ADMISSION CONTROL FOR A POLLING SYSTEM WITH MARKOV-MODULATED ARRIVAL PROCESSES AND BERNOULLI SERVICE SCHEDULE

Wei Feng[†], Kouichi Adachi[†] and Masashi Kowada[‡]

Received February 5, 2002; revised April 22, 2002

ABSTRACT. Admission control is an important part of modern high-speed network control and has received extensive attention in the recent literature. In this paper we consider an admission control problem for a discrete-time polling system consisting of two queues and a single server. The arrival process in each queue is a superposition of mutually independent Markov-modulated processes and the server serves the two queues according to a Bernoulli service schedule. Basing on the theory of effective bandwidths and the buffer upper bound results on the overflow probability obtained by large deviation techniques, we derive an admission control criterion for the polling system under which *Quality of Service* (QoS) requirement by each queue is guaranteed.

1. Introduction

In the emerging high-speed networks using asynchronous transfer mode (ATM) technology, each traffic-source is described by its stochastic characteristics, and is assured a quality of service, as measured by the cell loss probability due to the buffer overflows. Generally, the cell loss probability is desired to be controlled below very small level, e.g., in the order of 10^{-9} . Therefore, providing QoS guarantees is an important and challenging issue in the design of high-speed networks. Admission control using the concept of effective bandwidth is an integral part of this challenge and has received extensive attention in the recent literature (see Kelly[21], Gibbens and Hunt[15], Kesidis et al.[22], Elwalid and Mitra[10], Berger and Whitt[1], Liu et al. [19], Gautam and Kulkarni [14], [23], Whitt[28] and Zhang et al.[25]). The main aim of the admission control is to control acceptance of a new call that arrives to a network under the condition without violating existing QoS guarantees made to on on-going calls. Most of the previous work have been devoted to single queueing systems with a single class of traffic ([4], [7], [19]), or with multiple classes of traffic with and without priority structure ([21], [10], [15], [22], [28] and [1], [14], [23]). For a network consisting of two-parallel queues and a single server, Zhang et al. [25] construct a theoretical framework of the call admission control schemes with multiple statistical QoS guarantee under the Generalized Processor Sharing (GPS) schedule discipline.

In the present paper, we consider an admission control problem for a polling system. As known, polling systems have been used for modeling distributed multiqueue systems sharing a single server and extensively studied in the literature under various service schedules such as the exhaustive, gated, K-limited and Bernoulli service schedules (see [11], [18]). Especially, polling systems consisting of two-parallel queues and a single server have an important application in modeling communication systems with two different types of traffic:

²⁰⁰⁰ Mathematics Subject Classification. 60K25, 68M20, 90G22.

Key words and phrases. Polling systems; Bernoulli service schedule; Effective bandwidths; Quality of service; Large deviations; Tail distribution; Markov-modulated processes; Admission control.

the real-time traffic(i.e., voice and video) and the non-real-time traffic(i.e., data). In general, the real-time traffic has a more stringent delay requirement but can tolerate higher cell-loss; while the non-real-time traffic can tolerate higher delay but demands much smaller cell loss. They have different requirements for QoS. Therefore, deriving an admission control criterion for such polling systems is an important work in network control. To the best of our knowledge, such a research has not been done before. The reason can be considered to be that performance analysis such as the delay, cell loss and buffer overflow probabilities for polling systems with general arrival processes is extremely difficult. Recently, applying large deviation technique, we have obtained in [12] and [13] the upper and lower bounds of the overflow probability for polling systems under the Bernoulli service schedule. This sheds some light on the admission control problem of polling systems. Large deviation technique is an asymptotic technique for analyzing rare events and estimating rare event probabilities (see Bucklew[5], Dembo and Zeitouni[8]). In this decade, it has been extensively used to estimate tail probabilities for queueing systems (see Botvich and Duffield[4], Chang[6] and [7], Weiss[24], Glynn and Whitt[16], Duffield[9], Bertsimas et al.[2] and [3], and Zhang et al.[26] and [27].

The model considered in this paper is a discrete-time polling system consisting of two parallel queues and a single server. The arrival process in the *i*th queue is a superposition of mutually independent Markov-modulated processes. A single server serves the two queues according to the Bernoulli service schedule described as follows: At the beginning of each discrete time, the server who just completed the service in the *i*th queue makes a random decision: with the probability $p_i(0 \le p_i < 1)$, it continues to sever the packets of the *i*th queue in the next slot, with the probability $q_i = 1 - p_i$, it switches to the other queue. Further, when the queue being served becomes empty, the server switches its service to the other queue immediately. The server is assumed not to take switching times in its transition from one queue to the other. The service rate in Q_i is assumed to be c_i . Note that the Bernoulli service schedule constitutes a generalization to both the exhaustive and 1-limited service schedules. The main purpose of this paper is to present an admission control criterion under which the cell-loss probability requirement is satisfied for each queue.

The paper is organized as follows. In Section 2, we describe the model, define the potential service processes, and view some results on the large deviation and concept of effective bandwidth. In Section 3, we give the large deviation upper and lower bounds, and then, derive the admission control criterion for the polling system. In Section 4, we present an algorithm for a special case where the arrival processes are the superposition processes of mutually independent Markov on/off sources. Some conclusions are included in Section 5.

2. Model and Preliminaries

We denote the two queues by Q_1 and Q_2 . Throughout the paper, all time indices t, τ , etc., are always integers and $\mathbf{N} = \{0, 1, 2, \cdots\}$. On a notational remark, we denote by $S_{\tau,t}^X = \sum_{k=\tau}^{t-1} X_k, \tau < t$ and $S_t^X = \sum_{k=0}^{t-1} X_k$ the partial sums of the random sequence $X = \{X_t; t \in \mathbf{N}\}$, and by $S_t^X(s) = \sum_{k=0}^{\lfloor ts \rfloor} X_k/t, 0 \le s \le 1$ the scaled partial sum of X, respectively. We also denote by $\Lambda_X(\theta)$ and $\Lambda_X^*(\alpha)$ the limit logarithmic moment generating function of the partial sum process of X, and the Legendre-Fenchel transform of $\Lambda_X(\theta)$, namely,

$$\Lambda_X(\theta) = \lim_{t \to \infty} \frac{1}{t} \log E[e^{\theta S_t^X}], \quad \theta \in \mathbf{R},$$
(1)

$$\Lambda_X^*(\alpha) = \sup_{\theta \in \mathbf{R}} \{ \theta \alpha - \Lambda_X(\theta) \}, \quad \alpha \in \mathbf{R}.$$
 (2)

A. Arrival processes

The arrival process $\{A_t^i, t \in \mathbf{N}\}$ in Q_i is a superposition of a number of independent Markov Modulated Arrival Processes (MMAP's). There are N_i classes of traffic in Q_i and the n_{ij} streams of the class j. The mth stream of the jth class is a MMAP defined by $A_t^{ijm} = H_t^{ijm}(a_t^{ijm})$, where $\{a_t^{ijm}, t \in \mathbf{N}\}$ is an irreducible, aperiodic, stationary Markov chain on the finite state space S^{ij} with transition matrix $\mathbf{P}^{ij} = (p_{kl}^{ij})_{k,l\in S^{ij}}$, and stationary distribution π^{ij} , and $\{H_t^{ijm}(k), t \in \mathbf{N}\}$ is a renewal process with the mean $h_k^{ij} = E[H_0^{ij}(k)]$ and the moment generating function $\psi_k^{ij}(\theta) = E[exp(\theta H_0^{ij}(k))]$ for fixed $k \in S^{ij}$. In general, the source state a_t^{ijm} can be thought of as modeling the burstiness of the arrival process at time t, and the Markov structure models correlation in the arrival process. Assume that all the Markov chains $\{a_t^{ijm}, t \in \mathbf{N}\}$ and all the renewal processes $\{H_t^{ijm}(k), t \in \mathbf{N}\}$ are mutually independent, and these arrival processes in Q_1 and Q_2 are also mutually independent. Further assume that all the underlying Markov chains have reached their steady state. Let $\mathbf{a}_t^i = (a_t^{i11}, \cdots, a_t^{i1n_1}, \cdots, a_t^{iN_{i1}1}, \cdots, a_t^{iN_{i1}N_{i1}})$, then $\{\mathbf{a}_t^i; t \in \mathbf{N}\}$ is a Markov chain with state space $S^i = \prod_{n=1}^{N_i} (S^{ij})^{n_{ij}}$, transition matrix $\mathbf{P}^i = \bigotimes_{n=1}^{N_i} (\mathbf{P}^{ij})^{\otimes n_{ij}}$ and stationary distribution $\pi^i = \bigotimes_{n=1}^{n_i} (\pi^{ij})^{\otimes n_{ij}}$, where \otimes denotes the Kronecker product. Define $H_t^i(\mathbf{k}^i) = \sum_{n=1}^{N_i} \sum_{m=1}^{n_{ij}} (k^{ij})$ for $\mathbf{k}^i = (k^{i11}, \cdots, k^{i1n_{i1}}, \cdots, k^{iN_{in}N_i}) \in S^i$. Then, the aggregate arrival process in Q_i can be denoted as $A_t^i = H_t^i(\mathbf{a}_t^i)$. We have $\mathcal{A}^i \equiv E[A_t^i] = \sum_{j=1}^{N_i} \sum_{m=1}^{n_{ij}} \sum_{k^{ijm} \in S^{ij}} \pi_{k^{ijm}}^{ijm} h_{kijm}^{ij}$.

