LARGE DEVIATIONS BOUNDS FOR A POLLING SYSTEM WITH MARKOVIAN ON/OFF SOURCES AND BERNOULLI SERVICE SCHEDULE

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

Received September 10, 2004; revised May 8, 2006

Abstract. In this paper we consider a large deviations problem for a discrete-time polling system consisting of two-parallel queues and a single server. The arrival process of each queue is a superposition of traffic streams generated by a number of mutually independent and identical Markovian on/off sources, and the single server serves the two queues according to the so-called Bernoulli service schedule. Using the large deviations techniques, we derive the upper and lower bounds of the probability that the queue length of each queue exceeds a certain level (i.e., the buffer overflow probability). These results have important implications for traffic management of high-speed communication networks such as call admission control and bandwidth allocation.

1. Introduction

Polling systems consisting of two-parallel queues and a single server have been extensively applied to modelling high-speed communication systems with two types of traffic: real-time traffic (e.g., voice and video) and non-real-time traffic (e.g., data). Various service policies such as the exhaustive, K-limited, Bernoulli and Markovian disciplines have been also proposed in order to meet the increasing demands for development of high-speed communication networks (see[4], [16], [17], [18], [21], [22], [25], [26],[23] and [24]).

In this paper, we model an ATM multiplexer transmitting two types of traffic as a discrete-time fluid polling system with two queues (Q_1 and Q_2) and a single server. The arrival process in Q_i is a superposition of traffic streams generated by N_i mutually independent and identical Markovian on/off sources. Each source behaves independently of other sources, and alternates between the on-state and the off-state by following a binary Markov chain with transition probabilities α_i, $1-\alpha_i$ and $1-\beta_i$, $\beta_i(\,0<\alpha_i,\beta_i<1)$. A source produces information of traffic at the constant rate r_i while in the on-state, and no information in the off-state. The single server serves Q_1 and Q_2 according to the Bernoulli service schedule described as follows: at each discrete-time, the server (suppose that it just completed the service at Q_i) makes a random decision: with probability p_i, $0<p_i<1$, it continues to deal with packets of Q_i in the next slot, and with probability $q_i=1-p_i$, it switches to $Q_j(j\neq i)$ and deals with packets there in the next slot. The service rate at Q_i is c_i. The service is assumed to be work-conserving, that is, in each slot, the server can devote its residual service capacity to another queue whenever the present queue becomes empty. Further, the server is assumed not to take switching times in its transition from one queue to the other. All arrival processes and service processes are mutually independent.

We are motivated to consider such a discrete-time polling system by the following twofold. The first is its application-oriented. With the development of high-speed communication networks employing ATM digital technology, discrete-time queueing models become

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

Key words and phrases. exponential bound, polling system, Markov fluid input, on-off source, effective bandwidth, large deviations, tail distribution.

more and more important and a lot of research work has been done (see [1], [2], [5], [6], [28] and [29]). The second is the interesting feature that our model, in fact, is obtained from discretization of a continuous-time Markovian fluid polling system, which also consists of two queues (Q_1 and Q_2) and a single server. There have N_i independent sources emitting information of traffic into Q_i. Each source alternates between the on-state and the off-state according to a two-state ({on, off}) Markov process with the infinitesimal generator matrix

$$\mathcal{Q}_i = \begin{pmatrix} -\lambda_i^{on} & \lambda_i^{on} \\ \lambda_i^{off} & -\lambda_i^{off} \end{pmatrix}.$$

Then, $1/\lambda_i^{on}$ (resp. $1/\lambda_i^{off}$) is the mean duration of the on-state (resp. the off-state). The emitting rate of each source while in the on-state is r_i. The single server deals with the two queues by following to a two-state ({1, 2}) Markov process with the infinitesimal generator matrix

$$\mathcal{Q} = \begin{pmatrix} -\mu_1 & \mu_1 \\ \mu_2 & -\mu_2 \end{pmatrix}.$$

Then, $1/\mu_i$ is the mean duration of service spent in Q_i. The service rate at Q_i is c_i. The service is assumed to be work-conserving. That is, the server is permitted to devote its residual service capacity to the other queue when the total input rate of the present queue is less than its service rate. Here, assume that all arrival processes and service processes are observed at time $n\triangle$ ($n \in \mathbf{N}$, $\triangle > 0$), and interpret the amount of traffic emitted in the interval $[n\triangle, (n+1)\triangle)$ as the amount of arrival at time $(n+1)\triangle$, and the amount of traffic dealt with in the interval $[n\triangle, (n+1)\triangle)$ as the amount of service at time $(n+1)\triangle$. Then, the resulting discrete-time arrival processes and service process in Q_i are respectively two-state Markov chains with the transition matrices,

$$P_i = e^{\Delta \mathcal{Q}_i} = \begin{pmatrix} e^{-\lambda_i^{on}\triangle} & 1 - e^{-\lambda_i^{on}\triangle} \\ 1 - e^{-\lambda_i^{off}\triangle} & e^{-\lambda_i^{off}\triangle} \end{pmatrix}, \ P = e^{\triangle \mathcal{Q}} = \begin{pmatrix} e^{-\mu_1\triangle} & 1 - e^{-\mu_1\triangle} \\ 1 - e^{-\mu_2\triangle} & e^{-\mu_2\triangle} \end{pmatrix}.$$

Taking $\alpha_i = e^{-\lambda_i^{on}\triangle}$, $\beta_i = e^{-\lambda_i^{off}\triangle}$ and $p_i = e^{-\mu_i\triangle}$, $q_i = 1 - e^{-\mu_i\triangle}$, we obtain the discrete-time fluid polling system. Up to now, most of work for continuous-time fluid queueing models is mainly devoted to performance analysis of single queueing systems, very little attention has been paid for continuous-time fluid queueing networks.

In high-speed communication networks, as known, packet loss probabilities due to buffer overflows are often taken as criteria of *quality of service* (*QoS*), and desired to be controlled below very small level, e.g. in the order of 10^{-9}. Therefore, estimating the delay and buffer overflow probability is an important work for traffic management of high-speed communication networks. The aim of the paper is to derive the buffer overflow probability for each queue of the discrete-time polling system. However, the autocorrelation structure in the arrival processes and the service processes makes it extremely difficult to get the exact results for these probabilities (even in the case of i.i.d. arrival processes, the exact results are also very complicated, see Lee [22] and Feng *et al.* [18]). Here we utilize large deviations techniques to derive the upper and lower bounds of the buffer overflow probabilities. In the last decade, the theory of large deviations has been widely applied to problems of estimating the buffer overflow probability of queueing systems (see [3], [8], [11], [14], [19], [30] for single queueing systems, and [1], [2], [9], [10], [27], [28], [29] for queueing networks). For an Markovian polling system with a single server, Poisson arrival processes and exponentially distributed service times, Delcogne and Fortelle [12] presented a local rate function governing the sample path large deviations principle. To the best of our knowledge, the analysis of large deviations for discrete-time polling models with the autocorrelation arrival processes and service processes has not been carried out yet.

The paper is organized as follows. In Section 2, we first define exactly the arrival processes superposed by mutually independent Markovian on/off sources and the potential service processes by using Markov chains, and then give some large deviations results for these processes. In Section 3, we introduce a single $MAP/MSP/1$ queueing system, and derive the effective bandwidth functions of its departure processes. In Section 4, we prove the large deviations upper and lower bounds of the buffer overflow probabilities for the polling system, and In Section 5, some conclusions are included.

2. Preliminaries

In this section we define the arrival processes and the potential service processes of the polling system, and give some large deviations results for these processes. Throughout the paper, all time indices t, τ, etc. are always integers and $\mathbf{N} = \{0, 1, 2, \cdots\}$.

A. The arrival processes

The arrival process in Q_i is the superposition of traffic streams generated by N_i mutually independent and identical Markovian on/off sources. Each source alternates between the on-state and the off-state according to a binary Markov chain with the transition probability matrix

$$P_i = \begin{pmatrix} \alpha_i & 1-\alpha_i \\ 1-\beta_i & \beta_i \end{pmatrix}.$$

The definition implies that the lengths of on-state periods and off-state periods are mutually independent sequences of i.i.d. random variables with geometric distributions

$$\begin{aligned} P\{\text{on-state period contains } t \text{ slots}\} &= (1-\alpha_i)\alpha_i^{t-1}, \quad t \geq 1, \\ P\{\text{off-state period contains } t \text{ slots}\} &= (1-\beta_i)\beta_i^{t-1}, \quad t \geq 1. \end{aligned}$$

Let a_t^i be the number of sources in the on-state at time t. Then, we have

$$a_{t+1}^i = \sum_{j=1}^{a_t^i} \sigma_j^i + \sum_{j=1}^{N_i - a_t^i} \eta_j^i, \tag{1}$$

where $\{\sigma_j^i, j = 1, 2, \cdots, N_i\}$ and $\{\eta_j^i, j = 1, 2, \cdots, N_i\}$ are two mutually independent collections of i.i.d. Bernoulli random variables with probability distributions

$$P\{\sigma_j^i = 1\} = \alpha_i, \quad P\{\sigma_j^i = 0\} = 1 - \alpha_i \quad \text{and} \quad P\{\eta_j^i = 1\} = \beta_i, \quad P\{\eta_j^i = 0\} = 1 - \beta_i.$$

Let $a_0^i = 0$ $(i = 1, 2)$, i.e., both the queues start from empty. The following proposition can be obtained easily using the expression (1).

