A Temporal Network Calculus Approach to Service Guarantee Analysis of Stochastic Networks

Jing Xie
Department of Telematics
Norwegian University of Science and Technology

Yuming Jiang
Centre for Quantifiable Quality of Service in Communication Systems
Department of Telematics
Norwegian University of Science and Technology

ABSTRACT

Many computer networks such as wireless networks are stochastic in nature. In order to perform performance guarantee analysis of such networks, a theory, called *stochastic network calculus*, has evolved. In the stochastic network calculus literature, most results are based on *space-domain* traffic and service models where the arrival process and the service process are respectively characterized by the cumulative amount of arrival and the cumulative amount of service. Recently, a novel approach called *time-domain* approach to stochastic network calculus (SNC) has been proposed, where the traffic and service models are defined based on the cumulative inter-arrival times and the cumulative service times respectively.

In this paper, we concretize the time-domain SNC traffic and service models by linking some well-known stochastic processes to them. In addition, we exemplify the temporal analysis approach by investigating the delay performance of a Gilbert-Elliott channel. The results show that the delay bound can be improved under the independence condition. Furthermore, a comparison between the temporal and the spatial analysis results reveals that the two analytical approaches essentially yield close results.

1. INTRODUCTION

Nowadays computer networks are so widespread in our everyday life and provide a fast and convenient way to access various network services. One inevitable trend of network development is to deliver information with various traffic characteristics and diverse Quality of Service (QoS) requirements. In response to the continually growing demand for more bandwidth, it is necessary to improve existing technology and develop new ones. Performance analysis is needed to optimize the performance of existing technologies and evaluate the efficiency of new ones.

Performance models capture the behavior characteristics of networks. The behavior of a computer network is often subject to many irregularities and stochastic fluctuations.

The reason behind this phenomena is manifold. First of all, the diverse network applications incorporate some complicated dynamics which generate the varying traffic patterns, accordingly. Secondly, in many network systems, the services provided by the networks are non-deterministic. Such networks are stochastic in nature and are called *stochastic networks* in this paper. A representative type of such networks is wireless networks, where the error-prone nature of wireless channels causes data transmission inherently stochastic and influences the link capacity over time. In addition, aggregate multiplexing has been employed extensively in order to improve resource utilization. From the perspective of individual flows, the service received dynamically changes over time because new flows join or existing flows leave.

Stochastic network calculus is an analytical theory originated for stochastic performance guarantee analysis. The fundamental elements of this theory are *stochastic arrival curve* and *stochastic service curve*. The former is a traffic model characterizing the behavior of the traffic arrival process. The latter is a service model describing the service provided by a network element. The significant progress on modeling and analyzing network behavior from the spatial perspective is summarized in [3][7][9]. The study from a temporal perspective is, however, very limited.

Recently, a novel approach called *time-domain* approach to stochastic network calculus (SNC) has been proposed [12]. In this temporal approach, the traffic and service models are defined based on the cumulative interarrival times and the cumulative service times respectively, based on which delay and backlog bounds are derived [12]. However, a clear guidance of finding the *time-domain* arrival curve or service curve characterization of an arrival process or a service process is missing. An objective of this paper is to provide this missing link.

Another objective of the paper is to compare the temporal SNC approach and the spatial SNC approach. This is exemplified through delay analysis of a Gilbert-Elliott channel, for which, performance bounds derived under the temporal SNC approach are available in the literature. Particularly, the analysis is based on the *time-domain* traffic and service models defined in the temporal SNC approach, and delay bounds are obtained. For the analysis, a new service model called *time-domain stochastic strict service curve* is defined to help find the stochastic service curve of a network element. In addition, martingale inequalities are adopted when applicable. Furthermore, the analytical results from the temporal SNC approach are compared with those from the temporal

SNC approach. The comparison reveals that the two SNC approaches essentially give close results.

The rest of the paper is organized as follows. Sec. 2 introduces preliminary knowledge and relevant mathematical background. In Sec. 3, we investigate the stochastic arrive curve characterization of the arrival process from the temporal perspective. In Sec. 4, the service process is characterized through the time-domain stochastic strict service curve model and stochastic service curve model. The detailed procedure of analyzing the Gilbert-Elliott channel service characterization is given in Sec. 5. Sec. 6 demonstrates how to find the system delay bound and discusses the numerical results, based on the time-domain SNC approach. In addition, the delay bounds derived from the temporal analysis are compared to those derived from the spatial analysis. Conclusions are drawn in the last section.

2. PRELIMINARY AND BACKGROUND

This section first describes the system model and defines notations used throughout this paper, and then gives relevant mathematical background that is indispensable in the following analysis.

2.1 System Model and Notations

.

Consider packets arriving to a system. We use $P(n)$, $a(n)$, $d(n)$ and δ_n, to denote the nth packet entering the system, its arrival time to the system, departure time from the system and the service time provided by the system, respectively, where $n = 1, 2, ...$ The inter-arrival time and inter-departure time between $P(n-1)$ and $P(n)$ are denoted by τ_n and τ_n^*, respectively.

From the temporal perspective, an arrival process counts the cumulative inter-arrival time between two arbitrary packets and is denoted by $\Gamma(m, n) = \sum_{k=m+1}^{n} \tau_k$. A service process describes the cumulative service time received between two arbitrary packets and is denoted by $\Delta(m, n) = \sum_{k=m}^{n} \delta_k$. A departure process represents the cumulative inter-departure time between two arbitrary packets and is denoted by $\Gamma^*(m, n) = \sum_{k=m+1}^{n} \tau_k^*$. All processes are defined on $n \geq 1, m \geq 1$.

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

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

where for any function $f \in \mathcal{G}$, we set $f(x) = 0$ for $x < 0$.

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

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

where for any function $g \in \bar{\mathcal{G}}$, we set $g(x) = 1$ for $x < 0$.

2.2 Preliminaries

Min-plus algebra has been used to characterize and analyze network behavior from the spatial perspective [1][8][7]. Interestingly, modeling network behavior from the temporal perspective can be easily represented using max-plus algebra [8][12]. We review the basics of both min-plus and max-plus algebras used in this paper.

The min-plus convolution of single variate functions $F, G \in$

$\mathcal{F}$ denoted by $\otimes$ is defined as below:

$$F \otimes G(t) = \inf_{0 \leq s \leq t} \big\{ F(s) + G(t-s) \big\}$$

The max-plus convolution, denoted by $\bar{\otimes}$, of single variate functions $A, B \in \mathcal{F}$, is defined as:

$$A \bar{\otimes} B(n) = \sup_{0 \leq m \leq n} \big\{ A(m) + B(n-m) \big\}.$$

Note that the single variate min-plus and max-plus convolutions are commutative.

The moment generating function (MGF) $M_X(\eta)$ of a random variable X is defined for all values η by

$$M_X(\eta) = \mathbf{E}\big[e^{\eta X} \big],$$

where $\mathbf{E}$ is the expectation of its argument.

A martingale is a stochastic process, where the conditional expectation of an observation at time t, given all the observations up to some earlier time s, equals the observation at time s. Let A_n be a discrete-time process with finite mean, then it is a discrete-time *martingale* iff there holds for all n,

$$\mathbf{E}\big(A_{n+1} | A_1, A_2, ..., A_n \big) = A_n.$$

A_n is said to be a *submartingale* iff for all n,

$$\mathbf{E}\big(A_{n+1} | A_1, A_2, ..., A_n \big) \geq A_n.$$

A_n is said to be a *supermartingale* iff for all n,

$$\mathbf{E}\big(A_{n+1} | A_1, A_2, ..., A_n \big) \leq A_n.$$

Every martingale is a supermartingale and a submartingale.

Lemma 1 presents an inequality for supermartingale which is derived from the Doob's submartingale inequality [6].

LEMMA 1. *If $\big\{ U_k, 1 \leq k \leq n \big\}$ is a supermartingale and all U_k, $k = 1, ..., n$, are non-negative, then for any real value $x > 0$, there holds:*

$$P\Big\{ \sup_{1 \leq m \leq n} U_k \geq x \Big\} \leq \frac{\boldsymbol{E}[U_1]}{x}. \qquad (1)$$

3. ARRIVAL PROCESS CHARACTERIZATION

In this section, we first recall the definitions of time-domain stochastic arrival curve models and then obtain the stochastic arrival curve characterization of the arrival process.

