A Temporal Approach to Stochastic Network Calculus

Jing Xie\textsuperscript{a}, Yuming Jiang\textsuperscript{b}, Min Xie\textsuperscript{c}

\textsuperscript{a}Research and Innovation, Det Norske Veritas
Veritasveien 1, 1363, Høvik, Norway

\textsuperscript{b}Centre for Quantifiable Quality of Service in Communication Systems (Q2S) and
Department of Telematics
Norwegian University of Science and Technology (NTNU), Norway

\textsuperscript{c}University Centre at Blackburn College
University Close, United Kingdom

Abstract

Stochastic network calculus is a newly developed theory for stochastic service guarantee analysis of computer networks. In the current stochastic network calculus literature, its fundamental models are based on the cumulative amount of traffic or cumulative amount of service. However, there are network scenarios where direct application of such models is difficult. This paper presents a temporal approach to stochastic network calculus. The key idea is to develop models and derive results from the time perspective. Particularly, we define traffic models and service models based on the cumulative packet inter-arrival time and the cumulative packet service time, respectively. Relations among these models as well as with the existing models in the literature are established. In addition, we prove the basic properties of the proposed models, such as delay bound and backlog bound, output characterization, concatenation property and superposition property. These results form a temporal stochastic network calculus and compliment the existing results.

Keywords: Stochastic network calculus, max-plus algebra, min-plus algebra, stochastic arrival curve, stochastic service curve, performance guarantee analysis, delay bound, backlog bound

\textsuperscript{1}An early version of this paper was partially presented at MASCOTS 2009 [36].
1. Introduction

Stochastic network calculus is a theory dealing with queueing systems found in computer networks [8][13][20][22]. It is particularly useful for analyzing networks where service guarantees are provided stochastically. Such networks include wireless networks, multi-access networks and multimedia networks where applications can tolerate a certain level of violation of the desired performance [14].

Stochastic network calculus is based on properly defined traffic models [7][20][22][27][34][37] and service models [20][22]. In literature, it is typical to model the arrival process by a stochastic arrival curve and the service process by a stochastic service curve. The arrival curve provides probabilistic upper bounds on the cumulative amount of arrival traffic whereas the service curve lower bounds the cumulative amount of service. In this paper, we call such models space-domain models, from which extensive results have been derived. There are five most fundamental properties [20][22]: (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 and applications of these basic properties can be found [20][22].

However, there are many open challenges for stochastic network calculus, making its wide application difficult [22]. One is to analyze networks where users are served probabilistically. For example, in wireless networks, a wireless link is error-prone and consequently retransmission is often adopted to ensure reliability. In random multi-access networks, random backoff and retransmission are used to deal with contention and collision. To apply stochastic network calculus to analyze such networks, it is fundamental to find the stochastic characterization of the service time that provides successful transmissions for the user. However, direct application of existing space-domain models, which are built on the amount of cumulative service, is difficult.

This paper aims to rethink stochastic network calculus to address some of the challenges in current stochastic network calculus literature, such as the analysis of error-prone wireless channels and/or contention-based multi-access. To be specific, we present a temporal approach to stochastic network calculus. The key idea is to develop models and derive results from the time perspective. We define traffic models and service models based on the cumulative packet inter-arrival time and the cumulative packet service time, respectively. In this paper, we shall call such models time-domain models.

2
In addition to their easy use in network scenarios discussed above, the basic properties are also investigated based on the proposed time-domain models. Moreover, relations among the proposed time-domain models as well as with the corresponding space-domain models are established, which provide a tight link between the proposed temporal stochastic network calculus approach and the existing space domain stochastic network calculus approach. This gives increased flexibility in applying stochastic network calculus in challenging network scenarios.

The structure of the paper is as follows. In Section 2 we first introduce the notation and the system specification, followed by the review of the relevant results of stochastic network calculus. Section 3 defines the network calculus models in the time-domain and explores the model transformations. Four fundamental properties are thoroughly investigated in Section 4. The relevant discussion reveals the reasons of establishing the model transformations in Section 3. In Section 5 we conclude the paper and discuss the open issue.

2. Network Model and Related Work

This section specifies the network system and reviews mathematical preliminaries for the analysis in the following sections. A brief overview on stochastic network calculus of particular relevance to this paper is presented as well.

In this paper, we make the following assumptions unless stated otherwise.

- All packets have the same length.

- A packet is considered to be received by a network element when and only when its last bit has arrived to the network element.

- A packet can be served only when its last bit has arrived.

- A packet is considered out of a network element when and only when its last bit has been transmitted by the network element.

- Packets arriving to a network element are queued in the buffer and served in the FIFO order. All queues are empty at time 0.

- All network elements provide sufficient buffer space to store all incoming traffic and are lossless.
2.1. Notations and System Specification

We use $P(n)$, $a(n)$, $d(n)$ and $\delta_n$, to denote the $(n+1)$th packet entering the system, its arrival time to the system, its departure time from the system and its service time provided by the system, respectively, where $n = 0, 1, 2, \ldots$

- From the temporal perspective, an arrival process counts the cumulative inter-arrival time between two arbitrary packets and is denoted by $\Gamma(m, n) = a(n) - a(m)$ for any $0 \leq m \leq n$. Note $\Gamma(n, n) = 0$.

- A service process describes the cumulative service time received between two arbitrary packets and is denoted by $\Delta(m, n) = \sum_{k=m}^{n} \delta_k$ for any $0 \leq m \leq n$. Note $\Delta(n, n) = \delta_n$.

In the time-domain, the system backlog and system delay are defined below, respectively.

**Definition 1.** The system backlog at time $t \geq 0$ is denoted by $B(t)$:

$$B(t) \leq \inf \left\{ l \geq 0, \sup \{ n \geq 0 : a(n) \leq t \} : d(n - l) \leq a(n) \right\}. \quad (1)$$

The delay that packet $P(n)$ experiences in the system is denoted by $D(n)$:

$$D(n) = d(n) - a(n). \quad (2)$$

Moreover, the time that packet $P(n)$ waits in queue is denoted by $W(n)$:

$$W(n) = D(n) - \delta_n. \quad (3)$$

The following function sets are often used in this paper.

The set of non-negative wide-sense increasing functions is denoted by $\mathcal{F}$, where for each function $f(\cdot)$,

$$\mathcal{F} = \{ f(\cdot) : \forall 0 \leq x \leq y, 0 \leq f(x) \leq f(y) \}$$

and for any function $f(\cdot) \in \mathcal{F}$, we set $f(x) = 0$ for all $x < 0$.

We denote by $\bar{\mathcal{F}}$ the set of non-negative wide-sense decreasing functions where for each function $f(\cdot)$,

$$\bar{\mathcal{F}} = \{ f(\cdot) : \forall 0 \leq x \leq y, 0 \leq f(y) \leq f(x) \}$$

and for any function $f(\cdot) \in \bar{\mathcal{F}}$, we set $f(x) = 1$ for all $x < 0$. 

4
We denote by $\bar{G}$ a subset of $\bar{F}$, where for each function $f(\cdot) \in \bar{G}$, its $n$th-fold integration, denoted by $f^{(n)}(x) \equiv (\int_{x}^{\infty} dy)^{n}f(y)$, is bounded for any $x \geq 0$ and still belongs to $\bar{G}$ for any $n \geq 0$, i.e.,

$$\bar{G} = \{ f(\cdot) : \forall n \geq 0, (\int_{x}^{\infty} dy)^{n}f(y) \in \bar{G} \}.$$

For ease of exposition, we adopt the following notations in this paper:

$$[x]^+ \equiv \max[x, 0] \text{ and } [x]_1 \equiv \min[x, 1].$$

In addition, the ceiling and floor functions are used in this paper as well.

- The ceiling function $[x]$ returns the smallest integer not less than $x$.
- The floor function $\lfloor x \rfloor$ returns the largest integer not greater than $x$.

### 2.2. Mathematical Basis

An essential idea of (stochastic) network calculus is to use alternate algebras, particularly min-plus algebra and max-plus algebra [5], to transform complex non-linear network systems into analytically tractable linear systems. To the best of our knowledge, the existing models and results of stochastic network calculus mainly based on min-plus algebra that has basis operations suitable for characterizing the amount of cumulative traffic and service. As a result, these models focus on describing network behavior from the spatial perspective. Max-plus algebra is suitable for arithmetic operations with cumulative inter-arrival times and service times. Consequently, network modeling from the temporal perspective more relies on max-plus algebra. In the following, we review the basics of both min-plus algebra and max-plus algebra.

In min-plus algebra, the ‘addition’ operation represents infimum or minimum when it exists, and the ‘multiplication’ operation is $\oplus$. The min-plus convolution of functions $f, g \in \mathcal{F}$, denoted by $\otimes$, is defined as

$$(f \otimes g)(t) = \inf_{0 \leq s \leq t} \{ f(s) + g(t - s) \}$$

where, when it applies, ‘infimum’ should be interpreted as ‘minimum’. The min-plus deconvolution of functions $f, g \in \mathcal{F}$, denoted by $\oslash$, is defined as

$$(f \oslash g)(t) = \sup_{s \geq 0} \{ f(s + t) - g(s) \}$$
where, when it applies, ‘supremum’ should be interpreted as ‘maximum’.

In the max-plus algebra, the ‘addition’ operation represents supremum or maximum when it exists, and the ‘multiplication’ operation is +. The max-plus convolution of functions \( f, g \in \mathcal{F} \), denoted by \( \hat{\otimes} \), is defined as

\[
(f \hat{\otimes} g)(n) = \sup_{0 \leq m \leq n} \{ f(m) + g(n - m) \}
\]

where, when it applies, ‘supremum’ should be interpreted as ‘maximum’.

The max-plus deconvolution is associative and commutative \([5]\).

- **Associativity:** for any \( g_1, g_2, g_3 \in \mathcal{F} \), \((g_1 \hat{\otimes} g_2) \hat{\otimes} g_3 = g_1 \hat{\otimes} (g_2 \hat{\otimes} g_3)\).
- **Commutativity:** for any \( g_1, g_2 \in \mathcal{F} \), \(g_1 \hat{\otimes} g_2 = g_2 \hat{\otimes} g_1\).

### 2.3. State of The Art in Stochastic Network Calculus

The available literature on stochastic network calculus mainly focuses on modeling network behavior and analyzing network performance from the spatial perspective \([6][12][15][16][20][22][23][28][29][31]\). We call the corresponding models and results space-domain models and results in this paper.

In order to characterize the arrival process of a flow from the spatial perspective, let us consider the amount of traffic generated by this flow in a time interval \((s, t]\), denoted by \( A(s, t) \). In the context of stochastic network calculus, the arrival curve model is defined based on a stochastic upper bound on the cumulative amount of the arrival traffic. Here, we only review one relevant space-domain arrival curve model, virtual-backlog-centric (v.b.c) stochastic arrival curve (SAC) \([22]\).

**Definition 2. (v.b.c Stochastic Arrival Curve)**

A flow is said to have a virtual-backlog-centric (v.b.c) stochastic arrival curve \( \alpha(t) \in \mathcal{F} \) with bounding function \( f(x) \in \tilde{\mathcal{F}} \), if for all \( 0 \leq s \leq t \) and all \( x \geq 0 \), there holds

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

\([4]\)
In stochastic network calculus, the service curve model is defined as a stochastic lower bound on the cumulative amount of service provided by the system. Two space-domain service curve models [22] are reviewed here.

**Definition 3. (Weak Stochastic Service Curve)**

A network system is said to provide a weak stochastic service curve \( \beta(t) \in F \) with bounding function \( g(x) \in \bar{F} \) for the arrival process \( A(t) \), if for all \( t \geq 0 \) and all \( x \geq 0 \), there holds

\[
P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x),
\]

where \( A^*(t) \) denotes the cumulative amount of the departure traffic.