Proposition 2.1: $\{a_t^i;\ t \in \mathbf{N}\}$ is an irreducible Markov chain with state space $\{0, 1, 2, \cdots, N_i\}$ and transition probabilities:

$$\begin{aligned} p_{lk}^i &\equiv P\{a_{t+1}^i = k | a_t^i = l\} \\ &= \sum_{n=0}^{\min\{l,k\}} \binom{l}{n} \alpha_i^n (1-\alpha_i)^{l-n} \binom{N_i - l}{k-n} \beta_i^{k-n} (1-\beta_i)^{(N_i - l)-(k-n)}, \qquad l, k \in \{0, 1, 2, \cdots, N_i\}. \end{aligned} \tag{2}$$

Define $A_t^i = a_t^i r_i,\ t \in \mathbf{N}$. Then, $\{A_t^i, t \in \mathbf{N}\}$ is the input process of Q_i. Obviously, $\{A_t^i, t \in \mathbf{N}\}$ is also an irreducible Markov chain with state space $\mathcal{S}_{A^i} = \{0, r_i, 2r_i, \cdots, N_i r_i\}$ and transition matrix $(p^i_{lr_i,kr_i} = p^i_{lk})$. We denote its equilibrium distribution by $\boldsymbol{\pi}_A^i = (\pi_0^i, \pi_1^i, \cdots, \pi_{N_i}^i)$ and the mean by $\mathcal{A}^i = E[A_t^i] = \sum_{j=0}^{N_i} j r_i \pi_j^i$. Because of simplicity and capability to capture some of the correlation characteristics of ATM traffics, Markovian on/off source processes have been widely used in modeling high-speed communication network traffic (see [5], [6] and [15]).

B. The potential service processes

The Bernoulli service schedule describes that whenever both the Q_1 and Q_2 are not empty, the server switches its service between the two queues with probabilities $p_i, q_i (i = 1, 2)$. Denote by b_t^i the position of the server at time t, that is, $b_t^i = 1$ if the server is Q_i at time t, otherwise $b_t^i = 0$. Note that $b_t^2 = 1 - b_t^1$. Let $B_t^i = b_t^i c_i$, where c_i is the service rate at Q_i. Then $\{B_t^i,\ t \in \mathbf{N}\}$ is the service process devoted to Q_i by the server under the condition that both the queues are not empty. We call it *potential service process*. According to the Bernoulli service schedule, $\{B_t^i,\ \ t \in \mathbf{N}\}$ is an Markov chain with state space $\mathcal{S}_{B^i} = \{0, c_i\}$ and the transition matrix $\mathbf{P}_{b^i}$, where

$$\mathbf{P}_{b^1} = \begin{pmatrix} p_2 & q_2 \\ q_1 & p_1 \end{pmatrix}, \qquad \mathbf{P}_{b^2} = \begin{pmatrix} p_1 & q_1 \\ q_2 & p_2 \end{pmatrix}.$$

The equilibrium distributions of $\{B_t^1,\ t \in \mathbf{N}\}$ and $\{B_t^2,\ t \in \mathbf{N}\}$ are given by $\boldsymbol{\pi}_B^1 = (q_2/(q_1+q_2),\ q_1/(q_1+q_2))$ and $\boldsymbol{\pi}_B^2 = (q_1/(q_1+q_2),\ q_2/(q_1+q_2))$, respectively, and the means by $\mathcal{B}^i = E[B_t^i] = q_i c_i/(q_1+q_2), i = 1, 2$. Note that the sum process $\{B_t^1 + B_t^2,\ \ t \in \mathbf{N}\}$ is also an Markov chain with state space $\{c_1, c_2\}$ and the transition matrix $\mathbf{P}_{b^2}$. The equilibrium distribution and the mean are given by $\boldsymbol{\pi}_B^2 = (q_1/(q_1+q_2),\ q_2/(q_1+q_2))$ and $\mathcal{B}^1 + \mathcal{B}^2$, respectively.

C. The stability condition

Let L_t^i be the queue length of Q_i at time t and $L_t = L_t^1 + L_t^2$. Since no switching times are needed in the server transitions from one queue to another, $\{B_t^1 + B_t^2,\ \ t \in \mathbf{N}\}$ can be referred as the service process of the aggregate queue $\{L_t,\ t \in \mathbf{N}\}$. Then, it follows from Loynes's Stability Theorem 2 [20] that the polling system is stable if

$$\mathcal{A}^1 + \mathcal{A}^2 < \mathcal{B}^1 + \mathcal{B}^2. \tag{3}$$

Throughout the paper, we assume that the stability condition holds. Thus, the aggregate queue length process L_t converges in distribution to a finite random variable. As $L_t^i \le L_t$, L_t^i also converges in distribution to a finite random variable.

D. Large deviations results for the arrival processes and the potential service process

Here we present some large deviations results for the Markov arrival processes (MAP) $\{A_t^i, t \in \mathbf{N}\}$, and the Markov potential service processes (MSP) $\{B_t^i, t \in \mathbf{N}\}$.

Denote by $S_{\tau,t}^X = \sum_{k=\tau}^{t-1} X_k,\ \tau < t\ (S_t^X = S_{0,t}^X)$ and $S_t^X(s) = \sum_{k=0}^{\lceil ts \rceil} X_k/t,\ 0 \le s \le 1$ the partial sums and the scaled partial sums of the random sequence $X = \{X_t;\ t \in N\}$, respectively. 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 $\Lambda_X(\theta)$:

$$(4)\qquad \Lambda_X(\theta) = \lim_{t\to\infty} \frac{1}{t}\log E[e^{\theta S_t^X}], \quad \theta\in\mathbf{R}; \qquad \Lambda_X^*(\alpha) = \sup_{\theta\in\mathbf{R}}\{\theta\alpha - \Lambda_X(\theta)\}, \quad \alpha\in\mathbf{R}.$$

For $\theta\in\mathbf{R}$ and $i=1,2$, define $(N_i{+}1)\times(N_i{+}1)$ matrix $\Psi_{A^i}(\theta) = (p^i_{lr_i,kr_i}e^{\theta k r_i})_{1\le l,k\le N_i+1}$, and 2×2 matrices as follows:

$$\Psi_{B^1}(\theta) = \begin{pmatrix} p_2 & q_2 e^{\theta c_1} \\ q_1 & p_1 e^{\theta c_1}\end{pmatrix}, \qquad \Psi_{B^2}(\theta) = \begin{pmatrix} p_1 & q_1 e^{\theta c_2} \\ q_2 & p_2 e^{\theta c_2}\end{pmatrix}.$$

Let $\rho_{A^i}(\theta) = sp(\Psi_{A^i}(\theta))$ and $\rho_{B^i}(\theta) = sp(\Psi_{B^i}(\theta))$ be the spectral radii of $\Psi_{A^i}(\theta)$ and $\Psi_{B^i}(\theta)$, and let $\mathbf{x}^{A^i}(\theta) = (x_0^{A^i}(\theta), x_1^{A^i}(\theta),\cdots, x_{N_i}^{A^i}(\theta))^T$ and $\mathbf{x}^{B^i}(\theta) = (x_0^{B^i}(\theta), x_1^{B^i}(\theta))^T$ be the positive right eigenvector corresponding to $\rho_{A^i}(\theta)$ and $\rho_{B^i}(\theta)$. Further, let $\Gamma_{A^i}(\theta) = \max_{0\le k,j\le N_i} x_k^{A^i}(\theta)/x_j^{A^i}(\theta)$ and $\Gamma_{B^i}(\theta) = \max_{0\le k,j\le 1} x_k^{B^i}(\theta)/x_j^{B^i}(\theta)$. Then, we can directly calculate these eigenvalues and eigenvectors.

Proposition 2.2: (i)

$$\rho_{B^1}(\theta) = \frac{p_2 + p_1 e^{\theta c_1} + \sqrt{(p_2 - p_1 e^{\theta c_1})^2 + 4q_1q_2e^{\theta c_1}}}{2},$$

$$\rho_{B^2}(\theta) = \frac{p_1 + p_2 e^{\theta c_2} + \sqrt{(p_1 - p_2 e^{\theta c_2})^2 + 4q_1q_2e^{\theta c_2}}}{2}.$$

(ii) $$\mathbf{x}^{B^i}(\theta) = \Big(\frac{\rho_{B^i}(\theta) - p_i e^{\theta c_i}}{\rho_{B^i}(\theta) + q_i - p_i e^{\theta c_i}}, \quad \frac{q_i}{\rho_{B^i}(\theta) + q_i - p_i e^{\theta c_i}}\Big)^T, \qquad i=1,2.$$

(iii) $$\Gamma_{B^i}(\theta) = \max\Big\{\frac{q_i}{\rho_{B^i}(\theta) - p_i e^{\theta c_i}}, \quad \frac{\rho_{B^i}(\theta) - p_i e^{\theta c_i}}{q_i}\Big\}, \qquad i=1,2.$$

Applying the general results about the theory of large deviations for Markov chains (see [7], [8], [9] and [13]) to the arrival processes $\{A_t^i, t\in\mathbf{N}\}$ and the potential service processes $\{B_t^i, t\in\mathbf{N}\}$, we have the following theorem.