3.1 Moment Generating Function of Inter-arrival Time

When observing the traffic arrival process $\Gamma(m, n)$ from the temporal perspective, it is indeed formed from process $\{\tau_{m+1}\}$, where $\tau_{m+1} \equiv a(m+1) - a(m)$. In order to guarantee a certain level of QoS to this arrival process, the packet inter-arrival times should be constrained. For the arrival process $\Gamma(m, n)$ formed by identically distributed $\{\tau_{m+1}\}$, suppose there exists $\varphi(\eta, n - m)$ satisfying

$$\mathbf{E}[e^{\eta \Gamma(m,n)}] \geq e^{\eta(n-m)\varphi(\eta, n-m)}$$

which becomes the following expression

$$\frac{1}{\eta(n-m)} log \mathbf{E}\big[e^{\eta \Gamma(m,n)} \big] \geq \varphi(\eta, n-m).$$

If $\{\tau_{m+1}\}$ are independent and identically distributed (*i.i.d.*), then it is easily verified that $\frac{1}{\eta(n-m)}log\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big] = \frac{1}{\eta}log\mathbf{E}\big[e^{\eta\tau_1}\big]$, which is independent of $n-m$, and we hence adopt:

$$\frac{1}{\eta}log\mathbf{E}\big[e^{\eta\tau_1}\big] \geq \varphi(\eta).$$

3.2 i.a.t Stochastic Arrival Curve Characterization

The *i.a.t* stochastic arrival curve is defined as below [12]:

DEFINITION 1. *(i.a.t Stochastic Arrival Curve). A flow is said to have an inter-arrival-time (i.a.t) stochastic arrival curve $\lambda \in \mathcal{G}$ with bounding function $h \in \bar{\mathcal{G}}$, if for all $m, n \geq 1$ and $x \geq 0$, there holds*

$$P\Big\{\Gamma(m,n) < \big[\lambda(n-m) - x\big]^+\Big\} \leq h(x). \qquad (2)$$

Remark. We adopt $\lambda(0) = 0$ by convention.

Assume $\{\tau_{m+1}\}$ are identically distributed. Let $\varphi(\eta, n-m) \leq \frac{1}{\eta(n-m)}log\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big]$. Then (2) is rewritten as

$$
\begin{aligned}
& P\big\{\varphi(\eta, n-m) \cdot (n-m) - \Gamma(m,n) > x\big\} \\
=\ & P\big\{e^{\eta[\varphi(\eta,n-m)\cdot(n-m)-\Gamma(m,n)]} > e^{\eta x}\big\} \\
\leq\ & e^{-\eta x}\mathbf{E}\big[e^{\eta[\varphi(\eta,n-m)\cdot(n-m)-\Gamma(m,n)]}\big] \qquad (3) \\
=\ & e^{-\eta x}\frac{e^{\eta\varphi(\eta,n-m)\cdot(n-m)}}{\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big]} \\
\leq\ & e^{-\eta x} \qquad (4)
\end{aligned}
$$

for $\eta > 0$. Here (3) is known as the Chernoff bound, and (4) is obtained due to $\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big] \geq e^{\eta(n-m)\varphi(\eta,n-m)}$ by definition. The following lemma summarizes the above result.

LEMMA 2. *For an arrival process $\Gamma(m,n)$, if there exists $\varphi(\eta, n-m)$ which satisfies, for $m, n \geq 1$,*

$$\frac{1}{\eta(n-m)}log\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big] \geq \varphi(\eta, n-m),$$

then this process has an i.a.t. stochastic arrival curve $\lambda(n) = \varphi(\eta, n) \cdot n$ with bounding function $h(x) = e^{-\eta x}$ for $\eta > 0$.

If $\{\tau_{m+1}\}$ are *i.i.d.*, (2) is rewritten as follows:

$$
\begin{aligned}
& P\Big\{\varphi(\eta) \cdot (n-m) - \Gamma(m,n) > x\Big\} \\
\leq\ & e^{-\eta x}\mathbf{E}\big[e^{\eta[\varphi(\eta)-\tau_1]}\big],
\end{aligned}
$$

from which, we have the following lemma.

LEMMA 3. *If the inter-arrival times of arrival process $\Gamma(m,n)$ are i.i.d., then the arrival process has an i.a.t. stochastic arrival curve $\lambda(n) = \varphi(\eta) \cdot n$ with bounding function $h(x) = e^{-\eta x}\mathbf{E}\big[e^{\eta[\varphi(\eta)-\tau_1]}\big]$, where $\varphi(\eta) \leq \frac{1}{\eta}log\mathbf{E}\big[e^{\eta\tau_1}\big]$ for $\eta > 0$.*

Remark: Lemma 2 becomes Lemma 3 by taking into consideration the independence condition of inter-arrival times.

3.3 v.w.d Stochastic Arrival Curve Characterization

The *v.w.d* stochastic arrival curve is defined as below [12]:

DEFINITION 2. *(v.w.d Stochastic Arrival Curve). A flow is said to have a virtual-waiting-delay (v.w.d) stochastic*

arrival curve $\lambda \in \mathcal{G}$ with bounding function $h \in \bar{\mathcal{G}}$, if for any $m, n \geq 1$ and $x \geq 0$, there holds

$$P\Big\{\sup_{1 \leq m < n}\big\{\lambda(n-m) - \Gamma(m,n)\big\} > x\Big\} \leq h(x). \qquad (5)$$

The left-hand side of (5) represents an instantaneous property which is generally hard to calculate [10]. To address this difficulty, additional constraint on the bounding function is needed. Assume $\{\tau_n\}$ are identically distributed. Without loss of generality, assume when m takes m_0, the following holds

$$
\begin{aligned}
& \sup_{0 \leq m < n}\big\{\varphi(\eta, n-m) \cdot (n-m) - \Gamma(m,n)\big\} \\
=\ & \varphi(\eta, n-m_0) \cdot (n-m_0) - \Gamma(m_0, n).
\end{aligned}
$$

Then from (5), we can write, for any $x \geq 0$,

$$
\begin{aligned}
& P\Big\{\sup_{0 \leq m \leq n}\big\{\varphi(\eta, n-m) \cdot (n-m) - \Gamma(m,n)\big\} > x\Big\} \\
=\ & P\{\varphi(\eta, n-m_0) \cdot k - \Gamma(m_0, n) > x\} \leq e^{-\eta x},
\end{aligned}
$$

where, the last step is obtained from (4).

LEMMA 4. *For an arrival process $\Gamma(m,n)$, if there exists $\varphi(\eta, n-m)$ which satisfies, for any $1 \leq m < n$,*

$$\frac{1}{\eta(n-m)}log\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big] \geq \varphi(\eta, n-m)$$

then this process has a v.w.d stochastic arrival curve $\lambda(n) = \varphi(\eta, n) \cdot n$ with bounding function $h(x) = e^{-\eta x}$ for $\eta > 0$.

If the arrival process is formed by the *i.i.d.* inter-arrival times, it has a stochastic arrival curve given by Lemma 5.

