A Stochastic Network Calculus for Many Flows

Kazutomo Kobayashi*, Yukio Takahashi† and Hiroyuki Takada*

* Department of Computer and Information Sciences
Nagasaki University, Nagasaki 852-8131, Japan
Email: {kobayashi, htakada}@cis.nagasaki-u.ac.jp

† Department of Mathematical and Computing Sciences
Tokyo Institute of Technology, Tokyo 152-8552, Japan
Email: yukio@is.titech.ac.jp

***Abstract*—The stochastic network calculus receives much attention as a new methodology for end-to-end performance evaluation of networks, taking account of the effect of statistical multiplexing. In this paper, we present a new stochastic network calculus for many flows from an approach like large deviations techniques. In an n-node discrete-time tandem network with L flows, let $\overline{\mathcal{A}}^{\theta}(t,s)$ and $-\overline{\mathcal{S}}_i^{-\theta}(t,s)$ be the limits of the cumulant generating functions of $\overline{A}^{L}(t,s)$, arrivals to the network, and $\overline{S}_i^{L}(t,s)$, services at node i, during time interval $(s,t]$. Then, for the departures $\overline{D}^{L}(t,s)$ from the network during time interval $(s,t]$ and the backlog $Q^L(t)$ in the network at time t, we prove that the limits of the cumulant generating functions of them denoted by $\overline{\mathcal{D}}^{\theta}(t,s)$ and $\mathcal{Q}^{\theta}(t)$, respectively, satisfy an inequality $\overline{\mathcal{D}}^{\theta}(t,s) \le \overline{\mathcal{A}}^{\theta} \oslash (\overline{\mathcal{S}}_n^{\theta} * \overline{\mathcal{S}}_{n-1}^{\theta} * \cdots * \overline{\mathcal{S}}_1^{\theta})(t,s)$ and an equality $\mathcal{Q}^{\theta}(t) = \overline{\mathcal{A}}^{\theta} \oslash (\overline{\mathcal{S}}_n^{\theta} * \overline{\mathcal{S}}_{n-1}^{\theta} * \cdots * \overline{\mathcal{S}}_1^{\theta})(t,t)$, where $\oslash$ and $*$ are deconvolution and convolution operators. By using these results, we propose approximation formulas for the end-to-end evaluation of output burstness and backlog, and we apply the formula on backlog to a tandem network with cross traffic as an example.**

I. Introduction

The theory of network calculus has been developed in 1990's to give a deterministic methodology for a worst-case evaluation of packet networks [5]. It allows us to estimate the end-to-end backlog and delay bounds, and it has been used to calculate the end-to-end quality-of-service guarantees [10]. A merit of the network calculus is in its extendability where performance bound formulas for a single node can be easily extended to those for the end-to-end links by using *min-plus algebra* [1], [3]. More specifically, if we let $S_i(t)$ be a service curve, or a service guarantee, at node i along the route of a flow with n nodes, then $S(t) = S_1 * S_2 * \cdots * S_n(t)$ provides a service curve for the entire route of the flow, where $*$ is a convolution operator.

On the other hand, a drawback of the deterministic worst-case evaluation is in the overestimation for actually necessary network resources, especially when traffic load is low, the number of flows is large, and the number of nodes is large. It is because the effect of statistical multiplexing is disregarded. To overcome this weak point, a stochastic network calculus has been discussed [2], [3], [4], [7], [9]. Importing statistical evaluation methods to the network calculus, it takes account of the effect of statistical multiplexing. For example, in [7], a calculus with moment generating functions is proposed to provide closed-form, end-to-end, probabilistic performance bounds.

In this paper, we propose another new stochastic network calculus for many flows, which is much simpler than in [7], by applying a technique used in the large deviations [6], [11]. We consider a discrete-time tandem network with n nodes and L flows. Let $\overline{A}^{L}(t,s)$ be the total arrivals to the network during time interval $(s,t]$ and $\overline{S}_i^{L}(t,s)$ the total offered services at note i during $(s,t]$. Given these processes, the cumulative departures from the network $\overline{D}^{L}(t,s)$ during interval $(s,t]$ and the total backlog in the network $Q^L(t)$ at time t can be derived. For processes $\overline{A}^{L}(t,s)$, $\overline{S}_i^{L}(t,s)$, $\overline{D}^{L}(t,s)$ and $Q^L(t)$, we denote by $\overline{\mathcal{A}}^{\theta}(t,s)$, $-\overline{\mathcal{S}}_i^{-\theta}(t,s)$, $\overline{\mathcal{D}}^{\theta}(t,s)$ and $\mathcal{Q}^{\theta}(t)$, respectively, the limits of their cumulant generating functions when $L \to \infty$. See (35), (36), (41) and (42) in Section 3 for rigorous definitions. Then we will show that for θ in an interval

$$\begin{aligned} \overline{\mathcal{D}}^{\theta}(t,s) &\le \overline{\mathcal{A}}^{\theta} \oslash \overline{\mathcal{S}}^{\theta}(t,s) \quad \text{and} \\ \mathcal{Q}^{\theta}(t) &= \overline{\mathcal{A}}^{\theta} \oslash \overline{\mathcal{S}}^{\theta}(t,t), \end{aligned}$$

where $\overline{\mathcal{S}}^{\theta}(t,s) = \overline{\mathcal{S}}_n^{\theta} * \overline{\mathcal{S}}_{n-1}^{\theta} * \cdots * \overline{\mathcal{S}}_1^{\theta}(t,s)$ and operators $*$ and $\oslash$ are convolution and deconvolution ones defined in (16) and (17). From these results, for large L, we may evaluate the backlog in the network and the output burstiness as

$$P(Q^L(t) > Ly) \le B_Q(L,y) \exp\left(-L \sup_{\theta \in Z_Q} \{\theta y - \overline{\mathcal{A}}^{\theta} \oslash \overline{\mathcal{S}}^{\theta}(t,t)\}\right)$$

and

$$P(\overline{D}^{L}(t,s) > Ly) \le B_D(L,y) \exp\left(-L \sup_{\theta \in Z_D} \{\theta y - \overline{\mathcal{A}}^{\theta} \oslash \overline{\mathcal{S}}^{\theta}(t,s)\}\right)$$

with some functions $B_Q(L,y)$ and $B_D(L,y)$ varying slower than the exponential as L increases.

As an example, we apply the above evaluation to a tandem network with cross traffic. In the network, the amounts of the arrival traffic flows are limited by leaky buckets. We provide an evaluation formula for the upper bounds above and give a numerical result for a three node network. The result shows that the maximum backlog grows as the number of nodes

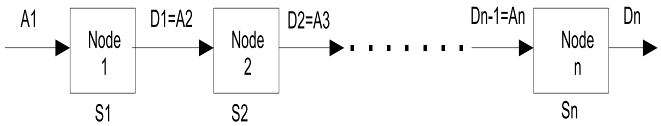

Fig. 1. Tandem network with n nodes

increases, but the probability $P(Q^L(t) > x)$ decreases faster than the exponential as the buffer threshold x increases in spite of the high link utilization of 96%.

The remainder of the paper is constructed as follows. In Section 2, we discuss a deterministic network calculus for a tandem network as preliminary. In Section 3, we derive a stochastic network calculus on the the limits of the cumulant generating functions of arrivals, services, departures and backlog. In Section 4, we discuss an application of the results to a tandem network with cross traffic, and give a numerical result. In Appendix, we present some proofs of the properties used in Section 4.