Theorem 2.3: (i) $\Lambda_{A^i}(\theta) = \log(\rho_{A^i}(\theta))$ and $\Lambda_{B^i}(\theta) = \log(\rho_{B^i}(\theta))$.
(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 deviations principle with the convex, good rate functions $\Lambda^*_{A^i}(\alpha) = \sup_{\theta\in\mathbf{R}}\{\theta\alpha - \Lambda_{A^i}(\theta)\}$ and $\Lambda^*_{B^i}(\alpha) = \sup_{\theta\in\mathbf{R}}\{\theta\alpha - \Lambda_{B^i}(\theta)\}$, respectively.
(iii) Let $\mathcal{F}_t^{A^i} = \sigma\{A_\tau^i;\ \tau\le t\}$ and $\mathcal{F}_t^{B^i} = \sigma\{B_\tau^i;\ \tau\le t\}$, then for all $\theta\in\mathbf{R}$ and $\tau, t\le 0$,

$$\Lambda_{A^i}(\theta)t - \Gamma_{A^i}(\theta) \le \log E[e^{\theta S^{A^i}_{\tau,\tau+t}}|\mathcal{F}_\tau^{A^i}] = \log E[e^{\theta S^{A^i}_{\tau,\tau+t}}|A_\tau^i] \le \Lambda_{A^i}(\theta)t + \Gamma_{A^i}(\theta), \qquad a.s.$$

$$\Lambda_{B^i}(\theta)t - \Gamma_{B^i}(\theta) \le \log E[e^{\theta S^{B^i}_{\tau,\tau+t}}|\mathcal{F}_\tau^{B^i}] = \log E[e^{\theta S^{B^i}_{\tau,\tau+t}}|B_\tau^i] \le \Lambda_{B^i}(\theta)t + \Gamma_{B^i}(\theta), \qquad a.s..$$

$\Lambda_{A^i}(\theta)$, $\Lambda_{B^i}(\theta)$ and $\Lambda^*_{A^i}(\alpha)$, $\Lambda^*_{B^i}(\alpha)$ have the similar properties to those given in [28]. In particular, the following proposition holds by the non-negative and bounded properties of $\{A_t^i,\ t\in\mathbf{N}\}$ and $\{B_t^i,\ t\in\mathbf{N}\}$(note that for any $t\in\mathbf{N}$, $0\le A_t^i\le N_i r_i$ and $0\le B_t^i\le c_i$, $i=1,2$).

Proposition 2.4:

$$(5)\qquad \Lambda^*_{A^i}(\alpha) = \sup_{\theta\in\mathbf{R}}\{\theta\alpha - \Lambda_{A^i}(\theta)\} = \begin{cases} \sup_{\theta\ge 0}\{\theta\alpha - \Lambda_{A^i}(\theta)\} & \text{if } \mathcal{A}_i < \alpha \le N_i r_i \\ \sup_{\theta<0}\{\theta\alpha - \Lambda_{A^i}(\theta)\} & \text{if } 0 < \alpha \le \mathcal{A}_i \\ \infty & \text{otherwise}\end{cases}$$

$$(6)\qquad \Lambda^*_{B^i}(\alpha) = \sup_{\theta\in\mathbf{R}}\{\theta\alpha - \Lambda_{B^i}(\theta)\} = \begin{cases} \sup_{\theta\geq 0}\{\theta\alpha - \Lambda_{B^i}(\theta)\} & \text{if } \mathcal{B}_i < \alpha \leq c_i \\ \sup_{\theta< 0}\{\theta\alpha - \Lambda_{B^i}(\theta)\} & \text{if } 0 < \alpha \leq \mathcal{B}_i \\ \infty & \text{otherwise.} \end{cases}$$

Since we are only concerned with the stationary version of the system, it is convenient to look backward in time and study the behavior of the system at time 0. The conclusions in Proposition 2.5 follows from the Markov property of the arrival processes and the potential service processes, which permit us to deal with the dependence of the stationary queue length L^i_τ at time τ and the further arrival process $\{A^i_t, t > \tau\}$.

Proposition 2.5: Let $\mathcal{F}^{A^i}_{(-\infty,k]} = \sigma\{A^i_t;\ -\infty < t \leq k\}$, $\mathcal{F}^{A^i}_{(k+n,\infty)} = \sigma\{A^i_t;\ k+n < t < \infty\}$, and $\mathcal{F}^{B^i}_{(-\infty,k]} = \sigma\{B^i_t;\ -\infty < t \leq k\}$, $\mathcal{F}^{B^i}_{(k+n,\infty)} = \sigma\{B^i_t;\ k+n < t < \infty\}$, and let

$$v^{A^i}(n) = \sup_{U\in\mathcal{F}^{A^i}_{(-\infty,k]},\, U'\in\mathcal{F}^{A^i}_{(k+n,\infty)},\, P\{U\}>0} \mid P(U'|U) - P(U') \mid,$$

$$v^{B^i}(n) = \sup_{U\in\mathcal{F}^{B^i}_{(-\infty,k]},\, U'\in\mathcal{F}^{B^i}_{(k+n,\infty)},\, P\{U\}>0} \mid P(U'|U) - P(U') \mid .$$

Then, $\lim_{n\to\infty} v^{A^i}(n) = 0$ and $\lim_{n\to\infty} v^{B^i}(n) = 0$.

3. Large deviations results for an $MAP/MSP/1$ queueing system

In order to establish the large deviations bounds for the polling system, we first consider a single $MAP/MSP/1$ queueing system with the arrival process $\{A_t = A^2_t;\ t \in \mathbf{N}\}$ and the service process $\{B_t = B^2_t;\ t \in \mathbf{N}\}$. For the convenience, write $r_2 = r, N_2 = N$ and $c_2 = c$. This system is stable if $\mathcal{A} < \mathcal{B}$, where $\mathcal{A} = \sum_{j=0}^N jr\pi_j$ and $\mathcal{B} = q_2/(q_1+q_2)c$ are respectively the mean arrival rate and the mean service rate. Let $\tilde{L}_t$ be the queue length at time t, then $\tilde{L}_t$ converges in distribution to a finite random variable $\tilde{L}_\infty$ under the stable condition. Look backward in time and assume that the queueing process has reached its steady state at time 0. The large deviations results for this $MAP/MSP/1$ queueing system can be easily obtained from the general discussions for $G/G/1$ queueing systems given in [3], [8], [11], [14], [19] and [30].

Theorem 3.1: Under $\mathcal{A} < \mathcal{B}$, the tail of the equilibrium distribution of the queue length L_0 is characterized by

$$(7)\qquad \lim_{x\to\infty} \frac{1}{x} \log P\{\tilde{L}_0 > x\} = -\delta^*$$

where $\delta^* > 0$ is the largest solution of the equation: $\Lambda_A(\theta) + \Lambda_B(-\theta) = 0$.

In the case $\delta^* = \infty$, the equality (7) holds trivially. To avoid such a case, we assume that δ^* is finite, which means that there exists a number $n_0 \in \{0, 1, \cdots, N\}$ such that $n_0 r > \mathcal{B}$. Let $\alpha_A(\theta) = \Lambda_A(\theta)/\theta$ and $\alpha_B(\theta) = \Lambda_B(\theta)/\theta$ be the effective bandwidths of the arrival process and the potential service process, respectively. First, we consider the stationary departure process $\{D_t, t \in \mathbf{N}\}$ from the $MAP/MSP/1$ queue, and calculate its effective bandwidth $\alpha_D(\theta) = \Lambda_D(\theta)/\theta$. D_t and its partial sum process S^D_t are governed by the following recursive equations:

$$(8)\qquad D_t = \min\{\tilde{L}_{t-1} + A_{t-1},\ \ B_{t-1}\}, \qquad S^D_t = \min\{\tilde{L}_0 + \inf_{0<\tau\leq t}\{S^A_\tau + S^B_{\tau,t}\},\ \ S^B_t\}.$$

Define the process S_t^M as follows:

$$S_t^M = \min\{\tilde{L}_0 + S_t^A, \quad S_t^B\}, \qquad t \in \mathbf{N}. \tag{9}$$