B. Potential service processes

According to the Bernoulli service schedule described above, the potential service process in each queue (i.e., the service processes when both queues are not empty) can be described by a Markov chain. Let $\{b_t^1, t \in \mathbf{N}\}$ be a Markov chain with state space $\{0, 1\}$ and transition matrix

$$\mathbf{P}_{b^1} = \begin{bmatrix} p_2 & q_2\\ q_1 & p_1 \end{bmatrix}.$$
(3)

Let $b_t^2 = 1 - b_t^1$, $t \in \mathbb{N}$. Then $\{b_t^2, t \in \mathbb{N}\}$ is also a Markov chain with state space $\{0, 1\}$ and transition matrix

$$\mathbf{P}_{b^2} = \begin{bmatrix} p_1 & q_1 \\ q_2 & p_2 \end{bmatrix}. \tag{4}$$

Furthermore, define $B_i^i = b_i^i c_i, t \in \mathbf{N}$. We have that $\{B_t^i, t \in \mathbf{N}\}$ is a Markov chain with state space $S_{B^i} = \{0, c_i\}$ and transition matrix \mathbf{P}_{b^i} . From the above definition, $\{B_t^i, t \in \mathbf{N}\}$, i = 1, 2 are independent of the arrival processes $\{A_t^1, t \in \mathbf{N}\}$ and $\{A_t^2, t \in \mathbf{N}\}$. The equilibrium distributions of $\{B_t^1, t \in \mathbf{N}\}$ and $\{B_t^2, t \in \mathbf{N}\}$ are $\pi_{B^1} = (q_2/(q_1 + q_2), q_1/(q_1 + q_2))$ and $\pi_{B^2} = (q_1/(q_1 + q_2), q_2/(q_1 + q_2))$, respectively. Let $\mathcal{B}^i \equiv E[B_t^i] = q_i/(q_1 + q_2)c_i, i = 1, 2$. Note that $\{B_t^1 + B_t^2, t \in \mathbf{N}\}$ is a Markov chain with state space $\{c_1, c_2\}$ and transition matrix \mathbf{P}_{b^2} . Its equilibrium distribution and mean are $\pi_B^2 = (q_1/(q_1 + q_2), q_2/(q_1 + q_2))$ and $\mathcal{B}^1 + \mathcal{B}^2$, respectively.

C. Stability condition

Since $\{B_t^1 + B_t^2, t \in \mathbf{N}\}$ can be referred as the service process of the aggregate arrival process $\{A_t^1 + A_t^2, t \in \mathbf{N}\}$, from Loynes's Stability Theorem[20], the stability condition of the polling system is given as follows:

$$\mathcal{A}^1 + \mathcal{A}^2 < \mathcal{B}^1 + \mathcal{B}^2. \tag{5}$$

Throughout the paper, we assume that the condition (5) holds.

D. Large deviations of the arrival processes and potential service processes

For $\mathbf{k}^i \in \mathcal{S}^i$, define $\psi_{\mathbf{k}^i}(\theta) = E[exp(\theta H_0^i(\mathbf{k}^i))]$. Then, by independent assumption we have $\psi_{\mathbf{k}^{i}}(\theta) = \prod_{j=1}^{N_{i}} \prod_{m=1}^{n_{ij}} \psi_{\mathbf{k}^{ijm}}^{ij}(\theta)$. Furthermore, let $\mathcal{D}^{i} = \{\theta > 0 : \psi_{\mathbf{k}^{i}}(\theta) < \infty, \ \mathbf{k}^{i} \in \mathcal{S}^{i}\}$ for i = 1, 2. We assume \mathcal{D}^i is non-empty and open. These technical assumptions are satisfied in most cases of practical interest which includes r.v.'s with phase-type distributions. For $\theta \in \mathcal{D}^i$, define the matrix

$$\mathbf{H}_{A}^{i}(\theta) = \left(\mathbf{P}^{i1}\boldsymbol{\Psi}^{i1}(\theta)\right)^{\otimes n_{i1}} \otimes \left(\mathbf{P}^{i2}\boldsymbol{\Psi}^{i2}(\theta)\right)^{\otimes n_{i2}} \otimes \cdots \otimes \left(\mathbf{P}^{iN_{i}}\boldsymbol{\Psi}^{iN_{i}}(\theta)\right)^{\otimes n_{iN_{i}}} \tag{6}$$

where $\Psi^{ij}(\theta) \equiv \text{diag}(\psi_k^{ij}(\theta), k \in S^{ij})$, and for any $\theta > 0$, define two-dimensional matrices

$$\mathbf{H}_{B}^{1}(\theta) = \begin{bmatrix} p_{2} & q_{2}e^{\theta c_{1}} \\ q_{1} & p_{1}e^{\theta c_{1}} \end{bmatrix}, \quad \mathbf{H}_{B}^{2}(\theta) = \begin{bmatrix} p_{1} & q_{1}e^{\theta c_{2}} \\ q_{2} & p_{2}e^{\theta c_{2}} \end{bmatrix}.$$
(7)

Furthermore, let $\rho_A^i(\theta) = sp(\mathbf{H}_A^i(\theta))$ and $\rho_B^i(\theta) = sp(\mathbf{H}_B^i(\theta))$ be the spectral radii of the matrices $\mathbf{H}_A^i(\theta)$ and $\mathbf{H}_B^i(\theta)$, $\mathbf{x}_A^i(\theta) = (x_A^{im}, 1 \le m \le \Pi_{j=1}^{N_i} |S^{ij}|^{n_{ij}})$ and $\mathbf{x}_B^i(\theta) = (x_A^{im}, 1 \le m \le \Pi_{j=1}^{N_i} |S^{ij}|^{n_{ij}})$ $(x_B^{i0}(\theta), x_B^{i1}(\theta))^T$ the positive right eigenvector corresponding to $\rho_A^{i}(\theta)$ and $\rho_B^{i}(\theta)$, where $|\mathcal{S}^{ij}|$ denotes the state number of state space \mathcal{S}^{ij} . Let $\Gamma_A^{i}(\theta) = \max_{0 \le k, l \le \prod_{i=1}^{N_i} |\mathcal{S}^{ij}|^{n_{ij}}} x_A^{ik}(\theta)/x_A^{il}(\theta)$, $\Gamma_B^i(\theta) = \max_{0 \le k, l \le 1} x_B^{ik}(\theta) / x_B^{il}(\theta).$ Then, the following properties hold (see Graham[17]).

Proposition 1: (i) $\rho_A^i(\theta) = \prod_{j=1}^{N_i} (\rho_A^{ij}(\theta))^{n_{ij}}$, where $\rho_A^{ij}(\theta) = sp(\mathbf{P}^{ij} \Psi^{ij}(\theta))$ is the spectral radii of matrix $\mathbf{P}^{ij} \Psi^{ij}(\theta)$ for $j = 1, \cdots, N_i$. (ii) $\mathbf{x}_A^i(\theta) = \bigotimes_{j=1}^{N_i} (\mathbf{x}_A^{ij}(\theta))^{\bigotimes n_{ij}}$, where $\mathbf{x}_A^{ij}(\theta) = (x_A^{ijm}, 1 \le m \le |\mathcal{S}^{ij}|)$ is the positive right

eigenvector corresponding to $\rho_A^{ij}(\theta)$ for $j = 1, \cdots, N_i$.

(iii)
$$\rho_B^i(\theta) = \frac{p_j + p_i e^{\theta c_i} + \sqrt{(p_j - p_i e^{\theta c_i})^2 + 4q_i q_j e^{\theta c_i}}}{2}, \quad i, j = 1, 2; i \neq j.$$

$$(\text{iv}) \quad \mathbf{x}_B^i(\theta) = (\frac{\rho_B^i(\theta) - p_i e^{\theta c_i}}{\rho_B^i(\theta) + q_i - p_i e^{\theta c_i}}, \quad \frac{q_i}{\rho_B^i(\theta) + q_i - p_i e^{\theta c_i}})^T, \quad i = 1, 2.$$

(v)
$$\Gamma_B^i(\theta) = \max\{\frac{q_i}{\rho_B^i(\theta) - p_i e^{\theta c_i}}, \frac{\rho_B^i(\theta) - p_i e^{\theta c_i}}{q_i}\}, \quad i = 1, 2$$

Applying the large deviation results on general Markov-modulated processes and chains (see Dembo and Zajic [7], and Chang [6]) to the processes $\{A_i^i, t \in \mathbf{N}\}$ and $\{B_i^i, t \in \mathbf{N}\}$, we get the following proposition.

