Asymptotic End-to-end Stochastic Evaluation for Tandem Networks with Many Flows

Kazutomo Kobayashi
Department of Computer and
Information Sciences
Nagasaki University
Nagasaki 852-8521, Japan
kobayashi@cis.nagasaki-u.ac.jp

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

Hiroyuki Takada
Department of Computer and
Information Sciences
Nagasaki University
Nagasaki 852-8521, Japan
htakada@cis.nagasaki-u.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 our previous paper, we proposed a new stochastic network calculus for many flows from an approach like large deviations techniques, and obtained asymptotic end-to-end evaluation formulas for output burstiness and backlog. However, we could not obtain the asymptotic evaluation formula for end-to-end delay in this framework.

In this paper, we enhance the calculation in the previous paper. Concretely we enlarge the domain of the deconvolution operator. Then in addition to for the backlog and the output burstiness in tandem networks, for the delay $V^L(t)$ of L flows at time t, using *min-plus algebra*, we obtain a function of $d > 0$ by which $\limsup_{L\to\infty} L^{-1} \log P(V^L(t) > d)$ are bounded from above. We then discuss an application of the result to a tandem network with cross traffic and give a numerical result.

Categories and Subject Descriptors

G.3 [**Probability and statistics**]: Queueing theory; C.2 [**Computer communication networks**]: Miscellaneous; B.8.2 [**Performance and reliability**]: Performance analysis and design aids

General Terms

Theory, Performance

Keywords

Stochastic network calculus, Queuing theory, Large deviations techniques

VALUETOOLS 2009, October 20-22, Pisa, Italy
Copyright © 2009 ICST 978-963-9799-70-7
DOI 10.4108/ICST.VALUETOOLS2009.7613

1. INTRODUCTION

The theory of network calculus has been developed since about 1990 to give a deterministic methodology for a worst-case evaluation of packet networks [7, 16, 2, 5]. It allows us estimating the end-to-end backlog and delay bounds, and it has been used to calculate the end-to-end quality-of-service guarantees. 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*. More definitely, 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. The min-plus algebra is an useful tool, so it has been studied for various purposes [9, 15].

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 [1, 4, 5, 6, 14]. Importing statistical evaluation methods to the network calculus, it takes account of the effect of statistical multiplexing. For example, in [6], an advanced calculus is proposed to derive a probabilistic evaluation of the cumulative departures from a given stochastic arrivals and stochastic service curves.

In [13], the authors proposed another new stochastic network calculus by applying a technique used in the large deviations [8]. Large deviations theory and techniques have been used in queueing systems with many flows (see, for example, [3, 10, 12, 17]). In the paper, to derive a new network calculus, the technique is applied to 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 $L \to \infty$. In [13], we showed that for a positive θ in an

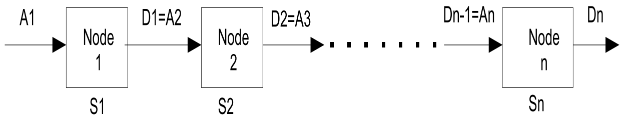

Figure 1: Tandem network with n nodes

interval

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

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 for bi-variate functions. From these results, for large L, we can do the asymptotic end-to-end evaluation for the backlog and the output burstiness. However, we cannot do the evaluation for the delay in this framework.

In this paper, we enhance the calculus in [13] to make an asymptotic evaluation possible for the delay as well as the backlog and the output burstiness. Concretely, we extend the domain of the deconvolution operator $\oslash$ from $\{(t,s)|0 \le s \le t\}$ to $\{(t,s)|t,s \ge 0\}$. This enable us to define $\overline{\mathcal{A}}^\theta \oslash \overline{\mathcal{S}}^\theta(t,s)$ for any $t,s \ge 0$, and then we obtain an evaluation formula for the end-to-end delay $V^L(t)$ at time t as

$$P(V^L(t) > d) \le \\ B_V(L,d)\exp\left(L \inf_{\theta\in\mathcal{Z}_V} \overline{\mathcal{A}}^\theta \oslash \overline{\mathcal{S}}^\theta(t-d,t)\}\right)$$

where $B_V(L,d)$ is a function varying slower than the exponential as L increases and d is a positive number.

As an example, we apply the 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 the end-to-end delay in a three node tandem network.

To be self-contained and easy to read, this paper includes the part that overlaps with the previous paper[13].

The remaining 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 and services, and the asymptotic evaluation formula for the end-to-end total queue, delay and output burstiness. 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.

2. 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 $A^{\rm net}(t)$ 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 $A^{\rm net}(t)$ and $S_i(t)$ are given ordinary (i.e., non-random) non-decreasing functions with $A^{\rm net}(0) = S_i(t) = 0$. 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 $Q_i(t)$ and $V_i(t)$ the backlog and the delay in node i at time t. Then, we have

$$\begin{aligned} A_1(t) &= A^{\rm net}(t), && (1)\\ Q_i(t) &= \max_{0\le\tau\le t}\{A_i(t)-A_i(\tau)-(S_i(t)-S_i(\tau))\} && (2)\\ D_i(t) &= A_i(t)-Q_i(t) \quad \text{and} && (3)\\ V_i(t) &= \min\{x\,|\,A_i(t-x)\le D_i(t),\, 0\le x\le t\} && (4) \end{aligned}$$

for $i=1,2,\ldots,n$, and

$$A_i(t) = D_{i-1}(t) \quad \text{for} \quad i=2,3,\ldots,n. \qquad (5)$$

The cumulative departures from the network during $(0,t]$ is given by $D^{\rm net}(t)=D_n(t)$

Equations (1) $\sim$ (5) determine functions $A_i(t)$, $D_i(t)$, $Q_i(t)$ and $V_i(t)$ for $i=1,2,\ldots,n$, uniquely. From the definitions, it is clear that $A_i(t)$, $Q_i(t)$, $D_i(t)$ and $V_i(t)$ are nonnegative with $A_i(0)=Q_i(0)=D_i(0)=V_i(0)=0$, and $A_i(t)$ and $D_i(t)$ are non-decreasing.