Unlike the arrival curve, it is difficult to identify the service curve from Eq. (5) because it couples the arrival process, the service curve and the departure process. Thus we need a more explicit model to directly reveal the relation between the service process and its service curve such as the following model [20].

**Definition 4. (Stochastic Strict Service Curve)**

A network system is said to provide stochastic strict service curve \( \beta(t) \in F \) with bounding function \( g(x) \in \bar{F} \), if during any period \( (s, t] \), the amount of service \( S(s, t) \) provided by this system satisfies, for any \( x \geq 0 \),

\[
P\{S(s, t) < \beta(t - s) - x\} \leq g(x).
\]

Definition 4 is applied to ‘any period’ which implies both worst-case scenario and other scenarios. If we could determine a function \( \beta(t) \) which makes Eq.(6) hold under the worst-case scenario, then Eq.(6) automatically holds under other scenarios as well.

Based on the arrival curve and service curve models, five fundamental properties have been proved to facilitate tractable analysis. For example, they can be used to derive service guarantees including delay bound and backlog bound, characterize the behavior of traffic departing from a server, describe the service provided along a multi-node path, determine the arrival curve for the aggregate flow, and compute the service provided to each constituent flow.
• **P.1: Service Guarantees (single-node)**
  Under the condition that the traffic arrival process has an arrival curve \( \alpha(t) \) with bounding function \( f(x) \) and the network node provides service with a service curve model \( \beta(t) \) with bounding function \( g(x) \), the stochastic delay bound and stochastic backlog bound can be derived. Particularly, the backlog bound is related to the maximal vertical distance between \( \alpha(t) \) and \( \beta(t) \); the delay bound is relevant to the maximal horizontal distance between \( \alpha(t) \) and \( \beta(t) \).

• **P.2: Output Characterization**
  To analyze the end-to-end performance of a multi-hop path, one option is the node-by-node analysis approach. This approach requires being able to characterize the traffic behavior after the traffic has been served and leaves the previous node. The output process of a flow from a node can also be characterized by an arrival curve which is determined by both the arrival curve of the arrival process and the service curve of the service process.

• **P.3: Concatenation Property (multi-node)**
  Network calculus possesses an unique property, concatenation property, which is also used to analyze the end-to-end performance but improves the results obtained from the node-by-node analysis. The essence of the concatenation property is to represent a series of nodes in tandem as a ‘black box’ which can be treated as a single node. The service curve of this equivalent system is determined by the service curve of all individual nodes along this path.

• **P.4: Superposition Property (aggregate flow)**
  Flow aggregation is very common in packet-switched networks. If multiple flows are aggregated into a single flow under the FIFO order, the aggregate flow also has an arrival curve which is the summation of the arrival curve of all constituent flows.

• **P.5: Leftover Service Characterization (per-flow)**
  The leftover service characterization makes per-flow performance analysis feasible under FIFO aggregate scheduling. The crucial concept is to represent all other constituent flows as an ‘aggregate cross flow’ which
can be characterized using an arrival curve. Then the service provided to the constituent flow of interest can also be described by a service curve which is determined by the service curve provided to all arrival flows and the arrival curve of the ‘aggregate cross flow’.

The superposition property of the v.b.c SAC [22] is reviewed here because it is relevant to the model transformation in the following content.

**Theorem 1.** Consider $N$ flows with arrival processes $A_i$, $i = 1, ..., N$, respectively. If each arrival process has a v.b.c SAC $\alpha_i(t) \in \mathcal{F}$ with bounding function $f_i(x) \in \mathcal{F}$, then the aggregate arrival process has a v.b.c SAC $\alpha(t) \in \mathcal{F}$ with bounding function $f(x) \in \mathcal{F}$, where

$$\alpha(t) = \sum_{i=1}^{N} \alpha_i(t) \quad \text{and} \quad f(x) = f_1 \otimes \cdots \otimes f_N(x).$$

### 3. Time-domain Modeling and Transformations

This section defines traffic and service models in the *time-domain*. Particularly, traffic models are defined based on probabilistic lower bounds on the *cumulative inter-arrival time* between two arbitrary packets. Service models are defined in terms of the virtual time function and probabilistic upper bounds on the *cumulative service time* between two arbitrary packets. Moreover, we establish the transformations among these models as well as the transformation between the time-domain model and the space-domain model.

#### 3.1. Time-domain Traffic Models

Consider an arrival process that specifies packets arriving to a network system at time $a(n)$, $n = 0, 1, 2, ...$. In order to stochastically guarantee a certain level of QoS to this arrival process, this arrival process should be constrained. By characterizing the constrained arrival traffic from the temporal perspective, we define an inter-arrival-time (i.a.t) stochastic arrival curve model.

**Definition 5. (i.a.t Stochastic Arrival Curve)**

A flow is said to have an inter-arrival-time (i.a.t) stochastic arrival curve $\lambda(n) \in \mathcal{F}$ with bounding function $h(x) \in \mathcal{F}$, if for any $m, n \geq 0$ and $x \geq 0$, there holds

$$P\left\{a(m+n)-a(n) < [\lambda(n)-x]^+\right\} \leq h(x).$$

(7)
Eq.(7) indicates that function $\lambda(n)$ is a probabilistic lower bound on the cumulative inter-arrival time. The violation probability that the cumulative inter-arrival time is smaller than $\lambda(n)$ is bounded above by function $h(x)$. If $h(x) = 0$ for all $x \geq 0$, Eq.(7) represents a time-domain deterministic arrival curve \[10\] which is a special case of the $i.a.t$ SAC.

Queueing theory typically characterizes the arrival process using the probability distribution of the inter-arrival time between two consecutive customers:

$$P\{a(n) - a(n-1) \leq x\} = F(x).$$

Comparing $F(x)$ with Eq.(7), we notice that Eq.(7) gives a more general probability expression of the inter-arrival time between two arbitrary packets. From this viewpoint, $F(x)$ is a special case of Eq.(7).

**Example 1.**

Consider a flow of packets with fixed packet size. Suppose that packet inter-arrival times follow an exponential distribution with mean $1/\mu$. Then, the packet arrival time has an Erlang distribution with parameter $(n, \mu)$ \[1\], where $n$ denotes the number of arrival packets. For any two packets $P(m)$ and $P(m+n)$, their inter-arrival time satisfies, for $x \geq 0$,

$$P\{a(m+n) - a(m) < \frac{n}{\mu} - x\} \leq P\{a(m+n) - a(m) \leq \left[\frac{n}{\mu} - x\right]^+\}$$

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

where $y = \frac{n}{\mu} - x$. Thus, the flow has an $i.a.t$ SAC $\lambda(n) = \frac{n}{\mu}$.

The $i.a.t$ SAC is simple but has limited applications. For example, consider a virtual single server queue (SSQ) fed with the arrival traffic which has an $i.a.t$ SAC $\lambda(n)$ with bounding function $h(x)$. Suppose that the virtual SSQ provides a constant service time $\lambda(1)$ for each packet. From Eq.(3), the waiting delay of $P(n)$ experienced in the virtual SSQ is

$$W(n) = d(n) - a(n) - \lambda(1)$$

$$= \sup_{0 \leq m \leq n} \left[a(m) + \lambda(n - m + 1)\right] - a(n) - \lambda(1) \quad (8)$$

$$= \sup_{0 \leq m \leq n} \left\{\lambda(n - m) - [a(n) - a(m)]\right\} \quad (9)$$

10
where $a(m)$ is the beginning of the backlogged period within which packet $P(n)$ is transmitted. Eq. (8) is derived from the departure time given in Eq. (13). Eq. (9) is called the virtual-waiting-delay property. It is difficult to compute the virtual-waiting-delay from Eq. (7). When investigating the performance guarantees such as delay bound and backlog bound in Section 4.1, we face the similar difficulty.

In order to deal with the difficulty of computing the virtual-waiting-delay, we define another stochastic arrival curve model based on Eq. (9).

**Definition 6. (v.w.d Stochastic Arrival Curve)**

A flow is said to have a virtual-waiting-delay (v.w.d) stochastic arrival curve $\lambda(n) \in \mathcal{F}$ with bounding function $h(x) \in \bar{\mathcal{F}}$, if for any $0 \leq m \leq n$ and $x \geq 0$, there holds

$$\mathbb{P}\left\{ \sup_{0 \leq m \leq n} \left\{ \lambda(n - m) - [a(n) - a(m)] \right\} > x \right\} \leq h(x). \quad (10)$$

Through some manipulations, Eq. (10) can be expressed as the max-plus convolution:

$$\mathbb{P}\{a(n) < a \otimes \lambda(n) - x\} \leq h(x). \quad (11)$$

Here, $a \otimes \lambda(n)$ can be considered as the expected time that the packet would arrive to the head-of-line (HOL) if the flow has passed through a virtual SSQ with the (deterministic) service curve $\lambda(n)$. The packet is expected to arrive not earlier than the expected HOL time. Here $x$ represents the difference between the expected HOL time and the actual arrival time. The violation probability is bounded by the non-increasing function $h(x)$.

We use the v.w.d SAC to characterize the arrival traffic in Example 1.

**Example 2.**

Consider a flow that consists of packets having the fixed packet size. Suppose that 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 [33], we say that the flow has a v.w.d SAC $\lambda(n) = \overline{h} \cdot n$ with bounding function $h^{exp}$ for $0 < \overline{h} < \frac{1}{\mu}$. Let $\rho = \mu \cdot \overline{h}$. We obtain the bounding function of the probability that the waiting delay $W(n)$ exceeds $x (\geq 0)$

$$h^{exp}(x) = 1 - (1 - \rho) \sum_{i=0}^{\lfloor x \rfloor} e^{-\mu(i\overline{h} - x)} \frac{[\mu(i\overline{h} - x)]^i}{i!}$$
where, \([y]\) denotes the floor function.

The definition of \(v.w.d\) SAC is more strict than that of \(i.a.t\) SAC. As a result, it is not trivial to derive the \(v.w.d\) SAC for an arrival process even if it can be characterized by an \(i.a.t\) SAC. Thus, it is important to explore whether there exists some relationship between the \(i.a.t\) SAC and the \(v.w.d\) SAC.

**Theorem 2.**

1. If a flow has a \(v.w.d\) SAC \(\lambda(n) \in \mathcal{F}\) with bounding function \(h(x) \in \bar{\mathcal{F}}\), then the flow has an \(i.a.t\) SAC \(\lambda(n) \in \mathcal{F}\) with the same bounding function \(h(x) \in \bar{\mathcal{F}}\).

2. Conversely, if a flow has an \(i.a.t\) SAC \(\lambda(n) \in \mathcal{F}\) with bounding function \(h(x) \in \mathcal{G}\), it also has a \(v.w.d\) SAC \(\lambda_{-\eta}(n) \in \mathcal{F}\) with bounding function \(h_\eta(x) \in \bar{\mathcal{G}}\), where for \(\eta > 0\)

\[
\lambda_{-\eta}(n) = [\lambda(n) - \eta \cdot n]^+ \quad \text{and} \quad h_\eta(x) = \left[ h(x) + \frac{1}{\eta} \int_x^{\infty} h(y) dy \right].
\]

**Remark.** In the second part, \(h(x) \in \mathcal{G}\) while not \(h(x) \in \mathcal{F}\). If the requirement on the bounding function is relaxed to \(h(x) \in \mathcal{F}\), the second part may not hold in general.

Theorem 2 reveals that if an arrival process can be modeled by a \(v.w.d\) SAC \(\lambda(n)\), then \(\lambda(n)\) is also the \(i.a.t\) SAC of this arrival process. On the other hand, if an arrival process can be modeled by an \(i.a.t\) SAC with the associated bounding function in \(\mathcal{G}\), then this arrival process also has a \(v.w.d\) SAC which may be associated with a more loose bounding function.

It is worth highlighting that the \(v.w.d\) SAC looks similar to the \(v.b.c\) SAC (see Definition 2) defined in the space-domain. Since these two models play an important role in performance analysis in their respective domains, we establish their relationship in the following theorem.

