# Stochastic Service Guarantee Analysis Based on Time-Domain Models

Jing Xie*, and Yuming Jiang*†
*Department of Telematics
†Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology, Trondheim, Norway N-7491
E-mails: jingxie@item.ntnu.no, ymjiang@ieee.org



***Abstract*—Stochastic network calculus is a theory for stochastic service guarantee analysis of computer communication networks. In the current stochastic network calculus literature, its traffic and server models are typically based on the cumulative amount of traffic and cumulative amount of service respectively. However, there are network scenarios where the applicability of such models is limited, and hence new ways of modeling traffic and service are needed to address this limitation. This paper presents *time-domain* models and results for stochastic network calculus. Particularly, we define traffic models, which are based on probabilistic lower-bounds on *cumulative packet inter-arrival time*, and server models, which are based on probabilistic upper-bounds on *cumulative packet service time*. In addition, examples demonstrating the use of the proposed time-domain models are provided. On the basis of the proposed models, the five basic properties of stochastic network calculus are also proved, which implies broad applicability of the proposed *time-domain* approach.**

## I. Introduction

Stochastic network calculus is a theory dealing with queueing systems found in computer communication networks [4][9][11][13]. It is particularly for analyzing networks where service guarantees are provided stochastically. Such networks include wireless networks, multi-access networks and multimedia networks where applications can tolerate some certain violation of the desired performance [7][8].

Stochastic network calculus is based on properly defined traffic models [3][11][13][14][17][18] and server models [11][13]. In the existing models of stochastic network calculus, an arrival process and a service process are typically modeled by some stochastic arrival curve, which probabilistically upper-bounds the *cumulative amount of arrival*, and respectively by some stochastic service curve, which probabilistically lower-bounds the *cumulative amount of service*. In this paper, we call such models *space-domain* models. Based on the *space-domain* traffic and server models, a lot of results have been derived for stochastic network calculus. Among the others, the most fundamental ones are the five basic properties [11] [13]: (P.1) *Service Guarantees* including delay bound and backlog bound; (P.2) *Output Characterization*; (P.3) *Concatenation Property*; (P.4) *Leftover Service*; (P.5) *Superposition Property*. Examples demonstrating the necessity of having these basic properties and their use can be found [11] [13].

Nevertheless, there are still many open research challenges for stochastic network calculus, and a critical one is *time-domain modeling and analysis* [13]. Time-domain modeling for service guarantee analysis has its root from the deterministic Guaranteed Rate (GR) server model [10], where service guarantee is captured by comparing with a (deterministic) virtual time function in the time-domain. This time-domain model has been extended to design aggregate-scheduling networks to support per-flow (deterministic) service guarantees [6][12], while few such results are available from space-domain models. Other network scenarios where time-domain modeling may be preferable include wireless networks and multi-access networks. In wireless networks, the varying wireless link condition can cause the sender fail to send when the link condition is ‘bad’ and then the sender may have to hold until the link state becomes ‘good’. For such cases, characterizing the service process is difficult in the space-domain, while much easy in the time-domain. In contention-based multi-access networks, backoff schemes are often employed to reduce collision occuring. While it is quite cumbersome for a space-domain server model to characterize the service taking into account the backoff process, the time-domain server models well suit the need. Having said these, however, how to define a stochastic version of the virtual time function and how to perform the corresponding analysis are yet open [13].

The objective of this paper is to define traffic models and server models in the *time-domain* and derive the corresponding five basic properties for stochastic network calculus. Particularly, we define traffic models that are based on probabilistic lower bounds on *cumulative packet inter-arrival time*. Also, we define server models that are based on some virtual time function and probabilistic upper bounds on *cumulative packet service time*. In addition, we establish relationships among the proposed time-domain models, and the mappings between the proposed time-domain models and the existing space-domain models. Furthermore, we prove the five basic properties based on the proposed time-domain models.

The remainder is structured as follows. Sec. II introduces the mathematical background and fundamental space-domain models and relevant results of stochastic network calculus. In Sec. III, we first introduce the time-domain deterministic traffic and server models, and then extend them to stochastic versions.

In addition, the relationships among them as well as with some existing space-domain models are established. Sec. IV explores the five basic properties. Finally, Sec. V summarizes the work.

## II. Notation and Relevant Background

To ease expression, we assume networks with **fixed unit length**[1] packets. By convention, we assume that a packet is considered to be received by a network element when and only when its last bit has arrived to the network element, and a packet is considered out of a network element when and only when its last bit has been transmitted by the network element. A packet can be served only when its last bit has arrived. All queues are assumed to be empty at time $0$. Packets within a flow are served in the first-in-first-out (FIFO) order.

### A. Notation

Let $p^n$, $r(n)$, $a(n)$ and $d(n)$ $(n = 0, 1, 2, ...)$ denote the $n^{th}$ packet of a flow, its allocated service rate, its arrival time and its departure time, respectively. Let $\mathcal{A}(t)$ and $\mathcal{A}^*(t)$ respectively denote the number of cumulative arrival packets and the number of cumulative departure packets by time $t$. By convention, we assume $a(0) = 0$, $d(0) = 0$, $\mathcal{A}(0) = 0$ and $\mathcal{A}^*(0) = 0$. For any $0 \le s \le t$, we denote $\mathcal{A}(s,t) \equiv \mathcal{A}(t) - \mathcal{A}(s)$ and $\mathcal{A}^*(s,t) \equiv \mathcal{A}^*(t) - \mathcal{A}^*(s)$.

In this paper, $a(n)$ and $\mathcal{A}(t)$ will be used to represent an arrival process interchangeably. A departure process will be represented by $d(n)$ and $\mathcal{A}^*(t)$ interchangeably.

The following function sets are often used in this paper. Specifically, we use $\mathcal{G}$ to denote the set of non-negative wide-sense increasing functions as follows:

$$\mathcal{G} = \{g(\cdot) : \forall 0 \le x \le y, 0 \le g(x) \le g(y); g(0) = 0\}$$

We denote by $\bar{\mathcal{G}}$ the set of non-negative wide-sense decreasing functions:

$$\bar{\mathcal{G}} = \{g(\cdot) : \forall 0 \le x \le y, 0 \le g(y) \le g(x)\}$$

Let $\bar{\mathcal{F}}$ denote the set of functions in $\bar{\mathcal{G}}$, where for each function $f(\cdot) \in \bar{\mathcal{F}}$, its nth-fold integration, denoted by $f^{(n)}(x) \equiv \big(\int_x^\infty dy\big)^n f(y)$, is bounded for any $x \ge 0$ and still belongs to $\bar{\mathcal{F}}$ for any $n \ge 0$, or

$$\bar{\mathcal{F}} = \big\{f(\cdot) : \forall n \ge 0, \big(\int_x^\infty dy\big)^n f(y)\big\}.$$

For ease of exposition, we adopt

$$[x]^+ \equiv \max[0,x] \quad and \quad [x]_1 \equiv \min[1,x],$$

and assume that for any bounding function $f(x)$, $f(x) = 1$ for any $x < 0$.

[1]The results can also be extended to networks with variable-length packets while the expression and results will be more complicated.

### B. Max-plus and Min-plus Algebra Basics

An essential idea of (stochastic) network calculus is to use alternate algebras particularly the min-plus algebra and max-plus algebra [2] to transform complex non-linear network systems into analytically tractable linear systems [13]. To the best of our knowledge, the existing models and results of stochastic network calculus are mainly under the *space-domain* and based on min-plus algebra that has basic operations particularly suitable for characterizing cumulative arrival and cumulative service. For characterizing arrival and service processes in the *time-domain*, interestingly, the max-plus algebra has basic operations that well suit the need.

In this paper, the following *max-plus* and *min-plus* operations will often be used:

- *Max-Plus Convolution* of $g_1$ and $g_1$ is
$$(g_1 \overline{\otimes} g_2)(n) = \sup_{0 \le m \le n} \{g_1(m) + g_2(n-m)\}$$
- *Max-Plus Deconvolution* of $g_1$ and $g_1$ is
$$(g_1 \overline{\oslash} g_2)(n) = \inf_{m \ge 0} \{g_1(n+m) - g_2(m)\}$$
- *Min-Plus Convolution* of $g_1$ and $g_1$ is
$$(g_1 \otimes g_2)(n) = \inf_{0 \le m \le n} \{g_1(m) + g_2(n-m)\}$$
- *Min-Plus Deconvolution* of $g_1$ and $g_1$ is
$$(g_1 \oslash g_2)(n) = \sup_{m \ge 0} \{g_1(n+m) - g_2(m)\}$$

In this paper, when applying *supremum* and *infimum*, they may be interpreted as *maximum* and *minimum* whenever appropriate, respectively.