LEMMA 5. *If the inter-arrival times of arrival process $\Gamma(m,n)$ are i.i.d., then the arrival process has a v.w.d stochastic arrival curve $\lambda(n) = \varphi(\eta) \cdot n$ with bounding function $h(x) = e^{\eta\varphi(\eta)}\mathbf{E}[e^{-\eta\tau_1}]e^{-\eta x}$ for $\eta > 0$, where $\varphi(\eta) \leq \frac{1}{\eta}log\mathbf{E}\big[e^{\eta\tau_1}\big]$.*

PROOF. Consider a sequence of non-negative random variables $\{V_m\}$, $m = 1, 2, ..., n-1$, formed by

$$V_m = e^{\eta\varphi(\eta)\cdot m - \eta\Gamma(n-m,n)} = e^{\eta\varphi(\eta)\cdot m - \eta\sum_{k=n-m+1}^{n}\tau_k}.$$

Since $\{\tau_k\}$ are *i.i.d.*, we then have

$$
\begin{aligned}
V_{m+1} &= e^{\eta\varphi(\eta)\cdot(m+1) - \eta\Gamma(n-m-1,n)} \\
&= e^{\eta\varphi(\eta)\cdot(m+1) - \eta\sum_{k=n-m}^{n}\tau_k} \\
&= e^{\eta\varphi(\eta)\cdot m - \eta\sum_{k=n-m+1}^{n}\tau_k} \cdot e^{\eta\varphi(\eta) - \eta\tau_{n-m}} \\
&= V_m \cdot e^{\eta\varphi(\eta) - \eta\tau_{n-m}}.
\end{aligned}
$$

In addition, there holds:

$$
\begin{aligned}
& \mathbf{E}\big[V_{m+1}|V_1, ..., V_m\big] = \mathbf{E}\big[V_{m+1}|\tau_n, \tau_{n-1}, ..., \tau_{n-m+1}\big] \\
=\ & \mathbf{E}\big[V_m \cdot e^{\eta\varphi(\eta)-\eta\tau_{n-m}}|\tau_n, ..., \tau_{n-m+1}\big] \\
=\ & \mathbf{E}\big[V_m|\tau_n, ..., \tau_{n-m+1}\big] \cdot \mathbf{E}\big[e^{\eta\varphi(\eta)-\eta\tau_{n-m}}\big] \qquad (6) \\
=\ & V_m \cdot \frac{e^{\eta\varphi(\eta)}}{\mathbf{E}\big[e^{\eta\tau_1}\big]} \qquad (7) \\
\leq\ & V_m \qquad (8)
\end{aligned}
$$

where, step (6) is due to that τ_{n-m} is independent of $\{\tau_n, \tau_{n-1}, ..., \tau_{n-m+1}\}$, step (7) is because $\{\tau_1, \tau_2, ...\}$ are identically distributed and

$$\mathbf{E}\big[V_m(\tau_n, \tau_{n-1}, ..., \tau_{n-m+1})|\tau_n, \tau_{n-1}, ..., \tau_{n-m+1}\big] = V_m,$$

and step (8) holds since $\mathbf{E}\big[e^{\eta\tau_1}\big] = e^{\eta\varphi(\eta)}$ by definition.

Hence $V_1, V_2, ..., V_n$ form a non-negative supermartingale. From Lemma 1, there holds

$$P\Big\{ \sup_{1 \leq m < n} \big\{ \varphi(\eta) \cdot (n-m) - \Gamma(m,n) \big\} > x \Big\}$$
$$= P\Big\{ \sup_{1 \leq m < n} \big\{ e^{\varphi(\eta)\cdot(n-m)-\Gamma(m,n)} \big\} > e^x \Big\}$$
$$= P\Big\{ \sup_{1 \leq m < n} V_m > e^x \Big\}$$
$$\leq \mathbf{E}[V_1]e^{-\eta x} = e^{\eta\varphi(\eta)}\mathbf{E}[e^{-\eta\tau_1}]e^{-\eta x}$$

which ends the proof. $\square$

Remark. Lemma 4 becomes Lemma 5 by taking into account the independence condition. The bounding function in Lemma 5 contains a scaling factor $e^{\eta\varphi(\eta)}\mathbf{E}[e^{-\eta\tau_1}]$ with regard to $\varphi(\eta)$. This scaling factor will yield tighter arrival curves if it is smaller than 1.

EXAMPLE 1. *Exponential inter-arrival time distribution.*

Consider an arrival process of packets generated at times $\{a(n)\}$. Suppose the inter-arrival times $\{\tau_n\}$ are *i.i.d.* exponentially distributed random variables with mean $\frac{1}{\mu}$. Then the arrival process $\Gamma(m,n)$ follows gamma distribution with parameters $n-m$ and μ. We thus have

$$\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big] = \Big(\frac{\mu}{\mu-\eta}\Big)^{n-m}$$
$$\Rightarrow \quad \frac{1}{\eta(n-m)}log\mathbf{E}\big[e^{\eta\Gamma(m,n)}\big] = \frac{1}{\eta}log\frac{\mu}{\mu-\eta}$$

Let $\varphi(\eta) = \frac{1}{\eta}log\frac{\mu}{\mu-\eta}$. By applying Lemma 3 and Lemma 5, the *i.a.t* stochastic arrival curve and the *v.w.d* stochastic arrival curve of the arrival process can be obtained. Specifically, this arrival process has a *v.w.d* stochastic arrival curve $\lambda(n)$ with bounding function $h(x)$, where

$$\lambda(n) = \varphi(\eta) \cdot n = \frac{n}{\eta}log\frac{\mu}{\mu-\eta}$$
$$h(x) = e^{\eta\varphi(\eta)}\mathbf{E}[e^{-\eta\tau_1}]e^{-\eta x} = e^{-\eta x}.$$

4. SERVICE PROCESS CHARACTERIZATION

This section first reviews the *i.d* stochastic service curve model, from which we define a stochastic strict service curve model to facilitate obtaining the *i.d* stochastic service curve. Moreover, a time-domain *error process* is introduced to decouple the characterization of the cumulative impaired service time from the real service process.

4.1 Stochastic Strict Service Curve

Consider a stochastic system. The following definition [12] describes the service process of this system by comparing the packet actual departure time $d(n)$ with a virtual departure time $a\bar{\otimes}\gamma(n)$.

DEFINITION 3. *(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 $j \in \bar{\mathcal{G}}$, if for any $n \geq 1, x \geq 0$, there holds*

$$P\Big\{ d(n) - a\bar{\otimes}\gamma(n) > x \Big\} \leq j(x). \tag{9}$$

Remark. We adopt $\gamma(0) = 0$ by convention.

Although (9) explores the relationship between the arrival process and the departure process, it does not explicitly characterize the service process. To solve this problem, we expand (9) as follows [12]:

$$d(n) - a\bar{\otimes}\gamma(n) = \sup_{1 \leq m \leq n} \big[a(m) + \Delta(m,n)\big] - a\bar{\otimes}\gamma(n). \tag{10}$$

Without loss of generality, assume $a(m_0)$ $(1 \leq m_0 \leq n)$ is the beginning of the backlogged period in which packet $P(n)$ is served. Then, $\sup_{1 \leq m \leq n} \big[a(m) + \Delta(m,n)\big] = a(m_0) + \Delta(m_0,n)$ and $a\bar{\otimes}\gamma(n) \geq a(m_0) + \gamma(n - m_0 + 1)$. We rewrite the right-hand side of (10) as

$$a(m_0) + \Delta(m_0,n) - a\bar{\otimes}\gamma(n)$$
$$\leq a(m_0) + \Delta(m_0,n) - a(m_0) - \gamma(n - m_0 + 1)$$
$$= \Delta(m_0,n) - \gamma(n - m_0 + 1). \tag{11}$$

Note that (11) holds for arbitrary $m_0 \leq n$. Inspired by this, we define a new service curve model as below.

DEFINITION 4. **Stochastic Strict Service Curve.** *A system is said to provide stochastic strict service curve $\gamma(n)$ with bounding function $j(x)$, if the cumulative service time between two arbitrary packets $P(m)$ and $P(n)$[1] satisfies*

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

for any $x \geq 0$.

Moreover, (11) reveals a relationship between the *i.d* stochastic service curve and the stochastic strict service curve.

