PAPER

A Backlog Evaluation Formula for Admission Control Based on the Stochastic Network Calculus with Many Flows*

Kazutomo KOBAYASHI[†a)], **Yukio TAKAHASHI**[††], *and* **Hiroyuki TAKADA**[†b)], *Members*

SUMMARY Admission control is a procedure to guarantee a given level of Quality of Service (QoS) by accepting or rejecting arrival connection requests. There are many studies on backlog or loss rate evaluation formulas for admission control at a single node. However, there are few studies on end-to-end evaluation formulas suitable for admission control. In a previous paper, the authors proposed a new stochastic network calculus for many flows using an approach taken from large deviations techniques and obtained asymptotic end-to-end evaluation formulas for output burstiness and backlog. In this paper, we apply this stochastic network calculus to a heterogeneous tandem network with many forwarding flows and cross traffic flows constrained by leaky buckets, and obtain a simple evaluation formula for the end-to-end backlog. In this formula, the end-to-end backlog can be evaluated by the traffic load at the bottle neck node. This result leads us to a natural extension of the evaluation formula for a single node.

key words: *admission control, quality-of-service (QoS), stochastic network calculus*

1. Introduction

Admission control is a procedure to guarantee a given level of Quality of Service (QoS) by accepting or rejecting arrival connection requests. To judge whether an arriving connection request is accepted or rejected, some simple evaluation formula is needed, which can evaluate QoS under given traffic loads including that of the new connection. There are many studies on evaluation formulas for backlog or loss rate at a single node [1], however, studies on end-to-end evaluation have been almost limited to ones by *network calculus*.

Network calculus is a deterministic methodology for a worst-case evaluation of packet networks [2]. 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 [3]. 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 end-to-end links by using *min-plus algebra* [4], [5]. 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 an operator called convolution.

On the other hand, a drawback of the deterministic worst-case evaluation is in the overestimation for 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, stochastic network calculus has been discussed [4], [6]–[13]. Importing statistical evaluation methods to the network calculus, it takes account of the effect of statistical multiplexing. However, they are still complicated for admission control. We note that [12] provides a comprehensive summary of the stochastic network calculus literature by 2008.

There also exists another approach that does not use the stochastic calculus. In [14], [15], an evaluation formula was obtained for a loop-free network by assuming that all nodes in the network have so small buffers that overflow probability asymptotics depend only on instantaneous rates of the traffic. However, this assumption of small buffers makes us difficult to capture the burstiness character of the network traffic.

In [16], [17], the authors proposed a new stochastic network calculus for many flows from an approach like large deviations techniques [18]–[20], and obtained asymptotic end-to-end evaluation formulas for output burstiness, backlog and delay. In this paper, we apply this stochastic network calculus to a heterogeneous tandem network with many forwarding flows and cross traffic flows constrained by leaky buckets and obtain a simple evaluation formula for the end-to-end backlog without the assumption of small buffers. This result leads us to a natural extension of the evaluation formula for a single node.

The tandem network considered here consists of m nodes with J types of flows. Type j flows for $j = 1, 2, \ldots, K$ are forwarding flows, and type j flows for $j = K + 1, K + 2, \ldots, J$ are cross traffic flows. We assume that a flow of type j is limited by a leaky bucket with token rate ρ_j and token bucket size σ_j, all flows are mutually independent, and the cross traffic flows are served with higher priority than the forwarding flows. Let n_j, $j = 1, 2, \ldots, K$, be the number of forwarding flows of type j and n_{ij}^{cross}, $i = 1, 2, \ldots, m$, $j = K + 1, K + 2, \ldots, J$, be the number of cross traffic flows of type j at note i. The service rate (link capacity) at node i is constant in time and equal to C_i for $i = 1, 2, \ldots, m$. We denote by $Q(t)$ the total backlog in the network at time t. As-

Manuscript received April 27, 2010.
Manuscript revised October 8, 2010.

[†]The authors are with the Department of Computer and Information Sciences, Nagasaki University, Nagasaki-shi, 852-8521 Japan.
[††]The author is with Tokyo Institute of Technology, Tokyo, 152-8552 Japan.

*This research was supported in part by NEC Corporation.

a) E-mail: kobayashi@cis.nagasaki-u.ac.jp
b) E-mail: htakada@cis.nagasaki-u.ac.jp

DOI: 10.1587/transcom.E94.B.1288

suming that the numbers of flows n_j's and n_{ij}^{cross}'s are large except for n_{ij}^{cross}'s being equal to 0 and that time t is also large, the tail probability that $Q(t)$ is larger than a threshold B is evaluated as follows.

Evaluation formula for the backlog: *If*

$$\sum_{j=1}^{K} n_j \frac{\rho_j}{\hat{\theta}\sigma_j}\left(e^{\hat{\theta}\sigma_j}-1\right)+\sum_{j=K+1}^{J} n_{ij}^{\text{cross}} \frac{\rho_j}{\hat{\theta}\sigma_j}\left(e^{\hat{\theta}\sigma_j}-1\right) \le C_i$$

for any $i = 1, 2, \ldots, m$ *with*

$$\hat{\theta} = -\frac{\log \epsilon}{B},$$

then

$$P(Q(t) > B) \lesssim \epsilon.$$

The remainder of the paper is constructed as follows. In Sect. 2, we describe the outline of the stochastic network calculus with many flows as preliminaries. In Sect. 3, we apply the stochastic network calculus with many flows to a tandem network with many forwarding flows and cross traffic flows constrained by leaky buckets. In Sect. 4, we obtain a formula for admission control and give numerical examples. In Appendix, we present some proofs of the properties used in Sect. 3.