### C. Random Variables and Stochastic Process Basics

**Lemma 1.** *For the sum of a collection of random variables $Z = \sum_{i=1}^n X_i$, no matter whether they are independent or not, there holds for the complementary cumulative distribution function (CCDF) of $Z$: (See Lemma 1.5 in [13])*

$$\bar{F}_Z(z) \le \bar{F}_{X_1} \otimes \cdots \otimes \bar{F}_{X_n}(z) \tag{1}$$

*where $\bar{F}_Z = P\{Z > z\}$, $-\infty < z \le \infty$.*

In this paper, we need some transformation between the number of cumulative arrival packets by time $t$ (i.e., $\mathcal{A}(t)$) and the time of a packet arriving to the system (i.e., $a(n)$). The transformation can be expressed by the following way. Consider a stochastic process. Let N(t) $(t \ge 0)$ denote the number of events occuring by time $t$ and $X(n)$ be the time of the $n^{th}$ event occuring. By convention, we assume $N(0) = 0$ and $X(0) = 0$. There exists an important relationship between $N(t)$ and $X(n)$ as follows [15]:

$$N(t) \ge n \quad \Leftrightarrow \quad X(n) \le t \tag{2}$$

From Eq.(2), we obtain

$$P\{N(t) \ge n\} = P\{X(n) \le t\} \tag{3}$$

If $N(t)$ is (probabilistically) upper-bounded with respect to some function $\alpha(t) \in \mathcal{G}$, we have the following relationships between $X(n)$ and $N(t)$:

**Lemma 2.** *(i) For function $\alpha(t) \in \mathcal{G}$, if $N(s,t) \le \alpha(t-s)$ for any $0 \le s \le t$, there holds*

$$N \otimes \alpha(t) = n \quad \Rightarrow \quad X \overline{\otimes} \lambda(n) \ge t \tag{4}$$

*where, $\lambda(n)$ is the inverse function of $\alpha(t)$ and is defined as follows:*

$$\lambda(n) = \inf\{t : \alpha(t) \ge n\}. \tag{5}$$

*(ii) Inversely, for function $\lambda(n) \in \mathcal{G}$, if $X(n) - X(m) \ge \lambda(n-m)$ for any $0 \le m \le n$, there holds*

$$X \overline{\otimes} \lambda(n) = t \quad \Rightarrow \quad N \otimes \alpha(t) \ge n \tag{6}$$

*where, $\alpha(t)$ is the inverse function of $\lambda(n)$ and is defined as follows:*

$$\alpha(t) = \sup\{n : \lambda(n) \le t\}. \tag{7}$$

*Proof:* (i) The condition is equivalent to enforcing $N(t) \le N \otimes \alpha(t) = n$. From Eq.(2), we have $X(n) \ge t$. As $X(n) \le X \overline{\otimes} \alpha^{-1}(n)$, we conclude $X \overline{\otimes} \alpha^{-1}(n) \ge t$.

(ii) The condition is equivalent to enforcing $X(n) \le X \overline{\otimes} \lambda(n) = t$. From Eq.(2), we obtain $N(t) \ge n$. As $N(t) \le N \otimes \lambda^{-1}(t)$, we have $N \otimes \lambda^{-1}(t) \ge n$. ■

### D. Relevant Results under Min-Plus Algebra

This sub-section reviews the basic *space-domain* traffic and server models of stochastic network calculus under min-plus algebra [13]. It is worth highlighting that all these models are for discrete time systems with unit discretization step.

The space-domain stochastic traffic models include v.b.c stochastic arrival curve and m.b.c stochastic arrival curve defined as follows:

**Definition 1.** ***(v.b.c Stochastic Arrival Curve).*** *A flow is said to have a virtual-backlog-centric (v.b.c) stochastic arrival curve $\alpha(t) \in \mathcal{G}$ with bounding function $f(x) \in \bar{\mathcal{G}}$, denoted by $\mathcal{A} \sim_{vb} \langle \alpha, f \rangle$, if for all $t \ge 0$ and all $x \ge 0$, there holds*

$$P\Big\{ \sup_{0 \le s \le t} \{\mathcal{A}(s,t) - \alpha(t-s)\} > x \Big\} \le f(x). \tag{8}$$

**Definition 2.** ***(m.b.c Stochastic Arrival Curve).*** *A flow is said to have a maximum-(virtual)-backlog-centric (m.b.c) stochastic arrival curve $\alpha(t) \in \mathcal{G}$ with bounding function $f_t(x) \in \bar{\mathcal{G}}$, denoted by $\mathcal{A}(t) \sim_{mb} \langle \alpha, f_t \rangle$, if for all $t \ge 0$ and all $x \ge 0$, there holds*

$$P\Big\{ \sup_{0 \le s \le t} \sup_{0 \le u \le s} \{\mathcal{A}(u,s) - \alpha(s-u)\} > x \Big\} \le f_t(x). \tag{9}$$

The space-domain stochastic server models include weak stochastic service curve and stochastic service curve defined as follows:

**Definition 3.** ***(Weak Stochastic Service Curve).*** *A system $\mathcal{S}$ is said to provide a weak stochastic service curve $\beta(t) \in \mathcal{G}$ with bounding function $g(x) \in \bar{\mathcal{G}}$, denoted by $S \sim_{ws} \langle \beta, g \rangle$, if for all $t \ge 0$ and all $x \ge 0$, there holds*

$$P\Big\{ \mathcal{A} \otimes \big(\beta(t) - x\big)^{+} - \mathcal{A}^*(t) > 0 \Big\} \le g(x). \tag{10}$$

**Definition 4.** ***(Stochastic Service Curve).*** *A system $\mathcal{S}$ is said to provide a stochastic service curve $\beta(t) \in \mathcal{G}$ with bounding function $g_t(x) \in \bar{\mathcal{G}}$, denoted by $S \sim_{sc} \langle \beta, g_t \rangle$, if for all $t \ge 0$ and all $x \ge 0$, there holds*

$$P\Big\{ \sup_{0 \le s \le t} [\mathcal{A} \otimes \big(\beta(s) - x\big)^{+} - \mathcal{A}^*(s)] > 0 \Big\} \le g_t(x) \tag{11}$$

Based on the above space-domain traffic and server models, a lot of results have been derived for stochastic network calculus which include the five basic properties [13]. In this paper, the following result will specifically be made use of in later analysis and is hence listed:

**Lemma 3.** ***(Superposition Property).*** *Consider $N$ flows with arrival processes $\mathcal{A}_i(t)$, $i = 1, ..., N$, respectively. Let $\mathcal{A}(t)$ denote the aggregate arrival process. If $\forall i$, $\mathcal{A}_i \backsim_{vb} \langle f_i, \alpha_i \rangle$, then $\mathcal{A} \backsim_{vb} \langle f, \alpha \rangle$ with $\alpha(t) = \sum_{i=1}^{N} \alpha_i(t)$, and $f(x) = f_1 \otimes \cdots \otimes f_N(x)$.*

## III. Time-Domain Models

This section first reviews the deterministic arrival curve and service curve models defined in the time-domain. Then, we generalize them and define *time-domain* stochastic arrival curve and stochastic service curve models.

### A. Deterministic Arrival Curve

Consider a flow of which packets arrive to a system at time $a(n)$. In order to deterministically guarantee a certain level of quality of service (QoS) to this flow, the traffic sent by this flow must be limited. The deterministic network calculus traffic model in the time-domain characterizes packet inter-arrival time using a lower-bound function, called arrival curve in this paper and defined as follows [5]:

**Definition 5.** ***(Arrival Curve).*** *A flow is said to have a (deterministic) arrival curve $\lambda(n) \in \mathcal{G}$, if its arrival process $a(n)$ satisfies, for all $0 \le m \le n$,*

$$a(n) - a(m) \ge \lambda(n-m). \tag{12}$$

The arrival curve model has the following triplicity principle which will be used as the basis in defining the stochastic arrival curve models in the subsequent subsections.

**Lemma 4.** *The following statements are equivalent:*

1) $\forall 0 \le m \le n$, $a(n) - a(m) \ge \lambda(n-m) - x$;
2) $\forall n \ge 0$, $\sup_{0 \le m \le n}\{\lambda(n-m) - [a(n) - a(m)]\} \le x$;
3) $\forall n \ge 0$, $\sup_{0 \le m \le n} \sup_{0 \le q \le m}\{\lambda(m-q) - [a(m) - a(q)]\} \le x$;

*for all $x \ge 0$, where $\lambda \in \mathcal{G}$.*

*Proof:* It is trivially true that $\lambda(n-m) - [a(n) - a(m)] \le \sup_{0 \le m \le n}\{\lambda(n-m) - [a(n) - a(m)]\}$, from which, (2) implies (1). In addition

$$\sup_{0 \le m \le n} \{\lambda(n-m) - [a(n) - a(m)]\}$$

$$\le \sup_{0\le m\le n}\sup_{m\le k\le n}\{\lambda(k-m)-[a(k)-a(m)]\}$$

$$=\sup_{0\le k\le n}\sup_{0\le m\le k}\{\lambda(k-m)-[a(k)-a(m)]\}$$

$$=\sup_{0\le m\le n}\sup_{0\le q\le m}\{\lambda(m-q)-[a(m)-a(q)]\}$$

with which, (3) implies (2). For (1)→(2), it holds since $a(n)-a(m)\ge\lambda(n-m)-x$ for all $0\le m\le n$. For (2)→(3), $\sup_{0\le m\le n}\sup_{0\le q\le m}\{\lambda(m-q)-[a(m)-a(q)]\}\le\sup_{0\le m\le n}[x]=x$. Thus (1), (2) and (3) are equivalent. ■

From Definition 5, the right-hand side of $a(n)-a(m)\ge\lambda(n-m)-x$ in Lemma 4.(1) defines an arrival curve $\lambda(n-m)-x$. In addition, we can construct a virtual single server queue (SSQ) system that is initially empty, fed with the same traffic flow, and have a service curve $\lambda$ which makes $d(n)\le a\overline{\otimes}\lambda(n)$ (see Definition 9). Then, the delay in the virtual SSQ system is upper-bounded by $d(n)-a(n)\le\sup_{0\le m\le n}[\lambda(n-m)-(a(n)-a(m))]\le x$, and the maximum system delay for the first $n$ packets is upper-bounded by $\sup_{0\le m\le n}\{d(m)-a(m)\}\le\sup_{0\le m\le n}\sup_{0\le q\le m}\{\lambda(m-q)-[a(m)-a(q)]\}\le x$.