Theorem 3.2: Under the stability assumption that $\mathcal{A} < \mathcal{B}$, for any $\alpha \in \mathbf{R}$,

$$\lim_{t\to\infty} \frac{1}{t} \log P\{S_t^D > \alpha t\} = \lim_{t\to\infty} \frac{1}{t} \log P\{S_t^M > \alpha t\} = -\inf_{x\geq\alpha} \Lambda_D^*(x), \tag{10}$$

$$\lim_{t\to\infty} \frac{1}{t} \log E[e^{\theta S_t^D}] = \lim_{t\to\infty} \frac{1}{t} \log E[e^{\theta S_t^M}] = \Lambda_D(\theta), \qquad \theta \geq 0, \tag{11}$$

where,

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

$$= \begin{cases} 0 & \text{if } \alpha < \mathcal{A} \\ \Lambda_A^*(\alpha) & \text{if } \alpha \leq \Lambda_A'(\delta^*) \text{ and } \mathcal{A} < \alpha \leq \mathcal{B} \\ \Lambda_A^*(\alpha) + \Lambda_B^*(\alpha) & \text{if } \alpha \leq \Lambda_A'(\delta^*) \text{ and } \mathcal{B} < \alpha \leq \min\{c, Nr\} \\ \delta^*\alpha - \Lambda_A(\delta^*) & \text{if } \alpha > \Lambda_A'(\delta^*) \text{ and } \mathcal{A} < \alpha \leq \mathcal{B} \\ \delta^*\alpha - \Lambda_A(\delta^*) + \Lambda_B^*(\alpha) & \text{if } \alpha > \Lambda_A'(\delta^*) \text{ and } \mathcal{B} < \alpha \leq \min\{c, Nr\} \\ \infty & \text{if } \alpha > \min\{c, Nr\}, \end{cases}$$

here, δ^* is the largest solution of the equation $\Lambda_A(\theta)+\Lambda_B(-\theta) = 0$, $\Lambda_A'(\delta^*) = \rho_A'(\delta^*)/\rho_A(\delta^*)$, and

$$\Lambda_D(\theta) = \sup_{\mathcal{A}\leq\alpha}\{\theta\alpha - \Lambda_D^*(\alpha)\} = \sup_{\mathcal{A}\leq\alpha\leq\min\{c,Nr\}}\{\theta\alpha - \Lambda_D^*(\alpha)\}. \tag{12}$$

For the proof, see Theorem 2 in Chang and Zajic [11].

Theorem 3.3: For any $\theta \geq 0$,

$$\Lambda_D(\theta) = \begin{cases} CASE1.\ \mathcal{A} < \Lambda_A'(\delta^*) \leq \mathcal{B} < \min\{c, Nr\} & \\ \quad \Lambda_A(\theta) & \text{if } \theta \leq \delta^* \\ \quad \Lambda_A(\delta^*) + \Lambda_B(\theta - \delta^*) & \text{if } \delta^* < \theta \text{ and} \\ & \mathcal{B} \leq \Lambda_B'(\theta - \delta^*) \leq \min\{c, Nr\} \\ \quad \Lambda_A(\delta^*) + (\theta - \delta^*)\min\{c, Nr\} & \text{if } \delta^* < \theta, \text{ and} \\ \qquad\qquad -\Lambda_B^*(\min\{c, Nr\}) & \quad \min\{c, Nr\} < \Lambda_B'(\theta - \delta^*) \\ CASE2.\ \mathcal{A} < \mathcal{B} < \Lambda_A'(\delta^*) \leq \min\{c, Nr\} & \\ \quad \Lambda_A(\theta) & \text{if } \theta : \Lambda_A'(\theta) \leq \mathcal{B} \\ \quad J(\theta) & \text{if } \theta : \Lambda_A'(\theta) > \mathcal{B},\ \theta \leq \delta^* \text{ or} \\ & \quad \Lambda_A'(\theta) > \mathcal{B}, \theta > \delta^*; \\ & \quad \Lambda_B'(\theta - \delta^*) \leq \Lambda_A'(\delta^*) \\ \quad \max\{J(\theta), \Lambda_A(\delta^*) + \Lambda_B(\theta - \delta^*)\} & \text{if } \theta : \Lambda_A'(\theta) > \mathcal{B}, \theta > \delta^* \text{ and} \\ & \Lambda_A'(\delta^*) < \Lambda_B'(\theta - \delta^*) \leq \min\{c, Nr\} \\ \quad \max\{J(\theta), \Lambda_A(\delta^*) + (\theta - \delta^*)\min\{c, Nr\} & \text{if } \theta : \Lambda_A'(\theta) > \mathcal{B}, \theta > \delta^* \text{ and} \\ \qquad\qquad -\Lambda_B^*(\min\{c, Nr\})\} & \quad \Lambda_B'(\theta - \delta^*) > \min\{c, Nr\} \end{cases} \tag{13}$$

where, $J(\theta) = (\theta - \tilde{\theta}_A^*(\theta) - \tilde{\theta}_B^*(\theta))\eta^{AB}(\theta) + \Lambda_A(\tilde{\theta}_A^*(\theta)) + \Lambda_B(\tilde{\theta}_B^*(\theta))$, here $\eta^{AB}(\theta)$ is the maximum point of the function $\theta\alpha - \Lambda_A^*(\alpha) - \Lambda_B^*(\alpha)$ in the interval $[\mathcal{B}, \Lambda_A'(\delta^*)]$, and for θ fixed, $\tilde{\theta}_A^*(\theta)$ and $\tilde{\theta}_B^*(\theta)$ are the unique solution of the equations $\Lambda_A'(\tilde{\theta}) = \eta^{AB}(\theta)$ and

$\Lambda'_B(\tilde{\theta}) = \eta^{AB}(\theta)$, respectively.

We first verify the following fact that will be used repeatedly in the proof of Theorem 3.3.

Lemma 3.4: For $X \in \{A, B\}$ and any real numbers $y < z$,

$$\sup_{y \le \alpha \le z} \{\theta\alpha - \Lambda^*_X(\alpha)\} = \begin{cases} \Lambda_X(\theta) & \text{if} \quad \theta : \Lambda'_X(\theta) \in [y, z] \\ \theta y - \Lambda^*_X(y) & \text{if} \quad \theta : \Lambda'_X(\theta) \le y \\ \theta z - \Lambda^*_X(z) & \text{if} \quad \theta : \Lambda'_X(\theta) \ge z \ . \end{cases} \tag{14}$$

Proof. For θ such as $\Lambda'_X(\theta) \in [y, z]$, we have

$$\theta\Lambda'_X(\theta) - \Lambda^*_X(\Lambda'_X(\theta))) \le \sup_{y \le \alpha \le z} \{\theta\alpha - \Lambda^*_X(\alpha)\} \le \sup_{\alpha \in \mathbf{R}} \{\theta\alpha - \Lambda^*_X(\alpha)\} = \Lambda_X(\theta).$$

It follows from the convex and differentiable properties of $\Lambda_X(\theta)$ that

$$\Lambda_X(\theta) = \theta\Lambda'_X(\theta) - \Lambda^*_A(\Lambda'_X(\theta)).$$

This implies that in the case $\Lambda'_X(\theta) \in [y, z]$, $\sup_{y \le \alpha \le z}\{\theta\alpha - \Lambda^*_X(\alpha)\} = \Lambda_X(\theta)$, i.e., the sup is achieved at $\Lambda'_X(\theta)$. In the case $\Lambda'_X(\theta) \notin [y, z]$, since the function $\theta\alpha - \Lambda^*_X(\alpha)$ is concave, it increases if $\Lambda'_X(\theta) \ge z$, and decreases if $\Lambda'_X(\theta) \le y$ in the interval $[y, z]$. Thus, the sup over the interval $[y, z]$ is achieved at the right end point z if $\Lambda'_X(\theta) \ge z$, and at the left end point y if $\Lambda'_X(\theta) \le y$. These complete the proof. □

Proof of Theorem 3.3. Since Nr is the maximum input rate, $S^A_t < Nrt$ for all $t \in \mathbf{N}()$. Thus, $d(\log E[\exp(\theta S^A_t)])/d\theta = E[S^A_t \exp(\theta S^A_t)]/E[\exp(\theta S^A_t)] < Nrt$. We have that $\Lambda'_A(\theta) = \lim_{t\to\infty} \{d(\log E[\exp(\theta S^A_t)])/d\theta\}/t \le Nr$ for $\theta \ge 0$. Similarly, by the bounded property of the service process, we have that $\Lambda'_B(\theta) = \lim_{t\to\infty} \{d(\log E[\exp(\theta S^B_t)])/d\theta\}/t \le c$ for $\theta \ge 0$. Moreover, since $\mathcal{B} < c$, we only need to distinguish the following two cases for $\Lambda'_A(\delta^*)$:

$CASE1.$ $\mathcal{A} < \Lambda'_A(\delta^*) \le \mathcal{B} < \min\{c, Nr\}$,
$CASE2.$ $\mathcal{A} < \mathcal{B} < \Lambda'_A(\delta^*) \le \min\{c, Nr\}$.

