohlfeld, "On Computing Bounds on Average Backlogs and Delays with Network Calculus," Proc. *ICC*, 2010.

[22] F. Ciucu and J. Schmitt, "Perspectives on Network Calculus C No Free Lunch, but Still Good Value," Proc. *ACM SIGCOMM*, 2012.

[23] Y. Jiang, P. Emstad, A. Nevin, V. Nicola, and M. Fidler, "Measurement-Based Admission Control for a Flow-Aware Network," Proc. *EuroNGI 1st Conference on Next Generation Internet Networks - Traffic Engineering*, 2005.

[24] Y. Liu, C.-K. Tham, and Y. Jiang, "Comformance Analysis in Networks with Service Level Agreements," *Computer Networks*, vol. 47, pp. 885-906, 2005.

[25] J. B. Schmitt and U. Roedig, "Sensor Network Calculus - A Framework for Worst Case Analysis," Proc. *DCOSS*, pp. 141-154, 2005.

[26] H. Kim and J.C. Hou, "Network Calculus Based Simulation for TCP Congestion Control: Theorems, Implementation, and Evaluation," Proc. *IEEE INFOCOM*, pp. 2844-2855, 2004.

[27] H. Kim, "Integrating Network-Calculus-Based Simulation with Packet-Level Simulation for TCP-Operated Networks," *Computer Networks*, vol. 50(12), pp. 1995-2012, 2006.

[28] J. Liebeherr, M. Fidler, and S. Valaee, "A Min-Plus System Interpretation of Bandwidth Estimation," Proc. *IEEE INFOCOM*, pp. 1127-1135, 2007.

[29] M. Bredel and M. Fidler, "Understanding Fairness and Its Impact on Quality of Service in IEEE 802.11," Proc. *IEEE INFOCOM*, pp. 1098-1106, 2009.

[30] F. Ciucu, O. Hohlfeld, and P. Hui, "Non-Asymptotic Throughput and Delay Distributions in Multi-Hop Wireless Networks," Proc. *Allerton*, 2010.

[31] A. Bose, X. Jiang, B. Liu, and G. Li, "Analysis of Manufactoring Blocking Systems with Network Calculus," Performance Evaluation, vol. 63(12), pp. 1216C1234, 2006.

Appendix A: Proof of Proposition 1

Proof:

For a fixed t, we construct a stochastic process $X(s) = e^{\theta(A(t-s,t)-rs)}$ $(0 \leq s \leq t)$ and we have $X(s+1) = X(s)e^{\theta(A(t-s-1,t-s)-r)}$. We will show that if $a(s) \equiv A(s) - A(s-1)$ is independent for each time slot s and $r \geq \rho(\theta) + \sigma(\theta)$, then $X(s)$ is supermartingale, i.e., $\mathrm{E}[X(s+1)|X(0),...,X(s)] \leq X(s)$.

Because $a(s)$ is time-independent, we have

$$\begin{aligned} &\mathrm{E}[X(s+1)|X(0),...,X(s)] = X(s) \cdot \mathrm{E}[e^{\theta(a(t-s)-r)}] \\ &= X(s) \cdot e^{-\theta r} \cdot \mathrm{E}e^{\theta a(t-s)}. \end{aligned} \tag{41}$$

Because $A(t)$ is $(\sigma(\theta), \rho(\theta))$-upper constrained, we have $\mathrm{E}e^{\theta a(s)} \leq e^{\rho(\theta)+\sigma(\theta)}$ for all $s \geq 0$. When $r \geq \rho(\theta)+\sigma(\theta)$ and by Eq. (41), we have

$$\mathrm{E}[X(s+1)|X(0),...,X(s)] \leq X(s). \tag{42}$$

Thus, $X(s)$ is a supermartingale.

Doob's martingale inequality says that $P\{\sup_{0\leq s\leq t} X(s) \geq k\} \leq \frac{\mathrm{E}X(0)}{k}$ when $X(s)$ is a supermartingale (note: $\mathrm{E}X(0) = 1$ here) for any constant k. Let $k = e^x$, we have

$$\begin{aligned} &P\{\sup_{0\leq s\leq t}[(A(s,t) - r(t-s)] > x\} \\ &= P\{\sup_{0\leq s\leq t}[X(s)] \geq e^x\} \leq e^{-x}. \end{aligned} \tag{43}$$

■