Combining (2) and (3), we have

$$D_i(t) = \min_{0\le\tau\le t}\{A_i(\tau)+S_i(t)-S_i(\tau)\}. \qquad (6)$$

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

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

If we set for any t and s

$$\begin{aligned} \overline{A}_i(t,s) &= A_i(t)-A_i(s), && (7)\\ \overline{S}_i(t,s) &= S_i(t)-S_i(s), \quad \text{and} && (8)\\ \overline{D}_i(t,s) &= D_i(t)-D_i(s), && (9) \end{aligned}$$

then the above inequality is rewritten as

$$\overline{D}_i(t,s) \le \max_{0\le\tau\le s}\{\overline{A}_i(t,\tau)-\overline{S}_i(s,\tau)\}. \qquad (10)$$

This is a little bit simpler expression than before. Using (7)$\sim$(9), the relations (2), (5) and (4) are also rewritten as

$$\begin{aligned} Q_i(t) &= \max_{0\le\tau\le t}\{\overline{A}_i(t,\tau)-\overline{S}_i(t,\tau)\} && (11)\\ \overline{A}_i(t,s) &= \overline{D}_{i-1}(t,s), \quad \text{and} && (12)\\ V_i(t) &= \min\{x\,|\,A_i(t-x)-D_i(t)\le 0,\, 0\le x\le t\}\\ &= \min\left\{x\,\middle|\,\max_{0\le\tau\le t}\{A_i(t-x)-A_i(\tau)\right.\\ &\qquad \left.-S_i(t)+S_i(\tau)\}\le 0,\, 0\le x\le t\right\}\\ &= \min\left\{x\,\middle|\,\max_{0\le\tau\le t}\{\overline{A}_i(t-x,\tau)-\overline{S}_i(t,\tau)\}\le 0,\right.\\ &\qquad \left. 0\le x\le t\right\} && (13) \end{aligned}$$

respectively. For the cumulative arrivals to the network and the cumulative departures from the network, we introduce

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7613
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7613

similar expressions to (7)∼(9) as

$$\begin{aligned} \overline{A}^{\rm net}(t,s) &= A^{\rm net}(t) - A^{\rm net}(s) \quad \text{and} & (14)\\ \overline{D}^{\rm net}(t,s) &= D^{\rm net}(t) - D^{\rm net}(s), & (15)\end{aligned}$$

and write the total backlogs and the total delay of the network at time t as

$$\begin{aligned} Q^{\rm net}(t) &= A^{\rm net}(t) - D^{\rm net}(t), & \\ &= \overline{A}^{\rm net}(t,0) - \overline{D}^{\rm net}(t,0) \quad \text{and} & (16)\\ V^{\rm net}(t) &= \min\{x \,|\, \overline{A}^{\rm net}(t-x,0) \le \overline{D}^{\rm net}(t,0),\ 0 \le x \le t\}. & \\ & & (17)\end{aligned}$$

To develop a new network calculus, we introduce operators $*$ and $\oslash$ for functions $f(t,s)$ and $g(t,s)$ of two variables t and s, as follows:

$$f * g(t,s) = \min_{s \le \tau \le t}\{f(t,\tau) + g(\tau,s)\} \quad \text{for } 0 \le s \le t, \tag{18}$$

and

$$f \oslash g(t,s) = \max_{0 \le \tau \le s}\{f(t,\tau) - g(s,\tau)\} \quad \text{for any } t,s \ge 0. \tag{19}$$

Note that the domains of the two operators are different. We call $*$ the *convolution* operator and $\oslash$ the *deconvolution* one since these definitions are similar to those of the convolution and deconvolution ones used in [1, 2, 4, 5, 6, 7, 14]. In the previous paper [13], we defined another deconvolution operator on the same domain as the convolution operator. Here, to deal with the delay, we enlarge the domain and this requires some changes in our discussions.

The convolution operator $*$ is associative in the sense that for three functions $f(t,s)$, $g(t,s)$ and $h(t,s)$

$$(f * g) * h(t,s) = f * (g * h)(t,s).$$

Hence, $(f * g) * h(t,s)$ or $f * (g * h)(t,s)$ can be written as $f * g * h(t,s)$. If the function $f(t,s)$ has an *incremental property*, i.e. it is written as $f(t,s) = f(t,0) - f(s,0)$ for any $t,s \ge 0$, then

$$\begin{aligned} f(t,0) - g * f(s,0) &= f(t,0) - \min_{0 \le \tau \le s}\{g(s,\tau) + f(\tau,0)\} & \\ &= \max_{0 \le \tau \le s}\{f(t,\tau) - g(s,\tau)\} & \\ &= f \oslash g(t,s). & (20)\end{aligned}$$