$CASE1.$ $\mathcal{A} < \Lambda'_A(\delta^*) \le \mathcal{B} < \min\{c, Nr\}$: By the definition of $\Lambda^*_D(\alpha)$, we can divide the sup of (12) into three parts:

$$\begin{aligned} \Lambda_D(\theta) &= \sup_{\mathcal{A} \le \alpha \le \min\{c,Nr\}} \{\theta\alpha - \Lambda^*_D(\alpha)\} \\ &= \max\{ \sup_{\mathcal{A} \le \alpha \le \Lambda'_A(\delta^*)} \{\theta\alpha - \Lambda^*_A(\alpha)\}, \quad \sup_{\Lambda'_A(\delta^*) \le \alpha \le \mathcal{B}} \{\theta\alpha - (\delta^*\alpha - \Lambda_A(\delta^*))\}, \\ &\qquad \sup_{\mathcal{B} \le \alpha \le \min\{c,Nr\}} \{\theta\alpha - (\delta^*\alpha - \Lambda_A(\delta^*) + \Lambda^*_B(\alpha))\} \ \} \\ &\equiv \max\{ \ Z^1_1(\theta), \ Z^1_2(\theta), \ Z^1_3(\theta) \ \}. \end{aligned}$$

(i) $0 \le \theta \le \delta^*$ (i.e. $\mathcal{A} \le \Lambda'_A(\theta) \le \Lambda'_A(\delta^*)$):

In this case, $Z^1_1(\theta) = \sup_{\mathcal{A} \le \alpha \le \Lambda'_A(\delta^*)} \{\theta\alpha - \Lambda^*_A(\alpha)\} = \Lambda_A(\theta)$. Since $\theta \le \delta^*$ and that $\Lambda_B(\alpha)$ is an increasing function of α,

$$\begin{aligned} Z^1_2(\theta) &= \sup_{\Lambda'_A(\delta^*) \le \alpha \le \mathcal{B}} \{\theta\alpha - (\delta^*\alpha - \Lambda_A(\delta^*))\} = \sup_{\Lambda'_A(\delta^*) \le \alpha \le \mathcal{B}} \{(\theta - \delta^*)\alpha\} + \Lambda_A(\delta^*) \\ &= (\theta - \delta^*)\Lambda'_A(\delta^*) + \Lambda_A(\delta^*), \end{aligned}$$

and

$$
\begin{aligned}
Z_3^1(\theta) &= \sup_{\mathcal{B}\le\alpha\le\min\{c,Nr\}} \{\theta\alpha - (\delta^*\alpha - \Lambda_A(\delta^*) + \Lambda_B^*(\alpha))\} \\
&= \sup_{\mathcal{B}\le\alpha\le\min\{c,Nr\}} \{(\theta-\delta^*)\alpha - \Lambda_B^*(\alpha)\} + \Lambda_A(\delta^*) \\
&= (\theta-\delta^*)\mathcal{B} - \Lambda_B^*(\mathcal{B}) + \Lambda_A(\delta^*) = (\theta-\delta^*)\mathcal{B} + \Lambda_A(\delta^*),
\end{aligned}
$$

where the last equality follows from $\Lambda_B^*(\mathcal{B}) = 0$. Since $\theta - \delta^* \le 0$ and $\Lambda_A'(\delta^*) \le \mathcal{B}$, it holds that $Z_2^1(\theta) \ge Z_3^1(\theta)$. Furthermore, from the convex property of $\Lambda_A(\cdot)$, we have that for any $\theta \le \delta^*$,

$$
\frac{\Lambda_A(\delta^*) - \Lambda_A(\theta)}{\delta^* - \theta} \le \Lambda_A'(\delta^*).
$$

That is, $\Lambda_A(\theta) \ge (\theta - \delta^*)\Lambda_A'(\delta^*) + \Lambda_A(\delta^*)$, which implicates that $Z_1^1(\theta) \ge Z_2^1(\theta)$. Hence, $\Lambda_D(\theta) = \max\{Z_1^1(\theta), Z_2^1(\theta), Z_3^1(\theta)\} = Z_1^1(\theta)$.

(ii) $\theta > \delta^*$ (i.e. $\Lambda_A'(\theta) > \Lambda_A'(\delta^*)$):

From Lemma 3.4, we know that the sup restricted to $\mathcal{A} \le \alpha \le \Lambda_A'(\delta^*)$ is achieved at $\Lambda_A'(\delta^*)$. Thus, $Z_1^1(\theta) = \sup_{\mathcal{A}\le\alpha\le\Lambda_A'(\delta^*)}\{\theta\alpha - \Lambda_A^*(\alpha)\} = \theta\Lambda_A'(\delta^*) - \Lambda_A^*(\Lambda_A'(\delta^*)) = \Lambda_A(\delta^*)$. As $\theta > \delta^*$, we have that

$$
Z_2^1(\theta) = \sup_{\Lambda_A'(\delta^*)\le\alpha\le\mathcal{B}} \{\theta\alpha - (\delta^*\alpha - \Lambda_A(\delta^*))\} = (\theta-\delta^*)\mathcal{B} + \Lambda_A(\delta^*).
$$

Clearly, $Z_2^1(\theta) \ge Z_1^1(\theta)$ because $\theta - \delta^* \ge 0$. Applying Lemma 3.4 to $\Lambda_B^*(\cdot)$, furthermore, we have that $\Lambda_B(\theta-\delta^*) = \sup_{\alpha\in\mathbf{R}}\{(\theta-\delta^*)\alpha - \Lambda_B^*(\alpha)\} = (\theta-\delta^*)\Lambda_B'(\theta-\delta^*) - \Lambda_B^*(\Lambda_B'(\theta-\delta^*))$, i.e., sup is achieved at $\Lambda_B'(\theta-\delta^*)$. Hence,

$$
\begin{aligned}
Z_3^1(\theta) &= \sup_{\mathcal{B}\le\alpha\le\min\{c,Nr\}} \{(\theta-\delta^*)\alpha - \Lambda_B^*(\theta)\} + \Lambda_A(\delta^*) \\
&= \begin{cases} \Lambda_A(\delta^*) + \Lambda_B(\theta-\delta^*) & \text{if} \quad \mathcal{B} \le \Lambda_B'(\theta-\delta^*) \le \min\{c,Nr\} \\ \Lambda_A(\delta^*) + (\theta-\delta^*)\min\{c,Nr\} - \Lambda_B^*(\min\{c,Nr\}) & \text{if} \quad \Lambda_B'(\theta-\delta^*) > \min\{c,Nr\}. \end{cases}
\end{aligned}
$$

Since $\Lambda_B'(\cdot)$ is increasing, $\theta > \delta^*$ and $\Lambda_B(0) = 0$, we have that $\Lambda_B'(\theta-\delta^*) \ge \Lambda_B'(0) = \mathcal{B}$. Thus, $\Lambda_B(\theta-\delta^*) \ge (\theta-\delta^*)\mathcal{B}$ in the case that $\mathcal{B} \le \Lambda_B'(\theta-\delta^*) \le \min\{c,Nr\}$. On the other hand, since $(\theta-\delta^*)\alpha - \Lambda_B^*(\alpha)$ is increasing in the interval $[\mathcal{B}, \min\{c,Nr\}]$, we have that $(\theta-\delta^*))\mathcal{B} = (\theta-\delta^*))\mathcal{B} - \Lambda_B^*(\mathcal{B}) \le (\theta-\delta^*))\min\{c,Nr\} - \Lambda_B^*(\min\{c,Nr\}))$ when $\Lambda_B'(\theta-\delta^*) > \min\{c,Nr\}$. It follows that $Z_3^1(\theta) \ge Z_2^1(\theta)$. Hence, $\Lambda_D(\theta) = Z_3^1(\theta)$.

The proof of $CASE2$ is similar, we omit it here. These complete the proof. □

Next, we consider the transient departure process $\{E_t, t \in \mathbf{N}\}$ from the $MAP/MSP/1$ queue (i.e., a departure process started from an empty queue at $t = 0$) and derive its effective bandwidth $\alpha_E(\theta) = \Lambda_E(\theta)/\theta$. E_t and its partial sum process S_t^E are governed by the following recursive equations:

$$
E_t = \min\{\tilde{L}_{t-1} + A_{t-1}, \;\; B_{t-1}\}, \qquad S_t^E = \min\{\inf_{0<\tau\le t}\{S_\tau^A + S_{\tau,t}^B\}, \;\; S_t^B\}. \tag{15}
$$

Similarly, define the process $\tilde{S}_t^M$ as follows:

$$
\tilde{S}_t^M = \min\{S_t^A, \;\; S_t^B\}.
$$

Then, we derive the large deviations results for these processes by the approach used in Theorem 3.2 and Theorem 3.3.