In addition, Definition 5 is equivalent to enforcing that for all $n\ge 0$, there holds

$$a(n)\le\sup_{0\le m\le n}\Big\{a(m)+\lambda(n-m)\Big\}=a\overline{\otimes}\lambda(n). \tag{13}$$

**Example 1**. The Generic Cell Rate Algorithm (GCRA) [1] with parameter $(T,\tau)$ is a parallel algorithm to the Leaky Bucket algorithm and has been used in fixed-length packet networks such as Asynchronous Transfer Mode (ATM) networks. Here, $T$ is an ideal inter-arrival between packets and $\tau$ is the maximum acceptable excursion that quantifies how early packets may arrive with respect to $T$. It can be verified that if a flow is GCRA$(T,\tau)$-constrained, it has an arrival curve

$$\lambda(n)=T\cdot n-\tau.$$

### B. i.t Stochastic Arrival Curve

Lemma 4.(1) defines a deterministic arrival curve $\lambda(n)-x$ which lower-bounds the inter-arrival time between any two packets. Based on this, we define its probabilistic counterpart as follows:

**Definition 6.** ***(i.t Stochastic Arrival Curve).*** *A flow is said to have an inter-arrival-time (i.t) stochastic arrival curve $\lambda\in\mathcal{G}$ with bounding function $h\in\bar{\mathcal{G}}$, denoted by $a(n)\sim_{it}\langle\lambda,h\rangle$, if for all $n\ge 0$ and all $x\ge 0$, there holds*

$$P\Big\{\lambda(n-m)-[a(n)-a(m)]>x\Big\}\le h(x). \tag{14}$$

**Example 2.** Consider a flow with fixed unit packet size. Suppose its packet inter-arrival time follow an exponential distribution with mean $\frac{1}{\rho}$. Then, the packet arrival time has an Erlang distribution with parameter $(n,\rho)$. And, for any two packets $p^m$ and $p^n$, their inter-arrival time $a(n)-a(m)$ satisfies, for any $x\ge 0$,

$$P\Big\{\frac{1}{\rho}(n-m)-[a(n)-a(m)]>x\Big\}$$

$$\le 1-\sum_{k=0}^{n-m-1}\frac{e^{-\rho y}(\rho y)^k}{k!}-\rho\frac{e^{-\rho y}(\rho y)^{n-m-1}}{(n-m-1)!}$$

where $y=\frac{1}{\rho}(n-m)-x$.

The i.t stochastic arrival curve is intuitively simple, but it has limited use if no additional constraint if enforced. Let us consider a simple example to understand this problem. Consider a single node with constant per packet service time $T$ and with its input flow $F$ satisfying $a(n)\sim_{it}\langle\tau\cdot n,h\rangle$ where $\tau\ge T$. Suppose we are interested in the delay $D(n)$, where, by definition, $D(n)=d(n)-a(n)$. As the node has constant per packet service time $T$, it has a (deterministic) service curve $T\cdot n$. In other words, $d(n)\le\sup_{0\le m\le n}[a(m)+T\cdot(n-m)]$. Then we have

$$D(n)\le\sup_{0\le m\le n}\Big\{a(m)+T\cdot(n-m)\Big\}-a(n)$$

$$=\sup_{0\le m\le n}\Big\{a(m)+T\cdot(n-m)-a(n)\Big\}$$

$$\le\sup_{0\le m\le n}\Big\{\tau\cdot(n-m)-[a(n)-a(m)]\Big\} \tag{15}$$

From Eq.(15), we have difficulty in further deriving more results if no additional constraint is added because we only know $P\{\tau\cdot(n-m)-[a(n)-a(m)]\}\le h(x)$. When investigating the performance metrics such as delay bound and backlog bound in Section IV-A, we meet the similar difficulty.

### C. v.s.d Stochastic Arrival Curve

The previous subsection stated the difficulty of applying i.t stochastic arrival curve to service guarantee analysis. This subsection introduces another stochastic arrival curve model that can help avoid such difficulty. This model is called the *virtual-system-delay* ($v.s.d$) stochastic arrival curve. The model explores the *virtual system delay property* of deterministic arrival curve as implied by Lemma 4.(2), which is that the amount of time a packet spends in a virtual single server queue (SSQ) fed with the same flow with a deterministic arrival curve is lower-bounded.

For a flow having deterministic arrival curve, we can construct a virtual SSQ system fed with the flow, which has infinite buffer space and the buffer is initially empty. Then, suppose the virtual SSQ provides a deterministic service curve $\lambda$ to the flow or $d(n)=a\overline{\otimes}\lambda(n)$ for all $n\ge 0$. We now have that the amount of time packet $n$ spends in the virtual SSQ system is $W_s(n)=d(n)-a(n)=\sup_{0\le m\le n}\{\lambda(n-m)-[a(n)-a(m)]\}$. If the flow is constrained by arrival curve $\lambda(n)-x$ for all $n\ge 0$, it is obviously that $W_s$ is also lower-bounded by $x$.

Based on the virtual system time property, we define virtual-system-delay (v.s.d) stochastic arrival curve to characterize the arrival process as follows:

**Definition 7.** ***(v.s.d Stochastic Arrival Curve).*** *A flow is said to have a virtual-system-delay (v.s.d) stochastic arrival curve*

*$\lambda \in \mathcal{G}$ with bounding function $h \in \bar{\mathcal{G}}$, denoted by $a(n) \sim_{vd} \langle \lambda, h \rangle$, if for all $n \geq 0$ and all $x \geq 0$, there holds*

$$P\Big\{ \sup_{0\leq m\leq n} \{\lambda(n-m) - [a(n) - a(m)]\} > x \Big\} \leq h(x). \quad (16)$$

**Example 3.** Consider a flow with the same fixed packet size. Suppose all packet inter-arrival times are exponentially distributed with mean $\frac{1}{\mu}$. Based on the steady-state probability mass function (PMF) of the queue-waiting time for an M/D/1 queue [16], we say that the flow has a v.s.d stochastic arrival curve $a(n) \sim_{vd} \langle h^{exp}, D \cdot n \rangle$ for any $D < \frac{1}{\mu}$, with $\rho = \frac{\mu}{D}$ and

$$h^{exp}(x) = 1 - (1-\rho) \sum_{i=0}^{\lfloor x/D \rfloor + 1} e^{-\mu(-x)} \frac{[\mu(-x)]^i}{i!}$$

where, $\lfloor x/D \rfloor$ denotes the greatest integer less than or equal to $x/D$.

The following theorem establishes a relationship between i.t stochastic arrival curve and v.s.d stochastic arrival curve.

**Theorem 1.** 1) *If a flow has a v.s.d stochastic arrival curve $\lambda \in \mathcal{G}$ with bounding function $h \in \bar{\mathcal{G}}$, then the flow has an i.t stochastic arrival curve $\lambda \in \mathcal{G}$ with the same bounding function $h \in \bar{\mathcal{G}}$.*

2) *Conversely, if a flow has an i.t stochastic arrival curve $\lambda \in \mathcal{G}$ with bounding function $h \in \bar{\mathcal{F}}$, it also has a v.s.d stochastic arrival curve $\lambda_{-\eta} \in \mathcal{G}$ with bounding function $h^{\eta} \in \bar{\mathcal{G}}$ where*

$$\lambda_{\eta}(n) = \lambda(n) - \eta \cdot n$$

$$h^{\eta}(x) = \Big[h(x) + \frac{1}{\eta}\int_x^{\infty} h(y)dy\Big]_1$$

*for any $\eta > 0$.*

*Proof:* The first part follows easily from the fact that for any $0 \leq m \leq n$, $\lambda(n-m) - [a(n) - a(m)] \leq \sup_{0\leq m\leq n}\{\lambda(n-m) - [a(n) - a(m)]\}$. For the second part, there holds

$$\sup_{0\leq m\leq n} \{\lambda_{-\eta}(n-m) - [a(n) - a(m)]\}$$

$$\leq_{st} \sup_{0\leq m\leq n} \{\lambda_{-\eta}(n-m) - [a(n) - a(m)]\}^+$$

Since for any $x \geq 0$, $P\big\{\{\lambda(n-m) - \eta \cdot (n-m) - [a(n) - a(m)]\}^+ > x\big\} = P\big\{\{\lambda(n-m) - \eta \cdot (n-m) - [a(n) - a(m)]\} > x\big\} \leq h\big(x + \eta \cdot (n-m)\big)$, we have

$$P\Big\{ \sup_{0\leq m\leq n} \{\lambda_{-\eta}(n-m) - [a(n) - a(m)]\} > x \Big\}$$

$$\leq \sum_{m=0}^{n} P\Big\{ \{\lambda_{-\eta}(n-m) - [a(n) - a(m)]\}^+ > x \Big\}$$

$$\leq \sum_{m=0}^{n} h(x + \eta \cdot (n-m)) = \sum_{k=0}^{n} h(x + \eta \cdot k)$$

$$\leq \sum_{k=0}^{\infty} h(x + \eta \cdot k) = h(x) + \sum_{k=1}^{\infty} h(x + \eta \cdot k)$$

$$\leq h(x) + \frac{1}{\eta}\int_x^{\infty} h(y)dy. \quad (17)$$

which is meaningful only when Eq.(17) is upper-bounded by one. The 1-fold integration of $h(x)$ is bounded by one because the condition assumes $h \in \bar{\mathcal{F}}$ as for the [17]. Then the second part follows from Eq.(17). ■

Note that in the second part of the above theorem, $h(x) \in \bar{\mathcal{F}}$ while not $\in \bar{\mathcal{G}}$. If the requirement on the bounding function is relaxed to $h(x) \in \bar{\mathcal{G}}$, the above relationship may not hold in general.