Using these operators, we can rewrite (6), (11), (13) and (10) as

$$\begin{aligned} D_i(t) &= \overline{D}_i(t,0) = \overline{S}_i * \overline{A}_i(t,0), & (21)\\ Q_i(t) &= \overline{A}_i \oslash \overline{S}_i(t,t), & (22)\\ V_i(t) &= \min\{x \,|\, \overline{A}_i \oslash \overline{S}_i(t-x,t) \le 0,\ 0 \le x \le t\} & (23)\end{aligned}$$

and

$$\overline{D}_i(t,s) \le \overline{A}_i \oslash \overline{S}_i(t,s), \tag{24}$$

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 t and s such that $0 \le s \le t$, we have*

$$\begin{aligned} D^{\rm net}(t) &= \overline{D}^{\rm net}(t,0) = \overline{S}^{\rm net} * \overline{A}^{\rm net}(t,0), & (25)\\ Q^{\rm net}(t) &= \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t,t), & (26)\\ V^{\rm net}(t) &= \min\{x \,|\, \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t-x,t) \le 0,\ 0 \le x \le t\}, & \\ & & (27)\end{aligned}$$

and

$$\overline{D}^{\rm net}(t,s) \le \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t,s), \tag{28}$$

where

$$\begin{aligned} \overline{S}^{\rm net}(t,s) &= \overline{S}_n * \overline{S}_{n-1} * \cdots * \overline{S}_1(t,s) & (29)\\ &= \min_{s=s_0 \le s_1 \le \cdots \le s_n = t}\{\overline{S}_n(s_n,s_{n-1}) + \cdots + \overline{S}_1(s_1,s_0)\}. & \\ & & (30)\end{aligned}$$

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 (12)

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

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 (25). Applying the above representation to (16) and (17), and using relation (20), we have

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

and

$$\begin{aligned} V^{\rm net}(t) &= \min\{x \,|\, \overline{A}^{\rm net}(t-x,0) - \overline{S}^{\rm net} * \overline{A}^{\rm net}(t,0) \le 0,\\ &\qquad\qquad 0 \le x \le t\}\\ &= \min\{x \,|\, \overline{A}^{\rm net} \oslash \overline{S}^{\rm net}(t-x,t) \le 0,\ 0 \le x \le t\}.\end{aligned}$$

This proves (26) and (27). Using the inequality $D^{\rm net}(t) \le A^{\rm net}(t)$ and relations (25) and (20), 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 (28). The representation (30) is a direct consequence of the definition of the convolution operator (18). □

From the relation (25), $\overline{S}^{\rm net}(t,s)$ 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.

3. 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. Furthermore, let $D^L(t)$ be the total departures from the network of L flows during $(0,t]$, and $Q^L(t)$ and $V^L(t)$ the total backlog and the total delay of L flows in the network at time t. Sample paths of the processes $\{A^L(t)\}$, $\{S_i^L(t)\}$,

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7613
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7613

$\{Q^L(t)\}$, $\{V^L(t)\}$ and $\{D^L(t)\}$ correspond to $A^{\rm net}(t)$, $S_i(t)$, $Q^{\rm net}(t)$, $V^{\rm net}(t)$ and $D^{\rm net}(t)$, respectively, in the previous section.

For a pair of times t and s, we introduce bi-variate functions $\overline{A}^L(t,s)$, $\overline{S}_i^L(t,s)$ and $\overline{D}^L(t,s)$ as in (14), (8) and (15), namely 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

$$\begin{aligned} Q^L(t) &= \overline{A}^L \oslash \overline{S}^L(t,t), && (31)\\ V^L(t) &= \min\{x \,|\, \overline{A}^L \oslash \overline{S}^L(t-x,t) \le 0,\ 0 \le x \le t\} && (32) \end{aligned}$$

and

$$\overline{D}^L(t,s) \le \overline{A}^L \oslash \overline{S}^L(t,s) \qquad (33)$$

with probability one, where

$$\overline{S}^L(t,s) = \overline{S}_n^L * \cdots * \overline{S}_1^L(t,s) \quad \text{for } 0 \le s \le t. \qquad (34)$$

We let

$$\overline{W}^L(t,s) = \overline{A}^L \oslash \overline{S}^L(t,s). \qquad (35)$$

Then (31), (32) and (33) are rewritten as

$$\begin{aligned} Q^L(t) &= \overline{W}^L(t,t), && (36)\\ V^L(t) &= \min\{x \,|\, \overline{W}^L(t-x,t) \le 0,\ 0 \le x \le t\} \quad \text{and} && (37)\\ \overline{D}^L(t,s) &\le \overline{W}^L(t,s). && (38) \end{aligned}$$

Note that the random variable $\overline{W}^L(t-x,t)$ is non-increasing as a function of x. Then, if $\overline{W}^L(t-d,t) > 0$ then all x satisfying $\overline{W}^L(t-x,t) \le 0$, $0 \le x \le t$, is greater than d, namely, $V^L(t) > d$. If $V^L(t) > d$ then $\overline{W}^L(t-d,t) > 0$ since if $\overline{W}^L(t-d,t) \le 0$ then $V^L(t) \le d$ from the definition. Hence, for any t and d such that $0 < d \le t$, $V^L(t) > d$ if and only if $\overline{W}^L(t-d,t) > 0$. Thus, from (36), (37) and (38), we have the following relations for $y, d > 0$:

$$\begin{aligned} P(Q^L(t) > Ly) &= P(\overline{W}^L(t,t) > Ly), && (39)\\ P(V^L(t) > d) &= P(\overline{W}^L(t-d,t) > 0) \quad \text{and} && (40)\\ P(\overline{D}^L(t,s) > Ly) &\le P(\overline{W}^L(t,s) > Ly). && (41) \end{aligned}$$