2. Preliminaries

We consider a discrete-time tandem network with m nodes and L flows, as illustrated in Fig. 1. (Note that L is the number of forwarding flows and there are no cross traffic flows, here.) Time t takes discrete values $0, 1, 2, \ldots$. For time $t \ge 0$, let $A^L(t)$ and $S_i^L(t)$, $i = 1, 2, \ldots, m$, be random variables representing the total arrivals to the network and the total offered services at node i, respectively, during time interval $(0, t]$ for L flows, with a convention $A^L(0) = S_i^L(0) = 0$. Further, let $Q^L(t)$ be the total backlog of L flows in the whole network at time t. In this paper, an arrival means an arrival of one packet, and packets are assumed to be of the same size. So, all flows seem like discrete-time fluid flows.

For a pair of times s and t such that $0 \le s \le t$, we let $\overline{A}^L(s,t) = A^L(t) - A^L(s)$ and $\overline{S}_i^L(s,t) = S_i^L(t) - S_i^L(s)$. (Notice that notations of these bi-variate functions are changed from the previous paper [16] to match with the definitions in other works [4], [10], [12], [13].) We define a *convolution* operator $*$ and a *deconvolution* operator $\oslash$ for functions $f(s,t)$ and $g(s,t)$ of two variables t, s such that $0 \le s \le t$ as

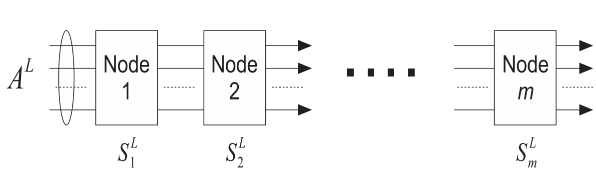

Fig. 1 Tandem network with m nodes and L flows.

$$f * g(s,t) = \min_{s \le \tau \le t}\{f(s,\tau) + g(\tau,t)\} \quad \text{and} \tag{1}$$

$$f \oslash g(s,t) = \max_{0 \le \tau \le s}\{f(\tau,t) - g(\tau,s)\}. \tag{2}$$

Then, as is shown in the previous works [4], [10], [13], [16], the backlog $Q^L(t)$ is given as

$$Q^L(t) = \overline{A}^L \oslash \overline{S}^{L,m}(t,t) \tag{3}$$

with probability one, where

$$\overline{S}^{L,m}(s,t) = \overline{S}_1^L * \overline{S}_2^L * \cdots * \overline{S}_m^L(s,t). \tag{4}$$

We discuss the limits of the cumulant generating functions of $\overline{A}^L(s,t), \overline{S}_i^L(s,t), i = 1, 2, \ldots, m$, and $Q^L(t)$, defined by

$$\overline{\mathcal{A}}^\theta(s,t) = \lim_{L\to\infty} L^{-1} \log E[e^{\theta \overline{A}^L(s,t)}], \tag{5}$$

$$\overline{\mathcal{S}}_i^\theta(s,t) = -\lim_{L\to\infty} L^{-1} \log E[e^{-\theta \overline{S}_i^L(s,t)}], \tag{6}$$

$$i = 1, 2, \ldots, m, \quad \text{and}$$

$$Q^\theta(t) = \lim_{L\to\infty} L^{-1} \log E[e^{\theta Q^L(t)}], \tag{7}$$

for $0 \le s \le t$, assuming their existence. Note that $\log E[e^{\theta X}]$ is called as *the cumulant generating function (cgf) of random variable X*.

We assume that for arbitrarily fixed $t > 0$ and any $s_0, s_1, \cdots, s_m$ such that $0 \le s_0 \le s_1 \le \cdots \le s_m = t$, $\overline{A}^L(s_0,t)$ and $\overline{S}_i^L(s_{i-1}, s_i)$, $i = 1, 2, \ldots, m$, are mutually independent, and there exists $\delta > 0$ such that, for $\theta \in (0,\delta)$, $\overline{\mathcal{A}}^\theta(s_0,t)$ and $\overline{\mathcal{S}}_i^\theta(s_{i-1}, s_i)$, $i = 1, 2, \ldots, m$, are finite. Under this assumption, in [16], it was proved that

$$Q^\theta(t) = \overline{\mathcal{A}}^\theta \oslash \overline{\mathcal{S}}^{\theta,m}(t,t) \tag{8}$$

where

$$\overline{\mathcal{S}}^{\theta,m}(s,t) = \overline{\mathcal{S}}_1^\theta * \overline{\mathcal{S}}_2^\theta * \cdots * \overline{\mathcal{S}}_m^\theta(s,t), \tag{9}$$

and for $b > 0$,

$$\limsup_{L\to\infty} L^{-1} \log P(Q^L(t) > Lb) \le -\sup_{\theta\in(0,\delta)} \{\theta b - \overline{\mathcal{A}}^\theta \oslash \overline{\mathcal{S}}^{\theta,m}(t,t)\}. \tag{10}$$