**Theorem 3.**

1. If a flow has a space-domain \(v.b.c\) SAC \(\alpha(t) \in \mathcal{F}\) with bounding function \(f(x) \in \bar{\mathcal{F}}\), the flow has a time-domain \(v.w.d\) SAC \(\lambda(n) \in \mathcal{F}\) with bounding function \(h(y) \in \mathcal{F}\), where

\[
\lambda(n) = \inf\{\tau : \alpha(\tau) \geq n\}, \quad \text{and} \quad h(y) = f(z^{-1}(y))
\]

with \(z^{-1}(y)\) denoting the inverse function of \(y\), where

\[
y = z(x) \equiv \sup_{k \geq 0} \{\lambda(k) - \lambda(k - x)\}.
\]

\(^2\)Note that \(\eta\) should not be greater than \(\lim_{n \to \infty} \frac{\lambda(n)}{n}\).
Specifically, if \( \lambda(\cdot) \) is sub-additive, \( z(x) = \lambda(x) \).

2. Conversely, if a flow has a time-domain v.w.d SAC \( \lambda(n) \in \mathcal{F} \) with bounding function \( h(y) \in \bar{\mathcal{F}} \), the flow has a space-domain v.b.c SAC \( \alpha(t) \in \mathcal{F} \) with bounding function \( f(x) \in \mathcal{F} \), where

\[
\alpha(t) = \sup \{ k : \lambda(k) \leq t \}, \quad \text{and} \quad f(x) = h(z^{-1}(x))
\]

with \( z^{-1}(x) \) denoting the inverse function of \( x \), where

\[
x = z(y) \equiv \sup_{\tau \geq 0} \{ \alpha(\tau + y) - \alpha(\tau) + 1 \}.
\]

Specifically, if \( \alpha(\cdot) \) is sub-additive\(^3\), \( z(y) = \alpha(y) + 1 \).

Note that in Theorem 3, the arrival curve \( \alpha(t) \) denotes the cumulative number of arrival packets while not the cumulative amount (in bits) of arrival traffic.

The generalized stochastically bounded burstiness (gSBB) \(^3\) is a special case of the space-domain v.b.c SAC. A summarization of some well-known traffic belonging to gSBB is given \(^2\), including both Gaussian self-similar processes \(^2\)[11][25][30], such as fractional Brownian motion, and non-Gaussian self-similar processes, such as \( \alpha \)-stable self-similar process \(^3\)[24], and the \((\sigma(\theta), \rho(\theta))\) stochastic traffic model \(^7\)[9]. With Theorem 3, the following example shows that gSBB can be readily represented using the time-domain v.w.d SAC.

**Example 3.**

If an arrival process \( A(t) \) can be described by gSBB with upper rate \( \rho \) and bounding function \( f(x) \in \mathcal{F} \), i.e., for any \( t, x \geq 0 \), there holds

\[
P \left\{ \sup_{0 \leq s \leq t} \left\{ A(s, t) - \rho \cdot (t - s) \right\} > x \right\} \leq f(x),
\]

then the process \( A(t) \) has a v.b.c SAC \( \alpha(t) = \rho \cdot t \) with the bounding function \( f(x) \). With Theorem 3 (1), the arrival process has a v.w.d SAC \( \lambda(n) = \frac{n}{\rho} \), which is sub-additive and the bounding function \( h(y) = f(\rho \cdot y) \), i.e.,

\[
P \left\{ \sup_{0 \leq m \leq n} \left\{ \frac{1}{\rho} \cdot (n - m) - [a(n) - a(m)] \right\} > y \right\} \leq f(\rho \cdot y).
\]

\(^3\) clarifies that \( \alpha(t) \) defines a meaningful constraint only if it is subadditive. If \( \alpha(t) \) is not subadditive, it can be replaced by its subadditive closure.
Remark. Theorem 3 allows us to readily utilize the results of gSBB traffic for time-domain models. If the traffic is more suitable for being characterized by the time-domain traffic models rather than the space-domain traffic models, then the transformation between two domains can facilitate the analysis.

3.2. Time-domain Service Models

Queueing theory characterizes the service process of a system based on the per customer service time. Like the arrival model, time-domain service models extend to the cumulative service time.

If packet $P(n)$ arrives to a network system after packet $P(n - 1)$ has departed from the system, the departure time of $P(n)$ is the arrival time $a(n)$ plus the service time $\delta_n$, i.e., $a(n) + \delta_n$. If $P(n)$ arrives to the system while $P(n - 1)$ is still in the system, then its departure time is $d(n - 1) + \delta_n$. The combination of both cases gives the departure time of $P(n)$

$$d(n) = \max[a(n), d(n - 1)] + \delta_n$$

with $d(0) = a(0) + \delta_0$. Applying Eq.(12) iteratively to its right-hand side results in

$$d(n) = \sup_{0 \leq m \leq n} [a(m) + \sum_{k=m}^{n} \delta_k].$$

(13)

The system usually allocates a minimum service rate to an arrival flow in order to meet its QoS requirements. The guaranteed minimum service rate is related to the guaranteed maximum service time for each packet of the flow. Accordingly, the time that the packet departs from the system is bounded. Denote the guaranteed maximum service time by $\hat{\delta}_n$. The Guaranteed Rate Clock (GRC) is defined based on $\hat{\delta}_n$ [17] [18]:

$$GRC(n) = \max[a(n), GRC(n - 1)] + \hat{\delta}_n.$$  

(14)

with $GRC(0) = a(0) + \hat{\delta}_0$. Applying Eq.(14) iteratively to its right-hand side yields

$$GRC(n) = \sup_{0 \leq m \leq n} [a(m) + \sum_{k=m}^{n} \hat{\delta}_k].$$  

(15)
Eq. (15) is similar to Eq. (13) except for that \( GRC(n) \) represents the guaranteed departure time\(^4\) while \( d(n) \) is the actual departure time.

If \( \sum_{i=m}^{n} \delta_i \) is denoted by a function \( \gamma(n - m + 1) \), then Eq. (15) becomes

\[
GRC(n) = \sup_{0 \leq m \leq n} \left[ a(m) + \gamma(n - m + 1) \right] = a(\otimes) \gamma(n) \quad (16)
\]

which is the basis for the time-domain (deterministic) service model\(^10\). For systems that only provide service guarantees stochastically or applications that require only stochastic QoS guarantees, the service time may not need to be deterministically guaranteed. In this case, we extend the (deterministic) service curve into a probabilistic one.

**Definition 7. (i.d Stochastic Service Curve)**

A system is said to provide an inter-departure time (i.d) stochastic service curve \( \gamma(n) \in \mathcal{F} \) with bounding function \( j(x) \in \bar{\mathcal{F}} \), if for any \( n, x \geq 0 \), there holds

\[
P\left\{d(n) - a(\otimes) \gamma(n) > x\right\} \leq j(x). \quad (17)
\]

Note that the stochastic service curve of a service process is not unique. Therefore optimization is needed to find the SSC of a specific system.

**Example 4.**

Consider two nodes, the transmitter and the receiver. They communicate through an error-prone wireless link which is modeled as a slotted system. The wireless link can be considered as a stochastic server. Packets have fixed-length and are served in a FIFO manner by the transmitter. To simplify the analysis, we assume that the length of time slot equals one packet transmission time\(^5\).

The transmitter sends packets only at the beginning of a time slot. Due to the error-prone nature of the wireless link, the probability that a packet is successfully transmitted is determined by packet error rate (PER). Here, we assume that packet errors happen independently in every transmission with a fixed PER denoted by \( P_e \). The successful transmission probability of

\(^4\)The guaranteed departure time is actually \( GRC(n) + \text{error term} \)\(^17\), where \text{error term} is determined by the employed service discipline. The underlying service discipline considered throughout this paper is FIFO, under which, the error term is zero.

\(^5\)It means we only compute the number of time slots in this example.
one packet is hence $1 - P_e$. If error happens, the unsuccessfully transmitted packet will be retransmitted in the next time slot immediately. In order to guarantee 100% reliability, the packet will be retransmitted until it is successfully received by the receiver.

The per-packet service time $\delta_n$ is a geometric random variable with parameter $1 - P_e$. The cumulative service time of successfully transmitting packets $P(m)$ to $P(n)$ is $\sum_{k=m}^{n} \delta_k$ which follows the negative binomial distribution with parameter $1 - P_e$. The mean service time denoted by $\bar{\delta}$ equals $\frac{1}{1 - P_e}$.

According to the complementary cumulative distribution function (CCDF) of the negative binomial distribution, the cumulative service time between two arbitrary packets $P(m)$ and $P(m + n)$ is given by

$$P\left\{ \sum_{k=m}^{m+n} \delta_k > \bar{\delta} \cdot (n + 1) + x \right\} \leq \sum_{i=[\gamma(n+1)+x]}^{\infty} \left( \frac{i - 1}{n} \right) (1 - P_e)^{n+1} P_e^{-i(n+1)} \quad (18)$$

for any $x \geq 0$, where $[\cdot]$ is the ceiling function.

The right-hand side of Eq.(18) represents the bound on the probability that the actual cumulative service time exceeds the cumulative mean service time. Let $\gamma\eta(n) = \bar{\delta} \cdot n + \eta \cdot n$ for $\eta > 0$ and $j(x)$ denote the right-hand side of Eq.(18). From Definition 7, we know

$$d(n) - a \bar{\delta} \gamma\eta(n) = \sup_{0 \leq m \leq n} \left[ a(m) + \sum_{k=m}^{n} \delta_k \right] - \sup_{0 \leq m \leq n} \left[ a(m) + (\bar{\delta} + \eta) \cdot (n - m + 1) \right]$$

$$\leq \sup_{0 \leq m \leq n} \left[ \sum_{k=m}^{n} \delta_k - \bar{\delta} \cdot (n - m + 1) - \eta \cdot (n - m + 1) \right],$$

from which, we have

$$P\left\{ \sup_{0 \leq m \leq n} \left[ \sum_{k=m}^{n} \delta_k - \bar{\delta} \cdot (n - m + 1) - \eta \cdot (n - m + 1) \right] > x \right\}$$

$$\leq \sum_{m=0}^{n} P\left\{ \sum_{k=m}^{n} \delta_k - \bar{\delta} \cdot (n - m + 1) > x + \eta \cdot (n - m + 1) \right\}$$

$$\leq \sum_{m=0}^{n} j(x + \eta \cdot (n - m + 1))$$

16
\[\sum_{k=1}^{n+1} j(x + \eta \cdot k) \leq \left[ \frac{1}{\eta} \int_x^{\infty} j(y) dy \right]_1.\]

Thus, we conclude that this error-prone wireless link provides an \textit{i.d} SSC \(\gamma_\eta(n)\) with the bounding function \(j_\eta(x)\) for \(\eta > 0\), where

\[
\gamma_\eta(n) = \delta \cdot n + \eta \cdot n \quad \text{and} \quad j_\eta(x) = \left[ \frac{1}{\eta} \int_x^{\infty} j(y) dy \right]_1.
\]

Since Eq. (18) is only relevant to the cumulative service time and does not involve the arrival process, it provides a method to find the \textit{i.d} SSC.

**Remark.** Example 4 demonstrates that we can obtain the \textit{i.d} SSC from analyzing per-packet service time. However, if applying the space-domain results to this case, we need an impairment process \[22\] to characterize the cumulative amount of service consumed by unsuccessful transmissions. In other words, we still need to compute the cumulative slots due to failed transmission and then convert it into the amount of service. Such conversion may introduce error or result in looser bounds whereas the time-domain model directly computes the service time and avoids the conversion error. This simple example thus illustrates the feasibility of the time-domain service curve model.

In Section 4, we show that many results can be derived from the \textit{i.d} SSC. However, without additional constraints, we have difficulty in proving the concatenation property for the \textit{i.d} SSC. To address this difficulty, we introduce another service curve model in the following.