Corollary 3.5: Under $\mathcal{A} < \mathcal{B}$, for any $\alpha \in \mathbf{R}$,

$$\lim_{t\to\infty} \frac{1}{t} \log P\{S_t^E > \alpha t\} = \lim_{t\to\infty} \frac{1}{t} \log P\{\tilde{S}_t^M > \alpha t\} = -\inf_{x\geq\alpha} \Lambda_E^*(x), \tag{16}$$

$$\lim_{t\to\infty} \frac{1}{t} \log E[e^{\theta S_t^E}] = \lim_{t\to\infty} \frac{1}{t} \log E[e^{\theta \tilde{S}_t^M}] = \Lambda_E(\theta), \qquad \theta \geq 0, \tag{17}$$

where,

$$\Lambda_E^*(\alpha) = \inf_{x\geq\alpha} \Lambda_A^*(x) + \inf_{x\geq\alpha} \Lambda_B^*(x) = \begin{cases} 0 & \text{if } \alpha < \mathcal{A} \\ \Lambda_A^*(\alpha) & \text{if } \alpha \leq \mathcal{B} \\ \Lambda_A^*(\alpha) + \Lambda_B^*(\alpha) & \text{if } \mathcal{B} < \alpha \leq \min\{c, Nr\} \\ \infty & \text{if } \alpha > \min\{c, Nr\} \end{cases}$$

and

$$\Lambda_E(\theta) = \sup_{\mathcal{A}\leq\alpha} \{\theta\alpha - \Lambda_E^*(\alpha)\} = \sup_{\mathcal{A}\leq\alpha\leq\min\{c,Nr\}} \{\theta\alpha - \Lambda_E^*(\alpha)\}.$$

Corollary 3.6: For any $\theta \geq 0$,

$$\Lambda_E(\theta) = \begin{cases} \Lambda_A(\theta) & \text{if } \theta : \Lambda_A'(\theta) \leq \mathcal{B} \\ K(\theta) & \text{if } \theta : \Lambda_A'(\theta) > \mathcal{B} \end{cases} \tag{18}$$

where, $K(\theta) = (\theta - \hat{\theta}_A^*(\theta) - \hat{\theta}_B^*(\theta))\xi^{AB}(\theta) + \Lambda_A(\hat{\theta}_A^*(\theta)) + \Lambda_B(\hat{\theta}_B^*(\theta))$, here $\xi^{AB}(\theta)$ is the maximum point of the function $\theta\alpha - \Lambda_A^*(\alpha) - \Lambda_B^*(\alpha)$ in the interval $[\mathcal{B}, \min\{c, Nr\}]$, and for θ fixed, $\hat{\theta}_A^*(\theta)$ and $\hat{\theta}_B^*(\theta)$ are the unique solution of the equations $\Lambda_A'(\hat{\theta}) = \xi^{AB}(\theta)$ and $\Lambda_B'(\hat{\theta}) = \xi^{AB}(\theta)$, respectively.

Corollary 3.7: (i) For any $\alpha \in \mathbf{R}$, $\Lambda_D^*(\alpha) \leq \Lambda_E^*(\alpha)$. In particular, $\Lambda_D^*(\alpha) = \Lambda_E^*(\alpha)$ if $\alpha \leq \Lambda_A'(\delta^*)$.
(ii) For any $\theta \geq 0$, $\Lambda_D(\theta) \geq \Lambda_E(\theta)$. In particular, $\Lambda_D(\theta) = \Lambda_E(\theta)$ if $\theta \leq \delta^*$.

Proof. Since $L_0 \geq 0$, $S_t^D \geq S_t^E$ for any $t \geq 0$. Thus, $P\{S^D \geq \alpha t\} \geq P\{S^E \geq \alpha t\}$. It follows from (10) and (16) that $-\inf_{x\geq\alpha} \Lambda_D^*(x) \geq -\inf_{x\geq\alpha} \Lambda_E^*(x)$. We get that $\Lambda_D^*(\alpha) \leq \Lambda_E^*(\alpha)$ from the convex properties of $\Lambda_D^*(\cdot)$ and $\Lambda_E^*(\cdot)$. Furthermore, we know from Theorem 2 in [11] that the effect of L_0 can be ignored when $\alpha \leq \Lambda_A'(\delta^*)$. Hence, $\Lambda_D^*(\alpha) = \Lambda_E^*(\alpha)$. For (ii), we have $\Lambda_D(\theta) = \sup_{\alpha\in\mathbf{R}}\{\theta\alpha - \Lambda_D^*(\alpha)\} \geq \sup_{\alpha\in\mathbf{R}}\{\theta\alpha - \Lambda_E^*(\alpha)\} = \Lambda_E(\theta)$. Comparing $\Lambda_D(\theta)$ with $\Lambda_E(\theta)$ and noting that $J(\theta) = K(\theta)$ when $\theta \leq \delta^*$, we obtain the assertion that $\Lambda_D(\theta) = \Lambda_E(\theta)$. □

4. Large deviations bounds for the polling system

In this section, we derive the upper and lower bounds of the buffer overflow probability for each queue in the polling system. Let $MAP^i/MSP^i/1$ be a single queueing system with the arrival process $\{A_t^i,\ t \in \mathbf{N}\}$ and the potential service process $\{B_t^i,\ t \in \mathbf{N}\}$, and denote their effective bandwidths by $\alpha_{A^i}(\theta) = \Lambda_{A^i}(\theta)/\theta = \log(\rho_{A^i}(\theta))/\theta$ and $\alpha_{B^i}(\theta) = \Lambda_{B^i}(\theta)/\theta = \log(\rho_{B^i}(\theta))/\theta$, respectively. Further, let $\{E_t^i,\ t \in \mathbf{N}\}$ and $\{D_t^i,\ t \in \mathbf{N}\}$ be the transient and stationary departure processes from the $MAP^i/MSP^i/1$ queue, and denote their effective bandwidths by $\alpha_{E^i}(\theta) = \Lambda_{E^i}(\theta)/\theta$ and $\alpha_{D^i}(\theta) = \Lambda_{D^i}(\theta)/\theta$. Note that under the stability condition (3) of the polling system, the situation that $\mathcal{A}^1 \geq \mathcal{B}^1$ or $\mathcal{A}^2 \geq \mathcal{B}^2$ might occur. Thus, taking account of these possibilities and using the large deviations results for the departure processes obtained in the previous section, we define

$\Lambda_{E^i}(\theta)$ and $\Lambda_{D^i}(\theta)$ as follows. For any $\theta \geq 0$,

$$\Lambda_{E^i}(\theta) = \begin{cases} CASE1.\ \mathcal{A}^i < \mathcal{B}^i & \\ \quad \Lambda_{A^i}(\theta) & \text{if } \theta : \Lambda'_{A^i}(\theta) \leq \mathcal{B}^i \\ \quad K_i(\theta) & \text{if } \theta : \Lambda'_{A^i}(\theta) > \mathcal{B}^i \\ CASE2.\ \mathcal{A}^i \geq \mathcal{B}^i & \\ \quad \mathcal{B}^i\theta & \end{cases} \tag{19}$$

where, $K_i(\theta) = (\theta - \hat{\theta}^*_{A^i}(\theta) - \hat{\theta}^*_{B^i}(\theta))\xi^{A^iB^i}(\theta) + \Lambda_{A^i}(\hat{\theta}^*_{A^i}(\theta)) + \Lambda_{B^i}(\hat{\theta}^*_{B^i}(\theta))$, 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 $[\mathcal{B}^i, \min\{c_i, N_ir_i\}]$, 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. And for any $\theta \geq 0$,

$$\Lambda_{D^i}(\theta) = \begin{cases} CASE1.\ \mathcal{A}^i < \Lambda'_{A^i}(\delta^*_i) \leq \mathcal{B}^i < \min\{c_i, N_ir_i\} & \\ \quad \Lambda_{A^i}(\theta) & \text{if } \theta \leq \delta^*_i \\ \quad \Lambda_{A^i}(\delta^*_i) + \Lambda_{B^i}(\theta - \delta^*_i) & \text{if } \delta^*_i < \theta \text{ and } \\ & \mathcal{B}^i \leq \Lambda'_{B^i}(\theta - \delta^*_i) \leq \min\{c_i, N_ir_i\} \\ \quad \Lambda_{A^i}(\delta^*_i) + (\theta - \delta^*_i)\min\{c_i, N_ir_i\} & \text{if } \delta^*_i < \theta, \text{ and} \\ \qquad -\Lambda^*_{B^i}(\min\{c_i, N_ir_i\}) & \quad \min\{c_i, N_ir_i\} < \Lambda'_{B^i}(\theta - \delta^*_i) \\ CASE2.\ \mathcal{A}^i < \mathcal{B}^i < \Lambda'_{A^i}(\delta^*_i) \leq \min\{c_i, N_ir_i\} & \\ \quad \Lambda_{A^i}(\theta) & \text{if } \theta : \Lambda'_{A^i}(\theta) \leq \mathcal{B}^i \\ \quad J_i(\theta) & \text{if } \theta : \Lambda'_{A^i}(\theta) > \mathcal{B}^i,\ \theta \leq \delta^*_i \text{ or} \\ & \quad \Lambda'_{A^i}(\theta) > \mathcal{B}^i, \theta > \delta^*_i; \\ & \quad \Lambda'_{B^i}(\theta - \delta^*_i) \leq \Lambda'_{A^i}(\delta^*_i) \\ \quad \max\{J_i(\theta), \Lambda_{A^i}(\delta^*_i) + \Lambda_{B^i}(\theta - \delta^*_i)\} & \text{if } \theta : \Lambda'_{A^i}(\theta) > \mathcal{B}^i, \theta > \delta^*_i \text{ and} \\ & \Lambda'_{A^i}(\delta^*_i) < \Lambda'_{B^i}(\theta - \delta^*_i) \\ & \qquad \leq \min\{c_i, N_ir_i\} \\ \quad \max\{J_i(\theta), \Lambda_{A^i}(\delta^*_i) + (\theta - \delta^*_i)\min\{c_i, N_ir_i\} & \text{if } \theta : \Lambda'_{A^i}(\theta) > \mathcal{B}^i, \theta > \delta^*_i \text{ and} \\ \qquad -\Lambda^*_{B^i}(\min\{c_i, N_ir_i\})\} & \quad \Lambda'_{B^i}(\theta - \delta^*_i) > \min\{c_i, N_ir_i\} \\ CASE3.\ \mathcal{A}^i \geq \mathcal{B}^i & \\ \quad \Lambda_{B^i}(\theta) & \end{cases} \tag{20}$$