From the definitions (1) and (2), the right hand side of (10) is rewritten as

$$\inf_{\theta\in(0,\delta)} \left\{-\theta b + \max_{0\le s_0 \le \cdots \le s_m = t} \left\{\overline{\mathcal{A}}^\theta(s_0,t) - \overline{\mathcal{S}}_1^\theta(s_0,s_1) - \cdots - \overline{\mathcal{S}}_m^\theta(s_{m-1},s_m)\right\}\right\}. \tag{11}$$

Now we consider a tandem network with cross traffic depicted in Fig. 2, in which cross traffic is served with higher priority than forwarding one at each node. We assume that the cross traffic at one node is independent of the others as

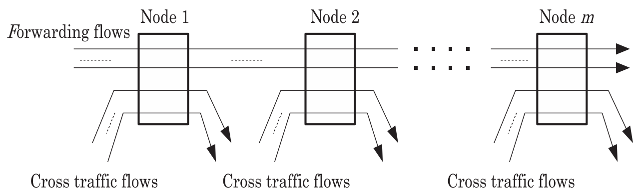

Fig. 2 Tandem network with cross traffic.

well as of the forward traffic. We let the link capacity, i.e., the offered services per unit time, at node i be constant in time and equal to Lc_i. Namely, c_i is the link capacity per one forwarding flow at node i. When we move L as in (5), (6) and (7), c_i is kept constant.

For the tandem network with cross traffic, which is served with higher priority than forwarding one at each node, it was shown in [16] that (11) can be evaluated as

$$\inf_{\theta\in(0,\delta)}\Bigg\{-\theta b+\max_{0\le s_0\le\cdots\le s_m=t}\Bigg\{\overline{\mathcal{A}}^{\theta}(s_0,t) - \left[c_1\theta(s_1-s_0)-\overline{\mathcal{A}}_1^{\theta,\mathrm{cross}}(s_0,s_1)\right]^+ - \cdots - \left[c_m\theta(s_m-s_{m-1})-\overline{\mathcal{A}}_m^{\theta,\mathrm{cross}}(s_{m-1},s_m)\right]^+\Bigg\}\Bigg\}, \qquad (12)$$

where $\overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(s,t)$ is the limit of the cumulant generating function of the cross traffic arrivals at node i during $(s,t]$ defined similarly to (5) and $[X]^+ = \max\{0,X\}$.

Remark 1: Strictly speaking, in [16], the evaluation (12) was derived under a more strict conditions that all flows were independent and subjecting to a common probabilistic law. However, by scrutinizing the proof, we can easily find that the evaluation is valid under the condition described here, namely, under the assumption of independence among forwarding traffic and cross traffics and the assumption of existence of the limits $\overline{\mathcal{A}}^{\theta}(s,t)$ and $\overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(s,t)$'s with a suitable interval $(0,\delta)$ of θ in which these limits of cumulant generating functions are finite.

3. A Tandem Network with Input Flows Limited by Leaky Buckets

We consider a tandem network consisting of m nodes with cross traffic flows depicted in Fig. 2. There are K types of forwarding flows and $\tilde{K}$ types of cross traffic flows. Let the flows of type 1 to type K be the forwarding flows and those of type $K+1$ to type $J = K+\tilde{K}$ be the cross traffic ones. Let the number of forwarding flows of type j be $L\alpha_j$, $j = 1,2,\ldots,K$, and the number of cross traffic flows of type j at node i be $L\beta_{ij}$, $j = K+1, K+2,\ldots,J$, $i = 1,2,\ldots,m$. If there exist no cross traffic flows of type j at node i, then we consider $\beta_{ij} = 0$. To meet the notations in the previous section, here L is set to the total number of forwarding flows, and hence we assume that $\sum_{j=1}^{K}\alpha_j = 1$. The total number of cross traffic flows at node i is given by $M_i = L\sum_{j=K+1}^{J}\beta_{ij}$.

We assume that the link capacity at node i, i.e., the offered services per unit time at node i, is constant in time and equal to Lc_i, $i = 1,2,\ldots,m$. When we move L (and M_i's), α_j's, β_{ij}'s, and c_i's are kept constant.

We let the forwarding flows of type j be $\{A_{j,1}(t)\}$, $\{A_{j,2}(t)\},\cdots,\{A_{j,L\alpha_j}(t)\}$, $j = 1,2,\ldots,K$, and the cross traffic flows of type j at node i be $\{A_{i,j,1}^{\mathrm{cross}}(t)\}$, $\{A_{i,j,2}^{\mathrm{cross}}(t)\}$, $\cdots$, $\{A_{i,j,L\alpha_j}^{\mathrm{cross}}\}$, $j = K+1, K+2,\ldots,J$. We assume that all flows (both forwarding flows and cross traffic flows) are mutually independent, and type j flows are subjecting to a common probabilistic law† Further we make the following assumptions denoting by $\{A_j(t)\}$ the arrival process of a typical flow of type j:

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

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

$$\overline{A}_j(s,t) \equiv A_j(t)-A_j(s) \le \rho_j(t-s)+\sigma_j.$$

This inequality implies that the average rate and the burst size of $A_j(t)$ are limitted by the token rate ρ_j and the token bucket size σ_j, respectivily.