From the definitions (18) and (19) of the convolution and deconvolution operators, $\overline{W}^L(t,s)$ itself is written as

$$\begin{aligned} &\overline{W}^L(t,s)\\ &= \max_{0 \le s_0 \le s} \Big\{ \overline{A}^L(t,s_0)\\ &\qquad - \min_{s_0 \le s_1 \le \cdots \le s_n = s} \{\overline{S}_n^L(s_n,s_{n-1}) + \cdots + \overline{S}_1^L(s_1,s_0)\} \Big\}\\ &= \max_{0 \le s_0 \le s_1 \le \cdots \le s_n = s} \{\overline{A}^L(t,s_0)\\ &\qquad - \overline{S}_n^L(s_n,s_{n-1}) - \cdots - \overline{S}_1^L(s_1,s_0)\}. \qquad (42) \end{aligned}$$

Hereafter, in this section, we regard times t and s are arbitrarily chosen so that $t, s \ge 0$ 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

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

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

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

$$\begin{aligned} \overline{\mathcal{A}}^\theta(t,s_0) &= \lim_{L\to\infty} L^{-1}\log E[e^{\theta \overline{A}^L(t,s_0)}] \quad \text{and} && (43)\\ \overline{\mathcal{S}}_i^\theta(s_i,s_{i-1}) &= -\lim_{L\to\infty} L^{-1}\log E[e^{-\theta \overline{S}_i^L(s_i,s_{i-1})}],\\ &\qquad i = 1, \cdots, n. && (44) \end{aligned}$$

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

$$\mathcal{Z}_{\overline{\mathcal{W}}(t,s)} = \bigcap_{0 \le s_0 \le \cdots \le s_n = s} \mathcal{Z}_{\overline{\mathcal{A}}(t,s_0)} \bigcap_{1 \le i \le n} \mathcal{Z}_{\overline{\mathcal{S}}_i(s_i,s_{i-1})} \qquad (45)$$

is a non-empty interval, too.

We have the following lemma.

Lemma 2. *Under assumptions A1, A2 and A3, the sequence*

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

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*

$$\begin{aligned} \overline{\mathcal{W}}^\theta(t,s) &= \overline{\mathcal{A}}^\theta \oslash \overline{\mathcal{S}}^\theta(t,s), && (46)\\ &= \max_{0 \le s_0 \le \cdots \le s_n = s} \{\overline{\mathcal{A}}^\theta(t,s_0)\\ &\qquad - \overline{\mathcal{S}}_n^\theta(s_n,s_{n-1}) - \cdots - \overline{\mathcal{S}}_1^\theta(s_1,s_0)\} && (47) \end{aligned}$$

where

$$\overline{\mathcal{S}}^\theta(t,s) = \overline{\mathcal{S}}_n^\theta * \overline{\mathcal{S}}_{n-1}^\theta * \cdots * \overline{\mathcal{S}}_1^\theta(t,s). \qquad (48)$$

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 (47) and the inferior limit of it is bounded from below by the same quantity.

Since $\theta > 0$, from (42), using the monotonicity of the exponential function and the inequality $\max(x_1, x_2) \leq x_1 + x_2$ for $x_1, x_2 \geq 0$, we have

$$\begin{aligned} & e^{\theta \overline{W}^L(t,s)} \\ & = \max_{0 \leq s_0 \leq \cdots \leq s_n = s} e^{\theta(\overline{A}^L(t,s_0) - \overline{S}^L_n(s_n,s_{n-1}) - \cdots - \overline{S}^L_1(s_1,s_0))} \\ & \leq \sum_{0 \leq s_0 \leq \cdots \leq s_n = s} e^{\theta(\overline{A}^L(t,s_0) - \overline{S}^L_n(s_n,s_{n-1}) - \cdots - \overline{S}^L_1(s_1,s_0))} \end{aligned}$$

with probability one. Taking expectation on both sides,

$$\begin{aligned} & E[e^{\theta \overline{W}^L(t,s)}] \\ & \leq \sum_{0 \leq s_0 \leq \cdots \leq s_n = s} E[e^{\theta(\overline{A}^L(t,s_0) - \overline{S}^L_n(s_n,s_{n-1}) - \cdots - \overline{S}^L_1(s_1,s_0))}] \\ & \leq (s+1)^n \cdot \\ & \quad \max_{0 \leq s_0 \leq \cdots \leq s_n = s} E[e^{\theta(\overline{A}^L(t,s_0) - \overline{S}^L_n(s_n,s_{n-1}) - \cdots - \overline{S}^L_1(s_1,s_0))}]. \end{aligned}$$

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

$$E[e^{\theta \overline{A}^L(t,s_0)}]\, E[e^{-\theta \overline{S}^L_n(s_n,s_{n-1})}] \cdots E[e^{-\theta \overline{S}^L_1(s_1,s_0))}].$$

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

$$\begin{aligned} L^{-1} \log E[e^{\theta \overline{W}^L(t,s)}] \leq L^{-1} \log(s+1)^n \\ + \max_{0 \leq s_0 \leq \cdots \leq s_n = s} \Big\{ L^{-1} \log E[e^{\theta(\overline{A}^L(t,s_0)}] \\ + L^{-1} \log E[e^{-\theta(\overline{S}^L_n(s_n,s_{n-1})}] + \cdots \\ + L^{-1} \log E[e^{-\theta(\overline{S}^L_1(s_1,s_0)}] \Big\}. \end{aligned}$$

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

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

On the other hand, from (42), for arbitrarily chosen $(s_0, s_1, \cdots, s_n)$, we have

$$\overline{W}^L(t,s) \geq \overline{A}^L(t,s_0) - \overline{S}^L_n(s_n,s_{n-1}) - \cdots - \overline{S}^L_1(s_1,s_0)$$

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

$$E[e^{\theta \overline{W}^L(t,s)}] \geq E[e^{\theta(\overline{A}^L(t,s_0) - \overline{S}^L_n(s_n,s_{n-1}) - \cdots - \overline{S}^L_1(s_1,s_0))}].$$