II. Preliminaries

We consider a discrete-time tandem network with n nodes illustrated in Figure 1. Time t takes discrete values as $0, 1, 2, \cdots$. Let $A^{\mathrm{net}}(t)$ be the cumulative arrivals to the network during time interval $(0, t]$, and $S_i(t)$, $i = 1, 2, \ldots, n$, be the cumulative offered services at node i during $(0, t]$. In this section, we consider $A^{\mathrm{net}}(t)$ and $S_i(t)$ are given ordinary (i.e., non-random) non-decreasing functions with $A^{\mathrm{net}}(0) = 0$ and $S_i(0) = 0$. In the next section, we will interpret these functions as sample paths of corresponding stochastic processes in the network model. We denote by $A_i(t)$ and $D_i(t)$ the cumulative arrivals at and departures from node i during $(0, t]$, and by $Q_i(t)$ the backlog in node i at time t. Then, for $t > 0$, we have

$$\begin{aligned} A_1(t) &= A^{\mathrm{net}}(t), && (1)\\ Q_i(t) &= \max_{0\le\tau\le t}\{A_i(t) - A_i(\tau) - (S_i(t) - S_i(\tau))\}, \text{ and} && (2)\\ D_i(t) &= A_i(t) - Q_i(t) \quad \text{for} \quad i = 1, 2, \ldots, n, && (3) \end{aligned}$$

and

$$A_i(t) = D_{i-1}(t) \quad \text{for} \quad i = 2, 3, \ldots, n. \qquad (4)$$

The cumulative departures from the network during $(0, t]$ is given by $D^{\mathrm{net}}(t) = D_n(t)$. These equations (1) $\sim$ (4) determine functions $A_i(t)$, $D_i(t)$ and $Q_i(t)$ for $i = 1, 2, \ldots, n$, uniquely. From the definitions, it is clear that $A_i(t)$, $Q_i(t)$ and $D_i(t)$ are nonnegative with $A_i(0) = Q_i(0) = D_i(0) = 0$, and $A_i(t)$ and $D_i(t)$ are non-decreasing.

Combining (2) and (3), we have

$$D_i(t) = \min_{0\le\tau\le t}\{A_i(\tau) + S_i(t) - S_i(\tau)\}. \qquad (5)$$

Since $D_i(t) \le A_i(t)$, we have

$$\begin{aligned} D_i(t) - D_i(s) &\le A_i(t) - D_i(s)\\ &= A_i(t) - \min_{0\le\tau\le s}\{A_i(\tau) + S_i(s) - S_i(\tau)\}\\ &= \max_{0\le\tau\le s}\{A_i(t) - A_i(\tau) - S_i(s) + S_i(\tau)\} \end{aligned}$$

for $0 \le s \le t$. If we set

$$\begin{aligned} \overline{A}_i(t,s) &= A_i(t) - A_i(s), && (6)\\ \overline{S}_i(t,s) &= S_i(t) - S_i(s), \quad \text{and} && (7)\\ \overline{D}_i(t,s) &= D_i(t) - D_i(s) && (8) \end{aligned}$$

for t and s such that $0 \le s \le t$, then the above inequality is rewritten as

$$\overline{D}_i(t,s) \le \max_{0\le\tau\le s}\{\overline{A}_i(t,\tau) - \overline{S}_i(s,\tau)\}. \qquad (9)$$

Using (6)$\sim$(8), the relations (2) and (4) are also written as

$$\begin{aligned} Q_i(t) &= \max_{0\le\tau\le t}\{\overline{A}_i(t,\tau) - \overline{S}_i(t,\tau)\} \quad \text{and} && (10)\\ \overline{A}_i(t,s) &= \overline{D}_{i-1}(t,s), && (11) \end{aligned}$$

respectively. For the cumulative arrivals to the network and the cumulative departures from the network, we introduce similar expressions to (6)$\sim$(8) as

$$\begin{aligned} \overline{A}^{\mathrm{net}}(t,s) &= A^{\mathrm{net}}(t) - A^{\mathrm{net}}(s) \quad \text{and} && (12)\\ \overline{D}^{\mathrm{net}}(t,s) &= D^{\mathrm{net}}(t) - D^{\mathrm{net}}(s), && (13) \end{aligned}$$

and write the total backlog of the network at time t as

$$Q^{\mathrm{net}}(t) = \sum_{i=1}^{n} Q_i(t). \qquad (14)$$

Using relations (3) and (4), it is easily checked that

$$Q^{\mathrm{net}}(t) = A^{\mathrm{net}}(t) - D^{\mathrm{net}}(t) = \overline{A}^{\mathrm{net}}(t,0) - \overline{D}^{\mathrm{net}}(t,0). \qquad (15)$$

Since $Q^{\mathrm{net}}(t)$ is nonnegative, this implies that $D^{\mathrm{net}}(t) \le A^{\mathrm{net}}(t)$.

To develop a new network calculus, we introduce the *convolution* operator $*$ and the *deconvolution* operator $\oslash$ for functions $f(t,s)$ and $g(t,s)$ of two variables t, s such that $0 \le s \le t$ as follows :

$$\begin{aligned} f * g(t,s) &= \min_{s\le\tau\le t}\{f(t,\tau) + g(\tau,s)\} \quad \text{and} && (16)\\ f \oslash g(t,s) &= \max_{0\le\tau\le s}\{f(t,\tau) - g(s,\tau)\}. && (17) \end{aligned}$$

These operators for non-random functions are the same as the ones introduced in [7]. However, their extensions to the stochastic case are different as will be shown in the next section.

Note that, for three functions $f(t,s)$, $g(t,s)$ and $h(t,s)$, we have $(f*g)*h(t,s) = f*(g*h)(t,s)$ for $0 \le s \le t$. Hence the operator $*$ is associative, and $(f*g)*h(t,s)$ or $f*(g*h)(t,s)$ can be written as $f*g*h(t,s)$. For the deconvolution operator, we also have a relation $f \oslash (h*g)(t,s) = (f \oslash g) \oslash h(t,s)$ for $0 \le s \le t$. If $f(t,s)$ has an *incremental property*, i.e. it satisfies a property $f(t,s) = f(t,0) - f(s,0)$ for any s,t such that $0 \le s \le t$, then

$$\begin{aligned} f(t,0) - g * f(s,0) &= f(t,0) - \min_{0\le\tau\le s}\{g(t,\tau) + f(\tau,0)\} \\ &= \max_{0\le\tau\le s}\{f(t,\tau) - g(s,\tau)\} = f \oslash g(t,s). \end{aligned} \quad (18)$$

From the definitions (16) and (17), the following inequalities hold. If $f_1(t,\tau) \le f_2(t,\tau)$ for any τ such that $s \le \tau \le t$, then

$$f_1 * g(t,s) \le f_2 * g(t,s) \text{ and } g * f_1(t,s) \le g * f_2(t,s). \quad (19)$$

If $f_1(t,\tau) \le f_2(t,\tau)$ for any τ such that $0 \le \tau \le s$, then

$$f_1 \oslash g(t,s) \le f_2 \oslash g(t,s) \text{ and } g \oslash f_2(t,s) \le g \oslash f_1(t,s). \quad (20)$$

Using these operators, we can rewrite (5), (10) and (9) as

$$\begin{aligned} D_i(t) &= \overline{D}_i(t,0) = \overline{S}_i * \overline{A}_i(t,0), && (21) \\ Q_i(t) &= \overline{A}_i \oslash \overline{S}_i(t,t), \quad \text{and} && (22) \\ \overline{D}_i(t,s) &\le \overline{A}_i \oslash \overline{S}_i(t,s), && (23) \end{aligned}$$

respectively. These equalities and inequality for node i can be extended to those for the whole network as stated in the following lemma.