LEMMA 6. *If a system provides stochastic strict service curve $\gamma(n)$ with bounding function $j(x)$, then it provides an i.d stochastic service curve $\gamma(n)$ with the same bounding function $j(x)$.*

With Lemma 6, we can further explore the stochastic service curve characterization. If the service times $\{\delta_n\}$ are identically distributed, we have

$$P\Big\{ d(n) - a\bar{\otimes}\gamma(n) > x \Big\}$$
$$\leq P\Big\{ e^{\eta[\Delta(m,n)-\gamma(n-m+1)]} > e^{\eta x} \Big\}$$
$$\leq e^{-\eta x}\mathbf{E}\big[e^{\eta[\Delta(m,n)-(\nu(\eta,n-m+1)+\eta_\gamma)\cdot(n-m+1)]}\big]$$
$$\leq e^{-\eta x}e^{-\eta\eta_\gamma(n-m+1)} \leq e^{-\eta x}e^{-\eta\eta_\gamma}$$

where $\gamma(n-m+1) = (\nu(\eta, n-m+1) + \eta_\gamma) \cdot (n-m+1)$ with $\nu(\eta, n-m+1)$ satisfying

$$\nu(\eta, n-m+1) \geq \frac{1}{\eta(n-m+1)}log\mathbf{E}\big[e^{\eta\Delta(m,n)}\big].$$

If the service times $\{\delta_n\}$ are *i.i.d.*, we have

$$P\Big\{ d(n) - a\bar{\otimes}\gamma(n) > x \Big\}$$
$$\leq e^{-\eta x}\mathbf{E}\big[e^{\eta[\Delta(n,n)-(\nu(\eta)+\eta_\gamma)]^{n-m+1}}\big] \leq e^{-\eta x}e^{-\eta\eta_\gamma}$$

where $\gamma(n-m+1) = (\nu(\eta) + \eta_\gamma) \cdot (n-m+1)$ with $\nu(\eta)$ satisfying

$$\nu(\eta) = \frac{1}{\eta}log\mathbf{E}\big[e^{\eta\delta_1}\big].$$

The following lemma summarizes the above discussion.

___

[1] If $P(m)$ and $P(n)$ are in the same backlogged period, $\Delta(m,n) = d(n) - d(m-1) = \Gamma^*(m-1,n)$.

411

LEMMA 7. *If the system provides stochastic strict service curve $\gamma(n)$ with bounding function $j(x)$ and*

- *the service times $\{\delta_n\}$ are identically distributed, then $\gamma(n - m + 1) = (\nu(\eta, n - m + 1) + \eta_\gamma) \cdot (n - m + 1)$ and $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$ for all $\eta > 0, \eta_\gamma \geq 0$, where*

$$\nu(\eta, n - m + 1) \geq \frac{1}{\eta(n - m + 1)} log\boldsymbol{E}\big[e^{\eta\Delta(m,n)}\big];$$

- *the service times $\{\delta_n\}$ are i.i.d., then $\gamma(n - m + 1) = (\nu(\eta) + \eta_\gamma) \cdot (n - m + 1)$ and $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$ for $\eta > 0, \eta_\gamma \geq 0$, where*

$$\nu(\eta) = \frac{1}{\eta}log\boldsymbol{E}\big[e^{\eta\delta_1}\big].$$

4.2 Error Server Due to Impairment

The error server model consists of an ideal service process and an error process [11]. The former describes the service process when there would be no error. The latter characterizes the errors in the service. We investigate a typical example of error server in this subsection.

Consider a wireless channel providing service with stochastic nature which is due to some random impairment process. The impairment degrades the network performance because packets may not be transmitted successfully when the channel condition is 'bad' or packets are queued in the buffer until the channel condition becomes 'good'. Either dropping/retransmitting the unsuccessfully delivered packets or holding the queued packets longer can be counted as errors in the service time. The cumulative error in service time can be explicitly described by an error process.

If the channel is always in 'good' condition, the service time of packet $P(n)$ equals the packet transmission time denoted by $\hat{\delta}_n$. However, the varying link condition may cause $P(n)$ to suffer additional delay denoted by ε_n. Then the actual service time equals the packet transmission time plus the additional delay, i.e., $\delta_n = \hat{\delta}_n + \varepsilon_n$.

Let $\hat{\Delta}(m,n) = \sum_{k=m}^{n} \hat{\delta}_k$ represent the ideal service process without errors, $\mathbb{E}(m,n) = \sum_{k=m}^{n} \varepsilon_k$ represent the error process, and $\Delta(m,n)$ represent the actual service process. The cumulative actual service time satisfies, for any $1 \leq m \leq n$,

$$\Delta(m,n) = \hat{\Delta}(m,n) + \mathbb{E}(m,n). \qquad (12)$$

If the ideal service process $\hat{\Delta}(m,n)$ has a (deterministic) strict service curve[2] $\hat{\gamma}(n)$, i.e., $\hat{\Delta}(m,n) \leq \hat{\gamma}(n - m + 1)$ for any $1 \leq m \leq n$, then from (12), there holds

$$\Delta(m,n) \leq \hat{\gamma}(n - m + 1) + \mathbb{E}(m,n).$$

Furthermore, if the error process $\mathbb{E}(m,n)$ has a stochastic strict service curve $\gamma_\mathbb{E}(n)$, we get a further result

$$\Delta(m,n) - \hat{\gamma}(n - m + 1) - \gamma_\mathbb{E}(n - m + 1)$$
$$\leq \mathbb{E}(m,n) - \gamma_\mathbb{E}(n - m + 1).$$

The following lemma illustrates that the above-mentioned error server provides a stochastic strict service curve.

[2]The deterministic strict service curve is a special case of the stochastic strict service curve with bounding function $j(x) = 0$.

LEMMA 8. *Consider an error server consists of an ideal service process $\hat{\Delta}(m,n)$ having a (deterministic) strict service curve $\hat{\gamma}(n)$ and an error process $\mathbb{E}(m,n)$ having a stochastic strict service curve $\gamma_\mathbb{E}(n)$ with bounding function $j_\mathbb{E}(x)$. Then, the error server provides a stochastic strict service curve $\gamma(n)$ with bounding function $j_\mathbb{E}(x)$, where*

$$\gamma(n) = \hat{\gamma}(n) + \gamma_\mathbb{E}(n).$$

Remark. With Lemma 7, if $\{\varepsilon_n\}$ are identically distributed, then $\gamma_\mathbb{E}(n) = (\nu_\mathbb{E}(\eta, n) + \eta_\gamma) \cdot n$ with bounding function $j_\mathbb{E}(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$ where

$$\nu_\mathbb{E}(\eta, n - m + 1) \geq \frac{1}{\eta(n - m + 1)} log\boldsymbol{E}\big[e^{\eta\mathbb{E}(m,n)}\big].$$

If $\{\varepsilon_n\}$ are i.i.d., then $\gamma_\mathbb{E}(n) = (\nu_\mathbb{E}(\eta) + \eta_\gamma) \cdot n$ with the same bounding function $j_\mathbb{E}(x)$, where $\nu_\mathbb{E}(\eta) = \frac{1}{\eta}log\boldsymbol{E}\big[e^{\eta\varepsilon_1}\big]$.

5. SERVICE CURVE EXAMPLE

This section gives an example to demonstrate how to obtain the stochastic service curve characterization of a Gilbert-Elliott channel. We first analyze the constant rate server which can be considered as the ideal service process of an error server. Then we investigate the Gilbert-Elliott channel in detail.

5.1 Constant Rate Server

Consider a server with constant service rate C. Let L_n denote the packet length of packet $P(n)$. If all packets have the fixed-length L, then the server provides a deterministic strict service curve $\gamma(n) = \frac{L}{C} \cdot n$. If the packet lengths are i.i.d. random variables with the MGF $M_L(\eta) = \boldsymbol{E}\big[e^{\eta L_1}\big]$, we have the following result.

LEMMA 9. *Consider a server with constant service rate C. If the arrival packets have i.i.d. packet lengths, then the server provides stochastic strict service curve $\gamma(n) = (\nu(\eta) + \eta_\gamma) \cdot n$ with bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$ for $\eta > 0, \eta_\gamma \geq 0$, where $\nu(\eta) = \frac{1}{\eta C}logM_L(\eta)$.*

EXAMPLE 2.

If the packet lengths follow the exponential distribution with parameter λ, then the server provides the stochastic strict service curve $(\nu(\eta) + \eta_\gamma) \cdot n$ with bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$, where

$$\nu(\eta) = \frac{1}{\eta C}log\frac{\lambda}{\lambda - \eta}.$$

EXAMPLE 3.