It is shown in [1] that, if we put

$$\eta_j(t,\theta) = \log\left[1+\frac{\rho_j t}{\rho_j t+\sigma_j}(e^{\theta(\rho_j t+\sigma_j)}-1)\right], \qquad (13)$$

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

$$\overline{\mathcal{A}}_j^{\theta}(s,t) = \overline{\mathcal{A}}_j^{\theta}(t-s,0) \le \eta_j(t-s,\theta) \quad \text{for } \theta\in(0,\infty). (14)$$

We denote the increment of arrivals in the forwarding traffic during $(s,t]$ as $\overline{A}^L(s,t) = A^L(t)-A^L(s)$. Then we have

$$\overline{A}^L(s,t) = \sum_{j=1}^{K}\sum_{k=1}^{L\alpha_j}\{A_{j,k}(t)-A_{j,k}(s)\},$$

and denoting by $\overline{\mathcal{A}}_{j,k}^{\theta}(s,t)$ the cgf of $\overline{A}_{j,k}(s,t) = A_{j,k}(t) - A_{j,k}(s)$, its cgf is given by $\sum_{j=1}^{K}\sum_{k=1}^{L\alpha_j}\overline{\mathcal{A}}_{j,k}^{\theta}(s,t)$. Since $\overline{\mathcal{A}}_{j,k}^{\theta}(s,t)$ is evaluated as in (14), the limit (5) is evaluated as

$$\overline{\mathcal{A}}^{\theta}(s,t) \le \sum_{j=1}^{K}\alpha_j\eta_j(t-s,\theta). \qquad (15)$$

Similarly, we denote the increment of arrivals in the cross traffic at node i during $(s,t]$ as $\overline{A}_i^{M_i,\mathrm{cross}}(s,t) = A_i^{M_i,\mathrm{cross}}(t) -$

†This assumption for type j flows to be subjecting to a common probabilistic law is too restrictive. We can relax somehow this assumption by assuming the existence of the limits $\overline{\mathcal{A}}^{\theta}(s,t)$ and $\overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(s,t)$'s.

$A_i^{M_i,\mathrm{cross}}(s)$. Then we have

$$\overline{A}_i^{M_i,\mathrm{cross}}(s,t) = \sum_{j=K+1}^{J} \sum_{k=1}^{L\beta_{ij}} \{A_{i,j,k}^{\mathrm{cross}}(t) - A_{i,j,k}^{\mathrm{cross}}(s)\}$$

and its cgf is given by $\sum_{j=K+1}^{J} \sum_{k=1}^{L\beta_{ij}} \overline{\mathcal{A}}_{i,j,k}^{\theta,\mathrm{cross}}(s,t)$, where $\overline{\mathcal{A}}_{i,j,k}^{\theta,\mathrm{cross}}(s,t)$ is the cgf of $\overline{A}_{i,j,k}^{\mathrm{cross}}(s,t) = A_{i,j,k}^{\mathrm{cross}}(t) - A_{i,j,k}^{\mathrm{cross}}(s)$. Since $\overline{\mathcal{A}}_{i,j,k}^{\theta,\mathrm{cross}}(s,t)$ is evaluated as in (14), the limit functions

$$\overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(s,t) \equiv \lim_{L\to\infty} L^{-1} \log E[e^{\theta \overline{A}_i^{M_i,\mathrm{cross}}(s,t)}],$$

are evaluated as

$$\overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(s,t) \leq \sum_{j=K+1}^{J} \beta_{ij}\eta_j(t-s,\theta), \tag{16}$$

for $i = 1, 2, \ldots, m$.

Since, from (16),

$$\left[c\theta(s_i - s_{i-1}) - \overline{\mathcal{A}}_i^{\theta,\mathrm{cross}}(s_{i-1}, s_i)\right]^+ \geq \left[c\theta(s_i - s_{i-1}) - \sum_{j=K+1}^{J} \beta_{ij}\eta_j(s_i - s_{i-1})\right]^+,$$

(12) can be reevaluated from above by using $\eta_j(t,\theta)$'s instead of $\overline{\mathcal{A}}^{\theta}(s,t)$ and $\overline{\mathcal{A}}^{\theta,\mathrm{cross}}(s,t)$'s as

$$\inf_{\theta\in(0,\infty)} \left\{-\theta b + \max_{0\leq s_0\leq\cdots\leq s_m=t} \left\{\sum_{j=1}^{K} \alpha_j\eta_j(t-s_0,\theta) - \left[c_1\theta(s_1-s_0) - \sum_{j=K+1}^{J} \beta_{1j}\eta_j(s_1-s_0)\right]^+ - \cdots - \left[c_m\theta(s_m-s_{m-1}) - \sum_{j=K+1}^{J} \beta_{mj}\eta_j(s_m-s_{m-1})\right]^+\right\}\right\}. \tag{17}$$

Here, we set $\delta = \infty$ in (12) since $\eta_j(t,\theta)$ is finite for any t and $\theta \in (-\infty, \infty)$.