Lemma 1: For s and t such that $0 \le s \le t$, we have

$$\begin{aligned} D^{\rm net}(t) &= \overline{D}^{\rm net}(t,0) = \overline{S}^{\rm net} * \overline{A}^{\rm net}(t,0), && (24) \\ Q^{\rm net}(t) &= \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t,t), \quad \text{and} && (25) \\ \overline{D}^{\rm net}(t,s) &\le \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t,s), && (26) \end{aligned}$$

where

$$\begin{aligned} &\overline{S}^{\rm net}(t,s) = \overline{S}_n * \overline{S}_{n-1} * \cdots * \overline{S}_1(t,s) && (27) \\ &= \min_{s=s_0\le s_1\le\cdots\le s_n=t}\{\overline{S}_n(s_n,s_{n-1}) + \cdots + \overline{S}_1(s_1,s_0)\}. && (28) \end{aligned}$$

Proof. Since $D^{\rm net}(t) = D_n(t)$, we have $D^{\rm net}(t) = \overline{D}^{\rm net}(t,0) = \overline{D}_n(t,0)$. On the other hand, from (21) and (11)

$$\overline{D}_i(t,0) = \overline{S}_i * \overline{A}_i(t,0) = \overline{S}_i * \overline{D}_{i-1}(t,0).$$

Using this relation repeatedly, we have

$$\begin{aligned} D^{\rm net}(t) &= \overline{D}^{\rm net}(t,0) = \overline{D}_n(t,0) = \overline{S}_n * \overline{D}_{n-1}(t,0) \\ &= \overline{S}_n * \overline{S}_{n-1} * \overline{D}_{n-2}(t,0) = \cdots \\ &\quad \cdots = \overline{S}_n * \overline{S}_{n-1} * \cdots * \overline{S}_2 * \overline{D}_1(t,0) \\ &= \overline{S}_n * \overline{S}_{n-1} * \cdots * \overline{S}_1 * \overline{A}_1(t,0) = \overline{S}^{\rm net} * \overline{A}^{\rm net}(t,0). \end{aligned}$$

This proves (24). Applying the above representation to (15) and using relation (18), we have

$$\begin{aligned} Q^{\rm net}(t) &= \overline{A}^{\rm net}(t,0) - \overline{D}^{\rm net}(t,0) \\ &= \overline{A}^{\rm net}(t,0) - \overline{S}^{\rm net} * \overline{A}^{\rm net}(t,0) = \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t,t). \end{aligned}$$

This proves (25). Using the inequality $D^{\rm net}(t) \le A^{\rm net}(t)$ and relations (24) and (18), we have

$$\begin{aligned} \overline{D}^{\rm net}(t,s) &= D^{\rm net}(t) - D^{\rm net}(s) \\ &\le A^{\rm net}(t) - D^{\rm net}(s) = \overline{A}^{\rm net}(t,0) - \overline{S}^{\rm net} * \overline{A}^{\rm net}(s,0) \\ &= \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t,s). \end{aligned}$$

This proves (26). The representation (28) is a direct consequence of the definition of the convolution operator (16). □

From the relation (24), $\overline{S}^{\rm net}(t,s)$ might be interpreted as the cumulative services through the network offered during $(s,t]$ though it does not have the incremental property, namely $\overline{S}^{\rm net}(t,s) \ne \overline{S}^{\rm net}(t,0) - \overline{S}^{\rm net}(s,0)$, in general.

III. Stochastic network calculus for many flows

We consider the same discrete-time tandem network with n nodes as in the previous section. In addition, we assume that the traffic through the network consists of L flows and that arrivals to the network and services at each node are not deterministic but stochastic. For time $t > 0$, let $A^L(t)$ and $S_i^L(t)$, $i = 1, 2, \ldots, n$, be random variables representing the total arrivals to the network and the total services at node i, respectively, during time interval $(0,t]$ for L flows. Further, let $D^L(t)$ be the total departures from the network of L flows during $(0,t]$ and $Q^L(t)$ the total backlog of L flows in the network at time t. Sample paths of the processes $\{A^L(t)\}$, $\{S_i^L(t)\}$, $\{Q^L(t)\}$ and $\{D^L(t)\}$ correspond to $A^{net}(t)$, $S_i(t)$, $Q^{net}(t)$ and $D^{net}(t)$, respectively, in the previous section.

For a pair of times t and s such that $0 \le s \le t$, we let $\overline{A}^L(t,s) = A^L(t) - A^L(s)$, $\overline{S}_i^L(t,s) = S_i^L(t) - S_i^L(s)$ and $\overline{D}^L(t,s) = D^L(t) - D^L(s)$. Then from Lemma 1, we have

$$\begin{aligned} Q^L(t) &= \overline{A}^L \oslash \overline{S}^L(t,t) \quad \text{and} && (29) \\ \overline{D}^L(t,s) &\le \overline{A}^L \oslash \overline{S}^L(t,s) && (30) \end{aligned}$$

with probability one, where

$$\overline{S}^L(t,s) = \overline{S}_n^L * \overline{S}_{n-1}^L * \cdots * \overline{S}_1^L(t,s). \quad (31)$$

We let $\overline{W}^L(t,s) = \overline{A}^L \oslash \overline{S}^L(t,s)$. Then (29) and (30) are rewritten as

$$\begin{aligned} Q^L(t) &= \overline{W}^L(t,t) \quad \text{and} && (32) \\ \overline{D}^L(t,s) &\le \overline{W}^L(t,s). && (33) \end{aligned}$$

From the definitions (16) and (17) of the convolution and deconvolution operators, $\overline{W}^L(t,s)$ itself is written as

$$\begin{aligned} &\overline{W}^L(t,s) \\ &= \max_{0\le s_0\le s}\Big\{\overline{A}^L(t,s_0) \\ &\quad - \min_{s_0\le s_1\le\cdots\le s_n=s}\{\overline{S}_n^L(s_n,s_{n-1}) + \cdots + \overline{S}_1^L(s_1,s_0)\}\Big\} \\ &= \max_{0\le s_0\le s_1\le\cdots\le s_n=s}\{\overline{A}^L(t,s_0) \\ &\qquad -\overline{S}_n^L(s_n,s_{n-1}) - \cdots - \overline{S}_1^L(s_1,s_0)\}. && (34) \end{aligned}$$

Hereafter, in this section, we regard times t and s are arbitrarily chosen so that $0 \le s \le t$ and fixed through discussions. Statements, equalities and inequalities including $s_0, s_1, \cdots, s_{n-1}$ or s_n should be understood to hold for any choice of $s_0, s_1, \cdots, s_{n-1}, s_n$ satisfying the relation $0 \le s_0 \le s_1 \le \cdots \le s_{n-1} \le s_n = s$, except for the cases stated otherwise.

For the sequences of random variables $\{\overline{A}^L(t, s_0)\}_{L=1,2,\cdots}$ and $\{\overline{S}_i^L(s_i, s_{i-1})\}_{L=1,2,\cdots}$, $i = 1, 2, \cdots, n$, we make the following assumptions. Note that $\log E[e^{\theta X}]$ is *the cumulant generating function* (cgf) of random variable X.

A1. The random variables $\overline{A}^L(t, s_0)$ and $\overline{S}_i^L(s_i, s_{i-1})$, $i = 1, \cdots, n$, are mutually independent.

A2. For each $\theta \in \boldsymbol{R}$, when $L \to \infty$, the sequences

$$\left\{L^{-1} \log E[e^{\theta \overline{A}^L(t,s_0)}]\right\}_{L=1,2,\cdots} \quad \text{and}$$

$$\left\{-L^{-1} \log E[e^{-\theta \overline{S}_i^L(s_i,s_{i-1})}]\right\}_{L=1,2,\cdots}, \ i = 1, \cdots, n,$$

have limits as extended real numbers (i.e., allowing $\pm\infty$). We denote the limits as

$$\overline{\mathcal{A}}^\theta(t, s_0) = \lim_{L\to\infty} L^{-1} \log E[e^{\theta \overline{A}^L(t,s_0)}] \quad \text{and} \quad (35)$$

$$\overline{\mathcal{S}}_i^\theta(s_i, s_{i-1}) = -\lim_{L\to\infty} L^{-1} \log E[e^{-\theta \overline{S}_i^L(s_i,s_{i-1})}], \quad i = 1, \cdots, n. \quad (36)$$

A3. The sets $\mathcal{Z}_{\overline{\mathcal{A}}(t,s_0)} \equiv \{\theta : |\overline{\mathcal{A}}^\theta(t, s_0)| < \infty\} \cap (0, \infty)$ and $\mathcal{Z}_{\overline{\mathcal{S}}_i(s_i,s_{i-1})} \equiv \{\theta : |\overline{\mathcal{S}}_i^\theta(s_i, s_{i-1})| < \infty\} \cap (0, \infty)$, $i = 1, \cdots, n$, are all non-empty.