**Definition 8. (\(\eta\)-Stochastic Service Curve)**

A system is said to provide an \(\eta\)-stochastic service curve \(\gamma(n) \in F\) with bounding function \(j_\eta(x) \in F\), if for any \(n, x \geq 0\), there holds

\[
P\left\{ \sup_{0 \leq m \leq n} \left[ d(m) - a \otimes \gamma(m) - \eta \cdot (n - m) \right] > x \right\} \leq j_\eta(x), \quad (19)
\]

for any small \(\eta > 0\).

Note that the left-hand side of Eq. (19) represents a property that is typically hard to calculate. It means that Definition 8 is more strict than Definition 7. Thus it is important to find the relationship between the \textit{i.d} SSC and the \(\eta\)-stochastic service curves.
Theorem 4. 1. If a system provides to its arrival process an $\eta$-stochastic service curve $\gamma(n)$ with bounding function $j_0(x) \in \bar{F}$, it provides to the arrival process an i.d SSC $\gamma(n)$ with the same bounding function $j_\eta(x) \in \bar{F}$; 

2. If a system provides to its arrival process an i.d SSC $\gamma(n)$ with bounding function $j(x) \in \bar{G}$, it provides to the arrival process an $\eta$-stochastic service curve $\gamma(n)$ with bounding function $j_\eta(x) \in \bar{G}$ for $\eta > 0$, where

$$j_\eta(x) = \left[ j(x) + \frac{1}{\eta} \int_x^\infty j(y)dy \right]_0^1.$$ 

Again, in the second part of Theorem 4, $j(x) \in \bar{G}$ while not $j(x) \in \bar{F}$. If the requirement on the bounding function is relaxed to $j(x) \in \bar{F}$, the above relationship may not hold in general.

Definition 7 explores the relationship between the arrival process and the departure process, but it does not explicitly characterize the service process. From Eq.(17), it is not trivial to find the stochastic service curve $\gamma(n)$ for a specific system. Example 4 illustrates how to add some increment $\eta$ to the stochastic service curve $\gamma(n)$. To this end, we expand Eq.(16) to

$$d(n) - a \otimes \gamma(n) = \sup_{0 \leq m \leq n} \left[ a(m) + \Delta(m,n) \right] - a \otimes \gamma(n). \quad (20)$$

Without loss of generality, assume $a(m_0)$ ($0 \leq m_0 \leq n$) is the beginning of the backlogged period in which packet $P(n)$ is served. Then,

$$\sup_{0 \leq m \leq n} \left[ a(m) + \Delta(m,n) \right] = a(m_0) + \Delta(m_0,n)$$

and $a \otimes \gamma(n) \geq a(m_0) + \gamma(n - m_0 + 1)$.

We rewrite the right-hand side of Eq.(20) as

$$a(m_0) + \Delta(m_0,n) - a \otimes \gamma(n) \leq a(m_0) + \Delta(m_0,n) - a(m_0) - \gamma(m_0 + 1) = \Delta(m_0,n) - \gamma(n - m_0 + 1). \quad (21)$$

Note that Eq.(21) holds for arbitrary $m_0 \leq n$. Inspired by this, we define a new service curve model.
Definition 9. (Stochastic Strict Service Curve)

A system is said to provide stochastic strict service curve $\gamma(n) \in \mathcal{F}$ with bounding function $j(x) \in \bar{\mathcal{F}}$, if the cumulative service time between two arbitrary packets $P(m)$ and $P(n)$ satisfies for any $x \geq 0$,

$$P\left\{\Delta(m, n) - \gamma(n - m + 1) > x\right\} \leq j(x).$$

(22)

Eq. (21) reveals a relationship between the i.d SSC and the stochastic strict service curve. Furthermore, in Theorem 4(2), the relationship between the stochastic strict service curve and the $\eta-$stochastic service curve is obtained.

Theorem 5. Consider a system providing stochastic strict service curve $\gamma(n) \in \mathcal{F}$ with bounding function $j(x) \in \bar{\mathcal{F}}$.

1. It provides an i.d SSC $\gamma(n)$ with the same bounding function $j(x)$.
2. If $j(x) \in \bar{\mathcal{G}}$, it provides an $\eta-$stochastic service curve $\gamma(n)$ with bounding function $j_\eta(x) \in \bar{\mathcal{G}}$, where

$$j_\eta(x) = \left[j(x) + \frac{1}{\eta} \int_x^\infty j(y)dy\right].$$

Note that the second part of Theorem 5 requires the bounding function $j(x) \in \bar{\mathcal{G}}$ while not $j(x) \in \bar{\mathcal{F}}$.

4. Fundamental Properties

In this section, we explore the four fundamental properties for time-domain models, i.e. service guarantees, output characterization, concatenation property and superposition property. Some properties can only be proved for the combination of a specific traffic model and a specific service mode. This is why we have established various transformations between models in Section 3. With these transformations, we can flexibly apply the corresponding models to specific network scenarios.

4.1. Service Guarantees

Suppose that the arrival process has a v.w.d SAC and the service process has an i.d SSC. Under this condition, we derive the delay bound and backlog bound.

---

6If $P(m)$ and $P(n)$ are in the same backlogged period, $\Delta(m, n) = d(n) - d(m - 1)$. 

19
4.1.1. Delay Bound

The system delay significantly impacts QoS and is an important performance metric.

**Theorem 6. (System Delay Bound).**

Consider that a system provides an i.d SSC \( \gamma(n) \in \mathcal{F} \) with bounding function \( j(x) \in \overline{\mathcal{F}} \) to the input which has a v.w.d SAC \( \lambda(n) \in \mathcal{F} \) with bounding function \( h(x) \in \overline{\mathcal{F}} \). Let \( D(n) = d(n) - a(n) \) be the system delay of packet \( P(n) \). For \( x \geq 0 \), \( D(n) \) is bounded by

\[
P\{D(n) > x\} \leq j \otimes h([x - \gamma \odot \lambda(1)])^+.
\]

(23)

If the arrival process and the service process are independent of each other, we obtain another system delay bound according to Lemma 6.1 [22].

**Lemma 1. (System delay bound: independent condition)**

Consider that a system provides an i.d SSC \( \gamma(n) \in \mathcal{F} \) with bounding function \( j(x) \in \overline{\mathcal{F}} \) to the arrival process which has a v.w.d SAC \( \lambda(n) \in \mathcal{F} \) with bounding function \( h(x) \in \overline{\mathcal{F}} \). Suppose that the arrival process and the service process are independent of each other. Then for \( x \geq 0 \), the system delay \( D(n) \) is bounded by

\[
P\{D(n) > x\} \leq 1 - \bar{j} * \bar{h}([x - \gamma \odot \lambda(1)])^+,
\]

(24)

where \( \bar{j}(x) = 1 - [j(x)]_1 \) and \( \bar{h}(x) = 1 - [h(x)]_1 \).

4.1.2. Backlog Bound

The system backlog represents the total number of packets in the system at time \( t \), including both the packets waiting in the buffer and the packet being served. It is determined by function [11]:

\[
B(t) \leq \inf \{ l \geq 0, \sup\{n \geq 0 : a(n \leq t) : d(n - l) \leq a(n)\} \}.
\]

The following theorem provides a probabilistic bound on the system backlog for the given arrival process and service process.

**Theorem 7. (Backlog Bound)**

Consider that a system provides an i.d SSC \( \gamma(n) \in \mathcal{F} \) with bounding function \( j(x) \in \overline{\mathcal{F}} \) to the arrival process which has a v.w.d SAC \( \lambda(n) \in \mathcal{F} \)
with bounding function \( h(x) \in \bar{F} \). The system backlog at time \( t \geq 0 \) is bounded by

\[
P\{B(t) > x\} \leq j \otimes h(\gamma \ominus \lambda([x - 1]^+))
\]

(25)

for \( x \geq 1 \).

Let

\[
H(\lambda, \gamma + x) = \sup_{m \geq 0} \left\{ \inf[k \geq 0 : \gamma(m) + x \leq \lambda(m + k)] \right\}
\]

represent the maximum horizontal distance between functions \( \lambda(n) \) and \( \gamma(n) + x \). The probability that \( B(t) \) exceeds \( H(\lambda, \gamma + x) \) is bounded by

\[
P\{B(t) > H(\lambda, \gamma + x) + 1\} \leq j \otimes h(x).
\]

(26)

Remark. \( H(\lambda, \gamma + x) \) can be considered as the maximum system backlog in a (deterministic) virtual system, where the arrival process is \( \lambda(n) \) and the service process is \( \gamma(n) + x \). Eq. (26) is thus a bound on this maximum system backlog.

If the arrival process and the service process are independent of each other, another backlog bound is derived according to Lemma 6.1 [22].

**Lemma 2. (Backlog Bound: independent condition)**

Consider that a system provides an i.d SSC \( \gamma(n) \in F \) with bounding function \( j(x) \in \bar{F} \) to the arrival process which has a v.w.d SAC \( \lambda(n) \in F \) with bounding function \( h(x) \in \bar{F} \). Suppose that the arrival process and the service process are independent of each other. Then the system backlog at time \( t \geq 0 \) is bounded by:

\[
P\{B(t) > x\} \leq 1 - \bar{j} \ast \bar{h}(\gamma \ominus \lambda([x - 1]^+))
\]

(27)

for \( x \geq 1 \).

The probability that \( B(t) \) exceeds \( H(\lambda, \gamma + x) \) is bounded by

\[
P\{B(t) > H(\lambda, \gamma + x) + 1\} \leq 1 - \bar{j} \ast \bar{h}(x).
\]

(28)

4.2. Output Characterization

The previous section has presented how to derive the service guarantees in a single node. Another common scenario with which performance analysis deals is the end-to-end performance. An intuitive and simple approach
is called node-by-node analysis \[19\] which requires characterization of the departure process from a single node.

Let us consider a simple network as shown in Figure 1. The departure process of Server 1 is the arrival process for Server 2.

The delay bound in Server 1 can be derived from the result of Section 4.1.1. To derive the delay bound in Server 2, we need to characterize the arrival process to Server 2, which is the departure process from Server 1. The problem is how to characterize the departure process from Server 1.

**Theorem 8. (Output Characterization)**

Consider that a system provides an i.d SSC \( \gamma(n) \) \( \in \mathcal{F} \) with bounding function \( j(x) \) \( \in \bar{\mathcal{F}} \) to its arrival process which has a v.w.d SAC \( \lambda(n) \) \( \in \mathcal{F} \) with bounding function \( h(x) \) \( \in \bar{\mathcal{F}} \). The output has an i.a.t SAC \( \lambda \ominus \gamma(n - m - 1) \) with bounding function \( j \otimes h(x) \) \( \in \bar{\mathcal{F}} \), i.e., for any \( 0 \leq m < n - 1 \), there holds

\[
P\left\{ \lambda \ominus \gamma(n - m - 1) - [d(n) - d(m)] > x \right\} \leq j \otimes h(x). \tag{29}
\]

**Remark.** In Theorem 8 the initial arrival process has a v.w.d SAC while the departure process has an i.a.t SAC. In order to derive the service guarantees in Server 2, we need Theorem 2 (2) to transform the i.a.t SAC into a v.w.d SAC. Such transformation introduces a loose bounding function. The node-by-node analysis thus generates a loose end-to-end delay bound. Network calculus possesses an attractive property, concatenation property, which is used to deal with the end-to-end performance analysis. The comparison between the node-by-node analysis and the concatenation analysis reveals that the latter yields a tighter end-to-end delay bound \[21\].

The output characterization property however is very useful when analyzing complicated network scenarios, such as Figure 2 where flows join or leave dynamically. In order to analyze the per-flow service guarantees, the departure process from each single node should be characterized using the arrival process to the node and the service process provided by the node.
Figure 2: Complicated network scenario

Moreover, if the arrival process and the service process are independent of each other, the following lemma depicts the departure process.