- **Proposition 2:** (i) $\Lambda_A^i(\theta) = \log(\rho_A^i(\theta)) = \sum_{j=1}^{N_i} n_{ij} \log(\rho_A^{ij}(\theta))$ and $\Lambda_B^i(\theta) = \log(\rho_B^i(\theta))$, and both $\Lambda_A^i(\theta)$ and $\Lambda_B^i(\theta)$ are convex function of θ . (ii) The processes $\{S_t^{A^i}/t; t \in \mathbf{N}\}$ and $\{S_t^{B^i}/t; t \in \mathbf{N}\}$ satisfy the large deviation principle with the convex, good rate functions $\Lambda_A^{i*}(\alpha) = \sup_{\theta \in \mathbf{R}} \{\theta \alpha \Lambda_A^i(\theta)\}$ and $\Lambda_B^{i*}(\alpha) = \int_{\Omega}^{\Omega} e^{-i\theta \alpha} \int_{\Omega}^{\Omega} e^{-i\theta \alpha} e^{-i\theta \alpha} \int_{\Omega}^{\Omega} e^{-i\theta \alpha} e^{-i\theta \alpha} \int_{\Omega}^{\Omega} e^{-i\theta \alpha} e^{-i\theta \alpha}$ $\sup_{\theta \in \mathbf{R}} \{\theta \alpha - \Lambda_B^i(\theta)\},$ respectively.
- (iii) Let $\mathcal{F}_{t}^{A^{i}} = \sigma\{A_{\tau}^{i}; \tau \leq t\}$ and $\mathcal{F}_{t}^{B^{i}} = \sigma\{B_{\tau}^{i}; \tau \leq t\}$, then for all $\theta \in \mathbf{R}$ and $\tau, t \leq 0$, $\Lambda_A^i(\theta)t - \Gamma_A^i(\theta) \le \log E[e^{\theta S_{\tau,\tau}^{A^i} + t} | \mathcal{F}_{\tau}^{A^i}] = \log E[e^{\theta S_{\tau,\tau}^{A^i} + t} | A_{\tau}^i] \le \Lambda_A^i(\theta)t + \Gamma_A^i(\theta),$ a.s. $\Lambda_B^i(\theta)t - \Gamma_B^i(\theta) \le \log E[e^{\theta S_{\tau,\tau+t}^{B^i}} | \mathcal{F}_{\tau}^{B^i}] = \log E[e^{\theta S_{\tau,\tau+t}^{B^i}} | B_{\tau}^i] \le \Lambda_B^i(\theta)t + \Gamma_B^i(\theta),$ *a.s.*.

Other basic properties of $\Lambda_A^i(\theta)$, $\Lambda_B^i(\theta)$, $\Lambda_A^{i*}(\alpha)$ and $\Lambda_B^{i*}(\alpha)$ can be found in Dembo and Zeitouni[7], and Zhang[19] and [20].

E. Admission control and effective bandwidth

The admission control for the statistical multiplexing of bursty sources aims at a admitting a new connection application into a network only if it can be guaranteed a minimal QoS without violating the QoS of other connection applications already in the system. The main task of the admission control is to construct an admissible set for the numbers of all connection types. The theory bases is the concept of effective bandwidths, which has been originally developed for a single server system with multiplexing input streams. Consider a single buffer fluid model with the service rate c and the arrival process which is a superposition of number of independent processes $\{A_t^{jm}, t \in \mathbf{N}\}, 1 \leq j \leq N, 1 \leq m \leq n_j$. Assume that n_j streams of class j are identical distribution. For θ ($\theta > 0$), the effective bandwidth function of $\{A_t^{jm}, t \in \mathbf{N}\}$ is defined as follows:

$$eb_A^j(\theta) = \lim_{t \to \infty} \frac{1}{\theta t} \log E[exp(\theta S_t^{A^j})] = \frac{\Lambda_A^j(\theta)}{\theta}.$$
(8)

Let ζ^* be the solution to the equation

$$\sum_{j=1}^{N} n_j e b_A^j(\theta) = c.$$
(9)

Basing on the result of the large-buffer asymptotics on the tail probability of the steady state queue length L:

$$\lim_{t \to \infty} P\{L_t > x\} = P\{L > x\} \approx e^{-\zeta^* x}$$
(10)

as $x \to \infty$, one has that the QoS criterion $P\{L > x\} < \epsilon$ for cell loss probability is satisfied if $e^{-x\zeta^*} < \epsilon$, as $x \to \infty$ and $\epsilon \to 0$ such that $-\log(\epsilon)/x \to \zeta^*$. Then, the effective bandwidth associated with $\{A_t^{jm}, t \in \mathbf{N}\}$ is assigned to be $eb_A^j = eb_A^j(\zeta^*)$, and an admissible set for the vector (n_1, \dots, n_N) is defined as follows.

$$\mathcal{N} = \{ (n_1, \cdots, n_N); \quad \sum_{j=1}^N n_j e b_A^j \le c \}.$$
(11)

Therefore, the effective bandwidth actually is a number associated with each connection such that if the sum of the effective bandwidths of all connection onto a buffer is less than the output rate of that buffer, then QoS is satisfied. Recently, the concept of effective bandwidths has been extended to networks with priority classes by Berger and Whitt[1], [2] and [3]. For the polling system considered here, we want to find a similar admissible set for which the cell-loss probability requirement is satisfied by each queue. Concretely, let L_t^i be the queue length of Q_i at time t. Since the server allocates its capacity randomly between the two queues, we know that $\{L_t^i, t \in \mathbf{N}\}$ is affected not only by itself arrival processes, but also by the arrival processes of $Q_j, j \neq i$. Namely, the behavior of $\{L_t^i, t \in \mathbf{N}\}$ depends on both $(n_{11}, \dots, n_{1N_1})$ and $(n_{21}, \dots, n_{2N_2})$. Let ϵ_i be the cell-loss probability target for the traffic of Q_i . The purpose of the admission control is to satisfy the QoS criterion for the each class:

$$G_i((n_{11}, \cdots, n_{1N_1}), (n_{21}, \cdots, n_{2N_2})) = \lim_{t \to \infty} P\{L_t^i > x_i\} \le \epsilon_i$$
(12)

on the asymptotic region: $x_i \to \infty$ and $\epsilon_i \to 0$ such that $-\log(\epsilon_i)/x_i \to \zeta_i^* > 0$. The main work is to identify the following feasible region:

3. Upper bounds and admission control

In this section, we first give the large deviation upper bounds on the overflow probability of each queue, and then basing on the upper bounds, present an admission control criterion for the polling system. By $\{D_t^i, t \in \mathbf{N}\}$ and $\{E_t^i, t \in \mathbf{N}\}$, we denote the stationary and transient departure processes from an MMAP/MSP/1 queueing system with the arrival process $\{A_t^i, t \in \mathbf{N}\}$ and the service process $\{B_t^i, t \in \mathbf{N}\}$, respectively. Here MMAP and MSP denote Markov-Modulated Arrival Process and Markov service process. Note that the stability condition of this system is $\mathcal{A}^i < \mathcal{B}^i$. The effective bandwidths of $\{A_t^i, t \in \mathbf{N}\}$ and $\{B_t^i, t \in$ $\mathbf{N}\}$ are expressed by $eb_A^i(\theta)$ and $eb_B^i(\theta)$, respectively. Then, from Proposition 2, we have $eb_A^i(\theta) = \sum_{j=1}^{N_i} \sum_{m=1}^{n_{ij}} eb_A^{ijm}(\theta) = \sum_{j=1}^{N_i} \sum_{m=1}^{n_{ij}} \Lambda_A^{ijm}(\theta)/\theta = \sum_{j=1}^{N_i} n_{ij} \log(\rho_A^i(\theta))/\theta$ and $eb_B^i(\theta) = \Lambda_B^i(\theta)/\theta = \log(\rho_B^i(\theta))/\theta$. Then $\Lambda_A^i{}'(\theta) \equiv (\Lambda_A^i(\theta))' = \sum_{j=1}^{N_i} \sum_{m=1}^{n_{ij}} (\Lambda_A^{ijm}(\theta))' =$ $\sum_{j=1}^{N_i} n_{ij} (\rho_A^{ij}(\theta))'/\rho_A^{ij}(\theta)$ and $\Lambda_B^i{}'(\theta) \equiv (\Lambda_B^i(\theta))' = (\rho_B^i(\theta))'/\rho_B^i(\theta)$. Furthermore, we define $eb_D^i(\theta)$ as follows:

where, λ_r^i is the right end point of $dom\Lambda_A^{i*} = int(ran\Lambda_A^{i'})$, and δ_i^* is the largest solution to the equation $\Lambda_A^i(\theta) + \Lambda_B^i(-\theta) = 0$. $J_i(\theta) = (\theta - \tilde{\theta}_{A^i}^*(\theta) - \tilde{\theta}_{B^i}^*(\theta))\eta^{A^iB^i}(\theta) + \Lambda_A^i(\tilde{\theta}_{A^i}^*(\theta)) + \Lambda_B^i(\tilde{\theta}_{B^i}^*(\theta))$, where $\eta^{A^iB^i}(\theta)$ is the maximum point of the function $\theta\alpha - \Lambda_A^{i*}(\alpha) - \Lambda_B^{i*}(\alpha)$ in the interval $[\mathcal{B}^i, \Lambda_A^i'(\delta^*)]$, and for θ fixed, $\tilde{\theta}_{A^i}^*(\theta)$ and $\tilde{\theta}_{B^i}^*(\theta)$ are the unique solution of the equations $\Lambda_A^{i'}(\tilde{\theta}) = \eta^{A^iB^i}(\theta)$ and $\Lambda_B^{i'}(\tilde{\theta}) = \eta^{A^iB^i}(\theta)$, respectively. $K_i(\theta) = (\theta - \hat{\theta}_{A^i}^*(\theta) - \hat{\theta}_{B^i}^*(\theta))\xi^{A^iB^i}(\theta) + \Lambda_A^i(\hat{\theta}_{A^i}(\theta)) + \Lambda_B^i(\hat{\theta}_{B^i}^*(\theta))$, here $\xi^{A^iB^i}(\theta)$ is the maximum point of the function $\theta\alpha - \Lambda_A^{i*}(\alpha) - \Lambda_B^{i*}(\alpha)$ in the interval $[\mathcal{B}^i, \min\{\lambda_r^i, c_i\}]$, and for θ fixed, $\hat{\theta}_{A^i}^*(\theta)$ and $\hat{\theta}_{B^i}^*(\theta)$ are the unique solution of the equations $\Lambda_A^{i'}(\hat{\theta}) = \xi^{A^iB^i}(\theta)$ and $\Lambda_B^{i'}(\hat{\theta}) = \xi^{A^iB^i}(\theta)$, respectively.

As will be seen, $eb_D^i(\theta)$ actually is the effective bandwidths of the stationary departure $\{D_t^i, t \in \mathbf{N}\}$. Obviously, $eb_D^i(\theta)$ is a function of the integer number n_{ij} and the effective bandwidths $eb_A^{ij}(\theta)$ for $j = 1, 2, \dots, N_i$. Recall that L_t^i denotes the queue length of the queue Q_i at time t. Then under the stability condition (5), L_t^i converges in distribution to a finite random variable L_{∞}^i . Here we assume that the queue processes of the polling system have reached their steady state. Thus L_0^i has the same distribution as L_{∞}^i .

Theorem 1. Under the stability condition (5), the steady state queue length L_0^i of the queue Q_i satisfies the following upper bound: (i) For any $\theta \in \mathcal{D}^i$, if $eb_A^i(\theta) + v_i eb_D^j(v_i\theta) < c_i, j \neq i$, then

$$\limsup_{x \to \infty} \frac{1}{x} \log P\{L_0^i > x\} \le -\theta, \tag{15}$$

where, $v_i = c_i/c_j$, i, j = 1, 2.

(ii) The positive solution $\Theta_{ii}^*(v_i)$ of the equation:

$$eb_A^i(\theta) + v_i eb_D^j(v_i\theta) = c_i, \qquad j \neq i.$$
(16)

exists uniquely and $-\Theta_{ii}^*(v_i)$ is the tightest upper bound.

Before going to prove Theorem 3, we first present two lemmas on the effective bandwidth of the stationary departure $\{D_t^i, t \in \mathbf{N}\}$. First lemma is due to Theorem 2 of Chang and Zajic[7].

Lemma 1. Under $\mathcal{A}^i < \mathcal{B}^i$, for any $\alpha \in \mathbf{R}$,

$$\lim_{t \to \infty} \frac{1}{t} \log P\{S_t^{D^i} > \alpha t\} = -\inf_{x \ge \alpha} \Lambda_D^{i*}(x), \tag{17}$$

$$\lim_{t \to \infty} \frac{1}{t} \log E[e^{\theta S_t^{D^i}}] = \Lambda_D^i(\theta), \qquad \theta \ge 0,$$
(18)

where, $\Lambda_D^{i*}(\alpha) = \delta_i^* \alpha - \sup_{x \leq \alpha} \{ \delta_i^* x - \Lambda_A^{i*}(x) \} + \inf_{x \geq \alpha} \Lambda_B^{i*}(x)$

$$= \begin{cases} 0 & \text{if } \alpha < \mathcal{A}^{i} \\ \Lambda_{A}^{i*}(\alpha) & \text{if } \alpha \leq \Lambda_{A}^{i}'(\delta^{*}) \text{ and } \mathcal{A}^{i} < \alpha \leq \mathcal{B}^{i} \\ \Lambda_{A}^{i*}(\alpha) + \Lambda_{B}^{i*}(\alpha) & \text{if } \alpha \leq \Lambda_{A}^{i}'(\delta^{*}_{i}) \text{ and } \mathcal{B}^{i} < \alpha \leq \min\{\lambda_{r}^{i}, c_{i}\} \\ \delta_{i}^{*}\alpha - \Lambda_{A}^{i}(\delta_{i}^{*}) & \text{if } \alpha > \Lambda_{A}^{i}'(\delta^{*}_{i}) \text{ and } \mathcal{A}^{i} < \alpha \leq \mathcal{B}^{i} \\ \delta_{i}^{*}\alpha - \Lambda_{A}(\delta_{i}^{*}) + \Lambda_{B}^{*}(\alpha) & \text{if } \alpha > \Lambda_{A}^{i}'(\delta_{i}^{*}) \text{ and } \mathcal{B}^{i} < \alpha \leq \min\{\lambda_{r}^{i}, c_{i}\} \\ \infty & \text{if } \alpha > \min\{\lambda_{r}^{i}, c_{i}\}, \end{cases}$$
(19)

here, δ_i^* is the largest solution of the equation, $\Lambda_A^i(\theta) + \Lambda_B^i(-\theta) = 0$, and

$$\Lambda_D^i(\theta) = \sup_{\mathcal{A}^i \le \alpha} \{\theta \alpha - \Lambda_D^{i*}(\alpha)\} = \sup_{\mathcal{A}^i \le \alpha \le \min\{\lambda_r^i, c_i\}} \{\theta \alpha - \Lambda_D^{i*}(\alpha)\}.$$
 (20)

For the MMAP/MAP/1 queueing system, we can directly derive $\Lambda_D^i(\theta)$ as follows. The proof is similar to that of Theorem 3.3 in Feng et al.[12].

Lemma 2. Under $\mathcal{A}^i < \mathcal{B}^i$, for any $\theta > 0$, $\Lambda_D^i(\theta) = \theta e b_D^i(\theta)$, where $e b_D^i(\theta)$ is given in the formula (13).

The proof of Theorem 1: For the sake of convenience, we look backwards in time from time 0. As the Bernoulli service schedule is work-conserving, the evolutions of the two queues are governed by the following recursive equations:

$$L_{-t}^{1} = \max\{L_{-t-1}^{1} + A_{-t-1}^{1} - \max\{B_{-t-1}^{1}, c_{1} - v_{1}(L_{-t-1}^{2} + A_{-t-1}^{2})\}, 0\},$$
(21)

$$L_{-t}^{2} = \max\{L_{-t-1}^{2} + A_{-t-1}^{2} - \max\{B_{-t-1}^{2}, c_{2} - v_{2}(L_{-t-1}^{1} + A_{-t-1}^{1})\}, 0\}.$$
 (22)

Define $R_{-t}^i = \max\{B_{-t}^i, c_i - v_i(L_{-t}^j + A_{-t}^j)\}, i, j = 1, 2; i \neq j$. Then, R_{-t}^i denotes the amount of the service actually received by Q_i at time -t. Expanding (21) and (22) recursively we have

$$L_0^i = \max_{t \in \mathbf{N}} \{ S_{-t}^{A^i} - S_{-t}^{R^i} \}, \quad i = 1, 2,$$
(23)

where, $S_{-t}^{R^i} = \sum_{\tau=-t}^{-1} R_{\tau}^i$ is the total amount of the service actually received by Q_i in [-t, 0). Observing that

$$S_{-t}^{R^{i}} = L_{-t}^{i} + S_{-t}^{A^{i}} - L_{0}^{i}, \quad i = 1, 2,$$
(24)

we have $S_t^{R^i} \geq S_t^{A^i} - L_0^i$. Without loss of generality, we consider here the case that i = 1, j = 2. Then, the maximum in (22) for i = 1 must be achieved at the time when $L_t^1 = 0$. Let $-t \leq 0$ be the first time such that $L_{-t}^1 = 0$ and $L_{-\tau}^1 > 0$ for $\tau \in (0, t)$. Since the queue Q_1 is busy during the interval (-t, 0] and the Bernoulli service schedule is a work-conserving policy, the queue Q_1 gets at least $S_{-t}^{B^1}$ amount of the service (by considering the situation that the queue Q_2 may become empty between -t and 0). Thus, $S_{-t}^{R^1} \geq S_{-t}^{B^1}$. On the other hand, since $S_{-t}^{R^2}$ is the amount of the service actually received by Q_2 during the interval [-t, 0) and the service rate is c_2 , $S_{-t}^{R^2}/c_2$ is the time that the server spends in Q_2 during the interval [-t, 0). We have that $c_1(t - S_{-t}^{R^2}/c_2) = c_1t - v_1S_{-t}^{R^2}$ is the amount of the service received by Q_1 , where $v_1 = c_1/c_2$. Hence,

$$S_{-t}^{R^{1}} = \max\{ c_{1}t - v_{1}S_{-t}^{R^{2}}, c_{1}t - v_{1}S_{-t}^{R^{2}} \} = c_{1}t - v_{1}\min\{ S_{-t}^{R^{2}}, S_{-t}^{B^{2}} \}.$$
 (25)