From the monotonicity of the logarithmic and exponential functions, it is easily checked that the set $\mathcal{Z}_{\overline{\mathcal{A}}(t,s_0)}$ is an open or semi-closed interval of the form $(0, \delta_{\overline{\mathcal{A}}(t,s_0)})$ or $(0, \delta_{\overline{\mathcal{A}}(t,s_0)}]$ with some positive number $\delta_{\overline{\mathcal{A}}(t,s_0)}$ or with $\delta_{\overline{\mathcal{A}}(t,s_0)} = \infty$. The set $\mathcal{Z}_{\overline{\mathcal{S}}_i(s_i,s_{i-1})}$ is also such an interval. Hence, under assumptions A2 and A3, the intersection

$$\mathcal{Z}_{\overline{\mathcal{W}}(t,s)} = \bigcap_{0\le s_0\le\cdots\le s_n=s} \mathcal{Z}_{\overline{\mathcal{A}}(t,s_0)} \bigcap_{1\le i\le n} \mathcal{Z}_{\overline{\mathcal{S}}_i(s_i,s_{i-1})} \quad (37)$$

is a non-empty interval, too.

We have the following lemma, in which the stochastic calculus is the same as the deterministic one in Lemma 1 and thus is much simpler than in [7].

Lemma 2: Under assumptions A1, A2 and A3, the sequence

$$\left\{L^{-1} \log E[e^{\theta \overline{W}^L(t,s)}]\right\}_{L=1,2,\ldots}$$

with $\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,s)}$ has a finite limit $\overline{\mathcal{W}}^\theta(t, s)$ as $L \to \infty$, and the limit is given by

$$\begin{aligned} \overline{\mathcal{W}}^\theta(t, s) &= \overline{\mathcal{A}}^\theta \oslash \overline{\mathcal{S}}^\theta(t, s), \quad (38)\\ &= \max_{0\le s_0\le\cdots\le s_n=s} \{\overline{\mathcal{A}}^\theta(t, s_0) \\ &\quad -\overline{\mathcal{S}}_n^\theta(s_n, s_{n-1}) - \cdots - \overline{\mathcal{S}}_1^\theta(s_1, s_0)\} \quad (39) \end{aligned}$$

where

$$\overline{\mathcal{S}}^\theta(t, s) = \overline{\mathcal{S}}_n^\theta * \overline{\mathcal{S}}_{n-1}^\theta * \cdots * \overline{\mathcal{S}}_1^\theta(t, s). \quad (40)$$

Proof. First we show that, when $L \to \infty$, the superior limit of $L^{-1} \log E[e^{\theta \overline{W}^L(t,s)}]$ is bounded from above by the right hand side of (39) and the inferior limit of it is bounded from below by the same quantity.

Since $\theta > 0$, from (34), using the monotonicity of the exponential function and the inequality $\max(x_1, x_2) \le x_1 + x_2$ for $x_1, x_2 \ge 0$, we have

$$\begin{aligned} &e^{\theta \overline{W}^L(t,s)} \\ &= \max_{0\le s_0\le\cdots\le s_n=s} e^{\theta(\overline{A}^L(t,s_0) - \overline{S}_n^L(s_n,s_{n-1}) - \cdots - \overline{S}_1^L(s_1,s_0))} \\ &\le \sum_{0\le s_0\le\cdots\le s_n=s} e^{\theta(\overline{A}^L(t,s_0) - \overline{S}_n^L(s_n,s_{n-1}) - \cdots - \overline{S}_1^L(s_1,s_0))} \end{aligned}$$

with probability one. Taking expectation on both sides,

$$\begin{aligned} &E[e^{\theta \overline{W}^L(t,s)}] \\ &\le \sum_{0\le s_0\le\cdots\le s_n=s} E[e^{\theta(\overline{A}^L(t,s_0) - \overline{S}_n^L(s_n,s_{n-1}) - \cdots - \overline{S}_1^L(s_1,s_0))}] \\ &\le (s+1)^n \cdot \\ &\quad \max_{0\le s_0\le\cdots\le s_n=s} E[e^{\theta(\overline{A}^L(t,s_0) - \overline{S}_n^L(s_n,s_{n-1}) - \cdots - \overline{S}_1^L(s_1,s_0))}]. \end{aligned}$$

From assumption A1, the expectation in the right hand side above can be written in a product form

$$E[e^{\theta \overline{A}^L(t,s_0)}]\, E[e^{-\theta \overline{S}_n^L(s_n,s_{n-1})}] \cdots E[e^{-\theta \overline{S}_1^L(s_1,s_0))}].$$

Then using the monotonicity of the logarithmic function, the above inequality is rewritten as

$$\begin{aligned} L^{-1} \log E[e^{\theta \overline{W}^L(t,s)}] &\le L^{-1} \log (s+1)^n \\ &+ \max_{0\le s_0\le\cdots\le s_n=s} \Big\{L^{-1} \log E[e^{\theta(\overline{A}^L(t,s_0)}] \\ &+ L^{-1} \log E[e^{-\theta(\overline{S}_n^L(s_n,s_{n-1})}] + \cdots \\ &+ L^{-1} \log E[e^{-\theta(\overline{S}_1^L(s_1,s_0)}]\Big\}. \end{aligned}$$

If we let $L \to \infty$, from assumptions A2 and A3, each term in the braces above converges to a finite limit. Hence by taking superior limit on both sides we have

$$\begin{aligned} &\limsup_{L\to\infty} L^{-1} \log E[e^{\theta \overline{W}^L(t,s)}] \le \\ &\max_{0\le s_0\le\cdots\le s_n=s} \{\overline{\mathcal{A}}^\theta(t, s_0) - \overline{\mathcal{S}}_n^\theta(s_n, s_{n-1}) - \cdots - \overline{\mathcal{S}}_1^\theta(s_1, s_0)\}. \end{aligned}$$

On the other hand, from (34), for arbitrarily chosen $(s_0, s_1, \cdots, s_n)$, we have

$$\overline{W}^L(t, s) \ge \overline{A}^L(t, s_0) - \overline{S}_n^L(s_n, s_{n-1}) - \cdots - \overline{S}_1^L(s_1, s_0)$$

with probability one. Then for $\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,s)}$, we have

$$E[e^{\theta \overline{W}^L(t,s)}] \ge E[e^{\theta(\overline{A}^L(t,s_0) - \overline{S}_n^L(s_n,s_{n-1}) - \cdots - \overline{S}_1^L(s_1,s_0))}].$$

From assumption A1, the right hand side is written in a product form. Then, taking logarithm and dividing by L we have

$$\begin{aligned} L^{-1} \log E[e^{\theta \overline{W}^L(t,s)}] &\ge L^{-1} \log E[e^{\theta \overline{A}^L(t,s_0)}] \\ &+ L^{-1} \log E[e^{-\theta \overline{S}_n^L(s_n,s_{n-1})}] + \cdots \\ &+ L^{-1} \log E[e^{-\theta \overline{S}_1^L(s_1,s_0)}]. \end{aligned}$$

As $L \to \infty$, from assumptions A2 and A3, each term in the right hand side converges to a finite limit. Then by taking inferior limit on both sides, we have

$$\liminf_{L\to\infty} L^{-1}\log E[e^{\theta \overline{W}^L(t,s)}] \geq$$
$$\overline{\mathcal{A}}^\theta(t,s_0) - \overline{\mathcal{S}}^\theta_n(s_n,s_{n-1}) - \cdots - \overline{\mathcal{S}}^\theta_1(s_1,s_0).$$

Since the inequality holds for any choice of $(s_0, s_1, \cdots, s_n)$, we have

$$\liminf_{L\to\infty} L^{-1}\log E[e^{\theta \overline{W}^L(t,s)}] \geq$$
$$\max_{0\leq s_0\leq\cdots\leq s_n=s}\{\overline{\mathcal{A}}^\theta(t,s_0) - \overline{\mathcal{S}}^\theta_n(s_n,s_{n-1}) - \cdots - \overline{\mathcal{S}}^\theta_1(s_1,s_0)\}.$$

Thus, both the superior limit and the inferior limit of $L^{-1}\log E[e^{\theta \overline{W}^L(t,s)}]$ are bounded by the right hand side of (39). This proves (39).