If the packet lengths are uniformly distributed over the range $[\alpha, \beta]$, then the server provides the stochastic strict service curve $(\nu(\eta) + \eta_\gamma) \cdot n$ with bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$, where

$$\nu(\eta) = \frac{1}{\eta C}log\frac{e^{\eta\beta} - e^{\eta\alpha}}{\eta(\beta - \alpha)}.$$

5.2 Gilbert-Elliott Channel: Markov Chain Modeling

Consider a time-slotted[3] Gilbert-Elliott ON-OFF communication channel [2][4] which is modeled by a two-state homogeneous Markov chain. The time (number of time slots)

[3]As the slot length approaches *zero*, the service process is approximately continuous.

between state transitions is a random variable with a memory-less distribution[4].
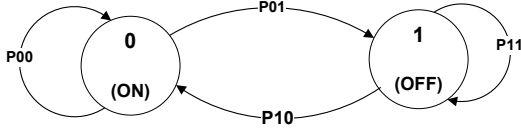

Figure 1: G-E channel model

In state ON, the channel transmits packets with a constant rate C. In state OFF, the channel does not transmit any packet and thus has the transmission rate 0. Here, we assume that when the packet is being transmitted, the channel is always in ON state and does not change to OFF state, i.e., the packet transmission will not be interrupted due to the channel state change.

As shown in Fig.(1), the transition probability from state i to j is denoted by p_{ij}, $i, j = 0, 1$ where 0 represents the 'ON' state and 1 represents the 'OFF' state. The transition probability matrix $\mathbb{P}$ is as follows:

$$\mathbb{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}.$$

For such ON-OFF service process, we define the *packet service time* as the interval between the time when a packet reaches (HOL) and the time when the last bit of this packet has been successfully transmitted. Let $v(n)$ denote the time when a packet reaches the head of line (HOL) and be called the *virtual start time* defined by:

$$v(n) = \max[a(n), d(n-1)]. \tag{13}$$

The service time δ_n is

$$\delta_n = d(n) - v(n).$$

If the packet reaches the HOL when the channel is in ON state, the packet is transmitted immediately. Otherwise, the packet has to wait until the channel state becomes ON. Let T_n^{Off} denote the OFF interval before packet $P(n)$ can be successfully transmitted. The service time is computed by

$$\delta_n = T_n^{Off} + t_n^{tx}, \tag{14}$$

where t_n^{tx} is the time (number of time slots) of transmitting a packet with length L_n. Assume that the packet lengths are *i.i.d.* and follow some general distribution.
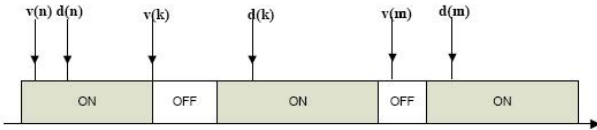

Figure 2: On-off model

Fig.(2) shows three scenarios of the packet service time.

- Scenario I: $v(n)$ is within the ON interval, thus

$$\delta_n = t_n^{tx}.$$

[4]Strictly speaking, the intervals between state transition are conditionally independent and follow geometric distribution.

- Scenario II: $v(k)$ is the boundary between the ON interval and the OFF interval, thus

$$\delta_k = T_k^{Off} + t_k^{tx}.$$

- Scenario III: $v(m)$ is within the OFF interval, thus

$$\delta_m = T_m^{Off} + t_m^{tx}.$$

For any packet $P(n)$ served in a backlogged period, Scenario III will not happen because $v(n) = d(n-1) + \epsilon$ ($\epsilon \to 0$) and $d(n-1)$ is always in the ON interval. Thus, the service time of packets which are served in backlogged periods equals either t_n^{tx} or $T_n^{Off} + t_n^{tx}$. Since Scenario III takes on value between Scenario I and Scenario II, it is suffice to analyze only Scenario I and Scenario II.

The cumulative service process $\Delta(m, n)$ (in any backlogged period) is formed from process $\{\delta_n^Y\}$ that takes on values $Y \cdot T^{off} + \frac{L_n}{C}$, where $Y = 1$ if $v(n)$ is exactly the boundary between the ON interval and the OFF interval; $Y = 0$ if $v(n)$ is within the ON interval. The MGF of the *i.i.d.* random variables δ_n^Y are $M_{\delta^Y}(\eta) = \mathbf{E}[e^{\eta \delta^Y}]$. Let $\mathbb{M}$ be the diagonal matrix

$$\mathbb{M} = \begin{bmatrix} M_{\delta^0}(\eta) & 0 \\ 0 & M_{\delta^1}(\eta) \end{bmatrix}.$$

Given the initial condition $v(1) = i$ ($i = 0, 1$), from Kolmogorov backward equation, we have

$$
\begin{aligned}
&\mathbf{E}[e^{\eta \Delta(1,n)} | v(1) = i] \\
=\ & \mathbf{E}[e^{\eta \delta^Y} | v(1) = i] \mathbf{E}[e^{\eta(\Delta(1,n) - \delta^Y)} | v(1) = i] \\
=\ & M_{\delta^i}(\eta) \sum_{j=0}^{1} \mathbf{E}[e^{\eta(\Delta(2,n))} | v(2) = j, v(1) = i] \cdot \\
& P(v(2) = j | v(1) = i) \\
=\ & M_{\delta^i}(\eta) \sum_{j=0}^{1} \mathbf{E}[e^{\eta(\Delta(2,n) - \delta_1)} | v(2) = j] p_{ij} \\
=\ & M_{\delta^i}(\eta) \sum_{j=0}^{1} \mathbf{E}[e^{\eta \Delta(2,n)} | v(2) = j] p_{ij} \tag{15}
\end{aligned}
$$

Let

$$\Phi(\eta, n) = \left(\mathbf{E}[e^{\eta \Delta(1,n)} | v(1) = 0], \mathbf{E}[e^{\eta \Delta(1,n)} | v(1) = 1] \right)$$

and $\Phi(\eta, n)^T$ be its transpose. We then rewrite (15) in matrix form:

$$\Phi(\eta, n)^T = \mathbb{M}\mathbb{P}\Phi(\eta, n-1)^T. \tag{16}$$

Applying (16) to its right-hand side iteratively results in

$$\Phi(\eta, n)^T = (\mathbb{M}\mathbb{P})^{n-1} \Phi(\eta, 1)^T. \tag{17}$$

The initial condition can be obtained by

$$\Phi(\eta, 1)^T = \mathbb{M}\mathbf{1}^T$$

where $\mathbf{1} = [1 \ 1]$ is a vector with two entries being one.

Let π_i be the steady probability at state i and $\psi = [\pi_0 \ \pi_1]$ which is computed by

$$\pi_0 = \frac{p_{10}}{2 - p_{00} - p_{11}}, \quad \pi_1 = \frac{p_{01}}{2 - p_{00} - p_{11}}.$$

Then we have

$$\mathbf{E}\left[e^{\eta\Delta(1,n)}\right] = \psi\Phi(\eta,n)^T$$
$$= \psi\left(\mathbb{M}\mathbb{P}\right)^{n-1}\mathbb{M}\mathbf{1}^T. \qquad (18)$$

Let $\rho(\cdot)$ denote the spectral radius of a matrix:

$$\rho(\cdot) = \sup\{|\alpha| : \alpha \in \sigma(\cdot)\}$$

where $|\cdot|$ denotes the absolute value of α, and $\sigma(\cdot)$ represents the set of all eigenvalues of a matrix. Then the spectral radius of matrix $\mathbb{M}\mathbb{P}$ is denoted by $\rho(\mathbb{M}\mathbb{P})$. Note that $\mathbb{M}\mathbb{P}$ is a non-negative matrix. Having known the transition probability matrix $\mathbb{P}$, the spectral radius of $\mathbb{M}\mathbb{P}$ is

$$\rho(\mathbb{M}\mathbb{P}) = \frac{p_{00}M_{\delta^0}(\eta) + p_{11}M_{\delta^1}(\eta) + \sqrt{Z}}{2} \qquad (19)$$

where

$$Z = \left(p_{00}M_{\delta^0}(\eta) - p_{11}M_{\delta^1}(\eta)\right)^2 + 4p_{01}p_{10}M_{\delta^0}(\eta)M_{\delta^1}(\eta).$$

A useful corollary (see Corollary 5.6.13 [5]) is introduced here to facilitate the following analysis.

COROLLARY 1. *Let $\mathbb{A}$ be an $k \times k$ matrix and $\epsilon > 0$ be given. There is a constant σ_ϵ[5] such that*

$$|(\mathbb{A}^m)_{ij}| \le \sigma_\epsilon(\rho(\mathbb{A}) + \epsilon)^m \qquad (20)$$

for all $m = 1, 2, 3, \ldots$ and $i, j = 1, 2, \ldots, k$.