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

$$\begin{aligned} L^{-1} \log E[e^{\theta \overline{W}^L(t,s)}] \geq L^{-1} \log E[e^{\theta \overline{A}^L(t,s_0)}] \\ + L^{-1} \log E[e^{-\theta \overline{S}^L_n(s_n,s_{n-1})}] + \cdots \\ + L^{-1} \log E[e^{-\theta \overline{S}^L_1(s_1,s_0)}]. \end{aligned}$$

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

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

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

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

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 (47). This proves (47). The representation (46) is led from (47) in the reverse way of (42). □

Theorem 1. *Under assumptions A1, A2 and A3 for $s = t > 0$, we have, for $y > 0$,*

$$\begin{aligned} & \limsup_{L \to \infty} L^{-1} \log P(Q^L(t) > Ly) \\ & \qquad \leq \inf_{\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,t)}} \{ -\theta y + \overline{\mathcal{W}}^\theta(t,t) \}. \qquad (49) \end{aligned}$$

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

$$\begin{aligned} & \limsup_{L \to \infty} L^{-1} \log P(V^L(t) > d) \\ & \qquad \leq \inf_{\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t-d,t)}} \overline{\mathcal{W}}^\theta(t-d,t). \qquad (50) \end{aligned}$$

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

$$\begin{aligned} & \limsup_{L \to \infty} L^{-1} \log P(\overline{D}^L(t,s) > Ly) \\ & \qquad \leq \inf_{\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,s)}} \{ -\theta y + \overline{\mathcal{W}}^\theta(t,s) \}. \qquad (51) \end{aligned}$$

Proof. We apply the Chernoff's bound (or the Markov inequality, see p.240 of [5]) to the tail probability $P(\overline{W}^L(t,s) > Ly)$. For any $\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,s)}$ and $y \geq 0$, we have

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

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

$$\begin{aligned} & \limsup_{L \to \infty} L^{-1} \log P(\overline{W}^L(t,s) > Ly) \\ & \qquad \leq \inf_{\theta \in \mathcal{Z}_{\overline{\mathcal{W}}(t,s)}} \{ -\theta y + \overline{\mathcal{W}}^\theta(t,s) \}. \qquad (52) \end{aligned}$$

We know that $P(Q^L(t) > Ly) = P(\overline{W}^L(t,t) > Ly)$ and $P(V^L(t) > d) = P(\overline{W}^L(t-d,t) > 0)$ from (39) and (40). Hence, if considering the case of $\overline{W}^L(t,t)$, the inequality (52) proves (49), and if considering the case of $\overline{W}^L(t-d,t)$ and $y = 0$, the inequality (52) proves (50). Similarly, from (41) we see that $P(\overline{D}^L(t,s) > Ly) \leq P(\overline{W}^L(t,s) > Ly)$. Hence the inequality (52) also implies (51). □

Theorem 1 suggests that the tail probabilities of $Q^L(t)$, $V^L(t)$ and $\overline{D}^L(t,s)$ might be evaluated by the form

$$P(Q^L(t) > Ly) \leq \\ B_Q(L,y)\exp\left(L \inf_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t,t)}} \{-\theta y + \overline{\mathcal{W}}^\theta(t,t)\}\right), \quad (53)$$

$$P(V^L(t) > d) \leq \\ B_V(L,d)\exp\left(L \inf_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t-d,t)}} \overline{\mathcal{W}}^\theta(t-d,t)\}\right) \quad (54)$$

and

$$P(\overline{D}^L(t,s) > Ly) \leq \\ B_D(L,y)\exp\left(L \inf_{\theta\in\mathcal{Z}_{\overline{\mathcal{W}}(t,s)}} \{-\theta y + \overline{\mathcal{W}}^\theta(t,s)\}\right), \quad (55)$$

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

4. APPLICATION TO A NETWORK WITH CROSS TRAFFIC

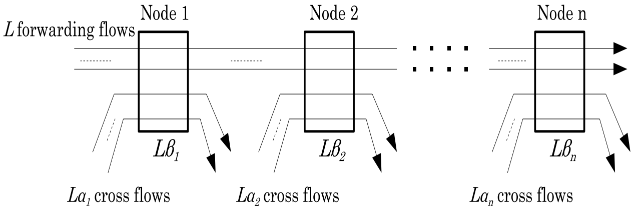

Figure 2: Tandem network with cross traffic

We apply the bound (54) 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 $\{A_{i1}^{\text{cross}}(t)\}, \{A_{i2}^{\text{cross}}(t)\}, \cdots, \{A_{iM_i}^{\text{cross}}(t)\}$. 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 subjecting 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 a greedy process which is limited by a leaky bucket, namely, the cgf $\mathcal{A}^\theta(t) = \log E[e^{\theta A(t)}]$ of $A(t)$ is given by

$$\eta^+(t,\theta) \equiv \log\left[1 + \frac{\rho t}{\rho t + \sigma}(e^{\theta(\rho t+\sigma)} - 1)\right], \quad (56)$$

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

If the arrival process $\{A(t)\}$ is just limited by a leaky bucket, namely, $\overline{A}(t,s) \equiv A(t) - A(s) \leq \rho(t-s) + \sigma$ holds with probability one for any t, s such that $0 \leq s \leq t$, then the cgf $\mathcal{A}^\theta(t)$ of $\{A(t)\}$ satisfies the inequality $\mathcal{A}^\theta(t) \leq \eta^+(t,\theta)$ from [11]. The assumption C2 is that the equation $\mathcal{A}^\theta(t) = \eta^+(t,\theta)$ holds.