The representation (38) is led from (39) in the reverse way of (34). □

Since $Q^L(t) = \overline{W}^L(t,t)$ from (32), as a special case of Lemma 2 with $s = t$, we see that for $\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,t)}$ the limit

$$\mathcal{Q}^\theta(t) \;=\; \lim_{L\to\infty} L^{-1}\log E[e^{\theta Q^L(t)}] \tag{41}$$

exists and is given by $\overline{\mathcal{W}}^\theta(t,t)$. This proves the first statement of the following theorem. For the second statement we let

$$\overline{\mathcal{D}}^\theta(t,s) \;=\; \limsup_{L\to\infty} L^{-1}\log E[e^{\theta \overline{D}^L(t,s)}]. \tag{42}$$

Theorem 1: Under assumptions A1, A2 and A3 for $s = t \geq 0$, we have for $\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,t)}$

$$\mathcal{Q}^\theta(t) = \overline{\mathcal{W}}^\theta(t,t). \tag{43}$$

Under assumptions A1, A2 and A3 for t and s such that $0 \leq s \leq t$, we have for $\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,s)}$

$$\overline{\mathcal{D}}^\theta(t,s) \leq \overline{\mathcal{W}}^\theta(t,s). \tag{44}$$

Proof. From (33), $\overline{D}^L(t,s) \leq \overline{W}^L(t,s)$ with probability one. Hence (44) is a direct consequence of Lemma 2. □

On the tail probabilities of $\overline{D}^L(t,s)$ and $Q^L(t)$, we have the following theorem.

Theorem 2: Under assumptions A1, A2 and A3 for $s = t > 0$, we have, for $y > 0$,

$$\limsup_{L\to\infty} L^{-1}\log P(Q^L(t) > Ly)$$
$$\leq \; -\sup_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t,t)}}\{\theta y - \overline{\mathcal{W}}^\theta(t,t)\}. \tag{45}$$

Under assumptions A1, A2 and A3 for $0 < s \leq t$, we have, for $y > 0$,

$$\limsup_{L\to\infty} L^{-1}\log P(\overline{D}^L(t,s) > Ly)$$
$$\leq \; -\sup_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t,s)}}\{\theta y - \overline{\mathcal{W}}^\theta(t,s)\}. \tag{46}$$

Proof. We apply the Chernoff's bound (or the Markov inequality, see p.240 of [3]) to the tail probability $P(\overline{W}^L(t,s) > Ly)$. For any $\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,s)}$ and $y > 0$, we have

$$P(\overline{W}^L(t,s) > Ly) \;=\; P(e^{\theta \overline{W}^L(t,s)} > e^{\theta Ly})$$
$$\leq \; e^{-\theta Ly}E[e^{\theta \overline{W}^L(t,s)}].$$

Taking the logarithm, dividing by L, and then taking the superior limit on both sides, we have

$$\limsup_{L\to\infty} L^{-1}\log P(\overline{W}^L(t,s) > Ly) \leq -\theta y + \overline{\mathcal{W}}^\theta(t,s).$$

Since the parameter θ does not appear in the left hand side, it follows that

$$\limsup_{L\to\infty} L^{-1}\log P(\overline{W}^L(t,s) > Ly)$$
$$\leq \; -\sup_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t,s)}}\{\theta y - \overline{\mathcal{W}}^\theta(t,s)\}.$$

We know that $Q^L(t) = \overline{W}^L(t,t)$ from (32). Hence, considering the case $s = t$, this inequality proves (45). Similarly, from (33) we see that $P(\overline{D}^L(t,s) > Ly) \leq P(\overline{W}^L(t,s) > Ly)$. Hence the above inequality also implies (46). □

Theorem 2 suggests that the tail probabilities of $Q^L(t)$ and $\overline{D}^L(t,s)$ might be evaluated by the form

$$P(Q^L(t) > Ly) \leq$$
$$B_Q(L,y)\exp\left(-L\sup_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t,t)}}\{\theta y - \overline{\mathcal{W}}^\theta(t,t)\}\right) \tag{47}$$

and

$$P(\overline{D}^L(t,s) > Ly) \leq$$
$$B_D(L,y)\exp\left(-L\sup_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t,s)}}\{\theta y - \overline{\mathcal{W}}^\theta(t,s)\}\right), \tag{48}$$

where $B_D(L,y)$ and $B_Q(L,y)$ are some functions varying slower than the exponential as L increases.

IV. Application to a network with cross traffic

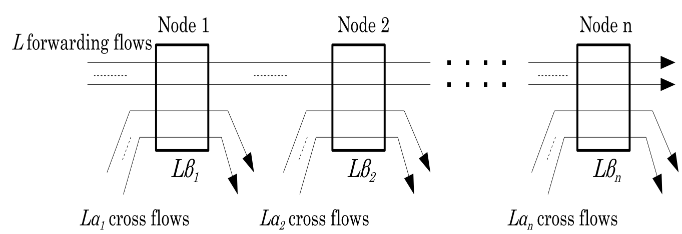

Fig. 2. Tandem network with cross traffic

We apply the bound (47) to a network with cross traffic depicted in Figure 2. In the network, there are L forwarding flows and M_i cross traffic flows at node i. We denote the L forwarding flows as $\{A_1(t)\}, \{A_2(t)\}, \cdots \cdots, \{A_L(t)\}$ and the M_i cross traffic flows at node i as $\{A^{\text{cross}}_{i1}(t)\}, \{A^{\text{cross}}_{i2}(t)\}, \cdots, \{A^{\text{cross}}_{iM_i}(t)\}$. We set $\alpha_i = M_i/L$, the ratio of the number of the cross traffic flows at node i to that of the forwarding traffic flows, and is kept constant when we move L (and M_i) to infinity later. The link capacity, i.e., the offered service per unit time, at node i is constant in time and equal to $C_i = \beta_i L$. When we move L later, β_i is kept constant. At the service, the cross traffic is served with higher priority than the forwarding traffic.

Here, for brevity of the model, we assume that flows (both forwarding flows and cross traffic flows) are mutually independent and subject to a common probabilistic law. We denote by $\{A(t)\}$ the arrival process of a typical flow, and make the following assumptions.

C1. The arrival process $\{A(t)\}$ is nondecreasing and has stationary increments with probability one.

C2. The arrival process $\{A(t)\}$ is limited by a leaky bucket. Namely, the following inequality holds with probability one for any t, s such that $0 \le s \le t$

$$\overline{A}(t,s) \equiv A(t) - A(s) \le \rho(t-s) + \sigma,$$

where ρ is the average flow rate and σ is the burst size.