From Corollary 1, we know that the every entry of matrix $\left(\mathbb{M}\mathbb{P}\right)^n$ is bounded above by $\sigma_\epsilon\left(\rho(\mathbb{M}\mathbb{P}) + \epsilon\right)^n$ for any $\epsilon > 0$ and some constant $\sigma_\epsilon > 0$.

Then (18) is bounded by

$$\mathbf{E}\left[e^{\eta\Delta(1,n)}\right] \le \psi\sigma_\epsilon\left(\rho(\mathbb{M}\mathbb{P}) + \epsilon\right)^{n-1}\mathbb{M}\mathbf{1}^T$$
$$= \sigma_\epsilon\left(\rho(\mathbb{M}\mathbb{P}) + \epsilon\right)^{n-1}\psi\mathbb{M}\mathbf{1}^T$$
$$= \sigma_\epsilon\left(\rho(\mathbb{M}\mathbb{P}) + \epsilon\right)^{n-1} \qquad (21)$$
$$\le \left(\rho(\mathbb{M}\mathbb{P}) + \epsilon\right)^n$$

for $\sigma_\epsilon \le \rho(\mathbb{M}\mathbb{P}) + \epsilon$. Here step (21) is obtained because $\psi\mathbb{M} = \psi$ and then $\psi\mathbf{1}^T = 1$. Hence, we have

$$\frac{1}{\eta n}\log\mathbf{E}\left[e^{\eta\Delta(1,n)}\right] \le \frac{1}{\eta}\log(\rho(\mathbb{M}\mathbb{P}) + \epsilon).$$

As ϵ is arbitrary, letting $\epsilon \to 0$ results in the stochastic strict service curve $(\nu(\eta) + \eta_\gamma) \cdot (n - m + 1)$ for the service process $\Delta(m, n)$ with the bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$, where

$$\nu(\eta) = \frac{1}{\eta}\log\rho(\mathbb{M}\mathbb{P}). \qquad (22)$$

EXAMPLE 4.

Consider a flow of variable-length packets. Suppose the packet lengths are *i.i.d.* variables with moment generating function $M_L(\eta)$. The OFF intervals follow geometric distribution with parameter π_0. Then $M_{\delta^0}(\eta) = \left(M_L(\eta)\right)^{\frac{1}{C}}$ and $M_{\delta^1}(\eta) = M_{\delta^0}(\eta)\frac{\pi_0 e^\eta}{1 - \pi_1 e^\eta}$ for $\pi_1 e^\eta < 1$. Inserting $M_{\delta^0}(\eta)$ and $M_{\delta^1}(\eta)$ into (19), we obtain the stochastic service curve $(\nu(\eta) + \eta_\gamma) \cdot n$ with bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$ for such flow, where

$$\nu(\eta) = \frac{1}{\eta}\log\frac{M_{\delta^0}(\eta)\left[p_{00} + \frac{p_{11}\pi_0 e^\eta}{1 - \pi_1 e^\eta} + \Upsilon\right]}{2} \qquad (23)$$

---

[5]This parameter is relevant to ϵ.

with

$$\Upsilon = \sqrt{\left(p_{00} - \frac{p_{11}\pi_0 e^\eta}{1 - \pi_1 e^\eta}\right)^2 + \frac{4p_{10}p_{01}\pi_0 e^\eta}{1 - \pi_1 e^\eta}}. \qquad (24)$$

EXAMPLE 5.

Consider a flow consisting of fixed-length packets. The OFF intervals follow geometric distribution with parameter π_0. Let the packet transmission time be T time slots. Then $M_{\delta^0}(\eta) = e^{\eta T}$ and $M_{\delta^1}(\eta) = e^{\eta T}\frac{\pi_0 e^\eta}{1 - \pi_1 e^\eta}$. Inserting $M_{\delta^0}(\eta)$ and $M_{\delta^1}(\eta)$ into (19), we have the stochastic service curve $(\nu(\eta) + \eta_\gamma) \cdot n$ with bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$ for such flow, where

$$\nu(\eta) = \frac{1}{\eta}\log\frac{e^{\eta T}\left[p_{00} + \frac{p_{11}\pi_0 e^\eta}{1 - \pi_1 e^\eta} + \Upsilon\right]}{2}$$
$$= T + \frac{1}{\eta}\log\frac{p_{00} + \frac{p_{11}\pi_0 e^\eta}{1 - \pi_1 e^\eta} + \Upsilon}{2} \qquad (25)$$

where Υ is given in (24).

5.3 Gilbert-Elliott Channel: Error Process Analysis

The previous subsection is based on directly analyzing backlogged periods. In this subsection, we adopt an intuitive way which models the channel OFF intervals as an error process. The channel is treated as an error server consisting of an ideal service process and an error process. The ideal service process provides service at a constant rate C. The error process is described as an ON-OFF process. The channel ON state corresponds to the error OFF state and the channel OFF state corresponds to the error ON state. When the channel is in error ON state, the error process provides service time ε_n following geometric distribution with parameter π_0. When the channel is in error OFF state, the error process does not provide service, i.e., the service time is zero. The transition probability matrix is the same as $\mathbb{P}$. However, the diagonal matrix $\mathbb{M}_\mathbb{E}$ is $diag\{M_{\mathbb{E},0}(\eta), M_{\mathbb{E},1}(\eta)\}$ where $M_{\mathbb{E},0}(\eta) = 1$ and $M_{\mathbb{E},1}(\eta) = \frac{\pi_0 e^\eta}{1 - \pi_1 e^\eta}$.

The error process provides a stochastic strict service curve $\gamma_\mathbb{E}(n) = (\frac{1}{\eta}\log\rho(\mathbb{M}_\mathbb{E}\mathbb{P}) + \eta_\gamma) \cdot n$ with bounding function $j_\mathbb{E}(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$, where $\rho(\mathbb{M}_\mathbb{E}\mathbb{P})$ is obtained from (19):

$$\rho(\mathbb{M}_\mathbb{E}\mathbb{P}) = \frac{p_{00} + \frac{p_{11}\pi_0 e^\eta}{1 - \pi_1 e^\eta} + \Upsilon}{2} \qquad (26)$$

where Υ is given in (24).

The error server thus provides a stochastic strict service curve $\gamma(n) = \hat{\gamma}(n) + \gamma_\mathbb{E}(n)$.

EXAMPLE 6.

Consider the same flow given in Example 4. Then the Gilbert-Elliott channel provides a stochastic strict service curve $(\nu(\eta) + \eta_\gamma) \cdot n$ with bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$, where

$$\nu(\eta) = \left(\frac{1}{\eta}\log\mathbf{E}\left[e^{\eta\frac{L_0}{C}}\right] + \frac{1}{\eta}\log\rho(\mathbb{M}_\mathbb{E}\mathbb{P})\right)$$
$$= \frac{1}{\eta}\log\frac{\mathbf{E}\left[e^{\eta\frac{L_0}{C}}\right]\left(p_{00} + \frac{p_{11}\pi_0 e^\eta}{1 - \pi_1 e^\eta} + \Upsilon\right)}{2}$$

which matches the result of Example 4.

Example 7.

Consider the same flow given in Example 5. Then the Gilbert-Elliott channel provides a stochastic strict service curve $(\nu(\eta)+\eta_\gamma) \cdot n$ with bounding function $j(x) = e^{-\eta x}e^{-\eta\eta_\gamma}$, where

$$\nu(\eta) \quad = \quad T + \frac{1}{\eta}log\frac{p_{00} + \frac{p_{11}\pi_0 e^{\eta}}{1-\pi_1 e^{\eta}} + \Upsilon}{2} \qquad (27)$$

which matches the result of Example 5.

Remark. By comparing the above two examples to Example 4 and Example 5, we notice that the backlogged period analysis and the error process analysis yield the same results for analyzing the two-state Gilbert-Elliott channel.