Substituting (23) into (22) for i = 1 yields

$$L_0^1 = \max_{t \in \mathbf{N}} \{ S_{-t}^{A^1} + v_1 \min\{ S_{-t}^{R^2}, S_{-t}^{B^2} \} - c_1 t \}.$$
(26)

Note that under the stability condition (5) of the polling system, it is possible that for some i = 1 or 2, $\mathcal{A}^i \geq \mathcal{B}^i$. Thus, it is necessary to distinguish theses two cases in the proof.

CASE1: $\mathcal{A}^2 < \mathcal{B}^2$.

We introduce a MMAP/MSP/1 queueing system with the arrival process $\{A_t^2, t \in \mathbf{N}\}$ and the service process $\{B_t^2, t \in \mathbf{N}\}$. Let \tilde{L}_{-t}^2 be its queue length at time -t. Since this

$$L_0^2 \le \max_{t \in \mathbf{N}} \{ S_{-t}^{A^1} + v_1 \min\{ \tilde{L}_{-t} + S_{-t}^{A^2}, S_{-t}^{B^2} \} - c_1 t \} = \max_{t \in \mathbf{N}} \{ S_{-t}^{A^1} + v_1 S_{-t}^{M^2} - c_1 t \},$$
(27)

where $S_{-t}^{M^2} = \min\{\tilde{L}_{-t}^2 + S_{-t}^{A^2}, S_{-t}^{B^2}\}$. From Proposition 2, Lemma 1 and 2, we have for any $\epsilon > 0$,

$$\begin{split} E[e^{\theta L_0^2}] &\leq E[e^{\theta \max_{t \in \mathbf{N}} \{S_{-t}^{A^2} + v_1 S_{-t}^{M^2} - c_1 t\}}] \leq \sum_{t \in \mathbf{N}} E[e^{\theta (S_{-t}^{A^2} + v_1 S_{-t}^{M^2} - c_1 t)}] \\ &\leq C_{\epsilon} + \sum_{t \geq t_{\epsilon}} e^{(\Lambda_A^1(\theta) + \Lambda_D^2(v_1\theta) + 2\epsilon - c_1\theta)t} \end{split}$$

where t_{ϵ} is sufficient large and C_{ϵ} is a constant dependent on ϵ . Therefore, we have $E[e^{\theta L_{0}^{1}}] < \infty$ if $\Lambda_{A}^{1}(\theta) + \Lambda_{D}^{2}(v_{1}\theta) + 2\epsilon - c_{1}\theta < 0$. By Chebyshev's inequality, $P\{L_{0}^{1} > x\} \leq e^{-\theta x} E[e^{\theta L_{0}^{1}}]$ for any $x \geq 0$. Then, using the definition of $eb_{A}^{1}(\theta)$ and $eb_{D}^{2}(\theta)$, we have that if $eb_{A}^{1}(\theta) + v_{1}eb_{D}^{2}(v_{1}\theta) + 2\epsilon/\theta - c_{1} < 0$,

$$\limsup_{x \to \infty} \frac{1}{x} P\{L_0^1 > x\} \le -\theta.$$
(28)

Taking $\epsilon \to 0$, we get the upper bound (15) in this case.

CASE2: $\mathcal{A}^2 \geq \mathcal{B}^2$ By (23), we have

$$L_0^1 \le \max_{t \in \mathbf{N}} \{ S_{-t}^{A^1} + v_1 S_{-t}^{B^2} - c_1 t \}.$$
(29)

For any $\theta > 0$, similarly, if $\Lambda_{A^1}(\theta) + \Lambda_{B^2}(v_1\theta) + 2\epsilon - c_1\theta < 0$, then,

$$E[e^{\theta L_0^1}] \le \sum_{t \in \mathbf{N}} E[e^{\theta (S_{-t}^{A^2} + v_1 S_{-t}^{M^2} - c_1 t)}] < \infty.$$

Again by Chebyshev's inequality, if $eb_A^1(\theta) + v_1 eb_{B^2}(v_1\theta) + 2\epsilon/\theta - c_1 < 0$,

$$\limsup_{x \to \infty} \frac{1}{x} \log P\{L_0^1 > x\} \le -\theta.$$
(30)

Taking $\epsilon \to 0$ and noting that $eb_D^2(\theta) = eb_B^2(\theta)$ in this case, we establish (15).

Next consider the assertion (ii) of Theorem 1. Write the equation $eb_A^i(\theta) + v_i eb_D^j(v_i\theta) = c_i$ as $\Lambda_A^i(\theta) + \Lambda_D^j(v_i\theta) = c_i\theta$. Note that $\{D_t^i, t \in \mathbf{N}\}$ is the stationary departure process from the MMAP/MSP/1 queueing system with the arrival process $\{A_t^i, t \in \mathbf{N}\}$ and the service process $\{B_t^i, t \in \mathbf{N}\}$. Then, $D_t^i \leq B_t^i$ for $t \in \mathbf{N}$. It follows that $E[D_0^j] \leq \mathcal{B}^j$. Hence, if $\mathcal{A}^i < \mathcal{B}^i$, we have $(\Lambda_A^i{}'(\theta) + v_i\Lambda_D^j{}'(\theta))|_{\theta=0} = \mathcal{A}^i + v_iE[D_0^j] < \mathcal{B}^i + v_i\mathcal{B}^j = c_i$. Otherwise, we must have that $\mathcal{A}^j < \mathcal{B}^j, j \neq i$ from the stability condition (5). Thus, the MMAP/MSP/1 queueing system with the arrival process $\{A_t^j, t \in \mathbf{N}\}$ and the service process $\{B_t^j, t \in \mathbf{N}\}$ is stable. We have $E[D_0^j] = \mathcal{A}^j$. Again by the stability condition (5), $\mathcal{A}^i + v_i E[D_0^j] < \mathcal{A}^i + v_i \mathcal{A}^j < \mathcal{B}^i + v_i \mathcal{B}^j = c_i$. Finally, by the convexity of $\Lambda_A^i(\theta)$ and $\Lambda_D^j(\theta)$, we obtain the assertion (ii). These complete the proof. \Box

Remarks: (1) Note that $\Theta_{ij}(v_i)$ is the function of the integer numbers n_{ij} and the

effective bandwidths $eb_A^{ij}(\theta)$ and the effective bandwidth $eb_B^i(\theta)$. (2) In fact, by the similar method in [12] and [13], we can also derive the lower bounds on

(2) In fact, by the similar method in [12] and [13], we can also derive the lower bounds on the overflow probability. Since only the upper bounds of the overflow probability will be used to determine the admission control criterion, here we give the lower bounds without proof.