It is shown in [8] that, if we put

$$\eta(t,\theta) = \log\left[1 + \frac{\rho t}{\rho t + \sigma}(e^{\theta(\rho t + \sigma)} - 1)\right], \qquad (49)$$

then, under assumptions C1 and C2, the cgf $\overline{\mathcal{A}}^{\theta}(t,s) = \log E[e^{\theta \overline{A}(t,s)}]$ of $\overline{A}(t,s)$ can be evaluated as

$$\overline{\mathcal{A}}^{\theta}(t,s) \le \eta(t-s,\theta) \quad \text{for } \theta \in (0,\infty). \qquad (50)$$

It is not difficult to see the function $\eta(t,\theta)$ is concave on t for each fixed θ and convex on θ for each fixed t.

In this network model, the cumulative arrivals to the network is given by the sum of cumulative arrivals of the L forwarding traffic flows, $A^L(t) = A_1(t) + A_2(t) + \cdots + A_L(t)$, and the cumulative services at node i for the forwarding traffic is given by

$$S_i^L(t) = \max_{0 \le \tau \le t}\left\{C_i\,\tau - A_i^{M_i,\mathrm{cross}}(\tau)\right\}, \qquad (51)$$

where $A_i^{M_i,\mathrm{cross}}(t)$ is the cumulative arrivals of the cross traffic flows in node i at time t, i.e., $A_i^{M_i,\mathrm{cross}}(t) = A_{i1}^{\mathrm{cross}}(t) + A_{i2}^{\mathrm{cross}}(t) + \cdots + A_{iM_i}^{\mathrm{cross}}(t)$. The representation (51) is derived from the relation

$$S_i^L(t) = C_i\,t - \left\{A_i^{M_i,\mathrm{cross}}(t) - Q_i^{M_i,\mathrm{cross}}(t)\right\}$$

with the backlog $Q_i^{M_i,\mathrm{cross}}(t)$ of the cross traffic in node i at time t, which is given, analogously to (2), by

$$Q_i^{M_i,\mathrm{cross}}(t) = \max_{0 \le \tau \le t}\left\{A_i^{M_i,\mathrm{cross}}(t) - A_i^{M_i,\mathrm{cross}}(\tau) - C_i(t-\tau)\right\}.$$

We denote the increment of arrivals in forwarding traffic as $\overline{A}^L(t,s) = A^L(t) - A^L(s)$. Then its cgf is given by L times of $\overline{\mathcal{A}}^{\theta}(t,s)$, i.e. $L\overline{\mathcal{A}}^{\theta}(t,s)$, since L flows $\{A_1(t)\}, \{A_2(t)\}, \cdots, \{A_L(t)\}$ are mutually independent and subjecting to a common probabilistic law. So the limit (35) is also given by $\overline{\mathcal{A}}^{\theta}(t,s)$. Similarly, we denote the increment of arrivals in the cross traffic during $(s,t]$ at node i as $\overline{A}_i^{M_i,\mathrm{cross}}(t,s) = A_i^{M_i,\mathrm{cross}}(t) - A_i^{M_i,\mathrm{cross}}(s)$. Then its cgf is given by $M_i\,\overline{\mathcal{A}}^{\theta}(t,s)$. Hence, the limit function $\overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(t,s) \equiv \lim_{L\to\infty} L^{-1}\{M_i\,\overline{\mathcal{A}}^{\theta}(t,s)\}$ is given by $\alpha_i\,\overline{\mathcal{A}}^{\theta}(t,s)$, and under assumptions C1 and C2 it satisfies the inequality

$$\overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(t,s) \le \alpha_i\,\eta(t-s,\theta) \quad \text{for } \theta \in (0,\infty). \qquad (52)$$

From (51), the increment of services $\overline{S}_i^L(t,s) = S_i^L(t) - S_i^L(s)$ during $(s,t]$ at node i is given by

$$\overline{S}_i^L(t,s) = \max_{0 \le \tau_1 \le t}\{C_i\,\tau_1 - A_i^{M_i,\mathrm{cross}}(\tau_1)\} - \max_{0 \le \tau_2 \le s}\{C_i\,\tau_2 - A_i^{M_i,\mathrm{cross}}(\tau_2)\}. \qquad (53)$$

From the independence assumption of input flows and inequalities (50) and (52), we can easily see that assumptions A1, A2 and A3 are satisfied with $\mathcal{Z}_{\overline{\mathcal{A}}(t,s_0)} = (0,\infty)$ and $\mathcal{Z}_{\overline{\mathcal{S}}_i(s_i,s_{i-1})} = (0,\infty)$. Hence the results of the preceding section can be applied with $\mathcal{Z}_{\overline{\mathcal{W}}(t,s)} = (0,\infty)$.

However, the function $\overline{S}_i^L(t,s)$ in (53) seems too complicated to evaluate $\overline{\mathcal{W}}^{\theta}(t,t)$ in (47) or $\overline{\mathcal{W}}^{\theta}(t,s)$ in (48) from the representation (39), because two maximization operations prevent us from calculating its cgf. Here we introduce two alternatives.

$$\hat{S}_i^L(t,s) = \max_{s \le \tau \le t}\left\{C_i\,(\tau - s) - \overline{A}_i^{M_i,\mathrm{cross}}(\tau,s)\right\} \quad \text{and} \qquad (54)$$

$$\check{S}_i^L(t,s) = \left[C_i\,(t-s) - \overline{A}_i^{M_i,\mathrm{cross}}(t,s)\right]^+, \qquad (55)$$

where $[x]^+ = \max\{0,x\}$. Unfortunately, they do not satisfy the incremental property. So they are difficult to understand as increments of some single variable functions representing cumulative services. However, the function $\hat{S}_i^L(t,s)$ provides the same function $\overline{\mathcal{W}}^{\theta}(t,s)$ as $\overline{S}_i^L(t,s)$ as shown below, and the function $\check{S}_i^L(t,s)$, which is naturally led from $\hat{S}_i^L(t,s)$, provides a calculable substitute for $\overline{\mathcal{W}}^{\theta}(t,s)$. Their properties stated below will be proved in Appendix.

The three functions $\overline{S}_i^L(t,s)$, $\hat{S}_i^L(t,s)$ and $\check{S}_i^L(t,s)$ satisfy the inequalities

$$\overline{S}_i^L(t,s) \le \hat{S}_i^L(t,s) \quad \text{and} \quad \hat{S}_i^L(t,s) \ge \check{S}_i^L(t,s) \qquad (56)$$

with probability one (see Appendix), and by taking a convolution with $\overline{A}_i^L(t,s)$ they can represent the cumulative departures $D_i^L(t)$ at node i as

$$D_i^L(t) = \overline{S}_i^L * \overline{A}_i^L(t,0) = \hat{S}_i^L * \overline{A}_i^L(t,0) \ge \check{S}_i^L * \overline{A}_i^L(t,0) \qquad (57)$$

(see Appendix), where $\overline{A}_i^L(t,s) = A_i^L(t) - A_i^L(s)$, as usual. Using these properties we easily see that

$$\overline{W}^L(t,s) = \overline{A}^L \oslash \hat{S}^L(t,s) \le \overline{A}^L \oslash \check{S}^L(t,s) \qquad (58)$$

with probability one and that

$$\overline{\mathcal{W}}^{\theta}(t,s) = \overline{\mathcal{A}}^{\theta} \oslash \hat{\mathcal{S}}^{\theta}(t,s) \le \overline{\mathcal{A}}^{\theta} \oslash \check{\mathcal{S}}^{\theta}(t,s) \qquad (59)$$

(see Appendix), where $\hat{S}^L(t,s)$ and $\hat{\mathcal{S}}^\theta(t,s)$ are functions defined by (31) and by (40) via (36) using $\hat{S}_i^L(t,s)$ instead of $\overline{S}_i^L(t,s)$, and $\breve{S}^L(t,s)$ and $\breve{\mathcal{S}}^\theta(t,s)$ are functions similarly defined from $\breve{S}_i^L(t,s)$. From (55), it is easily checked that

$$\begin{aligned}\breve{\mathcal{S}}_i^\theta(t,s) &\equiv -\lim_{L\to\infty} L^{-1}\log E[e^{-\theta \breve{S}_i^L(s_i,s_{i-1})}] \\ &\geq \left[\beta_i\,\theta(t-s) - \overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(t,s)\right]^+ \end{aligned} \tag{60}$$