where, δ^*_i is the largest solution to the equation $\Lambda_{A^i}(\theta) + \Lambda_{B^i}(-\theta) = 0$, and $J_i(\theta) = (\theta - \tilde{\theta}^*_{A^i}(\theta) - \tilde{\theta}^*_{B^i}(\theta))\eta^{A^iB^i}(\theta) + \Lambda_{A^i}(\tilde{\theta}^*_{A^i}(\theta)) + \Lambda_{B^i}(\tilde{\theta}^*_{B^i}(\theta))$, here $\eta^{A^iB^i}(\theta)$ 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.

Furthermore, we define $\Lambda^*_{E^i}(\alpha)$ as follows:

$$\Lambda^*_{E^i}(\alpha) = \begin{cases} CASE1.\ \mathcal{A}^i < \mathcal{B}^i & \\ \quad 0 & \text{if } \alpha < \mathcal{A}^i \\ \quad \Lambda^*_{A^i}(\alpha) & \text{if } \alpha \leq \mathcal{B}^i \\ \quad \Lambda^*_{A^i}(\alpha) + \Lambda^*_{B^i}(\alpha) & \text{if } \mathcal{B}^i < \alpha \leq \min\{c_i, N_ir_i\} \\ \quad \infty & \text{if } \alpha > \min\{c_i, N_ir_i\} \\ CASE2.\ \mathcal{A}^i \geq \mathcal{B}^i & \\ \quad 0 & \text{if } \alpha = \mathcal{B}^i \\ \quad \infty & \text{otherwise} \end{cases} \tag{21}$$

It follows from Corollary 3.5 that $\Lambda_{E^i}(\cdot)$ and $\Lambda^*_{E^i}(\cdot)$ is convex conjugate.

Theorem 4.1: Under the stability condition (3), the steady state queue length L_0^i of the queue Q_i satisfies
(i) *upper bound:*

$$\limsup_{x\to\infty} \frac{1}{x}\log P\{L_0^i > x\} \le -\Theta^*_{ij}(v_i), \tag{22}$$

where, $\Theta^*_{ij}(v_i)$ is the unique solution of the equation:

$$\alpha_{A^i}(\theta) + v_i\alpha_{D^j}(v_i\theta) = c_i, \qquad i \ne j, \tag{23}$$

and $v_i = c_i/c_j,\ i, j = 1, 2$.
(ii) *lower bound:*

$$\liminf_{x\to\infty} \frac{1}{x}\log P\{L_0^i > x\} \ge -\theta^*_{ij}(l_i), \tag{24}$$

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

$$\alpha_{A^i}(\theta) + l_i\alpha_{E^j}(l_i\theta) = c_i, \qquad i \ne j, \tag{25}$$

and $l_i = \max\{v_i, 1\}\mathbf{1}_{\{\mathcal{A}^j<\mathcal{B}^j\}} + v_i\mathbf{1}_{\{\mathcal{A}^j\ge\mathcal{B}^j\}}$, here $\mathbf{1}_C$ denotes the indicator function of the set C.

We will use the following lemma given in [8] in proof of Theorem 4.1.

Lemma 4.2: For the convex conjugate $\Lambda^*_X(\cdot)$ and $\Lambda_X(\cdot)$, it holds that

$$\inf_{\alpha>c} \frac{\Lambda^*_X(\alpha)}{\alpha - c} = \theta^* \tag{26}$$

where θ^* is the unique positive root of the equation $\Lambda_X(\theta) = c\theta$, and $c > E[X]$ is a constant.

Proof of Theorem 4.1. Without loss of generality, we establish the upper (22) and the lower (24) for Q_1, i.e., for the case that $i = 1$, $j = 2$. Again, we look backwards in time from time 0, and assume that the system has reached its steady state. Thus, L_0^i has the same distribution as L_∞^i.

1. Upper bound: The work-conservation of the Bernoulli service schedule permits the server to devote its residual service capacity to another queue whenever the present queue becomes empty in each slot. Under such the discipline, we analyze the amount of service actually received by each queue at slot $-t-1$. First, assume that the server is in Q_2 at the beginning of slot $-t-1$, i.e. $B^2_{-t-1} = c_2$ and $B^1_{-t-1} = 0$. If $L^2_{-t-1} + A^2_{-t-1} < c_2$, then $(L^2_{-t-1} + A^2_{-t-1})/c_2\,(< 1)$ is the duration for the server to deal with the amount of traffic $L^2_{-t-1} + A^2_{-t-1}$. Thus, the amount of service received by Q_1 at slot $-t-1$ is $\max\{B^1_{-t-1}, c_1(1-(L^2_{-t-1} + A^2_{-t-1})/c_2)\} = c_1(1-(L^2_{-t-1} + A^2_{-t-1})/c_2)$. Otherwise, if $L^2_{-t-1} + A^2_{-t-1} \ge c_2$, then $\max\{B^1_{-t-1}, c_1(1-(L^2_{-t-1} + A^2_{-t-1})/c_2)\} = 0$. The amount of service received by Q_1 at slot $-t-1$ is 0. Next, assume that the server is in Q_1 at the beginning of slot $-t-1$, i.e. $B^2_{-t-1} = 0$ and $B^1_{-t-1} = c_1$. Since it always holds that $c_1(1-(L^2_{-t-1} + A^2_{-t-1})/c_2) \le c_1$, $\max\{B^1_{-t-1},\ c_1(1-(L^2_{-t-1} + A^2_{-t-1})/c_2)\} = c_1$, which means that Q_1 can receive the amount of service c_1 at slot $-t-1$. Therefore, $\max\{B^1_{-t-1}, c_1(1-(L^2_{-t-1} + A^2_{-t-1})/c_2)\}$ is the amount of service actually received by Q_1 at slot $-t-1$. The similar results hold for Q_2. We have that for $t \ge 0$,

$$L^1_{-t} = \max\{L^1_{-t-1} + A^1_{-t-1} - \max\{B^1_{-t-1},\ c_1 - v_1(L^2_{-t-1} + A^2_{-t-1})\},\ \ 0\}, \tag{27}$$

$$L^2_{-t} = \max\{L^2_{-t-1} + A^2_{-t-1} - \max\{B^2_{-t-1},\ c_2 - v_2(L^1_{-t-1} + A^1_{-t-1})\},\ \ 0\} \tag{28}$$

where $v_i = c_i/c_j$ for $i \neq j$. Let $R^i_{-t} = \max\{B^i_{-t},\ c_i - v_i(L^j_{-t} + A^j_{-t})\},\ i,j = 1,2; i \neq j$. Expanding (27) and (28) recursively, we have that

$$L^i_0 = \max_{t\in\mathbf{N}}\{\ S^{A^i}_{-t} - S^{R^i}_{-t}\ \}, \qquad i = 1,2, \tag{29}$$

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

$$S^{R^i}_{-t} = L^i_{-t} + S^{A^i}_{-t} - L^i_0. \qquad i = 1,2, \tag{30}$$

Thus, $S^{R^i}_t \geq S^{A^i}_t - L^i_0$. It follows that the maximum in (29) for $i = 1$ must be achieved at the time when $L^1_t = 0$. Let $-t \leq 0$ be the first time such that $L^1_{-t} = 0$, then $L^1_{-\tau} > 0$ for $-\tau \in (-t,0]$. Since Q_1 is busy during the interval $(-t,0]$ and the Bernoulli service schedule is work-conserving, Q_1 gets at least the amount of service $S^{B^1}_{-t}$ (by considering the situation that Q_2 may become empty during $(-t,0]$). Hence, $S^{R^1}_{-t} \geq S^{B^1}_{-t}$. On the other hand, since $S^{R^2}_{-t}$ is the amount of service actually received by Q_2 in the interval $(-t,0]$ and the rate of service at Q_2 is c_2, $S^{R^2}_{-t}/c_2$ is the duration that the server spent in Q_2 over the interval $(-t,0]$. Thus, $c_1(t - S^{R^2}_{-t}/c_2) = c_1 t - v_1 S^{R^2}_{-t}$ is the amount of service received by Q_1. We have that

$$S^{R^1}_{-t} = \max\{\ c_1 t - v_1 S^{R^2}_{-t},\ \ S^{B^1}_{-t}\ \}. \tag{31}$$