Theorem 2. Under the stability condition (5), the steady state queue length L_0^i of the queue Q_i satisfies the following lower bound:

$$\liminf_{x \to \infty} \frac{1}{x} \log P\{L_0^i > x\} \ge -\theta_{ij}^*(l_i),$$

where, $\theta_{ij}^*(l_i)$ is the unique solution of the equation:

$$eb_A^i(\theta) + l_i eb_E^j(l_i\theta) = c_i, \quad i \neq j,$$

here, $l_i = \max\{v_i, 1\} \mathbf{1}_{\{\mathcal{A}^j < \mathcal{B}^j\}} + v_i \mathbf{1}_{\{\mathcal{A}^j \ge \mathcal{B}^j\}}, \mathbf{1}_C$ is the indicator function of the set C, and $eb_E^i(\theta)$ is defined as follows: for any $\theta \ge 0$,

$$eb_{E}^{i}(\theta) = \begin{cases} CASE1. \ \mathcal{A}^{i} < \mathcal{B}^{i} \\ eb_{A}^{i}(\theta) & \text{if } \theta : \Lambda_{A}^{i} '(\theta) \leq \mathcal{B}^{i} \\ K_{i}(\theta)/\theta & \text{if } \theta : \Lambda_{A}^{i} '(\theta) > \mathcal{B}^{i} \\ CASE2. \ \mathcal{A}^{i} \geq \mathcal{B}^{i} \\ \mathcal{B}^{i}. \end{cases}$$

By the upper bounds (15), for x_i sufficiently large, we have $P\{L_0^i > x_i\} \leq e^{-\Theta_{ij}^* x_i}$. Therefore, if $e^{-\Theta_{ij}^* x_i} \leq \epsilon_i$, i.e., $\zeta_i^* \equiv -\log(\epsilon_i)/x_i \leq \Theta_{ij}^*$, then it holds that $P\{L_0^i > x_i\} \leq \epsilon_i$. From Theorem 1(ii), this means that

$$eb_A^i(\zeta_i^*) + eb_D^j(\zeta_i^*) \le c_i.$$
(31)

Note that $eb_A^i(\zeta_i^*) = \sum_{j=1}^{N_i} n_{ij}eb_a^{ij}(\zeta_i^*) \equiv eb_A^i(\zeta_i^*; n_{i1}, \cdots, n_{iN_i})$ is a linear function of n_{i1}, \cdots, n_{iN_i} , and $eb_D^j(\zeta_i^*) \equiv eb_D^j(\zeta_i^*; n_{j1}, \cdots, n_{jN_j})$ is a nonlinear function of n_{i1}, \cdots, n_{iN_i} . Thus, (31) is a nonlinear constraint on the integer numbers $n_{11}, \cdots, n_{1N_1}; n_{21}, \cdots, n_{2N_2}$. Let \mathcal{N} be the set of points $((n_{11}, \cdots, n_{1N_1}); (n_{21}, \cdots, n_{2N_2}))$ such that

$$\mathcal{N} = \{ ((n_{11}, \cdots, n_{1N_1}), (n_{21}, \cdots, n_{2N_2})) : \\ eb_A^1(\zeta_1^*; n_{11}, \cdots, n_{1N_1}) + eb_D^2(\zeta_1^*; n_{21}, \cdots, n_{2N_2}) \le c_1, \\ eb_A^2(\zeta_2^*; n_{21}, \cdots, n_{2N_2}) + eb_D^1(\zeta_2^*; n_{11}, \cdots, n_{1N_1}) \le c_2 \}.$$
(32)

We have the following theorem.

Theorem 3. $\mathcal{N} \subset \mathcal{K}$, i.e., the QoS criteria

$$G_i((n_{11}, \cdots, n_{1N_1}), (n_{21}, \cdots, n_{2N_2})) \le \epsilon_i, \quad i = 1, 2$$

are satisfied if

$$eb_{A}^{i}(\zeta_{i}^{*}; n_{i1}, \cdots, n_{iN_{i}}) + eb_{D}^{j}(\zeta_{i}^{*}; n_{j1}, \cdots, n_{jN_{j}}) \le c_{i}, \quad i, j = 1, 2; j \neq i$$

$$(33)$$

where $\zeta_i^*, i = 1, 2$ is given by

$$\zeta_i^* = -\frac{\log \epsilon_i}{x_i}.$$

4. Algorithm for two classes case

In this section we present an algorithm to calculate the admissible set for a polling model that each queue has only one input class of traffic, i.e., $N_1 = N_2 = 1$. Assume that n_i input streams $\{a_t^{im}, t \in \mathbf{N}\}$ of Q_i are independent and identical stationary Makovian on/off sources. That is, $\{a_t^{im}, t \in \mathbf{N}\}$ is a stationary Markov chain with the state space $\{0, 1\}$ and transition matrix

$$\mathbf{P}_{a}^{i} = \begin{bmatrix} 1 - \alpha^{i} & \alpha^{i} \\ \beta^{i} & 1 - \beta^{i} \end{bmatrix}.$$
(34)

While in the state on (denoted by 1), the source produces information of traffic at a constant r_i ; while in the state off (denoted by 0), it produces no information. The equilibrium distribution of $\{a_t^{im}, t \in \mathbf{N}\}$ is $\pi_a^i = (\alpha^i/(\alpha^i + \beta^i), \beta^i/(\alpha^i + \beta^i))$. The aggregate arrival process A_t^i of Q_i can be described as $A_t^i = \sum_{m=1}^{n_i} a_t^{im} r_i$, and $\mathcal{A}^i \equiv E[A_t^i] = \sum_{m=1}^{n_i} \beta^i/(\alpha^i + \beta^i)n_ir_i$. Since each input stream $\{a_t^{im}, t \in \mathbf{N}\}$ is a two-state Markov chain, we can easily calculate $\rho_a^i(\theta) = sp(\mathbf{P}_a^i \mathbf{\Psi}_a^i(\theta))$ -the spectral radii of the matrix $\mathbf{P}_a^i \mathbf{\Psi}_a^i(\theta)$ as follows:

$$\rho_a^i(\theta) = \frac{1 - \alpha^i + (1 - \beta^i)e^{\theta r_i} + \sqrt{((1 - \alpha^i) - (1 - \beta^i)e^{\theta r_i})^2 + 4\alpha^i\beta^i e^{\theta r_i}}}{2}.$$
 (35)

We have $\Lambda_A^i(\theta) = n_i \log(\rho_a^i(\theta))$ and the effective bandwidth $eb_A^i(\theta, n_i) = n_i \log(\rho_a^i(\theta))/\theta \equiv n_i eb_a^i(\theta)$, where $eb_a^i(\theta)$ denotes the effective bandwidth of $\{a_t^i r_i, t \in \mathbf{N}\}$. Similarly, the admission control policy for this special case can be described as follows: under the condition that n_1 class-1 streams and n_2 class-2 streams are transmitting, if a new stream arrives into the polling system, the admission control scheme decides whether or not to admit this stream. A simple admission control scheme is an admissible region such that all points within it denote the numbers of class-1 and class-2 streams for which QoS of each queue is satisfied. Namely,

$$\mathcal{K} = \{ (n_1, n_2) : G_1(n_1, n_2) \le \epsilon_1, G_2(n_1, n_2) \le \epsilon_1 \}.$$
(36)

Now the subset \mathcal{N} of \mathcal{K} obtained by using the upper bounds becomes to:

$$\mathcal{N} = \{ (n_1, n_2) : eb_A^1(\zeta_1^*; n_1) + eb_D^2(\zeta_1^*; n_2) \le c_1, eb_A^2(\zeta_2^*; n_2) + eb_D^1(\zeta_2^*; n_1) \le c_2 \}.$$
(37)

Since for any $\theta \geq 0$ and any $n_i \geq 0$, $eb_D^i(\theta, n_i) \geq 0$. We have that the maximum numbers n_1^* and n_2^* of input streams under which QoS of each queue is ensured should be the maximum numbers satisfying inequalities $eb_A^1(\zeta_1^*; n_1) \leq c_1$ and $eb_A^2(\zeta_2^*; n_2) \leq c_2$. Thus, we can take $n_i^* = \lfloor \zeta_i^* c_i / \log(\rho_a^i(\zeta_i^*)) \rfloor$ for i = 1, 2. In the following, we give an algorithm to determine the admissible set \mathcal{N} .

Algorithm.

Step 1. For
$$i = 1, 2$$
 and $\theta \ge 0$, put
(i) $\rho_a^i(\theta) = \frac{1 - \alpha^i + (1 - \beta^i)e^{\theta r_i} + \sqrt{((1 - \alpha^i) - (1 - \beta^i)e^{\theta r_i})^2 + 4\alpha^i\beta^i e^{\theta r_i}}}{2}$
 $\Lambda_A^i(\theta, n_i) = n_i \log(\rho_a^i(\theta))$ and $eb_A^i(\theta, n_i) = n_i \log(\rho_a^i(\theta))/\theta$.