Lemma 3. (Output Characterization: independent condition.)
Consider that a system provides an i.d SSC $\gamma(n) \in \mathcal{F}$ with bounding function $j(x) \in \mathcal{F}$ to its arrival process which has a v.w.d SAC $\lambda(n) \in \mathcal{F}$ with bounding function $h(x) \in \bar{\mathcal{F}}$. The output has an i.a.t SAC $\lambda^*(n) \in \mathcal{F}$ with bounding function $h^*(x) \in \bar{\mathcal{F}}$, where

$$\lambda^*(n) = \lambda \otimes \gamma(n-1) \quad \text{and} \quad h^*(x) = 1 - \bar{j} \ast \bar{h}(x).$$

(30)

4.3. Concatenation Property

The concatenation property aims to use an equivalent system to represent a system of multiple servers connected in tandem if each server provides a service curve to its input. Then this equivalent system can be considered as a ‘black box’ which also provides the initial input with a service curve.

In the following discussion, $\gamma^k$ and $j^k$ denote the stochastic service curve and bounding function of the $k$th server. For packet $P(n)$, the time arriving to the $k$th server is $a^k(n)$ and the time departing from the $k$th server is $d^k(n)$. For a network of $N$ tandem servers, the initial arrival is $a(n)$ and the final departure is $d(n)$.

Theorem 9. (Concatenation Property)

Consider a flow passing through a system of $N$ nodes connected in tandem. If each node $k (= 1, 2, \ldots, N)$ provides an i.d SSC $\gamma^k(n) \in \mathcal{F}$ with bounding function $j^k(x) \in \bar{\mathcal{G}}$ to its input, the system provides to the initial input $a(n)$ an i.d SSC $\gamma(n)$ with bounding function $j(x)$, where

$$\gamma(n) = \gamma^1 \otimes \gamma^2 \otimes \cdots \otimes \gamma^N(n),$$

$$j(x) = j^{1 \cdot n} \otimes j^{2 \cdot n} \otimes \cdots \otimes j^{N}(x),$$

with

$$\gamma^k_{(k-1)n}(n) = \gamma^k(n) + (k-1) \cdot \eta \cdot n$$

23
for $k = 2, \ldots, N$ and $\eta > 0$, and

$$j^{k, \eta_k}(x) = \left[ j^k(x) + \frac{1}{\eta_k} \int_x^\infty j^k(y) dy \right]_1$$

for $k = 1, \ldots, N - 1$ and $\eta_k > 0$.

The proof of Theorem 9 utilizes the relationship between the i.d SSC and the $\eta$–stochastic service curve. The following lemma directly describes the service characterization of a system of nodes connected in tandem, where each single node provides an $\eta$–stochastic service curve to its input.

**Lemma 4.** Consider a flow passing through a system of $N$ nodes connected in tandem. If each node $k (= 1, 2, \ldots, N)$ provides an $\eta$-stochastic service curve $\gamma^k(n) \in \mathcal{F}$ with bounding function $j^k(x) \in \bar{\mathcal{F}}$ to its input, i.e.,

$$P \left\{ \sup_{0 \leq m \leq n} \{ d^k(m) - a^k \otimes \gamma^k(m) - \eta \cdot (n - m) \} > x \right\} \leq j^k(x),$$

then the system provides to the initial arrival process an i.d SSC $\gamma(n)$ with bounding function $j(x)$:

$$\gamma(n) = \gamma^1 \otimes \gamma^2 \otimes \cdots \otimes \gamma^N \eta(n)$$

$$j(x) = j^1 \otimes j^2 \otimes \cdots \otimes j^N(x),$$

where $\gamma^k(n) = \gamma^k(n) + (k - 1) \cdot \eta \cdot n$, $k = 2, \ldots, N$, for any small $\eta > 0$.

**Remark.** The proof of the concatenation property reveals another reason of defining the $\eta$–stochastic service curve model.

**4.4. Superposition Property**

The superposition property can be applied for multiplexing individual flows into an aggregated flow under the FIFO aggregate scheduling. The arrival process of the aggregate flow can be characterized by a stochastic arrival curve if the arrival process of each individual flow can be stochastically characterized by a stochastic arrival curve. Then we only need to analyze the service guarantees for the aggregate flow since all constituent flows are served equally.
4.4.1. Superposition of Renewal Processes

The superposition of multiple flows essentially falls into the research issue - superposition of renewal processes. In queueing networks, an individual server may receive inputs from different sources. It is reasonable to assume that the arrival process to a server is a superposition of statistically independent constituent processes [26]. The individual constituent processes are typically considered as renewal processes. A renewal process is a counting process in which the times between successive events are independent and identically distributed possibly with an arbitrary distribution [32].

The superposition of renewal processes has been widely studied since the original investigation by Cox and Smith [13]. However, the renewal property is not preserved under superposition except for Poisson sources. More precisely, the inter-arrival times in the superposition process become statistically dependent. This property cannot be captured by the renewal model [35].

In the following, we introduce how to characterize the superposition processes of multiple flows from a network calculus viewpoint.

4.4.2. Arrival Time Determination

First, we only consider the superposition of two flows denoted by $F_1$ 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, respectively. As shown in Figure 3, $F_1$ and $F_2$ are aggregated in the FIFO manner. If two or more than two packets which belong to different flows arrive simultaneously, they are inserted into the FIFO queue arbitrarily.

Figure 3: Aggregation of two flows

Figure 4 depicts that the arrival process of the aggregate flow is dependent on the arrival process of two constituent flows.

Recall that $P(n)$ denotes the $(n+1)$th packet of the aggregate flow. The same notation is also used for constituent flows $F_1$ and $F_2$. Thus, packet $P(n)$ of the aggregate flow is either the $m$th packet of flow $F_1$ (i.e., $P_1(m-1)$) or
the \((n + 1 - m)\)th packet of flow \(F_2\) (i.e., \(P_2(n - m)\)), where \(0 \leq m \leq n + 1\). When \(m = 0\), it means no packet of flow \(F_1\) arriving yet. When \(m = n + 1\), it means no packet of flow \(F_2\) arriving yet. By convention, we adopt \(a_i(n) = 0\) for \(n < 0\). Since \(m\) takes value between 0 to \(n + 1\), there are \(n + 2\) combinations.

**Theorem 10.** Consider that two flows \(F_1\) and \(F_2\) arrive to a network system and are aggregated into one flow \(F_A\) in the FIFO manner. Let \(a_1(n)\), \(a_2(n)\) and \(a(n)\) be the arrival process of flows \(F_1\), \(F_2\) and \(F_A\), respectively. Then the packet arrival time of the aggregate flow is determined by

\[
a(n) = \min_{0 \leq m \leq n+1} \left\{ \max\left[a_1(m-1), a_2(n-m)\right]\right\} \quad (31)
\]

with

\[
a(0) = \min \left\{ \max[0, a_2(0)], \max[a_1(0), 0] \right\} = \min[a_1(0), a_2(0)].
\]

We use an example to explain the underlying concept of Theorem 10. In Figure 4, observe the arrival process of the aggregate flow at time \(t\). Packet \(P(4)\) (arrival time: \(a(4) < t\)) is the last arrival packet, which is either packet \(P_1(m-1)\) or packet \(P_2(4-m)\), depending on which packet’s arrival time is closer to time \(t\), i.e., \(a(4) = \max[a_1(m-1), a_2(4-m)]\) for \(m = 0, 1, 2, 3, 4, 5\), the arrival time of packet \(P(4)\) is one element of the following set denoted by \(A\)

\[
A = \left\{ a_1(4), a_2(4), \max[a_1(0), a_2(3)], \max[a_1(1), a_2(2)], \max[a_1(2), a_2(1)], \max[a_1(3), a_2(0)] \right\},
\]
i.e., \( a(4) = \min \{ A \} \). We notice that \( \min \{ A \} \) is actually the expansion of Eq.\((31)\). According to the packet arrival times of two constituent flows shown in Figure 4, we have

\[
a(4) = \min \left\{ a(4), a_{1}(4), a_{2}(4), \max[a_{1}(0), a_{2}(3)] = a(3), \max[a_{1}(1), a_{2}(2)] = a_{2}(2), \max[a_{1}(2), a_{2}(1)] = a_{1}(2), \max[a_{1}(3), a_{2}(0)] = a_{1}(3) \right\} = a_{1}(2),
\]

which is consistent with Figure 4.

Theorem 10 can be generalized to the aggregation of \( N(\geq 2) \) flows.

**Corollary 1.** Consider that \( N(\geq 2) \) flows \( F_{1}, F_{2}, \ldots, F_{N} \) arrive to a network system and are aggregated into one flow \( F_{A} \) in the FIFO manner. Let \( a_{1}(n), a_{2}(n), \ldots, a_{N}(n) \) and \( a(n) \) be the arrival process of the \( N \) constituent flows and the aggregate flow, respectively. Then the packet arrival time of the aggregate flow is determined by

\[
a(n) = \min_{\sum_{m_{i}=n+1, m_{i} \in [0,n+1]}} \left\{ \max[a_{1}(m_{1}-1), a_{2}(m_{2}-1), \ldots, a_{N}(n - \sum_{i=1}^{N-1} m_{i})] \right\}
\]

with

\[
a(0) = \min \{ a_{1}(0), a_{2}(0), \ldots, a_{N}(0) \}.
\]

### 4.4.3. Superposition Process Characterization

Eq.\((31)\) can compute the packet arrival time of the aggregate flow. However, we still have the difficulty in characterizing the packet inter-arrival time of the aggregate flow if the packet inter-arrival times of two constituent flows follow the general distribution. For this reason, it is difficult to directly characterize the arrival process of the aggregate flow from the temporal perspective. Alternatively, we rely on the available results of the superposition property explored in the space-domain (see Theorem 1).

In the space-domain, the traffic arrival process is characterized based on the cumulative amount of arrival traffic. In the following, we use \( A(t), A_{1}(t) \) and \( A_{2}(t) \) to denote the cumulative number of arrival packets of the aggregate flow up to time \( t \), the cumulative number of arrival packets of \( F_{1} \) up to time \( t \) and the cumulative number of arrival packets of \( F_{2} \) up to time \( t \), respectively. \( A(t) \) is the sum of \( A_{1}(t) \) and \( A_{2}(t) \), from which we can find the stochastic arrival curve for the aggregate flow.
As shown in Figure 5, the condition is that the time-domain stochastic arrival curve of all constituent flows are known, and the target is to verify that the aggregate flow also has a time-domain stochastic arrival curve.

If a flow has a time-domain \(v.w.d\) SAC, with Theorem 3(2), this flow has a space-domain \(v.b.c\) SAC, for which the superposition property holds (refer Theorem 1). Applying Theorem 3(1) gives rise to the \(v.w.d\) SAC for the aggregate flow.

If flow \(F_i\) has a \(v.w.d\) SAC \(\lambda_i(n)\) with bounding function \(h_i(x), i = 1, 2, ..., N,\) from Theorem 3(2), we can verify that flow \(F_i\) has a \(v.b.c\) SAC \(\alpha_i(t)\) with bounding function \(f_i(x) = h_i(z_i^{-1}(x)),\) where \(\alpha_i(t)\) and \(z_i^{-1}(x)\) are given in Theorem 3(2). Furthermore, according to Theorem 1, the aggregate flow has a \(v.b.c\) SAC \(\alpha(t) = \sum_{i=1}^{N} \alpha_i(t)\) with bounding function \(f(x) = f_1 \otimes \cdots \otimes f_N(x).\) Finally, we apply Theorem 3(1) and can verify that the aggregate flow also has a \(v.w.d\) SAC.