Under assumptions C1 and C2, the cgf $\overline{\mathcal{A}}^\theta(t,s)$ of $\overline{A}(t,s)$, is given as

$$\overline{\mathcal{A}}^\theta(t,s) = \eta(t-s) \quad \text{for } \theta \in (0,\infty), \quad (57)$$

where

$$\eta(t) = \begin{cases} \eta^+(t,\theta) & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ -\eta^+(-t,\theta) & \text{if } t < 0. \end{cases} \quad (58)$$

The function $\eta^+(t,\theta)$ is concave on t for each fixed θ and convex on θ for each fixed t (see Appendix).

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

$$S_i^L(t) = \max_{0\leq\tau\leq t}\left\{C_i\,\tau - A_i^{M_i,\text{cross}}(\tau)\right\}, \quad (59)$$

where $A_i^{M_i,\text{cross}}(t)$ is the cumulative arrivals of the cross traffic flows in node i at time t, i.e., $A_i^{M_i,\text{cross}}(t) = A_{i1}^{\text{cross}}(t) + A_{i2}^{\text{cross}}(t) + \cdots + A_{iM_i}^{\text{cross}}(t)$. The representation (59) is derived from the relation

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

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

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

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 (43) 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 $\overline{A}_i^{M_i,\text{cross}}(t,s) = A_i^{M_i,\text{cross}}(t)–A_i^{M_i,\text{cross}}(s)$. Then its cgf is given by $M_i\,\overline{\mathcal{A}}^\theta(t,s)$. Hence, the function $\overline{\mathcal{A}}_i^{\theta,\text{cross}}(t,s) \equiv \lim_{L\to\infty} L^{-1}\{M_i\,\overline{\mathcal{A}}^\theta(t,s)\}$ is given by $\alpha_i\,\overline{\mathcal{A}}^\theta(t,s)$, and then

$$\overline{\mathcal{A}}_i^{\theta,\text{cross}}(t,s) = \alpha_i\,\eta(t-s,\theta) \quad \text{for } \theta \in (0,\infty). \quad (60)$$

From (59), 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

$$\overline{S}_i^L(t,s) = \max_{0\leq\tau_1\leq t}\{C_i\,\tau_1 - A_i^{M_i,\text{cross}}(\tau_1)\} \\ - \max_{0\leq\tau_2\leq s}\{C_i\,\tau_2 - A_i^{M_i,\text{cross}}(\tau_2)\}. \quad (61)$$

From the independence assumption of input flows and the equations (57) and (60), we can easily see that assumptions

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7613
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7613

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 (61) seems too complicated to evaluate $\overline{\mathcal{W}}^\theta(t-d,t)$ in (54), because two maximization operations prevent us from calculating its cgf. Here we introduce two alternatives.

$$\begin{aligned} \hat{S}_i^L(t,s) &= \max_{s\le\tau\le t}\left\{C_i\,(\tau-s)-\overline{A}_i^{M_i,\mathrm{cross}}(\tau,s)\right\} \text{ and} && (62)\\ \breve{S}_i^L(t,s) &= \left[C_i\,(t-s)-\overline{A}_i^{M_i,\mathrm{cross}}(t,s)\right]^+, && (63)\end{aligned}$$

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 $\breve{S}_i^L(t,s)$, 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 $\breve{S}_i^L(t,s)$ satisfy the inequalities

$$\overline{S}_i^L(t,s) \le \hat{S}_i^L(t,s) \quad\text{and}\quad \hat{S}_i^L(t,s) \ge \breve{S}_i^L(t,s) \qquad (64)$$

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

$$D_i^L(t) = \overline{S}_i^L * \overline{A}_i^L(t,0) = \hat{S}_i^L * \overline{A}_i^L(t,0) \ge \breve{S}_i^L * \overline{A}_i^L(t,0) \quad (65)$$

(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

$$\overline{W}^L(t,s) = \overline{A}^L \oslash \hat{S}^L(t,s) \le \overline{A}^L \oslash \breve{S}^L(t,s) \qquad (66)$$

with probability one and that

$$\overline{\mathcal{W}}^\theta(t,s) = \overline{\mathcal{A}}^\theta \oslash \hat{\mathcal{S}}^\theta(t,s) \le \overline{\mathcal{A}}^\theta \oslash \breve{\mathcal{S}}^\theta(t,s) \qquad (67)$$

(see Appendix), where $\hat{S}^L(t,s)$ and $\hat{\mathcal{S}}^\theta(t,s)$ are functions defined by (34) and by (48) via (44) 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 (63), it is easily checked that

$$\begin{aligned} \breve{\mathcal{S}}_i^\theta(t,s) &\equiv -\lim_{L\to\infty} L^{-1}\log E[e^{-\theta \breve{S}_i^L(s_i,s_{i-1})}] \\ &\ge \left[\beta_i\,\theta(t-s) - \overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(t,s)\right]^+ && (68)\end{aligned}$$

(see Appendix). Hence, under conditions C1 and C2, from (60), it is given as

$$\breve{\mathcal{S}}_i^\theta(t,s) \ge [\beta_i\,\theta(t-s) - \alpha_i\eta(t-s,\theta)]^+. \qquad (69)$$