6. STOCHASTIC DELAY BOUND

In the previous sections, we have introduced applying the logarithmic moment generating function to finding the stochastic arrival curve and the stochastic service curve for several arrival processes and service processes. If the stochastic arrival curve of an arrival process is known and the service process provided to the arrival process can be characterized by a stochastic service curve, we readily obtain the delay bound based on results of the time-domain SNC [12].

6.1 Delay Bound Analysis

We recall the main steps[6] of deriving the delay bound given the stochastic arrival curve $\lambda(n)$ with bounding function $h(x)$ and the stochastic service curve $\gamma(n)$ with bounding function $j(x)$.

For any packet $P(n)$, its system delay $D(n)$ is:

$$
\begin{aligned}
D(n) &= d(n) - a(n) \\
&= \Big[d(n) - a\bar{\otimes}\gamma(n)\Big] + \Big[a\bar{\otimes}\gamma(n) - a(n)\Big] \\
&\leq d(n) - a\bar{\otimes}\gamma(n) + \sup_{1\leq m<n}\Big\{\lambda(n-m) - \Gamma(m,n)]\Big\} \\
&\quad + \sup_{k\geq 0}\Big\{\gamma(k+1) - \lambda(k)\Big\} \qquad (28)
\end{aligned}
$$

In order to ensure network stability, we shall assume

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

Since $\gamma(k)$ and $\lambda(k)$ are linear functions of the variable k, we obtain

$$\sup_{k\geq 0}\Big\{\gamma(k+1) - \lambda(k)\Big\} = \gamma(1).$$

From the given conditions, the system delay is bounded by:

$$P\{D(n) > x\} \leq j \otimes h\big([x - \gamma(1)]^+\big) \qquad (29)$$

where $[y]^+ \equiv \max\big[0, y\big]$.

If the arrival process $\Gamma(m,n)$ is independent of the service process $\Delta(m,n)$, the stochastic delay bound is given below

$$P\{D(n) > x\} \leq 1 - \int_0^{x^*}\big(1 - j(x^* - z)\big)d(1 - h(z)) \qquad (30)$$

where $x^* = [x - \gamma(1)]^+$.

Remark. Step (28) requires that the arrival process has a $v.w.d$ stochastic arrival curve and the service process has

---

6We refer the reader to [12] for the detailed derivation procedure.

---

an $i.d$ stochastic service curve. (29) and (30) are obtained from Lemma 1.5 and Lemma 6.1 [7], respectively. Moreover, (30) implies that at least one of $h(x)$ and $j(x)$ is integrable.

Example 8.

Consider a flow of fixed-length packets, of which the inter-arrival times follow the same exponential distribution as Example 1. Packets of this flow arrive to a wireless node and are queued in the buffer before they are transmitted over a Gilbert-Elliott On-Off channel as given in Example 7. Considering the assumption that when a packet is being transmitted, the channel state will not change to OFF, we set the time slot length to one packet transmission time.

Then according to (29), the delay that a packet experiences in this system is stochastically bounded by

$$P\{D(n) > x\} \leq \inf_{0<\eta\leq\eta^*}\inf_{0\leq z\leq x^*}\Big[e^{-\eta\eta_\gamma}e^{-\eta z} + e^{-\eta(x^*-z)}\Big] \qquad (31)$$

where $x^* = [x - \nu(\eta) - \eta_\gamma]^+$ with $\nu(\eta)$ given in (27), and η^* is the maximal value that η can take under $\pi_1 e^\eta < 1$.

If the arrival process is independent of the service process, according to (30), the stochastic delay bound is given by

$$P\{D(n) > x\} \leq \inf_{0<\eta\leq\eta^*}\Big[e^{-\eta x^*} + x^*\eta e^{-\eta(\eta_\gamma + x^*)})\Big], \qquad (32)$$

where x^* is the same as that in (31).

6.2 Comparison between Spatial and Temporal Analysis Approaches

As mentioned in Sec. 1, the available SNC literature for performance guarantee analysis focuses on the spatial perspective. In this subsection, we derive the system delay bound using the spatial analysis approach. The arrival process given in Example 1 and the network system given in Example 7 are used here for ease of exposition.

The spatial approach characterizes the arrival process based on the cumulative amount of arrival traffic and the service process based on the cumulative amount of service. Accordingly, the *space-domain* stochastic arrival curve and the *space-domain* stochastic service curve are the bounds on the cumulative amount of traffic and service, respectively. Let $A(t)$ and $\alpha(t)$ denote the arrival process and its arrival curve which is with the bounding function $f(y)$, and $S(t)$ and $\beta(t)$ denote the service process and its service curve which is with the bounding function $g(y)$.

Example 9.

The arrival process given in Example 1 is a compound Poisson process from the spatial perspective. Let L be the packet length for all packets. The arrival rate of the Poisson process is μ. This compound Poisson arrival process has the stochastic arrival curve as below [6], for $t \geq 0$ and $\theta > 0, \theta_\alpha \geq 0$,

$$\alpha(t) = \Big(\frac{\mu}{\theta}(e^{\theta L} - 1) + \theta_\alpha\Big) \cdot t, \quad f(y) = e^{-\theta\theta_\alpha}e^{-\theta y}.$$

For the Gilbert-Elliott ON-OFF channel, its space-domain stochastic service curve is the variation of the ON-OFF model's envelop process (see [1]), for all $t \geq 0$ and $\theta > 0$,

$$\beta(t) = \frac{t}{\theta}log\Big(\frac{p_{11} + p_{00}e^{\frac{1}{T}\theta} + \mathbf{Y}}{2}\Big), \quad g(y) = e^{-\theta y},$$

where

$$\mathbf{Y} = \sqrt{(p_{11} + p_{00} e^{\frac{1}{T}\theta})^2 - 4(p_{11} + p_{00} - 1)e^{\frac{1}{T}\theta}}.$$

Here $1/T$ represents the channel transmission rate (number of packets) in the 'ON' state due to the time-slotted channel with the slot length T.

We define the maximum horizontal distance between functions $\alpha(t)$ and $\beta(t)$, denoted by $H(\alpha, \beta)$ as follows:

$$H(\alpha, \beta) = \sup_{s \geq 0} \big\{ \inf\{\tau \geq 0 : \alpha(s) \leq \beta(s+\tau)\} \big\},$$

which can be considered as the maximal system delay of a virtual system, where the arrival process is $\alpha(t)$ and the service process is $\beta(t)$.

Let $D(t)$ denote the system delay of the traffic arriving at time $t \geq 0$. The system delay bound is known as (see Theorem 5.4 [7]):

$$P\Big\{ D(t) > H(\alpha + y, \beta) \Big\}$$
$$\leq \inf_{0 < \theta \leq \theta^*} \inf_{0 \leq z \leq y} \big[e^{-\theta\theta\alpha} e^{-\theta z} + e^{-\theta(y-z)} \big]. \quad (33)$$

Similar to (30), if the arrival process $A(t)$ is independent of the service process $S(t)$, the system delay is stochastically bounded by the Stieltjes convolution of $f(y)$ and $g(y)$:

$$P\Big\{ D(t) > H(\alpha + y, \beta) \Big\}$$
$$\leq \inf_{0 < \theta \leq \theta^*} e^{-\theta y} + \theta y e^{-\theta\theta\alpha} e^{-\theta y}. \quad (34)$$

Remark. Although the bounding functions obtained by applying the spatial approach (see (33) and (34)) look very similar as those obtained from the temporal analysis (see (31) and (32)), their arguments have different meanings. To compute the delay bound, the spatial approach uses the amount of traffic denoted by y as the argument, while the temporal approach uses the time denoted by x (or x^*) as the argument. Since the bounding functions of both approaches are negative exponential functions, a larger argument yields a smaller result and vice versa.

6.3   Numerical Results

The available literature [7] provides a simple example to illustrate that by considering the independence condition, the tightness of the delay bound may be improved, i.e., (30) may provide a tighter bound compared to (29). In order to intuitively illustrate these two bounds, we use Matlab to numerically compute the two bounds derived in Example 8 (see (31) and (32)). Then, we investigate how the optimal parameter η_γ impact the delay bound. Moreover, the system delay bounds obtained by the temporal and the spatial approaches are compared.