**Theorem 11. (Superposition property)**

Consider the aggregate of \(N\) flows. If the arrival process of each flow has a \(v.w.d\) SAC \(\lambda_i(n) \in \mathcal{F}\) for \(i = 1, 2, ..., N,\) i.e.,

\[
P\{a_i(n) < a_i \otimes \lambda_i(n) - y\} \leq h_i(y),
\]
which implies that every flow also has a \(v.b.c\) SAC

\[
\alpha_i(t) = \sup\{k : \lambda_i(k) \leq t\}
\]
with bounding function

\[
f_i(x) = h_i(z_i^{-1}(x))
\]
where \( z_i^{-1}(x) \) denote the inverse function of \( x \):

\[
x = z_i(y) \equiv \sup_{\tau \geq 0} \{ \alpha_i(\tau + y) - \alpha_i(\tau) + 1 \}.
\]

Then the aggregate arrival process \( a(n) \) has a v.w.d SAC \( \lambda(n) \) with bounding function \( h(y) \), where

\[
\lambda(n) = \inf \{ \tau : \sum_{i=1}^{N} \alpha_i(\tau) \geq n \}, \quad h(y) = f(z^{-1}(y)),
\]

with \( f(x) = f_1 \otimes \cdots \otimes f_N(x) \) and \( z^{-1}(y) \) denoting the inverse function of \( y \):

\[
y = z(x) \equiv \sup_{k \geq 0} \{ \lambda(k) - \lambda(k - x) \}.
\]

### 4.4.4. Special Case: Superposition of Poisson Processes

As we have mentioned in Section 4.4.1, the Poisson process is a special case of renewal processes because its renewal property is preserved under superposition. In addition, the superposition of multiple Poisson processes is still a Poisson process. From the temporal perspective, the inter-arrival time between two arbitrary events of a superposition of Poisson arrivals follows the Gamma distribution.

**Example 5.** Consider the superposition process of two independent Poisson arrival processes. Suppose that all packets of both arrival processes have the same size. The packet inter-arrival times of two Poisson processes follow exponential distributions with mean \( \frac{1}{\mu_1} \) and \( \frac{1}{\mu_2} \), respectively. Find the time-domain v.w.d SAC for the superposition process.

In Example 2 the v.w.d stochastic arrival curve for a Gamma process has been derived. We thus know that the superposition process has a v.w.d SAC \( \lambda_s(n) = T_s \cdot n \) (0 < \( T_s < \frac{1}{\mu_1 + \mu_2} \)) with bounding function \( h_s(x) \):

\[
h_s(x) = 1 - (1 - \rho_s) \sum_{i=0}^{\lfloor \frac{x}{T_s} \rfloor} e^{-\mu_1 + \mu_2) \cdot (iT_s - x)} \left[ \frac{[(\mu_1 + \mu_2)(iT_s - x)]^i}{i!} \right],
\]

where \( \rho_s = (\mu_1 + \mu_2) \cdot T_s \).

**Remark.** It is readily to generalize the above example into the superposition of multiple independent Poisson processes.
5. Conclusions and Open Issue

This paper presented a temporal network calculus to formulate queueing systems in communication networks where applications can tolerate a certain level of performance violation. The time-domain models make it feasible to characterize the temporal behavior of network traffic and capture the temporal nature of the network capacity perceived by individual packets.

The models are defined in such a way to compromise between simple models and complex models. The former may not be sufficient to explore the fundamental properties whereas the latter may be too difficult to build. In order to solve this dilemma, we propose a transformation method such that the appropriate models are selected to some specific scenario. Moreover, we also link the temporal network calculus and the existing space-domain network calculus results through connecting the time-domain $v.w.d$ arrival curve and the corresponding space-domain $v.b.c$ arrival curve.

Four properties investigated in the time-domain facilitate performance analysis of various network scenarios. In addition, the proof of the superposition property has given insights into the importance of model transformation. We believe that this temporal network calculus is applicable for analyzing networks where users are served probabilistically and compliments the current network calculus results.

The leftover service characterization is useful for per-flow performance analysis and has been proved in the space-domain. We attempted to tackle this property under the condition that the arrival process has a deterministic time-domain arrival curve and the service process provides an i.d SSC $^{36}$. We will expand the investigation to this property under the general condition. One challenge is to be able to decouple the constituent flow’s arrival process from the aggregate arrival process.

References

[1] NIST/SEMATECH e-Handbook of Statistical Methods. 2006.

[2] R. Addie, P. Mannsralo, and I. Norros. Most probable paths and performance formulae for buffers with gaussian input traffic. European Transactions on Telecommunications, 13(3):183–196, 2002.

[3] A. Karasaridis and D. Hatzinakos. Network heavy traffic modeling using $\alpha$-stable self-similar processes. IEEE Trans. Commun., 49(7):1203–1214, July 2001.
[4] J.-Y. L. Boudec. Application of network calculus to guaranteed service networks. IEEE Trans. Infor. Theory, 44(3):1087–1096, May 1998.

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

[6] A. Burchard, J. Liebeherr, and S. D. Patek. A min-plus calculus for end-to-end statistical service guarantees. IEEE Trans. Information Theory, 52(9):4105–4114, Sept. 2006.

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

[8] 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.

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

[10] 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.

[11] J. Cheo and N. B. Shroff. A central-limit-theorem-based approach for analyzing queue behavior in high-speed networks. IEEE/ACM Trans. Networking, 6(5):659–671, Oct. 1998.

[12] F. Ciucu, A. Burchard, and J. Liebeherr. Scaling properties of statistical end-to-end bounds in the network calculus. IEEE Trans. Information Theory, 52(6):2300–2312, June 2006.

[13] D. R. Cox and W. L. Smith. On the superposition of renewal processes. Biometrika, 41(1-2):91–99, 1954.

[14] D. Ferrari. Client requirements for real-time communication services. IEEE Commun. Magazine, 28(11):65–72, Nov. 1990.

[15] M. Fidler. An end-to-end probabilistic network calculus with moment generating functions. In Proc. IEEE IWQoS, 2006.
[16] M. Fidler. A survey of deterministic and stochastic service curve models in the network calculus. *IEEE Commun. Surveys and Tutorials*, 12(1):59–86, Feb. 2010.

[17] 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.

[18] P. Goyal and H. M. Vin. Generalized guaranteed rate scheduling algorithms: A framework. *IEEE/ACM Trans. Networking*, 5(4):561–571, Aug. 1997.

[19] Y. Jiang. Delay bounds for a network of guaranteed rate servers with FIFO aggregation. *Computer Networks*, 40(6):683–694.

[20] Y. Jiang. A basic stochastic network calculus. In *Proc. ACM SIGCOMM 2006*, 2006.

[21] Y. Jiang. Internet quality of service - architectures, approaches and analyses. *http://www.q2s.ntnu.no/ jiang/Notes.pdf*, 2006.

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

[23] Y. Jiang, Q. Yin, Y. Liu, and S. Jiang. Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks. *Computer Networks*, 53(12):2011–2021, Mar. 2009.

[24] A. Karasaridis and D. Hatzinakos. A non-Gaussian self-similar process for broadband heavy traffic modeling. In *Proc. IEEE GLOBECOM*, 1998.

[25] H. S. Kim and N. B. Shroff. Loss probability calculations and asymptotic analysis for finite buffer multiplexers. *IEEE/ACM Trans. Networking*, 9(6):755–768, Dec. 2001.

[26] C. Y. T. Lam and J. P. Lehoczky. Superposition of renewal processes. *Advances in Applied Probability*, 23(1):64–85, March 1991.

[27] C. Li, A. Burchard, and J. Liebeherr. A network calculus with effective bandwidth. *IEEE/ACM Trans. Networking*, 15(6):1442–1453, Dec. 2007.
[28] C. Li, A. Burchard, and J. Liebeherr. A network calculus with effective bandwidth. *IEEE/ACM Trans. Networking*, 15(6):1442–1453, Dec. 2007.

[29] Y. Liu, C.-K. Tham, and Y. Jiang. A calculus for stochastic QoS analysis. *Performance Evaluation*, 64(6):547–572, July 2007.

[30] P. Mannersalo and I. Norros. A most probable path approach to queueing systems with general gaussian input. *Computer Networks*, 40(3):399–412, Oct. 2002.

[31] S. Mao and S. S. Panwar. A survey of envelope processes and their applications in quality of service provisioning. *IEEE Commun. Surveys and Tutorials*, 8(3):2–19, 2006.

[32] S. M. Ross. *Introduction to Probability Models*. Elsevier, 2006.

[33] 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.

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

[35] P. Torab and E. W. Kamen. On approximate renewal models for the superposition of renewal processes. In *Proc. IEEE ICC*, 2001.

[36] J. Xie and Y. Jiang. Stochastic service guarantee analysis based on time-domain models. In *Proc. 17th IEEE/ACM International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2009.

[37] 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.

[38] Q. Yin, Y. Jiang, S. Jiang, and P. Y. Kong. Analysis on generalized stochastically bounded bursty traffic for communication networks. In *Proc. 27th IEEE Local Computer Networks*, 2002.
Appendix A. Proofs of theorems and lemmas

Proof of Theorem 2

Proof. The first part follows from that for any $0 \leq m \leq n$, there trivially holds

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

For the second part, there holds

$$P\left\{ \sup_{0 \leq m \leq n} \{\lambda_{\eta}(n - m) - [a(n) - a(m)]\} > x \right\} \leq P\left\{ \sup_{0 \leq m \leq n} \{\lambda_{\eta}(n - m) - [a(n) - a(m)]\}^+ > x \right\}.$$

For any $x \geq 0$,

$$P\left\{ \{\lambda(n - m) - \eta \cdot (n - m) - [a(n) - a(m)]\}^+ > x \right\} = P\left\{ \lambda(n - m) - \eta \cdot (n - m) - [a(n) - a(m)] > x \right\} = P\left\{ \lambda(n - m) - [a(n) - a(m)] > x + \eta \cdot (n - m) \right\} \leq h(x + \eta \cdot (n - m)).$$

Based on the above steps, we have

$$P\left\{ \sup_{0 \leq m \leq n} \{\lambda_{\eta}(n - m) - [a(n) - a(m)]\} > x \right\} \leq \sum_{m=0}^{n} P\left\{ \{\lambda_{\eta}(n - m) - [a(n) - a(m)]\}^+ > x \right\} \leq \sum_{m=0}^{n} h(x + \eta \cdot (n - m)) = \sum_{k=0}^{n} h(x + \eta \cdot k) \leq \sum_{k=0}^{\infty} h(x + \eta \cdot k)$$

34
\[
= h(x) + \sum_{k=1}^{\infty} h(x + \eta \cdot k)
\]
\[
\leq h(x) + \frac{1}{\eta} \int_{x}^{\infty} h(y) dy.
\]

The right-hand side of the last inequality still belongs to \( \hat{G} \). The second part follows from the above inequality and the fact that the probability is always not greater than 1.

Proof of Theorem 3

Proof. (1) From Lemma 2 \[36\], we know that for any \( t, x \geq 0 \), event

\[ \{ A(t) \leq A \otimes \alpha(t) + x \} \]

implies event

\[ \{ a(n) \geq a \otimes \lambda(n) - y \} \]

where

\[ y = \sup_{k \geq 0} \{ \lambda(k) - \lambda(k - x) \} \equiv z(x) \].

Thus, there holds

\[ P\{ A(t) \leq A \otimes \alpha(t) + x \} \leq P\{ a(n) \geq a \otimes \lambda(n) - y \} \]

\[ \Rightarrow P\{ a(n) < a \otimes \lambda(n) - y \} \leq P\{ A(t) > A \otimes \alpha(t) + x \} \]

\[ \leq f(x) \].

Particularly, if \( \lambda \) is sub-additive, i.e. \( \lambda(a + b) \leq \lambda(a) + \lambda(b) \) for any \( a \) and \( b \), we then have:

\[ P\{ a(n) < a \otimes \lambda(n) - \lambda(x) \} \]

\[ \leq P\{ a(n) < a \otimes \lambda(n) - \sup_{k \geq 0} [\lambda(k) - \lambda(k - x)] \} \]

\[ \leq f(x) . \]

Hence, the first part follows.