(see Appendix). Hence, under conditions C1 and C2, from (52), it is evaluated as

$$\breve{\mathcal{S}}_i^\theta(t,s) \geq [\beta_i\,\theta(t-s) - \alpha_i\eta(t-s,\theta)]^+. \tag{61}$$

This inequality together with (50) enables us to calculate an upper bound of $\overline{\mathcal{W}}^\theta(t,s)$ from the inequality (59). In fact, the right hand side of (46) can be evaluated as

$$\begin{aligned}&\inf_{\theta\in(0,\infty)}\Big\{-\theta y + \max_{0\leq s_0\leq\cdots\leq s_n=s}\Big\{\overline{\mathcal{A}}^\theta(t,s_0) \\ &\qquad -\breve{\mathcal{S}}_n^\theta(s_n,s_{n-1}) - \cdots - \breve{\mathcal{S}}_1^\theta(s_1,s_0)\Big\}\Big\} \\ &\leq \inf_{\theta\in(0,\infty)}\Big\{-\theta y + \max_{0\leq s_0\leq\cdots\leq s_n=s}\{\eta(t-s_0,\theta) \\ &\quad -[\beta_1\,\theta(s_1-s_0) - \alpha_1\eta(s_1-s_0,\theta)]^+ - \cdots \\ &\quad \cdots - [\beta_n\,\theta(s_n-s_{n-1}) - \alpha_n\eta(s_n-s_{n-1},\theta)]^+\}\Big\}.\end{aligned} \tag{62}$$

By setting $s=t$ above, the right hand side of (45) can also be evaluated.

Discussions on numerical calculations: First we note that, for a fixed θ, the function $\varphi_i(t) \equiv \beta_i\theta t - \alpha_i\eta(t,\theta)$ is a convex function of t since $\eta(t,\theta)$ is a concave function of t. Its positive part $[\varphi_i(t)]^+ = [\beta_i\theta t - \alpha_i\eta(t,\theta)]^+$ is also a convex function. At $t=0$, $\varphi_i(0)=0$, and $\varphi_i(t)$ approaches to the line $(\beta_i-\alpha_i\rho)\theta t - \alpha_i\theta\sigma$ as t becomes large. Here the coefficient $\beta_i - \alpha_i\rho$ is interpreted as the average link capacity for the forward traffic per flow, because $C_i - M_i\rho = L(\beta_i-\alpha_i\rho)$ is the average link capacity that is offered to the forward traffic at node i in the long run. Hence, for the network to be stable, $\beta_i-\alpha_i\rho$ must be greater than ρ for any i.

Exploiting the concavity of $\eta(t,\theta)$ and the convexity of $\varphi_i(t)$, for each given θ, the maximum in the maximization of (62) can be numerically calculated by an iteration using a standard technique such as the bisection method or the Newton method. On the other hand, for given $t, s_0, \cdots, s_n(=s)$, the function to be taken infimum on θ in (62) is not convex, in general, because of the existence of operations $[\,x\,]^+$. So, there might exist multiple local minima. However, the number of such local minima is at most $2n$. So, it is not very difficult to find infimum numerically.

Consider the special case where n nodes are homogeneous and α_i and β_i are common, namely $\beta_i=\beta$ and $\alpha_i=\alpha$ for $i=1,2,\cdots,n$. From the convexity of $\varphi_i(t)$, the maximum in (62) is attained by $s_0, s_1, \cdots, s_n(=s)$ such that

$$s_1 - s_0 = s_2 - s_1 = \cdots = s_n - s_{n-1} = \frac{s-s_0}{n}.$$

Hence the right hand side of (62) is reduced to

$$\begin{aligned}\inf_{\theta\in(0,\infty)}\Big\{-\theta y + \max_{0\leq s_0\leq s}\Big\{\eta(t-s_0,\theta) \\ -\left[(s-s_0)\,\beta\,\theta - n\,\alpha\,\eta\left(\frac{s-s_0}{n},\theta\right)\right]^+\Big\}\Big\}.\end{aligned} \tag{63}$$

This inf-max problem can be easily solved numerically by using a usual numerical technique.

A numerical example: Now we apply the inequality (47) to a specific case of the network depicted in Figure 2. We consider a homogeneous case where the parameters are set as $n=3$, $L=10$, $L\alpha=50$, $L\beta=2.5$Gbps, $\rho=40$Mbps and $\sigma=4$Mbits. The link utilization in each node is 96% ($((10+50)\times 40\times 10^6)/(2.5\times 10^9) = 0.96$). The time t of $P(Q(t)>x)$ is chosen to be sufficiently large so that $P(Q(t)>x)$ indicates a stationary probability.

Figure 3 shows a numerical result calculating the exponential part of (47). The ordinate is the logarithm of (47) to base 10 and the abscissa is the buffer threshold $x(=Ly)$. The three curves with letters "1 node", "2 nodes" and "3 nodes" attached correspond to $Q_1(t)>x$, $Q_1(t)+Q_2(t)>x$ and $Q_1(t)+Q_2(t)+Q_3(t)>x$, respectively. The three perpendicular broken lines show the maximum backlogs for i nodes, $i=1,2,3$. The result shows that the maximum backlog increases as the number of nodes increases, but the probability $P(Q(t)>x)$ decreases faster than the exponential as the buffer threshold x increases in spite of the high link utilization of 96%.

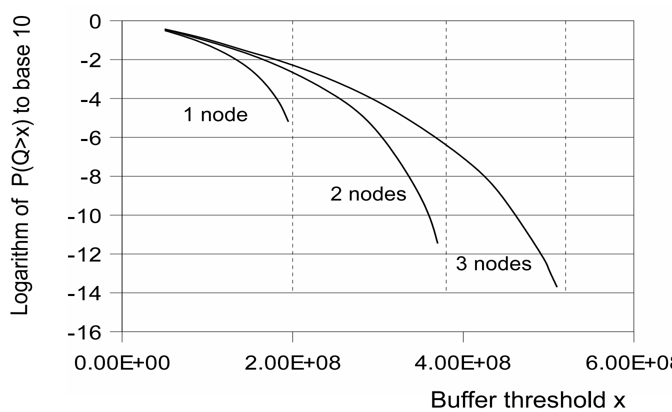

Fig. 3. Tail probability of backlog in the three-node tandem network

APPENDIX

Here we give proofs for the properties (56), (57), (58) and (59) of the functions $\hat{S}_i^L(t,s)$ and $\breve{S}_i^L(t,s)$ presented in Section 4. For brevity of notation, we set $\xi_i(\tau) = \beta_i L\tau - A_i^{M_i,\mathrm{cross}}(\tau)$ and $\overline{\xi}_i(\tau_1,\tau_2) = \xi(\tau_1) - \xi(\tau_2)$.

Proof of (56): The first inequality of (56) is proved as follows.

$$\begin{aligned}\overline{S}_i^L(t,s) &= \max_{0\le\tau_1\le t}\xi_i(\tau_1) - \max_{0\le\tau_2\le s}\xi_i(\tau_2)\\ &= \max\left[\max_{0\le\tau_1\le s}\xi_i(\tau_1), \max_{s\le\tau_1\le t}\xi(\tau_1)\right] - \max_{0\le\tau_2\le s}\xi_i(\tau_2)\\ &= \max\left[0, \max_{s\le\tau_1\le t}\xi_i(\tau_1) - \max_{0\le\tau_2\le s}\xi_i(\tau_2)\right]\\ &\le \max\left[0, \max_{s\le\tau_1\le t}\xi_i(\tau_1) - \xi_i(s)\right]\\ &= \max_{s\le\tau\le t}\overline{\xi}_i(\tau,s) = \hat{S}_i^L(t,s).\end{aligned}$$

Considering the cases $\tau = s$ and $\tau = t$ in the right hand side of (54), we have the inequality

$$\hat{S}_i^L(t,s) \ge \max\left[0, \overline{\xi}_i(t,s)\right] = \check{S}_i^L(t,s).$$

This proves the second inequality of (56).