This inequality together with (57) enables us to calculate an upper bound of $\overline{\mathcal{W}}^\theta(t,s)$ from the inequality (67). In fact, the right hand side of (51) can be evaluated as

$$\begin{aligned} &\inf_{\theta\in(0,\infty)}\Big\{-\theta y + \max_{0\le s_0\le\cdots\le s_n=s}\Big\{\overline{\mathcal{A}}^\theta(t,s_0) \\ &\qquad -\breve{\mathcal{S}}_n^\theta(s_n,s_{n-1}) - \cdots - \breve{\mathcal{S}}_1^\theta(s_1,s_0)\Big\}\Big\} \\ &\le \inf_{\theta\in(0,\infty)}\Big\{-\theta y + \max_{0\le s_0\le\cdots\le s_n=s}\{\eta(t-s_0,\theta) \\ &\qquad -[\,\beta_1\,\theta(s_1-s_0) - \alpha_1\eta(s_1-s_0,\theta)]^+ - \cdots \\ &\qquad \cdots - [\,\beta_n\,\theta(s_n-s_{n-1}) - \alpha_n\eta(s_n-s_{n-1},\theta)]^+\}\Big\}. \end{aligned} \qquad (70)$$

Discussions on numerical calculations: First we note that, for a fixed θ and $t>0$, 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 $t>0$ and each given θ, the maximum in the maximization of (70) can be numerically calculated by an iteration using a usual numerical 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 (70) 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 (70) 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 (70) is reduced to

$$\begin{aligned} &\inf_{\theta\in(0,\infty)}\left\{-\theta y + \max_{0\le s_0\le t}\left\{\eta(t-s_0,\theta)\right.\right. \\ &\qquad \left.\left. -\left[(s-s_0)\,\beta\,\theta - n\,\alpha\,\eta\left(\frac{s-s_0}{n},\theta\right)\right]^+\right\}\right\}. \end{aligned} \qquad (71)$$

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

When we evaluate the total delay by (50) and (67), the above formula is further reduced to

$$\begin{aligned} &\inf_{\theta\in(0,\infty)}\left\{\max_{0\le s_0\le t}\left\{\eta(t-d-s_0,\theta)\right.\right. \\ &\qquad \left.\left. -\left[(t-s_0)\,\beta\,\theta - n\,\alpha\,\eta\left(\frac{t-s_0}{n},\theta\right)\right]^+\right\}\right\}. \end{aligned} \qquad (72)$$

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7613
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7613

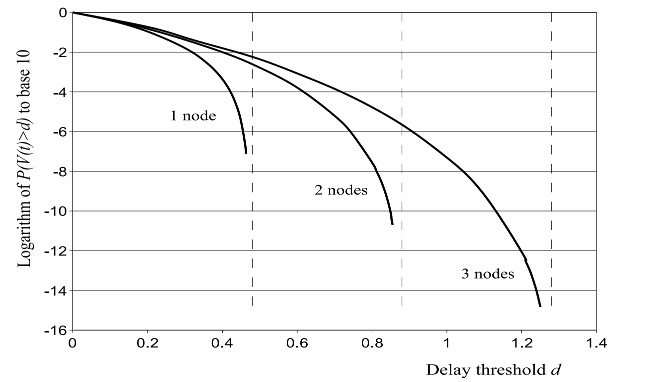

Figure 3: The upper bound of the tail probability of delay in the three-node tandem network

A numerical example: Now we apply the inequality (54) 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 $P(V(t) > d)$ is chosen to be sufficiently large so that $P(V(t) > d)$ indicates a stationary probability.

Figure 3 shows a numerical result calculating the exponential part of (54). The ordinate is the logarithm of (54) to base 10 and the abscissa is the delay threshold d. The three curves with letters "1 node", "2 nodes" and "3 nodes" attached correspond to the delay distribution through one node, that through two nodes and that through three nodes, respectively. The three perpendicular broken lines show the maximum delays for i nodes, $i = 1, 2, 3$ under the restriction by a leaky bucket.

In practical applications, the evaluation (54) may not be accurate in two reasons. The first is, as is usually seen in asymptotic evaluations, that the coefficient function $B_Q(L, y)$ is not known. We only know the exponential part. The second is that our model uses the greedy process which is limited by a leaky bucket as the input process. Usually, it is only known that the input process is limited by the leaky bucket, and the input process is not identified. So, the new findings we can obtain from the numerical results are somehow limited. However, at least, we can point out the following facts. In the greedy input case, the maximum delay increases as the number of nodes increases, but the upper bound of the probability $P(V(t) > d)$ decreases faster than the exponential as the delay threshold d increases in spite of the high link utilization of 96%.

5. REFERENCES

[1] R. Boorstyn, A. Burchard, J. Liebeherr and C. Oottamakorn. Statistical multiplexing gain of link scheduling algorithms in QoS networks (short version). Technical Report CS-99-23, University of Virginia, Computer Science Department, 1999.

[2] 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.

[3] D. D. Botvich and N. G. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems*, 20:293-320, 1995.

[4] 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.

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

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

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

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

[9] M. Draief, J. Mairesse and N. O'Connell. Queues, stores, and tableaux. *Journal of Applied Probability*, 42:1145-1167, 2005.

[10] A. Ganesh, N. O'Connell and D. Wischik. *Big Queues.* Springer, 2004.

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

[12] 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.

[13] K. Kobayashi, Y. Takahashi and H. Takada. *A Stochastic network calculus for many flow.* In *Proceedings of ITC21*, to appear.

[14] 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.

[15] N. O'Connel. A path-transformation for random walks and the Robinson-Schensted correspondence. *Transactions of the American Mathematical Society*, 355(9):3669-3697, 2003.

[16] 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.

[17] C. Walsh. Maximal effective bandwidth of constrained traffic. *Queueing Systems*, 44:161-182, 2003.