(2) From Lemma 3 \[36\], we know that for any \( n, y \geq 0 \), event

\[ \{ a(n) \geq a \otimes \lambda(n) - y \} \]

35
implies event
\[ \{ A(t) \leq A \otimes \alpha(t) + x \} \]
where
\[ x = \sup_{u \geq 0} \{ \alpha(u + y) - \alpha(u) + 1 \} \equiv z(y). \]
Thus, there holds
\[
\begin{align*}
P\{ a(n) \geq a \otimes \lambda(n) - y \} & \leq P\{ A(t) \leq A \otimes \alpha(t) + x \} \\
\implies P\{ A(t) > A \otimes \alpha(t) + x \} & \leq P\{ a(n) < a \otimes \lambda(n) - y \} \\
& \leq h(y).
\end{align*}
\]
Particularly, if \( \alpha \) is sub-additive, we have
\[
\begin{align*}
P\{ A(t) > A \otimes \alpha(t) + \alpha(y) + 1 \} & \leq P\{ A(t) > A \otimes \alpha(t) + \sup_{u \geq 0} [\alpha(u + y) - \alpha(u) + 1] \} \\
& \leq h(y),
\end{align*}
\]
which ends the proof. \( \square \)

**Proof of Theorem 4**

*Proof.* The first part follows since there always holds
\[ d(n) - a \otimes \gamma(n) \leq \sup_{0 \leq m \leq n} [d(m) - a \otimes \gamma(m) - \eta \cdot (n - m)] \]
by letting \( m = n \) on the right hand side.

For the second part, there holds
\[
\begin{align*}
\sup_{0 \leq m \leq n} [d(m) - a \otimes \gamma(m) - \eta \cdot (n - m)] & \leq \sup_{0 \leq m \leq n} \{d(m) - a \otimes \gamma(m) - \eta \cdot (n - m)\}^+.
\end{align*}
\]
Hence for any \( x \geq 0 \), there exists
\[
\begin{align*}
P\{ \sup_{0 \leq m \leq n} \{d(m) - a \otimes \gamma(m) - \eta \cdot (n - m)\} > x \} & \leq \sum_{m=0}^{n} P\{d(m) - a \otimes \gamma(m) - \eta \cdot (n - m) > x \}
\end{align*}
\]
\[
\leq \sum_{u=0}^{n} j(x + \eta \cdot u)
\]

\[
\leq \left[ j(x) + \frac{1}{\eta} \int_{x}^{\infty} j(y) \, dy \right].
\]

The right-hand side of the above inequality still belongs to \( \bar{G} \) and is always not greater than 1. The proof of the second part is completed. \( \square \)

**Proof of Theorem 6**

Proof. For any \( n \geq 0 \), according to the definition of \( D(n) \), there holds

\[
D(n) = d(n) - a(n)
\]

\[
= [d(n) - a \otimes \gamma(n)] + [a \otimes \gamma(n) - a(n)]
\]

\[
= [d(n) - a \otimes \gamma(n)] + \sup_{0 \leq m \leq n} \{a(m) + \gamma(n - m + 1) - a(n)\}
\]

\[
= [d(n) - a \otimes \gamma(n)] + \sup_{0 \leq m \leq n} \{\lambda(n - m) - [a(n) - a(m)] + \gamma(n - m + 1) - \lambda(n - m)\}
\]

\[
\leq d(n) - a \otimes \gamma(n) + \sup_{0 \leq m \leq n} \{\lambda(n - m) - [a(n) - a(m)]\}
\]

\[
+ \sup_{0 \leq m \leq n} \{\gamma(n - m + 1) - \lambda(n - m)\}
\]

\[
\leq d(n) - a \otimes \gamma(n) + \sup_{0 \leq m \leq n} \{\lambda(n - m) - [a(n) - a(m)]\}
\]

\[
+ \sup_{k \geq 0} \{\gamma(k + 1) - \lambda(k)\}.
\]

To ensure system stability, we require

\[
\lim_{k \to \infty} \frac{1}{k}[\gamma(k) - \lambda(k)] \leq 0. \quad (A.1)
\]

In the proofs of the following theorems, without explicitly stating, we shall assume Eq. (A.1) holds.

In addition, the following results are given

\[
P\{d(n) - a \otimes \gamma(n) > x\} \quad \text{and} \quad P\left\{ \sup_{0 \leq m \leq n} \{\lambda(n - m) - [a(n) - a(m)]\} > x \right\}.
\]

From Lemma 1.5 \[22\] and \( \sup_{k \geq 0} \{\gamma(k + 1) - \lambda(k)\} = \gamma \otimes \lambda(1) \), we can conclude

\[
P\{D(n) > x\} \leq j \otimes h(x - \gamma \otimes \lambda(1)).
\]

37
Proof of Theorem 7

Proof. According to the backlog definition

\[ B(t) \leq \inf \left\{ x \geq 0, \sup \{ n \geq 0 : a(n) \leq t \} : d(n - x) \leq a(n) \right\}, \]

we need to prove the bounding function on the violation probability, i.e., \( P\{B(t) > x\} \). For ease of exposition, let \( n = m + x \), then we have

\[
\begin{align*}
&d(m) - a(m + x) \\
= & \quad [d(m) - a\tilde{\otimes}\gamma(m)] + [a\tilde{\otimes}\gamma(m) - a(m + x)] \\
= & \quad [d(m) - a\tilde{\otimes}\gamma(m)] + \sup_{0 \leq k \leq m} \{a(k) + \gamma(m - k + 1)\} - a(m + x) \\
= & \quad [d(m) - a\tilde{\otimes}\gamma(m)] + \sup_{0 \leq k \leq m} \{\lambda(m + x - k) - [a(m + x) - a(k)]} \\
& \quad + \gamma(m - k + 1) - \lambda(m + x - k) \}
\end{align*}
\]

\[
\leq \quad [d(m) - a\tilde{\otimes}\gamma(m)] + \sup_{0 \leq k \leq m + x} \{\lambda(m + x - k) - [a(m + x) - a(k)]} \\
& \quad - \inf_{0 \leq k \leq m + x} \{\lambda(m - k + x) - \gamma(m - k + 1)\}
\]

Let \( v = m - k + 1 \). The above inequality is written as

\[
\begin{align*}
d(m) - a(m + x) \\
\leq & \quad [d(m) - a\tilde{\otimes}\gamma(m)] + \sup_{0 \leq k \leq m + x} \{\lambda(m + x - k) - [a(m + x) - a(k)]} \\
& \quad - \inf_{1 \leq v \leq m + 1} \{\lambda(v + x - 1) - \gamma(v)\}.
\end{align*}
\]

Because there holds

\[
\inf_{1 \leq v \leq m + 1} \{\lambda(v + x - 1) - \gamma(v)\} \geq \inf_{v \geq 1} \{\lambda(v + x - 1) - \gamma(v)\} = \lambda\tilde{\otimes}\gamma([x - 1]^+),
\]

with the same conditions as analyzing the delay, we obtain

\[ P\{B(t) > x\} \leq j \otimes h(\lambda\tilde{\otimes}\gamma([x - 1]^+)). \]
To prove Eq. (26), we replace \( x = H(\lambda, \gamma + y) + 1 \) in event \( \{B(t) > x\} \) and have

\[
\begin{align*}
    d(m) - a(m + H(\lambda, \gamma + y) + 1) \\
    \leq \left[ d(m) - a \otimes \gamma(m) \right] + a \otimes \lambda(m + H(\lambda, \gamma + y) + 1) \\
    - a(m + H(\lambda, \gamma + y) + 1) + \sup_{v \geq 0} \{ \gamma(v) - \lambda(v + H(\lambda, \gamma + y)) \}.
\end{align*}
\]

The definition of \( H(\lambda, \gamma + y) \) implies

\[
\gamma(v) + y \leq \lambda(v + H(\lambda, \gamma + y))
\]

for any \( v \geq 0 \), i.e.,

\[
\sup_{v \geq 0} \{ \gamma(v) - \lambda(v + H(\lambda, \gamma + y)) \} \leq -y.
\]

Then we conclude

\[
P\{B(t) > H(\lambda, \gamma + x) + 1\} \leq j \otimes h(x).
\]

\[\square\]

Proof of Theorem 8

Proof. For any two departure packets \( m < n \), there holds

\[
\begin{align*}
d(m) - d(n) & \leq d(m) - a(n) \\
& = d(m) - a(n) + a \otimes \gamma(m) - a \otimes \gamma(m) \\
& = \left[ d(m) - a \otimes \gamma(m) \right] + \sup_{0 \leq k \leq m} \{ a(k) + \gamma(m - k + 1) \} - a(n) \\
& = \left[ d(m) - a \otimes \gamma(m) \right] + \sup_{0 \leq k \leq m} \{ \gamma(m - k + 1) - [a(n) - a(k)] \} \\
& = \left[ d(m) - a \otimes \gamma(m) \right] + \sup_{0 \leq k \leq m} \left\{ \gamma(m - k + 1) - \lambda(n - k) \right. \\
& \quad \left. + \lambda(n - k) - [a(n) - a(k)] \right\} \\
& \leq \left[ d(m) - a \otimes \gamma(m) \right] + \sup_{0 \leq k \leq m} \left\{ \lambda(n - k) - [a(n) - a(k)] \right\} \\
& \quad + \sup_{0 \leq k \leq m} \{ \gamma(m - k + 1) - \lambda(n - k) \}.
\end{align*}
\]
Let \( v = m - k + 1 \). Then the above inequality is written as

\[
d(m) - d(n) \leq [d(m) - a \bar{\otimes} \gamma(m)] + \sup_{0 \leq k \leq n} \{ \lambda(n - k) - [a(n) - a(k)] \}
\]

\[
- \inf_{1 \leq v \leq m+1} \{ \lambda(n - m - 1 + v) - \gamma(v) \}
\]

\[
\leq [d(m) - a \bar{\otimes} \gamma(m)] + \sup_{0 \leq k \leq n} \{ \lambda(n - k) - [a(n) - a(k)] \}
\]

\[
- \inf_{0 \leq v \leq m+1} \{ \lambda(n - m - 1 + v) - \gamma(v) \}
\]

where the last step is because

\[
\inf_{0 \leq k \leq m+1} [f_k] \leq \inf_{1 \leq k \leq m+1} [f_k].
\]

Adding \( \inf_{0 \leq v \leq m+1} \{ \lambda(n - m - 1 + v) - \gamma(v) \} \) to both sides of the above inequality results in

\[
\inf_{0 \leq v \leq m+1} \{ \lambda(n - m - 1 + v) - \gamma(v) \} - [d(n) - d(m)]
\]

\[
\leq [d(m) - a \bar{\otimes} \gamma(m)] + \sup_{0 \leq k \leq n} \{ \lambda(n - k) - [a(n) - a(k)] \}.
\]

In addition, there holds

\[
\lambda \bar{\otimes} \gamma(n - m - 1) = \inf_{v \geq 0} \{ \lambda(n - m - 1 + v) - \gamma(v) \}
\]

\[
\leq \inf_{0 \leq v \leq m+1} \{ \lambda(n - m - 1 + v) - \gamma(v) \}.
\]