The Gilbert-Elliott channel provides $C = 2Mbps$ capacity when it is in the ON state. All packets have the same length 250 bytes. Hence the packet transmission time $T = 1msec$ which is the time slot length. The transition probabilities between ON and OFF states hold such relationship $p_{10}/p_{01} = 3$, from which, we calculate $\pi_0 = 0.75$, $\pi_1 = 0.25$ and $p_{00} = (2 + p_{11})/3$. If we set $p_{00} = 0.95$, the corresponding transition probabilities are $p_{01} = 0.05$, $p_{10} = 0.15$ and $p_{11} = 0.85$, respectively. Moreover, according to $\pi_1 = 0.25$, we obtain the maximal value of η, $\eta^* = 1.386$.

In the following figures, we use **Bound 1** and **Bound 2** to represent the bounds given in (31) and (32), respectively.
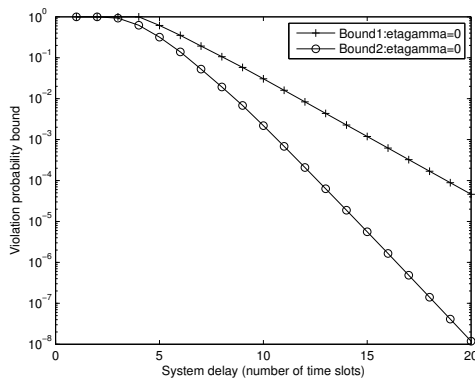

Figure 3: Bound Comparison: $\eta_\gamma = 0$

As shown in Fig. 3, Bound 1 is looser than Bound 2 under the same condition, i.e., the same arrival process and the same network system. This result implies that by considering the independence condition, the bound may be improved. The same phenomenon has been discussed in [7].
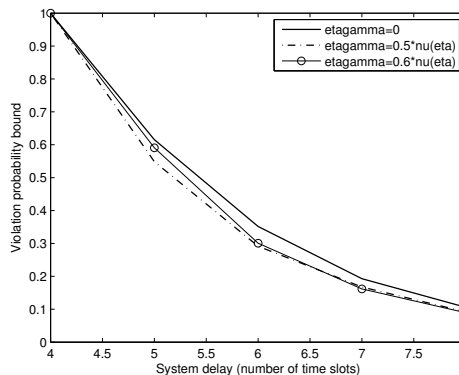

Figure 4: Bound 1 vs. Varying η_γ

Since both (31) and (32) contain the scaling factor, $e^{-\eta\eta_\gamma}$, how does this scaling parameter impact the delay bound?

From the definition of η_γ, it should be set in terms of $\nu(\eta)$. Fig. 4 shows Bound 1 when η_γ takes 0, $0.5\nu(\eta)$ and $0.6\nu(\eta)$. The bound obtained by setting $\eta_\gamma = 0.5\nu(\eta)$ is tighter than that obtained by setting $\eta_\gamma = 0$. However, when η_γ exceeds $0.5\nu(\eta)$ such as $0.6\nu(\eta)$, the bound becomes more loose than that obtained by setting $\eta_\gamma = 0.5\nu(\eta)$. Thus, $\eta_\gamma = 0.5\nu(\eta)$ is the optimal value.

Similarly, Fig. 5 shows Bound 2 against various η_γ. As η_γ increases from 0 to $0.3\nu(\eta)$ or $0.5\nu(\eta)$, the bound becomes looser. This is because the increase of η_γ results in the decrease of y in (32) for a fixed x. Moreover, taking the infimum impacts the final result as well.

Fig. 6 shows the bounds obtained from the spatial and the temporal approaches. When computing the space-domain delay bound according to (33), we need to first compute the space-domain arrival curve $\alpha(t)$ and the service curve $\beta(t)$ under an implicit relation, $\alpha(t) \leq \beta(t)$, which ensures system stability. With this condition and varying the arrival rate μ of the Poisson process, we can determine the maxi-
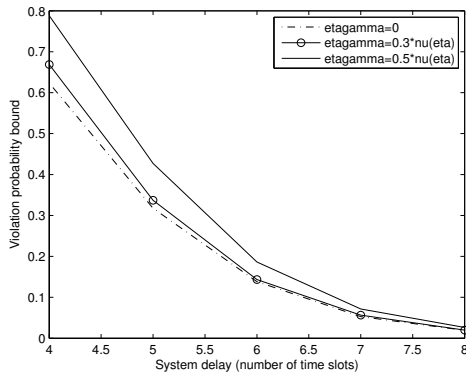
Figure 5: Bound 2 vs. Varying η_γ

mal θ, θ^*. As the amount of traffic y varies, the bound on the probability that the system delay exceeds $H(\alpha + y, \beta)$ can be obtained with θ^*. Let $\eta = \theta^*$ and $x = H(\alpha + y, \beta)$ when computing the time-domain delay bound according to (31). In Fig. 6, $\mu = 0.6$ or $\mu = 0.75$ means the arrival rate per time slot. As μ increases, more packets arrive and then the system delay becomes longer. Thus, fixing a certain time and observing the bounds obtained by setting $\mu = 0.6$ and $\mu = 0.75$, the trend is that a smaller μ causes a tighter bound. When $\mu = 0.75$, the space-domain bound and the time-domain bound are very close. However, when $\mu = 0.6$, the space-domain bound is tighter than the time-domain bound. The reason is that under the current parameter setting, the amount of traffic y used to compute the space-domain bound is larger than the time $H(\alpha + y, \beta) - \gamma(1)$ used to compute the time-domain bound. A larger argument yields a smaller result for the negative exponential functions as we have discussed in Sec. 6.2. From Fig. 6, we notice that the spatial and the temporal approaches give close results. However, how the individual parameters of bounding functions influence the final result still needs more investigation.
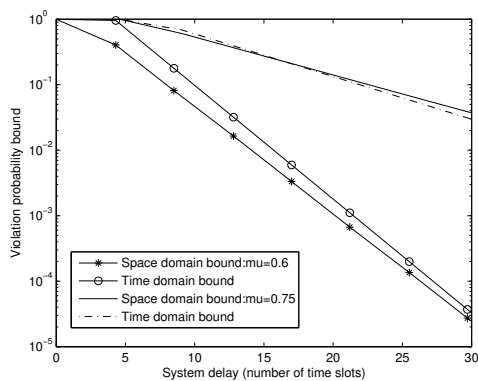

Figure 6: Space and time domain bound comparison vs. Varying μ ($\theta^* = \eta$, $\eta_\gamma = 0$ and $\theta_\alpha = 0$)

7. CONCLUSION

We introduced a temporal network calculus approach to performance guarantee analysis of stochastic networks. A key technique used in linking an arrival process or a service process to the time-domain stochastic arrival curve characterization or stochastic service curve characterization is MGF. Based on the arrival process characterization and the service process characterization, performance bounds such as delay bound can be further derived from temporal stochastic network calculus. In addition, the Gilbert-Elliott channel was particularly studied to demonstrate how to obtain the MGF of the service process. The numerical results showed that the delay bound was improved by taking into consideration the independence between the arrival process and the service process. Finally, we illustrated that the temporal and the spatial analysis approaches give close performance bounds under the appropriate match between the arguments used in both approaches.

8. REFERENCES

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

[2] E. O. Elliott. Estimates of error rates on bursty-noise channels. *Bell Systems Technical Journal*, 42(9):1977–1997, 1963.

[3] M. Fidler. A survey of deterministic and stochastic service curve models in the network calculus. *IEEE Commun. Surveys and Turotials*, 12(1):59–86, 2010.

[4] E. N. Gilbert. Capacity of a bursty-noise channel. *Bell Systems Technical Journal*, 39(5):1253–1265, 1960.

[5] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge, 1987.

[6] Y. Jiang. A note on applying stochastic network calculus. *http://q2s.ntnu.no/ jiang/publications.html*, 2010.

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

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

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

[10] K. Wu, Y. Jiang, and J. Li. On the model transform in stochastic network calculus. In *Proc. IEEE IWQoS*, 2010.

[11] J. Xie and Y. Jiang. An analysis on error servers for stochastic network calculus. In *Proc. IEEE LCN*, 2008.

[12] J. Xie and Y. Jiang. Stochastic service guarantee analysis based on time-domain models. In *17th IEEE/ACM MASCOTS*, 2009.