As will be proved in Appendix, the function $\eta_j(t,\theta)$ is increasing and concave on t for each fixed θ. Furthermore, $\eta_j(t,\theta)$ satisfies the inequality

$$\eta_j(t,\theta) \leq \theta t \cdot \phi_j(\theta) \tag{18}$$

for $t \geq 0$, where

$$\phi_j(\theta) = \frac{\rho_j}{\theta\sigma_j}\left(e^{\theta\sigma_j} - 1\right) > 0 \tag{19}$$

for $\theta > 0$. We put

$$\xi(\theta) = \sum_{j=1}^{K} \alpha_j\phi_j(\theta), \quad \text{and} \tag{20}$$

$$\psi_i(\theta) = \sum_{j=K+1}^{J} \beta_{ij}\phi_j(\theta), \quad i = 1, 2, \ldots, m. \tag{21}$$

Then, (17) is reevaluated from above as

$$\inf_{\theta\in(0,\infty)} \left\{-\theta b + \max_{0\leq s_0\leq\cdots\leq s_m=t} \left\{\theta(t-s_0)\cdot\xi(\theta) - \theta(s_1-s_0)\cdot[c_1 - \psi_1(\theta)]^+ - \cdots - \theta(s_m-s_{m-1})\cdot[c_m - \psi_m(\theta)]^+\right\}\right\}. \tag{22}$$

So far we have evaluated the right hand side of (10) for the tandem network in Fig. 2 with flows constrained by leaky buckets. Thus now we know

$$\limsup_{L\to\infty} L^{-1} \log P(Q^L(t) > Lb)$$

is bounded from above by (22). Then we have the following theorem.

Theorem 1: If

$$\sum_{j=1}^{K} \alpha_j\rho_j + \sum_{j=K+1}^{J} \beta_{ij}\rho_j < c_i, \tag{23}$$

for any $i = 1, 2, \ldots, m$, then for any $b > 0$ there exists $T(b)$ such that for any $t > T(b)$

$$\limsup_{L\to\infty} L^{-1} \log P(Q^L(t) > Lb) \leq -\theta^* b, \tag{24}$$

where θ^* is the largest $\theta > 0$ satisfying $\xi(\theta) + \psi_i(\theta) \leq c_i$, or equivalently

$$\sum_{j=1}^{K} \alpha_j \frac{\rho_j}{\theta\sigma_j}\left(e^{\theta\sigma_j} - 1\right) + \sum_{j=K+1}^{J} \beta_{ij}\frac{\rho_j}{\theta\sigma_j}\left(e^{\theta\sigma_j} - 1\right) \leq c_i, \tag{25}$$

for any $i = 1, 2, \ldots, m$.

Proof. We note that the function $\phi_j(\theta)$ is strictly increasing and convex, and $\lim_{\theta\downarrow 0}\phi_j(\theta) = \rho_j$ and $\lim_{\theta\to\infty}\phi_j(\theta) = \infty$, as will be proved in Appendix. Hence, under the condition (23), there exists some $\theta > 0$ satisfying (25), and θ^* is well defined. The condition (25) implies that $\xi(\theta) + \psi_i(\theta) - c_i \leq 0$ for any i and $\theta \leq \theta^*$. Also there exists an i_0 such that $\xi(\theta^*) + \psi_{i_0}(\theta^*) - c_{i_0} = 0$.

The right hand side of (22) is rewritten as

$$\inf_{\theta\in(0,\infty)} \left\{-\theta b + \max_{0\leq s_0\leq\cdots\leq s_m=t} \left\{\theta(s_1-s_0)\cdot\min\{\xi(\theta) + \psi_1(\theta) - c_1,\ \xi(\theta)\} + \cdots + \theta(s_m-s_{m-1}) \cdot \min\{\xi(\theta) + \psi_m(\theta) - c_m,\ \xi(\theta)\}\right\}\right\}. \tag{26}$$

When $0 < \theta \leq \theta^*$, as stated above, $\xi(\theta) + \psi_i(\theta) - c_i$ is less than or equal to 0, and hence

$$\min\{\xi(\theta) + \psi_i(\theta) - c_i,\ \xi(\theta)\} \leq 0.$$

So the maximum in (26) is attained at $s_0 = s_1 = \cdots = s_m = t$, and the maximum value is equal to zero. Thus the infimum of the quantity in the outermost braces of (26) on the restricted interval $(0, \theta^*]$ of θ is attained at $\theta = \theta^*$, and the infimum value is equal to $-\theta^* b$.