Proof of (57): The second equality in (57) is proved as follows. From (21) and (53),

$$\begin{aligned}D_i^L(t) &= \overline{S}_i^L * \overline{A}_i^L(t,0) = \min_{0\le s\le t}\left\{A_i^L(s) + \overline{S}_i^L(t,s)\right\}\\ &= \min_{0\le s\le t}\left\{A_i^L(s) + \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0<\tau_2\le s}\xi_i(\tau_2)\right\}\\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) + \min_{0\le s\le t}\max_{0<\tau_2\le s}\left\{A_i^L(s) - \xi_i(\tau_2)\right\}\\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0\le s\le t}\max_{0<\tau_2\le s}\left\{-A_i^L(s) + \xi_i(\tau_2)\right\}\\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0<\tau_2\le t}\max_{\tau_2\le s\le t}\left\{-A_i^L(s) + \xi_i(\tau_2)\right\}\\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0<\tau_2\le t}\left\{-A_i^L(\tau_2) + \xi_i(\tau_2)\right\} \qquad (64)\\ &\qquad\qquad (\text{since } A_i^L(\tau_2) \le A_i^L(s) \text{ for } \tau_2 \le s).\end{aligned}$$

Let τ_1^* be the value of τ_1 which attains the maximum in the first term, and τ_2^* be the minimum value of τ_2 which attains the maximum in the second term. Then we can show that $\tau_2^* \le \tau_1^*$. Because, if contrary, i.e. $\tau_2^* > \tau_1^*$, then $A_i^L(\tau_2^*) \ge A_i^L(\tau_1^*)$, and hence

$$-A_i^L(\tau_2^*) + \xi_i(\tau_2^*) \ge -A_i^L(\tau_1^*) + \xi_i(\tau_1^*) \ge -A_i^L(\tau_2^*) + \xi_i(\tau_2^*).$$

The first inequality comes from the definition of τ_2^* and the second inequality comes from the definition of τ_1^*. These inequalities imply that $\tau_2 = \tau_1^*$ $(< \tau_2^*)$ also attains the maximum in the second term of (64), and it contradicts with the minimum assumption of τ_2^*. Then (64) can be rewritten as

$$\begin{aligned}D_i^L(t) &= \min_{0<\tau_2\le t}\left\{\max_{\tau_2\le\tau_1\le t}\xi_i(\tau_1) + A_i^L(\tau_2) - \xi_i(\tau_2)\right\}\\ &= \min_{0<\tau_2\le t}\left\{A_i^L(\tau_2) + \max_{\tau_2\le\tau_1\le t}\overline{\xi}_i(\tau_1,\tau_2)\right\}\\ &= \min_{0<\tau_2\le t}\left\{A_i^L(\tau_2) + \hat{S}_i^L(\tau_1,\tau_2)\right\} = \hat{S}_i^L * \overline{A}_i^L(t,0).\end{aligned}$$

This proves the second equality in (57).

The inequality in (57) is derived from the second inequality of (56) and the first inequality of (19).

Proof of (58) and (59): We denote the increment of cumulative departures during $(s,t]$ at node i as $\overline{D}_i^L(t,s) = D_i^L(t) - D_i^L(s)$, as usual. Using the second equality $D_i^L(t) = \overline{D}_i^L(t,0) = \hat{S}_i^L * \overline{A}_i^L(t,0)$ in (57) repeatedly, we can show that $D^L(t) = \hat{S}^L * \overline{A}^L(t,0)$ as in the proof of (24) in Lemma 1. Since $D^L(t) = \overline{S}^L * \overline{A}^L(t,0)$, we see that $\overline{S}^L * \overline{A}^L(t,0) = \hat{S}^L * \overline{A}^L(t,0)$. Hence from the relation (18),

$$\begin{aligned}\overline{W}(t,s) &= \overline{A}^L \oslash \overline{S}^L(t,s) = \overline{A}^L(t,0) - \overline{S}^L * \overline{A}^L(s,0)\\ &= \overline{A}^L(t,0) - \hat{S}^L * \overline{A}^L(s,0) = \overline{A}^L \oslash \hat{S}^L(t,s).\end{aligned}$$

This proves the equality of (58).

Similarly, using the inequality $D_i^L(t) \ge \check{S}_i^L * \overline{A}_i^L(t,0) = \check{S}_i^L * \overline{D}_{i-1}^L(t,0)$ in (57) repeatedly, we can show that $D^L(t) \ge \check{S}^L * \overline{A}^L(t,0)$. Hence, $\overline{S}^L * \overline{A}^L(t,0) = D^L(t) \ge \check{S}^L * \overline{A}^L(t,0)$. Then from the relation (18),

$$\begin{aligned}\overline{W}(t,s) &= \overline{A}^L \oslash \overline{S}^L(t,s) = \overline{A}^L(t,0) - \overline{S}^L * \overline{A}^L(s,0)\\ &\le \overline{A}^L(t,0) - \check{S}^L * \overline{A}^L(s,0) = \overline{A}^L \oslash \check{S}^L(t,s).\end{aligned}$$

This proves the inequality of (58).

The relation (59) is easily derived from (58).

Proof of (60): The representation (60) is derived as follows.

$$\begin{aligned}\check{\mathcal{S}}_i^\theta(t,s) &= -\lim_{L\to\infty} L^{-1}\log E[e^{-\theta\check{S}_i^L(t,s)}]\\ &= -\lim_{L\to\infty} L^{-1}\log E\left[e^{-\theta\max\left[0,\beta_i L(t-s) - \overline{A}_i^{M_i,\mathrm{cross}}(t,s)\right]}\right]\\ &\ge \max\left\{0, -\lim_{L\to\infty} L^{-1}\log E\left[e^{-\theta\beta_i L(t-s) - \theta\overline{A}_i^{M_i,\mathrm{cross}}(t,s)}\right]\right\}\\ &= \max\left\{0, \theta\beta_i(t-s) - \overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(t,s)\right\}.\end{aligned}$$

References

[1] J. Y. Le Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queueing Systems for the Internet*, On line Version of the Book Springer Verlag, LNCS 2050, 2004.

[2] A. Burchard, J. Liebeherr and S. D. Patec, A calculus for end-to-end statistical service guarantees, Technical Report CS-2001-19, University of Virginia, Computer Science Department, 2002.

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

[4] F. Ciucu, A. Burchard and J. Liebeherr, A network service curve approach for the stochastic analysis of networks, Proceedings of ACM Sigmetrics '05, 2005.

[5] R. Cruz, A calculus for network delay, parts I and II, *IEEE Transactions on Information Theory* **37**(1), 114-141, 1991.

[6] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, 1993.

[7] M. Fidler, An end-to-end probabilistic network calculus with moment generating functions, Proceedings of IEEE IWQoS 2006, 261-270, 2006.

[8] F. Kelly, Notes on effective bandwidths, In *Stochastic Networks: Theory and Applications*, Oxford University Press, 1996.

[9] C. Li, A. Burchard and J. Liebeherr, A network calculus with effective bandwidth, Technical Report CS-2003-20, University of Virginia, Computer Science Department, 2003.

[10] A. K. Parekh and R. G. Gallager, A generalized processor sharing approach to flow control in integrated service networks: the multiple node case, *IEEE/ACM Transactions Networking* **2**, 137-150, 1994.

[11] K. Kobayashi and Y. Takahashi, Overflow probability for a discrete-time queue with non-stationary multiplexed input, *Telecommunication Systems* **15** (1,2), 157-166, 2000.