(ii)
$$\begin{split} \rho_b^i(\theta) &= \frac{p^j + p^i e^{\theta c_i} + \sqrt{(p^j - p^i e^{\theta c_i})^2 + 4q^i q^j e^{\theta c_i}}}{2}, \ j = 1, 2; j \neq i, \\ \Lambda_B^i(\theta) &= \log(\rho_b^i(\theta)) \text{ and } eb_B^i(\theta) = \log(\rho_b^i(\theta))/\theta. \end{split}$$

- Step 2. For the given ϵ_i and x_i , calculate $\zeta_i^* = -\log(\epsilon_i)/x_i$, and the maximum numbers of the input streams in each queue $n_i^* = \lfloor \zeta_i^* c_i / \log(\rho_a^i(\zeta_i^*)) \rfloor$.
- Step 3. For i = 1, 2, determine the set $\mathcal{N}_1 = \{(n_1, n_2), eb_A^1(\zeta_1^*; n_1) + eb_D^2(\zeta_1^*; n_2) \le c_1\}$ and $\mathcal{N}_2 = \{(n_1, n_2), eb_A^2(\zeta_2^*; n_2) + eb_D^1(\zeta_2^*; n_1) \le c_2\}$ as follows: first, set

$$\mathcal{N}_1 = \{(0,0), (1,0), \cdots, (n_1^*, 0)\}$$
 and $\mathcal{N}_2 = \{(0,0), (0,1), \cdots, (0, n_2^*)\}$

For i = 1 to 2,

For $n_i = 0$ to n_i^* ,

- For $n_j = 1$ to n_j^*
- (i) Find $\delta_j^*(n_j)$ —the largest root of the equation $\Lambda_A^j(\theta, n_j) + \Lambda_B^j(-\theta) = 0$, or equivalently, the largest root of the equation $(\rho_a^j(\theta))^{n_j} = \rho_b^j(-\theta)^{-1}$.
- (ii) calculate $\Lambda_A^{j'}(\delta_j^*(n_j), n_j)$, where

$$\Lambda_{A}^{j\,\prime}(\theta,n_{j}) = \frac{n_{j}r_{j}e^{\theta r_{j}}}{2\rho_{a}^{j}(\theta)} \left\{ (1-\beta^{j}) + \frac{2\alpha^{j}\beta^{j} - (1-\beta^{j})((1-\alpha^{j}) - (1-\beta^{j})e^{\theta r_{j}})}{\sqrt{((1-\alpha^{j}) - (1-\beta^{j})e^{\theta r_{j}})^{2} + 4\alpha^{j}\beta^{j}e^{\theta r_{j}}}} \right\}.$$

(iii) Distinguish the following four cases:

 $\begin{array}{ll} \text{CASE1. if} & \mathcal{A}^{j} < \Lambda_{A}^{j} \left(\delta_{j}^{*}(n_{j}), n_{j} \right) \leq \mathcal{B}^{j} < \min\{\lambda_{r}^{j}, c_{j}\}, \text{ then go to (1).} \\ \text{CASE2. if} & \mathcal{A}^{j} < \mathcal{B}^{j} < \Lambda_{A}^{j} \left(\delta_{j}^{*}(n_{j}, n_{j}) \leq \min\{\lambda_{r}^{j}, c_{j}\}, \text{ then go to (2).} \\ \text{CASE3. if} & \mathcal{A}^{j} < \mathcal{B}^{j} < \min\{\lambda_{r}^{j}, c_{j}\} < \Lambda_{A}^{j} \left(\delta_{j}^{*}(n_{j}), n_{j} \right), \text{ then go to (3).} \\ \text{CASE4. if} & \mathcal{A}^{j} \geq \mathcal{B}^{j}, \text{ then go to (4).} \end{array}$

(1) calculate $eb_D^j(\zeta_i^*; n_j)$ for CASE1.

$$eb_{D}^{j}(\zeta_{i}^{*};n_{j}) = \begin{cases} n_{j} eb_{a}^{j}(\zeta_{i}^{*}) & \text{if } \zeta_{i}^{*} \leq \delta_{j}^{*}(n_{j}) \\ \frac{\delta_{j}^{*}(n_{j})}{\zeta_{i}^{*}}n_{j} eb_{a}^{j}(\delta_{j}^{*}(n_{j}) + \frac{\zeta_{i}^{*} - \delta_{j}^{*}(n_{j})}{\zeta_{i}^{*}} & \text{if } \delta_{j}^{*}(n_{j}) < \zeta_{i}^{*} \text{ and} \\ \times eb_{b}^{j}(\zeta_{i}^{*} - \delta_{j}^{*}(n_{j})) & \mathcal{B}^{j} \leq \Lambda_{B}^{j}(\zeta_{i}^{*} - \delta_{j}^{*}(n_{j})) \\ \leq \min\{\lambda_{r}^{j}, c_{j}\} & \frac{\delta_{j}^{*}(n_{j})}{\zeta_{i}^{*}}n_{j} eb_{a}^{j}(\delta_{j}^{*}(n_{j})) + \frac{\zeta_{i}^{*} - \delta_{i}^{*}(n_{j})}{\zeta_{i}^{*}}\min\{\lambda_{r}^{j}, c_{j}\} & \text{if } \delta_{j}^{*}(n_{j}) < \zeta_{i}^{*} \text{ and} \\ -\frac{\min\{\lambda_{r}^{j}, c_{j}\}}{\zeta_{i}^{*}} eb_{b}^{j}(\min\{\lambda_{r}^{j}, c_{j}\}) & \min\{\lambda_{r}^{j}, c_{j}\} < \Lambda_{B}^{j'}(\zeta_{i}^{*} - \delta_{i}^{*}(n_{j})), \end{cases}$$

then go to (iv).

(2) calculate $eb_D^j(\zeta_i^*; n_j)$ for CASE2. (2.1) find $\eta^{A^j B^j}(\zeta_i^*)$ —the maximum point of the function $f(\alpha) = \zeta_i^* - \Lambda_A^{j*}(\alpha) - \Lambda_B^{j*}(\alpha)$ on the interval $[\mathcal{B}^j, \Lambda_A^{j'}(\delta_j^*(n_j))]$. (2.2) find $\tilde{\theta}_{A^j}^*(\zeta_i^*)$ and $\tilde{\theta}_{B^j}^*(\zeta_i^*)$ —the unique solution of the equations $\Lambda_A^{j'}(\tilde{\theta}, n_j) = \eta^{A^j B^j}(\zeta_i^*)$ and $\Lambda_B^{j'}(\tilde{\theta}) = \eta^{A^j B^j}(\zeta_i^*)$, where

$$\begin{split} \Lambda_{B}^{j\,\prime}(\theta) &= \frac{c_{j}e^{\theta c_{j}}}{2\rho_{b}^{j}(\theta)} \left\{ p^{j} + \frac{2q^{j}q^{i} - p^{j}(p^{i} - p^{j}e^{\theta c_{j}})}{\sqrt{(p^{i} - p^{j}e^{\theta c_{j}})^{2} + 4q^{j}q^{i}e^{\theta c_{j}}}} \right\}, \quad i \neq j. \end{split}$$

$$(2.3) \text{ put} \\ J_{j}(\zeta_{i}^{*}) &= (\zeta_{i}^{*} - \tilde{\theta}_{A^{j}}^{*}(\zeta_{i}^{*}) - \tilde{\theta}_{B^{j}}^{*}(\zeta_{i}^{*}))\eta^{A^{j}B^{j}}(\zeta_{i}^{*}) + \Lambda_{A}^{j}(\tilde{\theta}_{A^{j}}^{*}(\zeta_{i}^{*})) + \Lambda_{B}^{j}(\tilde{\theta}_{B^{j}}^{*}(\zeta_{i}^{*})). \\ (2.4) \text{ calculate } eb_{D}^{j}(\zeta_{i}^{*}; n_{j}): \\ eb_{D}^{j}(\zeta_{i}^{*}) &= (\Lambda_{A}^{j}/(\zeta_{i}^{*}) > B^{j}, \zeta_{i}^{*} \le \delta_{j}^{*}(n_{j}) \text{ or } \\ \Lambda_{A}^{j}/(\zeta_{i}^{*}) > B^{j}, \zeta_{i}^{*} \le \delta_{j}^{*}(n_{j}) \text{ or } \\ \Lambda_{B}^{j}'(\zeta_{i}^{*}) > B^{j}, \zeta_{i}^{*} > \delta_{j}^{*}(n_{j}) \text{ and } \\ \Lambda_{B}^{j}'(\zeta_{i}^{*}) > B^{j}, \zeta_{i}^{*} > \delta_{j}^{*}(n_{j}) \text{ and } \\ \Lambda_{B}^{j}'(\zeta_{i}^{*}) > B^{j}, \zeta_{i}^{*} > \delta_{j}^{*}(n_{j}) \text{ and } \\ + \frac{\zeta_{i}^{*} - \delta_{j}^{*}(n_{j})}{\zeta_{i}^{*}} eb_{b}^{j}(\zeta_{i}^{*} - \delta_{j}^{*}(n_{j}))\} \qquad \Lambda_{A}^{j}'(\delta_{i}^{*}(n_{j})) < \Lambda_{B}^{j}'(\zeta_{i}^{*} - \delta_{j}^{*}(n_{j})) \\ \max\{\frac{J_{j}(\zeta_{i}^{*})}{\zeta_{i}^{*}}, \quad \frac{\delta_{j}^{*}(n_{j})}{\zeta_{i}^{*}} eb_{b}^{j}(\delta_{j}^{*}(n_{j})) + \frac{\zeta_{i}^{*} - \delta_{j}^{*}(n_{j})}{\zeta_{i}^{*}} \text{ if } \Lambda_{A}^{j}'(\zeta_{i}^{*}) > B^{j}, \zeta_{i}^{*} > \delta_{j}^{*}(n_{j}) \text{ and } \\ \times \min\{\lambda_{r}^{j}, c_{j}\} - \frac{\min\{\lambda_{r}^{j}, c_{j}\}}{\zeta_{i}^{*}} eb_{B}^{j}(\min\{\lambda_{r}^{j}, c_{j}\})\} \qquad \min\{\lambda_{r}^{j}, c_{j}\} < \Lambda_{B}^{j}'(\zeta_{i}^{*} - \delta_{j}^{*}(n_{j})) \end{cases}$$

then go to (iv).