The v.s.d stochastic arrival curve has a counterpart defined under min-plus algebra, the v.b.c stochastic arrival curve as defined in Definition 1. The following result establishes a relationship between these two models.

**Theorem 2.** 1) *If a flow has a v.b.c stochastic arrival curve $\alpha(t) \in \mathcal{G}$ with bounding function $f(x) \in \bar{\mathcal{G}}$, the flow has a v.s.d stochastic arrival curve $\lambda(n) \in \mathcal{G}$ with bounding function $h(x) \in \bar{\mathcal{G}}$ where $h(x) = f\big([\alpha(x) - y]^+\big)$ with $y = A \otimes \alpha(t) + \alpha(x) - n + 1 - A(t-x,t)$, and $\lambda(n) = \inf\{t : \alpha(t) \geq t\}$.*

2) *If a flow has a v.s.d stochastic arrival curve $\lambda(n) \in \mathcal{G}$ with bounding function $h(x) \in \bar{\mathcal{G}}$, the flow has a v.b.c stochastic arrival curve $\alpha(t) \in \mathcal{G}$ with bounding function $f(x) \in \bar{\mathcal{G}}$ where $f(x) = h\big([\lambda(x) - \theta]^+\big)$ and $\alpha(t) = \sup\{n : \lambda(n) \leq t\}$.*

*Proof:* For the first part, Eq.(2) implies that for any $x \geq 0$, if $a(n) < a\overline{\otimes}\lambda(n) - x$ there must be $\mathcal{A}(a\overline{\otimes}\lambda(n) - x) \geq n$ since otherwise if $\mathcal{A}(a\overline{\otimes}\lambda(n) - x) < n$ then $a(n) > a\overline{\otimes}\lambda(n) - x$, that would contradict the condition $a(n) < a\overline{\otimes}\lambda(n) - x$. In other words, event $\{a(n) < a\overline{\otimes}\lambda(n) - x\}$ implies event $\{\mathcal{A}(a\overline{\otimes}\lambda(n) - x) \geq n\}$ and thus

$$P\{a(n) < a\overline{\otimes}\lambda(n) - x\} \leq P\{\mathcal{A}(a\overline{\otimes}\lambda(n) - x) \geq n\} \quad (18)$$

Due to the assumption of fixed unit packet size and a packet being counted when and only when its last bit has arrived, event $\{\mathcal{A}(a\overline{\otimes}\lambda(n) - x) \geq n\}$ implies event $\{\mathcal{A}(a\overline{\otimes}\lambda(n) - x) > n-1\}$ and hence,

$$P\{\mathcal{A}(a\overline{\otimes}\lambda(n) - x) \geq n\} \leq P\{\mathcal{A}(a\overline{\otimes}\lambda(n) - x) > n-1\} \quad (19)$$

Let $a\overline{\otimes}\lambda(n) = t$. From Eq.(6), we know $\mathcal{A} \otimes \alpha(t) \geq n$. The right-hand side of Eq.(19) can be rewritten as $\mathcal{A}(t-x) > n-1$. As $\mathcal{A}(t-x) = \mathcal{A}(t) - \mathcal{A}(t-x,t)$, we have

$$\mathcal{A}(t) > n - 1 + \mathcal{A}(t-x,t) = \mathcal{A} \otimes \alpha(t) + \alpha(x) - y \quad (20)$$

where $y = \mathcal{A} \otimes \alpha(t) + \alpha(x) - n + 1 - \mathcal{A}(t-x,t)$. As the flow has a v.b.c stochastic arrival curve $\alpha$ with bounding function $f(x)$, we rewrite Eq.(8) as follows

$$P\{\mathcal{A}(t) - \mathcal{A} \otimes \alpha(t) > x\} \leq f(x)$$

Then, from Eq.(20), we obtain

$$P\Big\{\mathcal{A}(t) - \mathcal{A} \otimes \alpha(t) > [\alpha(x) - y]^+\Big\} \leq f\big([\alpha(x) - y]^+\big)$$

where, because of the restriction $f(x) = 1$ for any $x < 0$. According to Eq.(18), we conclude

$$P\big\{a(n) < a\overline{\otimes}\lambda(n) - x\big\} \leq f\big([\alpha(x) - y]^+\big).$$

where $a(n) < a\overline{\otimes}\lambda(n) - x = \sup_{0\leq m\leq n}\{\lambda(n-m) - [a(n) - a(m)] > x$. Thue, the first part is proved.

For the second part, let $\tau(i)$, i=1,2,..., denote inter-arrival time between the $(i-1)^{th}$ and the $i^{th}$ packets, then $a(n) = \sum_{i=1}^{n}\tau(i)$. Eq.(2) implies that for any $x \geq 0$, if $\mathcal{A}(t) > \mathcal{A}\otimes\alpha(t) + x$ there must be $a(\mathcal{A}\otimes\alpha(t) + x) \leq t$ since otherwise if $a(\mathcal{A}\otimes\alpha(t) + x) > t$ and $\mathcal{A}(t) < \mathcal{A}\otimes\alpha(t) + x$, that would contradict the condition $\mathcal{A}(t) > \mathcal{A}\otimes\alpha(t) + x$. In other words, event $\{\mathcal{A}(t) > \mathcal{A}\otimes\alpha(t)+x\}$ implies event $\{a(\mathcal{A}\otimes\alpha(t)+x) \leq t\}$, and thus

$$P\{\mathcal{A}(t) > \mathcal{A}\otimes\alpha(t) + x\} \leq P\big\{a\big(\mathcal{A}\otimes\alpha(t) + x\big) \leq t\big\} \quad (21)$$

For any $\theta > 0$ and $\theta \to 0$, event $\big\{a\big(\mathcal{A}\otimes\alpha(t) + x\big) \leq t\big\}$ implies event $\big\{a\big(\mathcal{A}\otimes\alpha(t) + x\big) < t + \theta\big\}$ and hence

$$P\big\{a\big(\mathcal{A}\otimes\alpha(t)+x\big) \leq t\big\} \leq P\big\{a\big(\mathcal{A}\otimes\alpha(t)+x\big) < t+\theta\big\} \quad (22)$$

Let $\mathcal{A}\otimes\alpha(t) = n$. From Eq.(4), we know $a\overline{\otimes}\lambda(n) \geq t$. In addition, $a(n+x) - a(n) = \sum_{i=n+1}^{n+x}\tau(i) \geq \lambda(x)$. Then the right-hand side of Eq.(22) can be rewritten as follows:

$$a(n) < t + \theta - \sum_{i=n+1}^{n+x}\tau(i) \leq a\overline{\otimes}\lambda(n) + \theta - \lambda(x) \quad (23)$$

As the flow has a $v.s.d$ stochastic arrival curve $\lambda(n)$ with bounding function $h(x)$, we rewrite Eq.(16) as follows

$$P\big\{a\overline{\otimes}\lambda(n) - a(n) > x\big\} \leq h(x)$$

From Eq.(23), we obtain

$$P\big\{a\overline{\otimes}\lambda(n) - a(n) > \lambda(x) - \theta\big\} \leq h\big(\lambda(x) - \theta\big)$$

According to Eq.(21), we conclude

$$P\{\mathcal{A}(t) > \mathcal{A}\otimes\alpha(t) + x\} \leq h\big(\lambda(x) - \theta\big)$$

where $\mathcal{A}(t) > \mathcal{A}\otimes\alpha(t)+x = \sup_{0\leq s\leq t}[\mathcal{A}(s,t) - \alpha(s,t)] > x$. Thus, the second part is proved.

■

### D. m.s.d Stochastic Arrival Curve

The maximum-(virtual)-system-delay (m.s.d) stochastic arrival curve explores the *maximum virtual system delay property* of deterministic arrival curve implied by Lemma 4.(3), which is that the maximum system delay of a virtual SSQ fed with the same flow with a deterministic arrival curve is lower-bounded.

Similar to the discussion for v.s.d stochastic arrival curve, for a flow having arrival curve, we can construct a virtual SSQ system fed with the flow, which has infinite buffer space and the buffer is initially empty. Then, suppose the virtual SSQ provides a deterministic service curve $\lambda$ to the flow or $d(n) = a\overline{\otimes}\lambda(n)$ for all $n \geq 0$. We have the maximum system delay in the virtual SSQ system for the first $n$ arrival packets as $\sup_{0\leq m\leq n} W_s(m) = \sup_{0\leq m\leq n}\sup_{0\leq q\leq m}\{\lambda(m-q) - [a(m) - a(q)]\}$. If the flow is constrained by arrival curve $\lambda(n) - x$ for all $n \geq 0$, it is clear that the maximum system delay in the virtual SSQ is also upper-bounded by $x$.