APPENDIX

Here we give a proof for the concavity of the function $\eta^+(t, \theta)$ given in (56) and proofs for the properties (64), (65), (66) and (67) of the functions $\hat{S}_i^L(t, s)$ and $\check{S}_i^L(t, s)$ presented in Section 4. For brevity of notation, we set $\xi_i(\tau) = \beta_i L\tau - A_i^{M_i,\text{cross}}(\tau)$ and $\bar{\xi}_i(\tau_1, \tau_2) = \xi(\tau_1) - \xi(\tau_2)$.

Proof of the concavity of $\eta^+(t, \theta)$ in (56): The function $\eta^+(t, \theta)$ is rewritten as

$$\eta^+(t, \theta) \;=\; \theta(\rho t + \sigma) - \log\left[\rho t + \sigma\right] + \log\left[\rho t + \sigma e^{-\theta(\rho t+\sigma)}\right].$$

Then it is easily checked that $\eta^+(0, \theta) = 0$ and $\lim_{t\to\infty} \left\{\eta^+(t, \theta) - \theta(\rho t + \sigma)\right\} = 0$. So, roughly speaking, the func-

Digital Object Identifier: 10.4108/ICST.VALUETOOLS2009.7613
http://dx.doi.org/10.4108/ICST.VALUETOOLS2009.7613

tion $\eta^+(t,\theta)$ grows along a straight line $\theta(\rho t+\sigma)$ as $t \to \infty$. Its first and second derivatives on t are given as

$$\begin{aligned}\frac{\partial}{\partial t}\eta^+(t,\theta) &= \theta\rho - \frac{\rho}{\rho t+\sigma} + \frac{\rho - \sigma\theta\rho e^{-\theta(\rho t+\sigma)}}{\rho t + \sigma e^{-\theta(\rho t+\sigma)}}, \quad \text{and} \\ \frac{\partial^2}{\partial^2 t}\eta^+(t,\theta) &= -\frac{\rho^2}{(\rho t+\sigma)^2(\rho t+\sigma e^{-\theta(\rho t+\sigma)})^2} \cdot \\ &\left[\sigma^2\left\{1-2\theta(\rho t+\sigma)e^{-\theta(\rho t+\sigma)} - e^{-2\theta(\rho t+\sigma)}\right\}\right. \\ &+2\sigma\rho t\left\{1-e^{-\theta(\rho t+\sigma)} - \theta(\rho t+\sigma)e^{-\theta(\rho t+\sigma)}\right. \\ &\left.\left.-\frac{1}{2}\theta^2(\rho t+\sigma)^2e^{-\theta(\rho t+\sigma)}\right\}\right].\end{aligned}$$

It is easily checked that the first derivative is positive and the second derivative is negative for $t > 0$, because $h_1(x) \equiv 1 - 2xe^{-x} - e^{-2x} > 0$ and $h_2(x) \equiv e^x - 1 - x - \frac{1}{2}x^2 > 0$ for $x > 0$. Hence, as a function of t, $\eta(t,\theta)$ is increasing and concave. On the other hand, the first derivative on θ is given by

$$\frac{\partial}{\partial\theta}\eta^+(t,\theta) = \frac{\rho t(\rho t+\sigma)}{\rho t+\sigma e^{-\theta(\rho t+\sigma)}}.$$

It is positive and increasing. Hence, as a function of θ, $\eta^+(t,\theta)$ is increasing and convex.

Proof of (64): The first inequality of (64) 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 (62), we have the inequality

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

This proves the second inequality of (64).

Proof of (65): The second equality in (65) is proved as follows. From (21) and (61),

$$\begin{aligned}D_i^L(t) = \overline{S}_i^L * \overline{A}_i^L(t,0) &= \min_{0\le s\le t}\left\{A_i^L(s) + \overline{S}_i^L(t,s)\right\} \\ &= \min_{0\le s\le t}\left\{A_i^L(s) + \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0<\tau_2\le s}\xi_i(\tau_2)\right\} \\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) + \min_{0\le s\le t}\max_{0<\tau_2\le s}\left\{A_i^L(s) - \xi_i(\tau_2)\right\} \\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0\le s\le t}\max_{0<\tau_2\le s}\left\{-A_i^L(s) + \xi_i(\tau_2)\right\} \\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0<\tau_2\le t}\max_{\tau_2\le s\le t}\left\{-A_i^L(s) + \xi_i(\tau_2)\right\} \\ &= \max_{0<\tau_1\le t}\xi_i(\tau_1) - \max_{0<\tau_2\le t}\left\{-A_i^L(\tau_2) + \xi_i(\tau_2)\right\} \qquad (73) \\ &\qquad (\text{since } A_i^L(\tau_2) \le A_i^L(s) \text{ for } \tau_2 \le s).\end{aligned}$$

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 (73), and it contradicts with the minimum assumption of τ_2^*. Then (73) 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 (65).

The inequality in (65) is derived from the second inequality of (64) and the property of the convolution operator.

Proof of (66) and (67): 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 (65) repeatedly, we can show that $D^L(t) = \hat{S}^L * \overline{A}^L(t,0)$ as in the proof of (25) 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 (20),

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

This proves the equality of (66).

Similarly, using the inequality $D_i^L(t) \ge \breve{S}_i^L * \overline{A}_i^L(t,0) = \breve{S}_i^L * \overline{D}_{i-1}^L(t,0)$ in (65) repeatedly, we can show that $D^L(t) \ge \breve{S}^L * \overline{A}^L(t,0)$. Hence, $\overline{S}^L * \overline{A}^L(t,0) = D^L(t) \ge \breve{S}^L * \overline{A}^L(t,0)$. Then from the relation (20),

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

This proves the inequality of (66).

The relation (67) is easily derived from (66).

Proof of (68): The representation (68) is derived as follows.

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