=

(3) calculate $eb_D^j(\zeta_i^*; n_j)$ for CASE3. (3.1) find $\xi^{A^iB^i}(\zeta_j^*)$ —the maximum point of the function $f(\alpha) = \zeta_j^* - \Lambda_A^{i*}(\alpha) -\Lambda_B^{i*}(\alpha)$ on the interval $[\mathcal{B}^i, \min\{\lambda_r^j, c_j\}]$. (3.2) find $\hat{\theta}_{A^i}(\zeta_j^*)$ and $\hat{\theta}_{B^i}^*(\zeta_j^*)$ —the unique solution of the equations $\Lambda_A^{i}'(\hat{\theta}) = \xi^{A^iB^i}(\zeta_j^*)$ and $\Lambda_B^{i}'(\hat{\theta}) = \xi^{A^iB^i}(\zeta_j^*)$. (3.3) put $K_i(\zeta_j^*) = (\zeta_j^* - \hat{\theta}_{A^i}^*(\zeta_j^*) - \hat{\theta}_{B^i}^*(\zeta_j^*))\xi^{A^iB^i}(\zeta_j^*) + \Lambda_A^i(\hat{\theta}_{A^i}^*(\zeta_j^*)) + \Lambda_B^i(\hat{\theta}_{B^i}^*(\zeta_j^*))$. (3.4) calculate $eb_D^j(\zeta_i^*; n_i)$:

$$eb_D^j(\zeta_i^*; n_j) = \begin{cases} n_j \ eb_a^j(\zeta_i^*) & \text{if } \Lambda_A^{j'}(\zeta_i^*) \le \mathcal{B}^j, \\ \frac{K_j(\zeta_i^*)}{\zeta_i^*} & \text{if } \Lambda_A^{j'}(\zeta_i^*) > \mathcal{B}^j, \end{cases}$$

then go to (iv).

(4) calculate $eb_D^j(\zeta_i^*; n_j)$ for CASE4.

 $eb_D^j(\zeta_i^*; n_j) = eb_b^j(\zeta_i^*)$, then go to (iv).

(iv) If $eb_A^i(\zeta_i^*; n_i) + eb_D^j(\zeta_i^*; n_j) \le c_i$, then next n_j , else go to (v). (v) set $n_j^*(n_i) = n_j$, and

$$\mathcal{N}_{i} = \mathcal{N}_{i} \cup \{ (n_{i}, 0), (n_{i}, 1), \cdots, (n_{i}, n_{i}^{*}(n_{i})) \}.$$

Next n_i

Next i.

Step 4. Set $\mathcal{N} = \mathcal{N}_1 \cap \mathcal{N}_2$. Then, end.

5. Conclusion

In this paper we have considered an admission control problem for the polling system consisting of two-parallel queues and a single server, under the Bernoulli service schedule. Basing on the effective bandwidth theory and the large deviation bounds, we have derived an admission control criterion under which QoS of each queue is guaranteed, and presented an algorithm to find the admissible set. As known, the effective bandwidth approach based completely on large-buffer asymptotics often give a very conservative approximation, e.g., see [1], [2]. Moreover, the large deviation upper and lower bounds used here are not matching. This fact also affects the accuracy of approximation. Nevertheless, we believe that our results can be very useful because they identify an appropriate structure for the admissible set for the polling system. Furthermore, since $eb_D^j(\zeta_i^*; n_{i1}, \cdots, n_{iN_i})$ is non-linear function of n_{i1}, \dots, n_{iN_i} , the resulting admissible set (32) has nonlinear constraints for each class of each queue. These non-linear constraints make the notation of effective bandwidths lose much of its original meaning: i.e., assigning an effective bandwidth to each connection of each type. However, we can produce a smaller admissible set with linear constraint boundary by using the similar method in [2,3], e.g., approximating the admissible set (32) by a linear hyperplane chosen to be tangent to the admissible set at some point of typical operating region, say, n^* . With this admissible set, we can assign an effective bandwidth to each connection of each type in each queue. As a future work, it should be worthwhile to consider the problem that if taking the probabilities of routing servers, $p^i, i = 1, 2$ as control variables, whether or not there exist optimal values of p^i , i = 1, 2 such that they give a maximum admissible set.

Acknowledgements

The authors wish to thank the referee for useful suggestions.

References

- A.W. Berger and W. Whitt, Effective bandwidths with priorities, IEEE/ACM, Trans. on Networking, 6(4) (1998), 447-460.
- [2] D. Bertsimas, I.C. Paschalidis and J.N. Tsitsiklis, Asympotic buffer overflow probabilities in Multiclass multiplexers: An optimal control approach, IEEE Trans. Automat. Contr., 43(3) (1998), 315-335.
- [3] D. Bertsimas, I.C. Paschalidis and J.N. Tsitsiklis, Large deviations analysis of the generalized processor sharing policy, Queueing Systems, 32 (1999), 319-349.
- [4] 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 (1995), 293-320.
- [5] J.A. Bucklew, Large-Deviation Techniques in Decision, Simulation and Estimation. (Wiley, New York, 1990).

- [6] C.S. Chang, Sample path large deviations and intree network, Queueing Systems, 20 (1995), 45-66.
- [7] C.S. Chang and T. Zajic, Effective bandwidths of departure processes from queues with time varying capacities, in: Proc. IEEE INFOCOM'95, (1995) 1001-1009.
- [8] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications. (Jones and Bartlett Publishers, 1993).
- [9] N.G. Duffield, Exponential bounds for queues with Makovian arrivals, Queueing Systems, 17 (1994), 413-430.
- [10] A.I. Elwalid and D. Mitra, Effective bandwidth of general Markovian traffic sources and admission control of high-speed networks, IEEE/ACM Trans. on Networking 1(3) (1993), 329-343.
- [11] W. Feng, M. Kowada and K. Adachi, A two-queue model with Bernoulli service schedule and switching times, Queueing Systems, 30 (1998), 405-434.
- [12] W. Feng, K. Adachi and M. Kowada, Large deviation approximation for a polling system with Markovian on/off sources and Bernoulli service schedule, submitted to publication.
- [13] W. Feng, K. Adachi and M. Kowada, Large deviation analysis of a polling system with twoqueue and multiple-server, submitted to publication.
- [14] N. Gautam and V.G. Kulkarni, Applications of SMP bounds to multi-class traffic in high-speed networks, Queueing Systems, 36 (2000), 352-379.
- [15] R.J. Gibbens and P.J. Hunt, Effective bandwidth for Multi-type UAS channel, Queueing Systems, 9 (1991), 17-28.
- [16] P.W. Glynn and W. Whitt, Logarithmic asymptotics for steady-state tail probabilities in a single-server queue, J. Appl. Probab., 31 (1994), 131-156.
- [17] A. Graham, Kronecher Products and Matrix Calculus with Applications. (Chischester: Ellis Horwood, 1981).
- [18] D.S. Lee, Analysis of a cyclic server queue with Bernoulli schedules, J. Appl. Prob., 34 (1997), 176-191.
- [19] Z. Liu, P. Nain and D. Towsley, Exponential bounds with applications to call admission, J. Assoc. Comput. 44 (1997), 366-394.
- [20] R.M. Loynes, The stability of a queue with non-independent inter-arrival and service times, Proc. Cambridge Philos. Soc., 58 (1962), 497-520.
- [21] F.P. Kelly, Effective bandwidths for multi-classes queues, Queueing Systems, 9 (1991), 5-16.
- [22] G. Kesidis, J. Warand and C.S. Chang, Effective bandwidths for multiclass Markov fluids and other ATM sources, IEEE/ACM Trans. on Networking 1(4) (1993), 424-428.
- [23] V.G. Kulkarni and N. Gautam, Admission control of multi-class trffic with service priorities in high-speed networks, Queueing Systems, 27 (1997), 79-97.
- [24] A. Weiss, An introduction to large deviations for communication networks, IEEE J. Selec. Areas Commun., 13(6) (1993), 938-385.
- [25] Z.-L. Zhang, Z. Liu, J. Kurose and D. Towsley, Call admission control schedule under generalized proceesor sharing scheduling, Telecommunication Systems, 7 (1997), 125-152.
- [26] Z.-L. Zhang, Large deviations and the generalized processor sharing schedule for a two-queue system, Queueing Systems, 26 (1997), 229-254.
- [27] Z.-L. Zhang, Large deviations and the generalized processor sharing schedule for a multiplequeue system, Queueing Systems, 28 (1998), 349-276.
- [28] W. Whitt, Tail probabilities with statistical multiplexing and effective bandwidths in multiclass queue, Telecommunication Systems, 2 (1993), 71-107.

[†] Department of Systems Engineering, Nagoya Institute of Technology

[‡]School of Culture-Information Studies, Sugiyama Jogakuen University