On the other hand, when $\theta > \theta^*$, the value of $\xi(\theta) + \psi_i(\theta) - c_i$ may be positive. However, the function $\xi(\theta) + \psi_i(\theta) - c_i$ is increasing and convex since $\phi_j(\theta)$ is increasing and convex. So, especially for $i = i_0$, we have

$$\xi(\theta) + \psi_{i_0}(\theta) - c_{i_0} \geq \{\xi'(\theta^*) + \psi'_{i_0}(\theta^*)\}(\theta - \theta^*) > 0,$$

where $\xi'(\theta)$ and $\psi'_{i_0}(\theta)$ are derivatives of $\xi(\theta)$ and $\psi_{i_0}(\theta)$, respectively. Thus the maximum in (26) is larger than or equal to

$$\theta t \{\xi'(\theta^*) + \psi'_{i_0}(\theta^*)\}(\theta - \theta^*).$$

Then we see that the quantity in the outermost braces in (26), for which the infimum is taken on θ, is larger than $-\theta^* b$ for sufficiently large t, because

$$\begin{aligned} &-\theta b + \theta t \{\xi'(\theta^*) + \psi'_{i_0}(\theta^*)\}(\theta - \theta^*) - (-\theta^* b) \\ &\quad \geq \left\{\{\xi'(\theta^*) + \psi'_{i_0}(\theta^*)\}\theta t - b\right\}(\theta - \theta^*) > 0 \end{aligned}$$

for t larger than

$$T(b) = \frac{b}{\{\xi'(\theta^*) + \psi'_{i_0}(\theta^*)\}\theta^*}.$$

Combining the above results for $0 < \theta \leq \theta^*$ and for $\theta > \theta^*$, we see that, if $t > T(b)$, the infimum in (26) over $\theta \in (0, \infty)$ is attained by $\theta = \theta^*$ and the value is equal to $-\theta^* b$. This completes the proof. □

Remark 2: For a single node case $m = 1$ with J types of flows, [1] and [19] obtained a similar result

$$\lim_{L\to\infty} L^{-1} \log P(\hat{Q}^L > Lb) = \sup_{\tau\geq 0} \inf_{\theta} \left\{ \sum_j \alpha_j \eta_j(\tau, \theta) - \theta(b + c\tau) \right\}$$

for the backlog $\hat{Q}^L$ in the steady state by applying the large deviations techniques. Here cross traffic was not taken into consideration.

Using the inequality $\sup_x \inf_y f(x, y) \leq \inf_y \sup_x f(x, y)$ and the evaluation (18), we can easily derive a corresponding result as

$$\begin{aligned} &\lim_{L\to\infty} L^{-1} \log P(\hat{Q}^L > Lb) \\ &\leq \inf_{\theta} \left\{ -\theta b + \sup_{\tau\geq 0} \left\{ \sum_{j=1}^{J} \alpha_j \frac{\rho_j}{\sigma_j} \left(e^{\theta\sigma_j} - 1\right)\tau - c\theta\tau \right\} \right\}. \end{aligned}$$

A similar discussion to the proof of the theorem reveals that the right hand side can be reduced to (24).

Remark 3: The parameter value θ^* defined in the theorem satisfies the equality $\xi(\theta^*) + \psi_{i_0}(\theta^*) = c_{i_0}$ for some i_0, and for other $i \neq i_0$ it satisfies the inequality $\xi(\theta^*) + \psi_i(\theta^*) \leq c_i$. This implies that node i_0 is a bottle neck node of the network. Then the theorem says that the asymptotic tail probability of the end-to-end backlog can be determined by the traffic load at the bottle neck node. Furthermore, it can be done by the total traffic load of the forwarding flows and the cross ones.

4. A Formula for Admission Control and Numerical Examples

Theorem 1 in the preceding section says that, if L are t are large enough, $L^{-1} \log P(Q^L(t) > Lb)$ is less than or equal to $-\theta b$ for any positive θ satisfying (25). We will write this as

$$P(Q^L(t) > Lb) \lesssim e^{-L\theta b}. \tag{27}$$

For an application to an admission control, it is convenient to rewrite this result without using L. Let us denote the backlog threshold as $B = Lb$, the link capacity at node i as $C_i = Lc_i$, the number forwarding flows of type j as $n_j = L\alpha_j$ and the number of cross traffic flows of type j at node i as $n_{ij}^{\text{cross}} = L\beta_{ij}$. Then, for a given tail probability bound ϵ, we obtain the following backlog evaluation formula for admission control.

A formula for admission control: *If n_j's, n_{ij}^{cross}'s and t are large enough (except for cross traffic flows such that $n_{ij}^{\text{cross}} = 0$) and if*

$$\hat{\theta} = -\frac{\log \epsilon}{B} \tag{28}$$

satisfies the inequality

$$\sum_{j=1}^{K} n_j \frac{\rho_j}{\hat{\theta}\sigma_j}\left(e^{\hat{\theta}\sigma_j} - 1\right) + \sum_{j=K+1}^{J} n_{ij}^{\text{cross}} \frac{\rho_j}{\hat{\theta}\sigma_j}\left(e^{\hat{\theta}\sigma_j} - 1\right) \leq C_i \tag{29}$$

for any $i = 1, 2, \ldots, m$, then

$$P(Q^L(t) > B) \lesssim \epsilon. \tag{30}$$

From (29), we see that the quantity corresponding to the *effective bandwidth* [1] of a flow of type j, is given by

$$\frac{\rho_j}{\hat{\theta}\sigma_j}\left(e^{\hat{\theta}\sigma_j} - 1\right). \tag{31}$$

Also (29) for $i = i_0$, namely the inequality at a bottle neck node, gives us an *admissible region* for the numbers n_j's and $n_{i_0 j}^{\text{cross}}$'s of type j flows, $j = 1, 2, \ldots, J$.