To ensure that the right-hand side of the above inequality is meaningful, it requires \( n - m - 1 > 0 \). With the same conditions as analyzing delay, we conclude

\[
P\{ \lambda \bar{\otimes} \gamma(n - m - 1) - [d(n) - d(m)] > x \}
\]

\[
\leq P\{ [d(m) - a \bar{\otimes} \gamma(m)] + \sup_{0 \leq k \leq n} \{ \lambda(n - k) - [a(n) - a(k)] \} > x \}
\]

\[
\leq j \otimes h(x).
\]
Proof of Theorem 9

Proof. We shall only prove the three-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 $d^1(n) = a^2(n)$ and $d^2(n) = a^3(n)$. We then have,

$$d(n) - a \otimes \gamma^1 \otimes \gamma^2 \otimes \gamma^3(n)$$

$$= d(n) - \sup_{0 \leq m \leq n} \left\{ a \otimes \gamma^1(m) + \gamma^2 \otimes \gamma^3(n - m + 1) \right\} + d^1(m) - d^1(m)$$

$$\leq d(n) - \sup_{0 \leq m \leq n} \left\{ \gamma^2 \otimes \gamma^3(n - m + 1) + d^1(m) - \eta \cdot (n - m + 1) \right\}$$

$$- [d^1(m) - a \otimes \gamma^1(m) - \eta \cdot (n - m)]$$

$$\leq d(n) - \sup_{0 \leq m \leq n} \left\{ \gamma^2 \otimes \gamma^3(n - m + 1) + d^1(m) - \eta \cdot (n - m + 1) \right\}$$

$$+ \sup_{0 \leq m \leq n} \left\{ d^1(m) - a \otimes \gamma^1(m) - \eta \cdot (n - m) \right\}$$

$$= d(n) - \sup_{0 \leq m \leq n} \left\{ a^2(m) + \sup_{0 \leq k \leq n-m+1} [\gamma^2(k) + \eta \cdot k + \gamma^3(n - m + 1 - k) + 2 \eta \cdot (n - m + 1 - k)] - \eta \cdot (n - m + 1) \right\} +$$

$$+ \sup_{0 \leq m \leq n} \left\{ d^1(m) - a \otimes \gamma^1(m) - \eta \cdot (n - m) \right\}$$

$$= d(n) - a^2 \otimes \gamma^2 \otimes \gamma^3(n) + \sup_{0 \leq m \leq n} \left\{ d^1(m) - a \otimes \gamma^1(m) - \eta \cdot (n - m) \right\}$$

$$\leq d(n) - \sup_{0 \leq m \leq n} \left\{ a^2 \otimes \gamma^2(m) + \gamma^3(n - m + 1) \right\} - a^3(m) + \eta \cdot (n - m + 1)$$

$$+ d^2(m) - \eta \cdot (n - m) + \sup_{0 \leq m \leq n} \left\{ d^1(m) - a \otimes \gamma^1(m) - \eta \cdot (n - m) \right\}$$

$$\leq d(n) - \sup_{0 \leq m \leq n} \left\{ a^3(m) + \gamma^3(n - m + 1) - \eta \cdot (n - m + 1) \right\}$$

$$+ \sup_{0 \leq m \leq n} \left\{ d^2(m) - a^2 \otimes \gamma^2(m) - \eta \cdot (n - m) \right\}$$

$$+ \sup_{0 \leq m \leq n} \left\{ d^1(m) - a \otimes \gamma^1(m) - \eta \cdot (n - m) \right\}$$
Based on the relationship between the i.d SSC and the \( \eta \)-stochastic service curve presented in Theorem 4(2), the following inequality holds

\[
P\{d(n) - a \bar{\otimes} \gamma^1 \otimes \gamma^2(n) > x\} \leq j^3 \otimes j^2 \otimes j^1,\eta,
\]

which completes the proof.

Note that both the max-plus convolution and the min-plus convolution are associative and commutative.

**Proof of Lemma 4**

Proof. We shall only prove two-node case, from which, the proof can be extended to the \( N \)-node case. Keep in mind that \( a^2(n) = d^1(n) \). For the two-node case, we have

\[
d(n) - a \bar{\otimes} \gamma^1 \otimes \gamma^2(n) \\
= d(n) - \sup_{0 \leq m \leq n} \{a \bar{\otimes} \gamma^1(m) + \gamma^2(n - m + 1) + \eta \cdot (n - m + 1)\} \\
\leq d(n) - \sup_{0 \leq m \leq n} \{a \bar{\otimes} \gamma^1(m) + \gamma^2(n - m + 1) + \eta \cdot (n - m)\} \\
+ \sup_{0 \leq m \leq n} \{d^1(m) - a \bar{\otimes} \gamma^1(m) - \eta \cdot (n - m)\} \\
= d(n) - a^2 \bar{\otimes} \gamma^2(n) + \sup_{0 \leq m \leq n} \{d^1(m) - a \bar{\otimes} \gamma^1(m) - \eta \cdot (n - m)\} \\
\leq \sup_{0 \leq m \leq n} \{d(m) - a^2 \bar{\otimes} \gamma^2(m) - \eta \cdot (n - m)\} \\
+ \sup_{0 \leq m \leq n} \{d^1(m) - a \bar{\otimes} \gamma^1(m) - \eta \cdot (n - m)\}.
\]

The last step holds because of Theorem 4(1). From the condition, we conclude

\[
P\{d(n) - a \bar{\otimes} \gamma^1 \otimes \gamma^2(n) > x\} \leq j^1 \otimes j^2(x).
\]
Proof of Theorem 10

Proof. We use the induction way to prove this theorem.

Step (1) We start from \( n = 1 \) with the given condition \( a(0) = \min[a_1(0), a_2(0)] \).

If \( a(0) = a_1(0) \), then \( a(1) = \min\{a_1(1), a_2(0)\} \); if \( a(0) = a_2(0) \), then
\[
a(1) = \min\{a_1(0), a_2(1)\}.
\]

We expand Eq. (31) into the following expression:
\[
a(1) = \min\{a_1(1), a_2(1), \max[a_1(0), a_2(0)]\}
\]

Thus Eq. (31) holds for \( n = 1 \).

Step (2) Assume \( n = k \) holds for \( k > 1 \):
\[
a(k) = \min_{0 \leq m \leq k+1} \left\{ \max[a_1(m-1), a_2(k-m)] \right\} \quad \text{(induction hypothesis)},
\]
which has four solutions as below:
\[
a(k) = \begin{cases} 
a_1(k), & \text{if } a_1(k) < a_2(0); 
a_2(k), & \text{if } a_2(k) < a_1(0); 
a_1(m^* - 1), & \text{if } a_2(k - m^*) < a_1(m^* - 1) \text{ for } 0 < m^* < k + 1; 
a_2(k - m^*), & \text{if } a_1(m^* - 1) < a_2(k - m^*) \text{ for } 0 < m^* < k + 1.
\end{cases}
\]

Step (3) Prove \( n = k + 1 \) holds:
\[
a(k + 1) = \min_{0 \leq m \leq (k+1)+1} \left\{ \max[a_1(m-1), a_2(k+1-m)] \right\}
\]
which can be expanded into
\[
a(k + 1) = \min\left\{a_1(k + 1), a_2(k + 1), \max[a_1(0), a_2(k)], \max[a_1(1), a_2(k - 1)], 
\max[a_1(2), a_2(k - 2)], ..., \max[a_1(k), a_2(0)]\right\}. \quad (A.2)
\]

We prove Eq. (A.2) based on the four solutions of the induction hypothesis, respectively.
(I) If \(a(k) = a_1(k)\) which implies \(a_1(k) < a_2(0)\), then \(a(k + 1)\) is either \(a_1(k + 1)\) or \(a_2(0)\) depending on the minimum one, i.e., \(a(k + 1) = \min \{a_1(k + 1), a_2(0)\}\). Since the condition \(a_1(k) < a_2(0)\) implies

\[
a_1(0) < a_1(1) < ... < a_1(k) < a_2(0) < a_2(1) ... < a_2(k),
\]

the induction hypothesis given in Step (2) is expanded into

\[
a(k) = \min \{a_1(k), a_2(k), a_2(k - 1), ..., a_2(0)\} = a_1(k).
\]

From this, Eq. (A.2) becomes

\[
a(k + 1) = \min \{a_1(k + 1), a_2(k + 1), a_2(k), a_2(k - 1), ..., a_2(0)\} = \min \{a_1(k + 1), a_2(0)\},
\]

which proves that Eq. (A.2) holds.

(II) If \(a(k) = a_2(k)\) which implies \(a_2(k) < a_1(0)\), then \(a(k + 1)\) is either \(a_2(k + 1)\) or \(a_1(0)\) depending on the minimum one, i.e., \(a(k + 1) = \min \{a_2(k + 1), a_1(0)\}\). Since the condition \(a_2(k) < a_1(0)\) implies

\[
a_2(0) < a_2(1) < ... < a_2(k) < a_1(0) < a_1(1) ... < a_1(k),
\]

the induction hypothesis given in Step (2) is expanded into

\[
a(k) = \min \{a_2(k), a_1(k), a_1(k - 1), ..., a_1(0)\} = a_2(k).
\]

From this, Eq. (A.2) becomes

\[
a(k + 1) = \min \{a_2(k + 1), a_1(k + 1), a_1(k), a_1(k - 1), ..., a_1(0)\} = \min \{a_2(k + 1), a_1(0)\},
\]

which proves that Eq. (A.2) holds.

(III) Without loss of generality, if \(a(k) = a_1(m^* - 1)\) for \(0 < m^* < k + 1\), which implies \(a_2(k - m^*) < a_1(m^* - 1)\), then \(a(k + 1)\) is either \(a_1(m^*)\) or \(a_2(k - m^* + 1)\) depending on the minimum one, i.e., \(a(k + 1) = \min \{a_1(m^*), a_2(k - m^*)\}\). Since the condition \(a_2(k - m^*) < a_1(m^* - 1)\) implies \(a_2(k - m^*) < a_1(m^* - 1) < a_1(m^*) \leq a_1(k)\) and \(a_2(k - m^*) < a_1(m^* - 1) < a_2(k + 1 - m^*) \leq a_2(k)\) for \(0 < m^* < k + 1\), the induction hypothesis given in Step (2) is expanded into

\[
a(k) = \min \{a_1(k), a_2(k), a_2(k - 1), ..., a_2(k + 1 - m^*), a_1(m^* - 1), ..., a_1(k - 1)\} = a_1(m^* - 1)
\]

44
From this, Eq.(A.2) becomes

\[ a(k + 1) = \min \{a_1(k + 1), a_2(k + 1), a_2(k), a_2(k - 1), ..., a_2(k + 1 - m^*), a_1(m^*), a_1(m^* + 1), ..., a_1(k)\} \]

\[ = \min \{a_2(k + 1 - m^*), a_1(m^*)\} \]

which proves that Eq.(A.2) holds.

(IV) Without loss of generality, if \(a(k) = a_2(k - m^*)\) for \(0 < m^* < k + 1\), which implies \(a_1(m^* - 1) < a_2(k - m^*)\), then \(a(k + 1)\) is either \(a_1(m^*)\) or \(a_2(k - m^* + 1)\) depending on the minimum one, i.e., \(a(k + 1) = \min \{a_1(m^*), a_2(k + 1 - m^*)\}\). Since the condition \(a_1(m^* - 1) < a_2(k - m^*)\) implies \(a_1(m^* - 1) < a_2(k - m^*) < a_1(m^*) \leq a_1(k)\) and \(a_1(m^* - 1) < a_2(k - m^*) < a_2(k + 1 - m^*) \leq a_2(k)\) for \(0 < m^* < k + 1\), the induction hypothesis given in Step (2) is expanded into

\[ a(k) = \min \{a_1(k), a_2(k), a_2(k - 1), ..., a_2(k - m^*), a_1(m^*), ..., a_1(k - 1)\} = a_2(k - m^*). \]

From this, Eq.(A.2) becomes

\[ a(k + 1) = \min \{a_1(k + 1), a_2(k + 1), a_2(k), a_2(k - 1), ..., a_2(k + 1 - m^*), a_1(m^*), a_1(m^* + 1), ..., a_1(k)\} \]

\[ = \min \{a_2(k + 1 - m^*), a_1(m^*)\} \]

which proves that Eq.(A.2) holds.

Combining the above three steps concludes that Eq.(31) holds for all \(n \geq 0\). \(\square\)