Based on the maximum virtual system delay property, we define m.s.d stochastic arrival curve as follows:

**Definition 8.** ***(m.s.d Stochastic Arrival Curve).*** *A flow is said to have a maximum-(virtual)-system-delay (m.s.d) stochastic arrival curve $\lambda(n) \in \mathcal{G}$ with bounding function $h(x) \in \bar{\mathcal{G}}$, denoted by $a(n) \sim_{md} \langle\lambda, h\rangle$, if for all $n \geq 0$ and all $x \geq 0$, there holds*

$$P\Big\{\sup_{0\leq m\leq n}\sup_{0\leq q\leq m}\big\{\lambda(m-q) - [a(m) - a(q)]\big\} > x\Big\} \leq h(x). \quad (24)$$

### E. Deterministic Service Curve

To provide service guarantees to an arrival-constrained flow $F$, the system usually needs to allocate a minimum service rate to $F$. A guaranteed minimum service rate is equivalent to a guaranteed maximum service time for each packet of the flow, and accordingly the packet's departure time from the system is bounded. As packets of the same flow are served in FIFO manner, any packet $p^n$ from this flow will depart by $\hat{d}(n)$, with $\hat{d}(n)$ iteratively defined by

$$\hat{d}(0) = 0$$
$$\hat{d}(n) = \max[a(n), \hat{d}(n-1)] + \delta(n) \quad (25)$$

where $\delta(n)$ is the service time guaranteed to $p^n$. By applying Eq.(25) iteratively to its right-hand side, it becomes

$$\hat{d}(n) = \sup_{0\leq m\leq n}[a(m) + \sum_{i=m}^{n}\delta(i)] \quad (26)$$

where $\sum_{i=m}^{n}\delta(i)$ is the guaranteed cumulative service time for packet $p^m$ to $p^n$. Suppose we can use a function $\gamma(n-m)$ for $\sum_{i=m}^{n}\delta(i)$, i.e. $\gamma(n-m) = \sum_{i=m}^{n}\delta(i)$. Then, the above equation becomes

$$\hat{d}(n) = \sup_{0\leq m\leq n}[a(m) + \gamma(n-m)] = a\overline{\otimes}\gamma(n),$$

which provides a basis for the following *time-domain* (deterministic) server model that essentially charaterizes the service using an upper bound on the cumulative service time [5]:

**Definition 9.** ***(Service Curve).*** *Consider a system $\mathcal{S}$ with input process $a(n)$ and output process $d(n)$. The system is said to provide to the input a (deterministic) service curve $\gamma(n) \in \mathcal{G}$, if for all $n \geq 0$,*

$$d(n) \leq a\overline{\otimes}\gamma(n). \quad (27)$$

The (deterministic) service curve model has the following duality principle:

**Lemma 5.** *For any $x \geq 0$, $d(n) - a\overline{\otimes}\gamma(n) \leq x$ for all $n \geq 0$, if and only if $\sup_{0\leq m\leq n}[d(n) - a\overline{\otimes}\gamma(n)] \leq x$ for all $n \geq 0$, where $\gamma \in \mathcal{G}$.*

*Proof:* For the "if" part, it holds trivally since $d(n) - a\overline{\otimes}\gamma(n) \leq \sup_{0\leq m\leq n}[d(n) - a\overline{\otimes}\gamma(n)]$. For the "only if" part, since $d(n) - a\overline{\otimes}\gamma(n) \le x$ for all $n \ge 0$, $\sup_{0\le m\le n}[d(n) - a\overline{\otimes}\gamma(n)] \le \sup_{0\le m\le n}[x] = x$. ∎

By the definition of service curve, it is clear that the first part of Lemma 5 defines a service curve $\gamma(n) + x$. Lemma 5 states that if a server provides service curve $\gamma(n) + x$, then $\sup_{0\le m\le n}[d(m) - a\overline{\otimes}\gamma(m)] \le x$ holds, and vice versa. In this sense, we call Lemma 5 the *duality principle* of service curve.

### F. Stochastic Service Curve

For networks providing stochastic service guarantees, following the principle of Eq.(26), we have the following expression for the expected departure time of packet $p^n$

$$\hat{d}(n) = \sup_{0\le m\le n}[a(m) + \sum_{i=m}^{n}(\delta(i) + \epsilon(i))]$$

where we assume $\delta(i)$ is the deterministic part while $\epsilon(i)$ the random part in the total service time $\delta(i) + \epsilon(i)$ guaranteed to packet $p^i$. We call $\epsilon(i)$ *stochastic error term* associated to $\delta(i)$. Here, $\epsilon(n)$ is introduced to represent the additional delay of $p^n$ due to some randomness. For example, an error-prone wireless link is often considered to operate in two states. If the link is in 'good'condition, it can send and receive data correctly; if the link is in 'bad'condition due to errors, the data that should be sent immediately has to be queued longer until the channel changes to 'good'condition. Then, $\epsilon(n)$ in this case represents the time period in which the channel is in 'bad'condition between the time when $p^{n-1}$ has been sent correctly and the time when $p^n$ can be sent.

With the consideration of the stochastic error term, the (deterministic) service curve can be extended to a stochastic version as follows:

**Definition 10.** ***(i.d Stochastic Service Curve).*** *A system is said to provide an* inter-departure time (i.d) stochastic service curve $\gamma \in \mathcal{G}$ *with bounding function* $j \in \bar{\mathcal{G}}$, *denoted by* $\mathcal{S} \sim_{id} \langle j, \gamma\rangle$, *if for all* $n \ge 0$ *and all* $x \ge 0$, *there holds*

$$P\Big\{d(n) - a\overline{\otimes}\gamma(n) > x\Big\} \le j(x). \tag{28}$$

**Example 4.** Consider two nodes, the sender and the receiver, communicate through an error-prone wireless link. Packets have fixed-length. Packets arriving to the sender node are served in FIFO manner. Assume the guaranteed per-packet service time is $\delta$ without any error. To simplify the analysis, assume the time slot length equals $\delta$. The sender sends packets correctly only when the link is in 'good'condition. If the link is in 'bad'condition, no packets can be sent correctly. In addition, the sender can send the head-of-queue packet only at the beginning of a time slot, i.e., the time period during which the link is in 'bad'condition should be an integer times of $\delta$. The probability that a packet can be sent correctly is determined by packet error rate (PER). PER is determined by the packet length and the bit error rate (BER). Here, we assume packet errors happen independently and the same PER denoted by $P_e$ is applied to all packets. The successful transmission probability of one packet is hence $1 - P_e$.

Suppose $P\{\Delta(n) = i\} = P_e^{i-1}(1-P_e)$, $i \ge 1$, where $\Delta(n)$ represents the number of time slots necessary to successfully send the $n^{th}$ packet with respect to the successful transmission probability $1 - P_e$. The number of time slots necessary to successfully send $n$ packets is $\sum_{k=1}^{n}\Delta(k)$ which has the negative binomial distribution

$$P\Big\{\sum_{k=1}^{n}\Delta(k) = i\Big\} = \begin{cases} \binom{i-1}{n-1}(1-P_e)^n P_e^{i-n}, & i \ge n \\ 0, & i < n \end{cases}$$

Then the sender provides to its input a stochastic service curve $\gamma$ which has the following distribution

$$P\{\gamma(n) = \lceil\frac{\tau}{\delta}\rceil\} = \binom{\lceil\frac{\tau}{\delta}\rceil - 1}{n-1}(1-P_e)^n P_e^{\lceil\frac{\tau}{\delta}\rceil - n}$$

where $\tau$ is the guaranteed service time to successfully send $n$ packets and $\lceil x\rceil$ denotes the smallest integer greater than or equal to $x$.

We can find $n_0 \le n$ such that $a\overline{\otimes}\gamma(n)$ takes its maximum value, i.e., $a\overline{\otimes}\gamma(n) = a(n_0) + \gamma(n - n_0 + 1)$. From Eq.(28), we have

$$P\{\gamma(n - n_0 + 1) < d(n) - a(n_0) - x\} \le j(x)$$

where

$$j(x) = \sum_{i=n}^{\lceil\frac{d(n)-a(n_0)-x}{\delta}\rceil - 1}\binom{i-1}{n-1}(1-P_e)^n P_e^{i-n}$$

In Sec. IV, we will show that many results can be derived from i.d stochastic service curve model. However, without additional constraints, we have difficulty to prove the concatenation property for i.d stochastic service curve. To address this difficulty, we introduce a stronger definition in the following subsection.

### G. Constrained Stochastic Service Curve

The constrained stochastic service curve model is generalized from the (deterministic) service curve model based on its duality principle. From Lemma 5, we know that a system with input $a(n)$ and output $d(n)$ has a service curve $\gamma(n)$ if and only if for all $n \ge 0$,

$$\sup_{0\le m\le n}\{d(m) - a\overline{\otimes}\gamma(m)\} \le x. \tag{29}$$

Inequality (29) provides the basis to generalize the (deterministic) service curve model to the constrained stochastic service curve defined as follows:

**Definition 11.** ***(Constrained Stochastic Service Curve).*** *A system is said to provide a* constrained stochastic service curve *(c.s)* $\gamma \in \mathcal{G}$ *with bounding function* $j \in \bar{\mathcal{G}}$, *denoted by* $\mathcal{S} \sim_{cs} \langle j, \gamma\rangle$, *if for all* $n \ge 0$ *and all* $x \ge 0$, *there holds*

$$P\Big\{\sup_{0\le m\le n}[d(m) - a\overline{\otimes}\gamma(m)] > x\Big\} \le j(x). \tag{30}$$

The following theorem establishes a relationship between i.d stochastic service curve and c.s stochastic service curve.