Numerical examples: From Remark 3, we can evaluate the end-to-end backlog by the total traffic load of the forwarding flows and the cross ones at a bottle neck node. So, here we consider a bottle neck node having two types of flows, type 1 flows and type 2 flows. We set the token rate and the token bucket size of type 1 flows as $\rho_1 = 40$ Mbps and $\sigma_1 = 4$ Mbits, and those of type 2 flows as $\rho_2 = 20$ Mbps and $\sigma_2 = 10$ Mbits. We calculate effective bandwidths of type 1 and type 2 using (31) and admissible region from (29). Figure 3 shows the effective bandwidths of type 1 and type 2 with the buffer thresholds $B = 100$ Mbits and

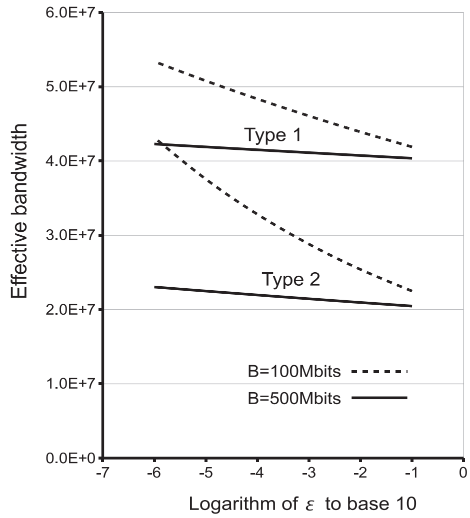

Fig. 3 Effective bandwidths of type 1 and type 2.

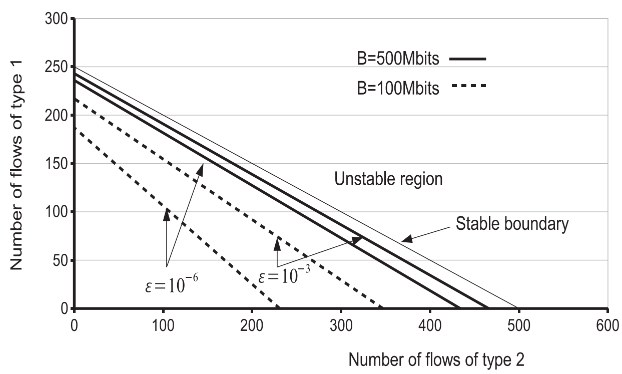

Fig. 4 Admissible region with $C = 10\,\text{Gbps}$.

$B = 500\,\text{Mbits}$. As ϵ decrease, the effective bandwidths increase. However, when the buffer threshold B is 500 Mbits, 50 times the token bucket size of type 2, they don't increase so much. Figure 4 shows the admissible region with the link capacity $C = 10\,\text{Gbps}$. As ϵ decrease, the admission region becomes smaller, but in case of $B = 500\,\text{Mbits}$, it doesn't so small.

5. Conclusion

For a heterogeneous tandem network with many forwarding flows and cross traffic flows constrained by leaky buckets, we obtained a simple evaluation formula for the end-to-end backlog using the proposed stochastic network calculus for many flows. In this formula, the end-to-end backlog can be evaluated by the total traffic load at the bottle neck node. This result leads us to a natural extension of the evaluation formula for a single node.

References

[1] F.P. Kelly, "Notes on effective bandwidths," in Stochastic Networks: Theory and Applications, ed. F.P. Kelly, S. Zachary, and I.B. Zledins, pp.141–168, Oxford University Press, 1996.
[2] R. Cruz, "A calculus for network delay, parts I and II," IEEE Trans. Inf. Theory, vol.37, no.1, pp.114–141, 1991.
[3] 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 Trans. Netw., vol.2, no.2, pp.137–150, 1994.
[4] C.S. Chang, Performance Guarantees in Communication Networks, Springer-Verlag, 2000.
[5] 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.
[6] D. Starobinski and M. Sidi, "Stochastically bounded burstiness for communication networks," IEEE Trans. Inf. Theory, vol.46, no.1, pp.206–212, 2000.
[7] Q. Yin, Y. Jiang, S. Jiang, and P.Y. Kong, "Analysis on generalized stochastically bounded bursty traffic for communication networks," Proc. IEEE LCN'2002, pp.141–149, 2002.
[8] 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.
[9] F. Ciucu, A. Burchard, and J. Liebeherr, "A network service curve approach for the stochastic analysis of networks," Proc. ACM Sigmetrics '05, 2005.
[10] M. Fidler, "An end-to-end probabilistic network calculus with moment generating functions," Proc. IEEE IWQoS 2006, pp.261–270, 2006.
[11] Y. Liu, C.-K. Tham, and Y. Jiang, "A calculus for stochastic QoS analysis," Performance Evaluation, vol.64, no.6, pp.547–572, 2007.
[12] Y. Jiang and Y. Liu, Stochastic Network Calculus, Springer, 2008.
[13] Y. Jiang, "Network calculus and queueing theory: Tow sides of one coin," Proc. Valuetools 2009, 2009.
[14] N.B. Likhanov, R.R. Mazumdar, and F. Theberge, "Providing QoS in large networks: Statistical mulitplexing and admission control," in Analysis, Control and Optimization of Complex Dynamic Systems, ed. E. Boukas and R. Malhame, pp.137–168, Springer, 2005.
[15] R.R. Mazumdar, "Statistical multiplexing, the stochastic knapsack and admission control," Proc. ITC21, Keynote V, 2009.
[16] K. Kobayashi, Y. Takahashi, and H. Takada, "A stochastic network calculus with many flows," Proc. ITC 21, 2009.
[17] K. Kobayashi, Y. Takahashi, and H. Takada, "Asymptotic end-to-end stochastic evaluation for tandem networks with many flows," Proc. Valuetools 2009, 2009.
[18] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Jones and Bartlett, 1993.
[19] A. Ganesh, N. O'Connell, and D. Wischik, Big Queues, Springer, 2004.
[20] K. Kobayashi and Y. Takahashi, "Overflow probability for a discrete-time queue with non-stationary multiplexed input," Telecommunication Systems, vol.15, no.1-2, pp.157–166, 2000.