**Theorem 3.** 1) *If a server $\mathcal{S}$ provides to its input $a(n)$ a c.s stochastic service curve $\gamma(n)$ with bounding function $j(x) \in \bar{\mathcal{G}}$, it provides to the input $a(n)$ an i.d stochastic service curve $\gamma(n)$ with the same bounding function $j(x) \in \bar{\mathcal{G}}$, i.e., $\mathcal{S} \sim_{id} \langle j, \gamma \rangle$;*

2) *If a server $\mathcal{S}$ provides to its input $a(n)$ an i.d stochastic service curve $\gamma(n)$ with bounding function $j(x) \in \bar{\mathcal{F}}$, it provides to the input $a(n)$ a c.s stochastic service curve $\gamma_{+\eta}(n) = \gamma(n) + \eta \cdot n$ with bounding function $j^{\eta}(x) \in \bar{\mathcal{F}}$ where*

$$j_{\eta}(x) = \left[\frac{1}{\eta}\int_{x-\eta\cdot n}^{n} j(y)dy\right]_1$$

*for any $\eta > 0$.*

*Proof:* The first part follows easily, since there always holds $d(n) - a\overline{\otimes}\gamma(n) \le \sup_{0\le m\le n}\{d(m) - a\overline{\otimes}\gamma(m)\}$.

For the second part, there holds for any $n \ge m$,

$$a\overline{\otimes}\gamma_{+\eta}(m) \ge a\overline{\otimes}\gamma(m) + \eta\cdot m - \eta\cdot n$$

and then

$$d(m) - a\overline{\otimes}\gamma_{+\eta}(m) \le d(m) - a\overline{\otimes}\gamma(m) - \eta\cdot m + \eta\cdot n$$

Thus, we obtain

$$\begin{aligned} &P\Big\{\sup_{0\le m\le n}\{d(m) - a\overline{\otimes}\gamma_{+\eta}(m)\} > x\Big\} \\ &\le P\Big\{\sup_{1\le m\le n}\big[d(m) - a\overline{\otimes}\gamma(m) - \eta(m)\big]^{+} > x - \eta\cdot n\Big\} \end{aligned}$$

for which when $x - \eta\cdot n < 0$, the right hand side is equal to 1. In the following, we assume $x - \eta\cdot n \le 0$ under which, there holds

$$\begin{aligned} &P\Big\{\sup_{0\le m\le n}\{d(m) - a\overline{\otimes}\gamma_{+\eta}(m)\} > x\Big\} \\ &\le \sum_{m=1}^{n} P\Big\{\big[d(m) - a\overline{\otimes}\gamma(m) - \eta(m)\big] > x - \eta\cdot n\Big\} \\ &\le \sum_{m=1}^{n} j(x - \eta\cdot n + \eta\cdot m) \le \frac{1}{\eta}\int_{x-\eta\cdot n}^{n} j(y)dy \end{aligned}$$

As the probability is always not greater than 1, the second part follows the above inequality. ■

Note that in the second part of the above theorem, $j(x) \in \bar{\mathcal{F}}$ while not $\in \bar{\mathcal{G}}$. If the requirement on the bounding function is relaxed to $j(x) \in \bar{\mathcal{G}}$, the above relationship may not hold in general.

## IV. Basic Properties

This section presents results derived from the time-domain traffic models and server models introduced in Sec III. Particularly, we investigate the five basic properties introduced in Sec. I, which are service guarantees including delay bound and backlog bound, output characterization, concatenation property and superposition property. However, some properties can directly be proved only for the combination of a specific traffic model and a specific server model. This explains why we need to establish the various relationships between models in Sec III. With these relationships, we can extend and obtain the corresponding results for models which we are interested in.

### A. Service Guarantees

This subsection investigates probabilistic bounds on delay and backlog under the combination of v.s.d stochastic arrival curve and i.d stochastic service curve.

We start with deriving the bound on delay that a packet would experience in a system.

**Theorem 4.** ***(Delay Bound).*** *Consider a system $\mathcal{S}$ providing an i.d stochastic service curve $\gamma \in \mathcal{G}$ with bounding function $j \in \bar{\mathcal{G}}$ to the input which has a v.s.d arrival curve $\lambda \in \mathcal{G}$ with bounding function $h \in \bar{\mathcal{G}}$. Let $D(n) = d(n) - a(n)$ be the delay in the system of the $n^{th}(\ge 0)$ packet. For any $x \ge 0$, $D(n)$ is bounded by*

$$P\{D(n) > x\} \le j \otimes h(x - \gamma \oslash \lambda(0)). \tag{31}$$

*Proof:* For any $n \ge 0$, there holds

$$\begin{aligned} &d(n) - a(n) = \big[d(n) - a\overline{\otimes}\gamma(n)\big] + \big[a\overline{\otimes}\gamma(n) - a(n)\big] \\ &= \big[d(n) - a\overline{\otimes}\gamma(n)\big] + \sup_{0\le m\le n}\big\{\lambda(n-m) - \big[a(n) - a(m)\big] \\ &\quad + \gamma(n-m) - \lambda(n-m)\big\} \\ &\le \big[d(n) - a\overline{\otimes}\gamma(n)\big] + \sup_{0\le m\le n}\big\{\lambda(n-m) - \big[a(n) - a(m)\big]\big\} \\ &\quad + \sup_{0\le m\le n}\big\{\gamma(n-m) - \lambda(n-m)\big\} \\ &\le \big[d(n) - a\overline{\otimes}\gamma(n)\big] + \sup_{0\le m\le n}\big\{\lambda(n-m) - \big[a(n) - a(m)\big]\big\} \\ &\quad + \sup_{k\ge 0}\big\{\gamma(k) - \lambda(k)\big\}. \end{aligned} \tag{32}$$

The right-hand side of Eq.(32) implies a sufficient condition to obtain $P\{D(n) > x\}$, which is that $P\big\{d(n) - a\overline{\otimes}\gamma(n) > x\big\}$ and $P\Big\{\sup_{0\le m\le n}\big\{\lambda(n-m) - \big[a(n) - a(m)\big]\big\} > x\Big\}$ are known. To ensure the system's stability, we should also have

$$\lim_{k\to\infty}\frac{1}{k}[\gamma(k) - \lambda(k)] \le 0. \tag{33}$$

In the rest of the paper, without explicitly stating, we shall assume inequality (33) holds. From Lemma 1 and $\sup_{k\ge 0}\big\{\gamma(k) - \lambda(k)\big\} = \gamma \oslash \lambda(0)$, we conclude

$$P\{D(n) > x\} \le j \otimes h(x - \gamma \oslash \lambda(0)).$$

■

Next, we consider backlog bound of a system. By definition, the backlog in the system at time $t \ge 0$ is $\mathcal{B}(t) = \mathcal{A}(t) - \mathcal{A}^*(t)$. If $a(n)$ is the arrival time of the latest packet arriving to the system by time $t$, then $\mathcal{B}(t)$ is

$$\mathcal{B}(t) \le \inf\big\{k \ge 0 : d(n-k) \le a(n)\big\}. \tag{34}$$

Eq.(34) implies that, for any $x \ge 0$, if $\mathcal{B}(t) > x$, there must be $a(n) < d(n-x)$, since otherwise if there would be $a(n) \ge d(n-x)$ and we would have $\mathcal{B}(t) \le x$ that contradicts the condition $\mathcal{B}(t) > x$. In other words, event $\{\mathcal{B}(t) > x\}$ implies event $\{a(n) < d(n-x)\}$, and hence

$$P\{\mathcal{B}(t) > x\} \le P\{a(n) < d(n-x)\}. \tag{35}$$

Then we have the following result for backlog.

**Theorem 5.** ***(Backlog Bound).*** *Consider a system $\mathcal{S}$ providing an i.d stochastic service curve $\gamma \in \mathcal{G}$ with bounding function $j \in \bar{\mathcal{G}}$ to the input which has a v.s.d stochastic arrival curve $\lambda \in \mathcal{G}$ with bounding function $h \in \bar{\mathcal{G}}$. The backlog at time $t$ ($t \geq 0$), $\mathcal{B}(t)$, is bounded by*

$$P\{\mathcal{B}(t) > H(\lambda, \gamma + x)\} \leq j \otimes h(x) \tag{36}$$

*for any $x \geq 0$, where, $H(\lambda, \gamma + x) = \sup_{n \geq 0}\big\{\inf[k \geq 0 : \gamma(n-k) + x \leq \lambda(n)]\big\}$ is the maximum horizontal distance between functions $\lambda(n)$ and $\gamma(n) + x$ for any $x \geq 0$.*

*Proof:* Similar to prove the delay bound, we have

$$d(n-x)-a(n) = [d(n-x)-a\overline{\otimes}\gamma(n-x)]+[a\overline{\otimes}\gamma(n-x)-a(n)]$$
$$= \big[d(n-x)-a\overline{\otimes}\gamma(n-x)\big]+\sup_{0\leq k\leq n-x}\big\{\lambda(n-k)-[a(n)-a(k)]$$
$$+\gamma(n-x-k)-\lambda(n-k)\big\}$$
$$\leq \big[d(n-x)-a\overline{\otimes}\gamma(n-x)\big]+\sup_{0\leq k\leq n-x}\big\{\lambda(n-k)-\big[a(n+x)$$
$$-a(k)\big]\big\}+\sup_{0\leq k\leq n-x}\big\{\gamma(n-x-k)-\lambda(n-k)\big\}$$

Let $v = n - k$. The above inequality is written as

$$d(n-x)-a(n) \leq \big[d(n-x)-a\overline{\otimes}\gamma(n-x)\big]+\sup_{0\leq k\leq n}\big\{\lambda(n-k)$$
$$-\big[a(n)-a(k)\big]\big\}+\sup_{x\leq v\leq n}\big\{\gamma(v-x)-\lambda(v)\big\}$$

Let $x = H(\lambda, \gamma + y)$, we have

$$d\big(n-h(\lambda,\gamma+y)\big)-a(n) \leq \big[d\big(n-h(\lambda,\gamma+y)\big)$$
$$-a\overline{\otimes}\gamma\big(n-h(\lambda,\gamma+y)\big)\big]+\sup_{0\leq k\leq n}\big\{\lambda(n-k)-[a(n)-a(k)]\big\}-y \tag{37}$$

Under the same conditions as analyzing the delay, we obtain

$$P\{\mathcal{B}(t) > H(\lambda, \gamma + x)\} \leq j \otimes h(x).$$

■

### B. Output Characterization

This subsection presents the result for characterizing the departure process from a system.

**Theorem 6.** ***(Output Characterization).*** *Consider a system $\mathcal{S}$ provides an i.d stochastic service curve $\gamma(n) \in \mathcal{G}$ with bounding function $j(x) \in \bar{\mathcal{G}}$ to its input which has a v.s.d stochastic arrival curve $\lambda(n) \in \mathcal{G}$ with bounding function $h(x) \in \bar{\mathcal{G}}$. The output has an i.t stochastic arrival curve $\lambda\overline{\oslash}\gamma(n-m)$ with bounding function $j(x) \otimes h \in \bar{\mathcal{G}}$.*

*Proof:* For any two departure packets $m < n$, there holds

$$d(n)-d(m) \geq a(n)-a\overline{\otimes}\gamma(m)+a\overline{\otimes}\gamma(m)-d(m)$$
$$-\big[d(n)-d(m)\big] \leq \big[d(m)-a\overline{\otimes}\gamma(m)\big]+a\overline{\otimes}\gamma(m)-a(n)$$
$$\leq \big[d(m)-a\overline{\otimes}\gamma(m)\big]+\sup_{0\leq k\leq n}\big\{\lambda(n-k)-[a(n)-a(k)]\big\}$$
$$+\sup_{0\leq v\leq m}\big\{\gamma(v)-\lambda(n-m+v)\big\}$$
$$= \big[d(m)-a\overline{\otimes}\gamma(m)\big]+\sup_{0\leq k\leq n}\big\{\lambda(n-k)-[a(n)-a(k)]\big\}$$
$$-\inf_{0\leq v\leq m}\big\{\lambda(n-m+v)-\gamma(v)\big\}$$

Adding $\inf_{0\leq v\leq m}\big\{\lambda(n-m+v)-\gamma(v)\big\}$ to both sides of the above inequality, we get

$$\inf_{0\leq v\leq m}\big\{\lambda(n-m+v)-\gamma(v)\big\}-[d(n)-d(m)]$$
$$\leq \big[d(m)-a\overline{\otimes}\gamma(m)\big]+\sup_{0\leq k\leq n}\big\{\lambda(n-k)-[a(n)-a(k)]\big\}$$

With the same conditions as analyzing delay, we conclude

$$P\Big\{\lambda\overline{\oslash}\gamma(n-m)-[d(n)-d(m)] > x\Big\} \leq j \otimes h(x)$$

■

### C. Concatenation Property

The concatenation property uses an equivalent system to represent a system of multiple servers connected in tandem, each of which provides stochastic service curve to the input. Then the equivalent system provides the input a stochastic service curve, which is derived from the stochastic service curve provided by all involved individual servers.

**Theorem 7.** ***(Concatenation Property).*** *Consider a flow passing through a network of $N$ systems in tandem. If each system $k(=1,2,...,N)$ provides a c.s stochastic service curve $\mathcal{S}^k \sim_{cs} \langle j^k, \gamma^k\rangle$ to its input, then the network guarantees to the flow a c.s stochastic service curve $\mathcal{S} \sim_{cs} \langle j, \gamma\rangle$ with*

$$\gamma(n) = \gamma^1\overline{\otimes}\gamma^2\overline{\otimes}\cdots\overline{\otimes}\gamma^N(n) \tag{38}$$

$$j(x) = j^1 \otimes j^2 \otimes \cdots \otimes j^N(x). \tag{39}$$

*Proof:* We shall only prove the two-node case, from which, the proof can be easily extended to the $N$-node case. The departure of the first node is the arrival to the second node, so $d^1(n) = a^2(n)$. In addition, the arrival to the network is the arrival to the first node, i.e., $a(n) = a^1(n)$, and the departure from the network is the departure from the second node, i.e., $d(n) = d^2(n)$, where, $a(n)$ and $d(n)$ denote the arrival process to and departure process from the network, respectively. We then have,

$$\sup_{0\leq m\leq n}\{d(m)-a\overline{\otimes}\gamma^1\overline{\otimes}\gamma^2(m)\}$$
$$= \sup_{0\leq m\leq n}\{d^2(m)-(a^1\overline{\otimes}\gamma^1)\overline{\otimes}\gamma^2(m)\} \tag{40}$$

Now let us consider any $m$, ($0 \leq m \leq n$), for which we get,

$$d^2(m)-(a^1\overline{\otimes}\gamma^1)\overline{\otimes}\gamma^2(m)$$
$$= d^2(m)-\sup_{0\leq k\leq m}\big\{a^1\overline{\otimes}\gamma^1(k)+\gamma^2(m-k)-d^1(k)+a^2(k)\big\}$$
$$= d^2(m)+\inf_{0\leq k\leq m}\big\{d^1(k)-a^1\overline{\otimes}\gamma^1(k)-\gamma^2(m-k)-a^2(k)\big\}$$

$$\leq \sup_{0\leq k\leq m}\{d^1(k)-a^1\overline{\otimes}\gamma^1(k)\}+d^2(m)$$

$$+\inf_{0\leq k\leq m}\big\{-[a^2(k)+\gamma^2(m-k)]\big\}$$

$$\leq \sup_{0\leq k\leq m}\{d^1(k)-a^1\overline{\otimes}\gamma^1(k)\}+[d^2(m)-a^2\overline{\otimes}\gamma^2(m)] \quad (41)$$

Applying Eq.(40) to Eq.(41), we obtain

$$\sup_{0\leq m\leq n}\{d^2(m)-(a^1\overline{\otimes}\gamma^1)\overline{\otimes}\gamma^2(m)\}$$

$$\leq \sup_{0\leq k\leq n}\{d^1(k)-a^1\overline{\otimes}\gamma^1(k)\}+\sup_{0\leq m\leq n}\{d^2(m)-a^2\overline{\otimes}\gamma^2(m)\} \quad (42)$$

with which, since both nodes provide c.s stochastic service curve to their input, the theorem follows from Lemma 1 and the definition of c.s stochastic service curve. ∎

### D. Superposition Property

The superposition property means that the superposition of flows can be represented using the same traffic model. With this property, the aggregate of multiple individual flows may be viewed as a single aggregate flow. Then the service guarantees for the aggregate flow can be derived in the same way as for a single flow.

First, we only consider the aggregate of two flows, $F_1$ and $F_2$. Let $a_1(n)$, $a_2(n)$ and $a(n)$ be the arrival process of $F_1$, $F_2$ and the aggregate flow $F_A$, respectively.

For any packet $p^n$ of the aggregate flow $F_A$, it is either the $m^{th}$ packet from flow $F_1$ or the $(n-m)^{th}$ packet from flow $F_2$, where $m\in[0,n]$, i.e.

$$a(n)=\max\{a_1(m),a_2(n-m)\}$$

For example, $a(1)$ is either $\max[a_1(0),a_2(1)]$ or $\max[a_1(1),a_2(0)]$ and is the minimum of these two possibilities,

$$a(1)=\inf\big\{\max[a_1(0),a_2(1)],\max[a_1(1),a_2(0)]\big\}.$$

We can see another example

$$a(2)=\inf\big\{\max[a_1(0),a_2(2)],\max[a_1(1),a_2(1)],\max[a_1(2),a_2(0)]\big\}.$$

Essentially, we have for any packet $n$ of the aggregate flow

$$a(n)=\inf_{0\leq m\leq n}\big\{\max[a_1(m),a_2(n-m)]\big\}. \quad (43)$$

We generalize the result to the superposition of $N(\geq 2)$ flows

$$a(n)=\inf_{\sum m_i=n}\big\{\max[a_1(m_1),a_2(m_2),...,a_N(n-\sum_{i=1}^{N-1}m_i)]\big\}. \quad (44)$$

Whereas, it is difficult to characterize the packet inter-arrival time of the aggregate flow directly from Eq.(44). We know that if a flow has v.s.d stochastic arrival curve, then with Theorem 2, this flow has a v.b.c stochastic arrival curve as well. It has been proved that the v.b.c stochastic arrival curve has the superposition property [13]. Thus, we can indirectly prove that the superposition property holds for the v.s.d stochastic arrival curve.