Appendix

Here we give a proof for the concavity of the function $\eta_j(t,\theta)$ given in (13) and the convexity of the function $\phi_j(\theta)$ given in (19).

Proof of the concavity of $\eta(t,\theta)$ in (13): For brevity of exposition, we omit the subscripts j used in (13). Then 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

$$\eta(0,\theta) = 0 \quad \text{and} \quad \lim_{t\to\infty}\{\eta(t,\theta) - \theta(\rho t + \sigma)\} = 0.$$

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

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

$$\frac{\partial^2}{\partial^2 t}\eta(t,\theta)=-\frac{\rho^2}{(\rho t+\sigma)^2(\rho t+\sigma e^{-\theta(\rho t+\sigma)})^2}\cdot\left[\sigma^2\left\{1-2\theta(\rho t+\sigma)e^{-\theta(\rho t+\sigma)}-e^{-2\theta(\rho t+\sigma)}\right\}\right.$$
$$\left.+2\sigma\rho t\left\{1-e^{-\theta(\rho t+\sigma)}-\theta(\rho t+\sigma)e^{-\theta(\rho t+\sigma)}-\frac{1}{2}\theta^2(\rho t+\sigma)^2e^{-\theta(\rho t+\sigma)}\right\}\right].$$

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.

At $t=0$, the first derivative reduces to $\phi(\theta)$ defined in (19). Hence the line in the right hand side of (18) is the tangential line of $\eta(t,\theta)$ at $t=0$, and from the concavity of $\eta(t,\theta)$, the inequality (18) holds. □

Proof of the convexity of $\phi_j(\theta)$ in (19): For brevity of notation, here we omit the subscripts j used in (19) as

$$\phi(\theta)=\frac{\rho}{\theta\sigma}\left(e^{\theta\sigma}-1\right).$$

Its first and second derivatives are given as

$$\frac{d}{d\theta}\phi(\theta)=\frac{\rho}{\theta^2\sigma}e^{\theta\sigma}\left\{e^{-\theta\sigma}-1+\theta\sigma\right\}\quad\text{and}$$
$$\frac{d^2}{d^2\theta}\phi(\theta)=-\frac{2\rho}{\theta^3\sigma}e^{\theta\sigma}\left\{e^{-\theta\sigma}-1+\theta\sigma-\frac{1}{2}\theta^2\sigma^2\right\}.$$

It is easily checked that both derivative are positive for $\theta>0$, since $g_1(x)\equiv e^{-x}-1+x>0$ and $g_2(x)\equiv e^{-x}-1+x-\frac{1}{2}x^2<0$ for $x>0$. Hence, $\phi(\theta)$ is increasing and convex.

It is clear that $\lim_{\theta\downarrow 0}\phi(\theta)=\rho$ and that $\lim_{\theta\to\infty}\phi(\theta)=\infty$. □

Kazutomo Kobayashi received the B.S. degree in Electrical Engineering from Saga University in 1975, the M.S. degree in Communication Engineering from Kyushu University in 1977 and the Ph.D. degree in Information Science from Tokyo Institute of Technology in 1996. He had been with NEC Corporation since 1977 until 2003. Since 2003, He has been Professor of Department of Computer and Information Sciences, Nagasaki University. His research interests include queueing theory and stochastic processes as applied to communication and computer networks. He is a member of ORSJ and IPSJ.

Yukio Takahashi is Professor Emeritus, Tokyo Institute of Technology. His research interests include the theory and applications of stochastic models and analyses of congestion phenomena.

Hiroyuki Takada received the B.S. degree in Applied Mathematics from Tokyo University of Science in 1995, the M.S. degree in Information Science from Tokyo University of Science in 1997 and the Ph.D. degree in Information Science from Tokyo University of Science in 2002. He had been Assistant Professor of Department of Information Sciences since 1997 until 2002. He had been Assistant Professor of Department of Electrical and Computer Engineering from Tokyo University of Science, Yamaguchi since 2002 until 2004. Since 2004, he has been Assistant Professor of Department of Computer and Information Sciences, Nagasaki University. His research interests include queueing theory and stochastic processes. He is a member of ORSJ and JSIAM.