If flow $i$ has v.s.d stochastic arrival curve $a_i(n)\sim_{vd}\langle h_i,\lambda_i\rangle$ $i=1,2,...,N$, from Theorem 2, flow $i$ has v.b.c stochastic arrival curve $\alpha_i(t)$ with bounding function $f_i(x)=h_i\big([\lambda_i(x)-\theta_i]^+\big)$ for any $\theta_i>0$, where $\alpha_i(t)=\sup\{n:\lambda_i(n)\leq t\}$. According to Lemma 3, the aggregate flow has a v.b.c stochastic arrival curve $\alpha(t)=\sum_{i=1}^{N}\alpha_i(t)$ with bounding function $f(x)=f_1\otimes\cdots\otimes f_N(x)$.

In summary, we can obtain the following result:

**Theorem 8.** *Consider $N$ flows with arrival processes $a_i(n)\sim_{vd}\langle h_i,\lambda_i\rangle$, $i=1,...,N$. For the aggregate of these flows, the following result holds:*

*$a(n)\sim_{vd}\langle h,\lambda\rangle$ with $\lambda(n)=\inf\{t:\alpha(t)\geq n\}$ and $h(x)=f[\alpha(x)-y]^+$, where $y=\mathcal{A}\otimes\alpha(t)+\alpha(x)-n+1-\mathcal{A}(t-x,t)$, $\alpha(t)=\sum_{i=1}^{n}\alpha_i(t)$ and $\alpha_i(t)=\sup\{n:\lambda_i(n)\leq t\}$.*

### E. Leftover Service Characterization

This subsection explores the leftover service characterization under aggregate scheduling. To ease the discussion, we consider the simplest case when there are two flows competing resource in a system under FIFO aggregation. Suppose that if packets arrive to the system simultaneously, they are inserted into the FIFO queue randomly. Consider a system fed with a flow $F_A$ which is the aggregation of two constituent flows $F_1$ and $F_2$. Suppose both the service characterization from the server and traffic characterization from $F_2$ are known. We are interested in characterizing the service time received by $F_1$, with which per-flow bounds for $F_1$ can be then easily obtained using earlier results derived in the previous subsections.

**Theorem 9.** *Consider a system $\mathcal{S}$ with input $F_A$ that is the aggregation of two constituent flows $F_1$ and $F_2$. Suppose $F_2$ has a (deterministic) arrival curve $\lambda_2(n)\in\mathcal{G}$, and the system provides to the input an i.d stochastic service curve $\gamma\in\mathcal{G}$ with bounding function $j(x)\in\bar{\mathcal{G}}$. Then if $\gamma(n+\sup[q:\lambda_2(q)\leq a_1(n)]\in\mathcal{G}$, $F_1$ receives an i.d stochastic service curve $\gamma(n+\sup[q:\lambda_2(q)\leq a_1(n)]$ with the same bounding function $j(x)$.*

*Proof:* Suppose packet $p_1^n$ is the $(n+m)^{th}$ packet of $F_A$, i.e., $a(n+m)=a_1(n)$, where $m$ represents the number of packets from $F_2$. As the system provides an i.d stochastic service curve $\gamma(n)$ to the aggregate flow $F_A$, there holds

$$P\{d(n+m)-a\overline{\otimes}\gamma(n+m)>x\}\leq j(x).$$

$a_1(n)=a(n+m)$ indicates $a_2(m)\leq a_1(n)$. Let $\bar{m}=\sup[q:\lambda_2(q)\leq a_1(n)]$. As $\lambda_2$ is the (deterministic) arrival curve of $F_2$, we have $\bar{m}\geq m$ because of $a_2(m)\geq\lambda_2(m)$. Then $\gamma(n+\bar{m})\geq\gamma(n+m)$. Let $\gamma_1(n)=\gamma(n+\bar{m})$. From $\gamma_1(n)\geq\gamma(n+m)$, we have $a_1\overline{\otimes}\gamma_1(n)\geq a\overline{\otimes}\gamma(n+m)$. As $d(n+m)=d_1(n)$, there holds

$$d_1(n)-a_1\overline{\otimes}\gamma_1(n)\leq d(n+m)-a\overline{\otimes}\gamma(n+m)$$

Thus, we conclude

$$P\{d_1(n)-a_1\overline{\otimes}\gamma_1(n)>x\}\leq j(x)$$

∎

### *F. Discussion*

In this section, we have presented the five basic properties of stochastic network calculus under various traffic models and server models defined in the time-domain and introduced some simple applications. For example, a GCRA-constrained flow has a deterministic arrival curve. If a flow's packet inter-arrival times are exponentially distributed, then this flow has a v.s.d stochastic arrival curve. The service process of an error-prone wireless link can be modeled by an i.d stochastic service curve.

For each basic property, we investigated one combination of a specific traffic model and a specific server model. Particularly, we proved that the service guarantees and the output characterization hold for the combination of v.s.d stochastic arrival curve and i.d stochastic service curve. For the concatenation property, we investigated the case that all servers provide the constrained service curve to their input but did not specify the type of arrival curve. In order to prove the superposition property, we used the transformation between v.s.d stochastic service curve and v.b.c stochastic service curve. The leftover service characterization was only proved for the combination of deterministic arrival curve and i.d stochastic service curve.

With the relationships and transformations among models established in Sec. III, these five properties may be directly or indirectly proved for other combinations of traffic models and server models. For example, it is easily to prove the service guarantees and output characterization for the combination of m.s.d stochastic arrival curve and i.d stochastic service curve. Due to space limitation, these results are not included. However, to prove the concatenation property and superposition property for other server models and traffic models, it will require additional transformations among models. For the leftover service characterization, we may need more constraints or transformations when proving it for other combinations of traffic models and server models. We leave these as our future work.

## V. Conclusion

For stochastic service guarantee analysis, we introduced several *time-domain* models for traffic and service modeling. The essential idea of them is to base the model on cumulative packet inter-arrival time for traffic and on cumulative service time for service. Simple examples have been given to demonstrate the use of them. Based on the proposed time-domain models, the five basic properties for stochastic network calculus were derived, with which, the results can be easily applied to both the single-node and the network cases. We believe, the proposed time-domain models and derived results can be particularly useful for analyzing stochastic service guarantees in systems where the behavior of a server involves some stochastic processes and due to this, it is difficult to characterize using the current *space-domain* server models. Such systems include wireless links and multi-access networks where backoff schemes may be employed in case of channel error or collision occuring. In this paper, however, we only analyzed a simple case of wireless network to illustrate how to apply the proposed server model to characterize the service process of a wireless node. The future work is to investigate the performance of some typical contention-based multi-access networks including IEEE 802.11 networks.

## References


[1] I.-T. S. G. 13. Recommendation I.371 traffic control and congestion control in B-ISDN. 1995.

[2] J.-Y. L. Boudec and P. Thiran. *Network calculus: A Theory of Deterministic Queueing Systems for the Internet*. Springer, LNCS, 2004.

[3] C.-S. Chang. Stability, queue length and delay of deterministic and stochastic queueing networks. *IEEE Trans. Auto. Control*, 39(5):913–931, May 1994.

[4] C.-S. Chang. On the exponentiality of stochastic linear systems under the max-plus algebra. *IEEE Trans. Automatic Control*, 41(8):1182–1188, Aug. 1996.

[5] C.-S. Chang and Y. H. Lin. A general framework for deterministic service guarantees in telecommunication networks with variable length packets. *IEEE/ACM Trans. Automatic Control*, 46(2):210–221, Feb. 2001.

[6] J. A. Cobb. Preserving quality of service guarantees in spite of flow aggregation. *IEEE/ACM Trans. Networking*, 10(1):43–53, Feb. 2002.

[7] D. Ferrari. Client requirements for real-time communication services. *IEEE Commun. Magazine*, pages 65–72, Nov. 1990.

[8] D. Ferrari. Multimedia network protocols: where are we? *Multimedia Systems*, 4(6):299–304, Dec. 1996.

[9] M. Fidler. An end-to-end probabilistic network calculus with moment generating functions. In *Proc. IEEE IWQoS 2006*, pages 261–270, 2006.

[10] P. Goyal, S. S. Lam, and H. M. Vin. Determining end-to-end delay bounds in heterogeneous networks. *Multimedia System*, 5(3):157–163, May 1997.

[11] Y. Jiang. A basic stochastic network calculus. In *Proc. ACM SIGCOMM 2006*, pages 123–134, 2006.

[12] Y. Jiang. Per-domain packet scale rate guarantee for expedited forwarding. *IEEE/ACM Trans. Networking*, 14(3):630–643, June 2006.

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

[14] C. Li, A. Burchard, and J. Liebeherr. A network calculus with effective bandwidth. *Technical Report, CS-2003-20, University of Virginia*, Nov. 2003.

[15] S. M. Ross. *Introduction to Probability Models*. 7th Edition, Academic Press, 2000.

[16] J. F. Shortle and P. H. Brill. Analytical distribution of waiting time in the M/iD/1 queue. *Queueing Systems*, 50(2):185–197, 2005.

[17] D. Starobinski and M. Sidi. Stochastically bounded burstiness for communication networks. *IEEE/ACM Trans. Information Theory*, 46(1):206–212, Jan. 2000.

[18] O. Yaron and M. Sidi. Performance and stability of communication network via robust exponential bounds. *IEEE/ACM Trans. Networking*, 1(3):372–385, June 1993.