In addition, it follows from the definition of $\{B^i_t, t \in \mathbf{N}\}$ that for any $t \geq 0$,

$$S^{B^i}_{-t} = c_i t - v_i S^{B^j}_{-t}, \qquad i,j = 1,2;\ i \neq j. \tag{32}$$

Therefore,

$$S^{R^1}_{-t} = \max\{\ c_1 t - v_1 S^{R^2}_{-t},\ \ c_1 t - v_1 S^{B^2}_{-t}\ \} = c_1 t - v_1 \min\{\ S^{R^2}_{-t},\ \ S^{B^2}_{-t}\ \}. \tag{33}$$

Substituting (33) into (29) for $i = 1$ yields

$$L^1_0 = \max_{t\in\mathbf{N}}\{\ S^{A^1}_{-t} + v_1 \min\{\ S^{R^2}_{-t},\ \ S^{B^2}_{-t}\ \} - c_1 t\ \}. \tag{34}$$

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

In this case we can bound L^2_t from the above by the queue length of the single $MAP^2/MSP^2/1$ queueing system. Let $\tilde{L}^2_{-t}$ be the queue length of the single queue at time $-t$. Since this queueing system does not receive extra service except $S^{B^2}_{-t}$, it always holds that $L^2_{-t} \leq \tilde{L}^2_{-t}$. We have that from (30)

$$S^{R^2}_{-t} \leq L^2_{-t} + S^{A^2}_{-t} \leq \tilde{L}^2_{-t} + S^{A^2}_{-t}. \tag{35}$$

Combining (34) with (35) yields

$$L^1_0 \leq \max_{t\in\mathbf{N}}\{\ S^{A^1}_{-t} + v_1 \min\{\ \tilde{L}^2_{-t} + S^{A^2}_{-t},\ \ S^{B^2}_{-t}\ \} - c_1 t\ \} = \max_{t\in\mathbf{N}}\{\ S^{A^1}_{-t} + v_1 S^{M^2}_{-t} - c_1 t\} \tag{36}$$

where $S^{M^2}_{-t} = \min\{\ \tilde{L}^2_{-t} + S^{A^2}_{-t},\ \ S^{B^2}_{-t}\ \}$. From Theorem 2.3 and Theorem 3.2, it follows that for any $\theta > 0$,

$$\lim_{t\to\infty} \frac{1}{t} \log E[e^{\theta S^{A^1}_{-t}}] = \Lambda_{A^1}(\theta),$$

$$\lim_{t\to\infty}\frac{1}{t}\log E[e^{\theta v_1 S^{M^2}_{-t}}] = \Lambda_{D^2}(v_1\theta).$$

Then, for any $\epsilon > 0$, there exists a sufficiently large t_ϵ such that for any $t \geq t_\epsilon$

$$E[e^{\theta S^{A^1}_{-t}}] \leq e^{(\Lambda_{A^1}(\theta)+\epsilon)t}, \quad \text{and} \quad E[e^{\theta v_1 S^{M^2}_{-t}}] \leq e^{(\Lambda_{D^2}(v_1\theta)+\epsilon)t}.$$

We have

$$E[e^{\theta L^1_0}] \leq E[e^{\theta \max_{t\in\mathbf{N}}\{S^{A^1}_{-t}+v_1 S^{M^2}_{-t}-c_1 t\}}] \leq \sum_{t\in\mathbf{N}} E[e^{\theta(S^{A^1}_{-t}+v_1 S^{M^2}_{-t}-c_1 t)}]$$

$$= \sum_{t\in\mathbf{N}} E[e^{\theta S^{A^2}_{-t}}]E[e^{\theta v_1 S^{M^2}_{-t}}]e^{-\theta c_1 t} \leq C_\epsilon + \sum_{t\geq t_\epsilon} e^{(\Lambda_{A^1}(\theta)+\Lambda_{D^2}(v_1\theta)+2\epsilon-c_1\theta)t},$$

where the last second equality follows from the independence of $S^{A^2}_{-t}$ and $S^{M^2}_{-t}$, and C_ϵ is a constant dependent on ϵ. It follows that $E[e^{\theta L^1_0}] < \infty$ if $\Lambda_{A^1}(\theta)+\Lambda_{D^2}(v_1\theta)+2\epsilon-c_1\theta < 0$ (i.e. $\alpha_{A^1}(\theta)+v_1\alpha_{D^2}(v_1\theta)+2\epsilon/\theta-c_1 < 0$). By Chebyshev's inequality, $P\{L^1_0 > x\} \leq e^{-\theta x}E[e^{\theta L^1_0}]$ for any $x \geq 0$. Thus, if $\alpha_{A^1}(\theta) + v_1\alpha_{D^2}(v_1\theta) + 2\epsilon/\theta - c_1 < 0$,

$$\limsup_{x\to\infty}\frac{1}{x}\log P\{L^1_0 > x\} \leq -\theta.$$

Taking $\epsilon \to 0$ and getting the tightest upper bound, we establish (22).

*CASE*2. : $\mathcal{A}^2 \geq \mathcal{B}^2$
According to (30), we have that

$$L^1_0 \leq \max_{t\in\mathbf{N}}\{\ S^{A^1}_{-t} + v_1 S^{B^2}_{-t} - c_1 t\ \}.$$

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

$$E[e^{\theta L^1_0}] \leq \sum_{t\in\mathbf{N}} E[e^{\theta(S^{A^2}_{-t}+v_1 S^{D^2}_{-t}-c_1 t)}] < \infty.$$

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

$$\limsup_{x\to\infty}\frac{1}{x}\log P\{L^1_0 > x\} \leq -\theta.$$

Taking $\epsilon \to 0$ and getting the tightest upper bound (note that $\alpha_{D^2}(\theta) = \Lambda_{B^2}(\theta)/\theta$ in this case), we establish (22).

2. Lower bound: In section 2, we have defined $L_t = L^1_t + L^2_t$ as the aggregate queue length of the two queues, and $R_t = R^1_t + R^2_t$ as the aggregate service process for L_t. Hence,

$$L_0 = \max_{t\in\mathbf{N}}\{\ S^{A^1}_{-t} + S^{A^2}_{-t} - (S^{R^1}_{-t} + S^{R^2}_{-t})\ \}. \tag{37}$$

This maximum must also be achieved at the time when $L_{-t} = 0$. Let $-t^* \leq 0$ be the first time such that $L_{-t^*} = 0$ (which implies that $L^1_{-t^*} = L^2_{t^*} = 0$) and $L_{-t} > 0$ for $t \in (0, t^*)$. In addition, we have a similar expression to (34) for L^2_0 :

$$L^2_0 = \max_{t\in\mathbf{N}}\{\ S^{A^2}_{-t} + v_2\min\{\ S^{R^1}_{-t},\ \ S^{B^1}_{-t}\ \} - c_2 t\ \}. \tag{38}$$

This maximum must also be achieved at the time when $L^2_{-t} = 0$. Let $-\tau^* \le 0$ be the first time such that $L^2_{-\tau^*} = 0$ and $L^2_{-t} > 0$ for $t \in (0, \tau^*)$. Then, we have $\tau^* \le t^*$, $S^{A^2}_{-t^*,-\tau^*} = S^{R^2}_{-t^*,-\tau^*}$ and $S^{B^2}_{-\tau^*} \le S^{R^2}_{-\tau^*}$. Utilizing these facts and the relations (37) and (38), we have that

$$
\begin{aligned}
(39)\quad L^1_0 &= L_0 - L^2_0 \\
&= \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + S^{A^2}_{-t} - (S^{R^1}_{-t} + S^{R^2}_{-t}) \} - \max_{t\in\mathbf{N}}\{ S^{A^2}_{-t} + v_2 \min\{ S^{R^1}_{-t}, \ S^{B^1}_{-t} \} - c_2 t \} \\
&= \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + S^{A^2}_{-t} - (S^{R^1}_{-t} + S^{R^2}_{-t}) - \max_{0\le\tau\le t}\{ S^{A^2}_{-\tau} + v_2 \min\{ S^{R^1}_{-\tau}, \ S^{B^1}_{-\tau} \} - c_2\tau \}\} \\
&= \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + S^{A^2}_{-t} - (S^{R^1}_{-t} + S^{R^2}_{-t}) - \max_{0\le\tau\le t}\{ S^{A^2}_{-\tau} + v_2 S^{B^1}_{-\tau} - c_2\tau \}\} \\
&= \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + \min_{0\le\tau\le t}\{S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau}\} - (S^{R^1}_{-t} + S^{R^2}_{-t})\} \\
&= \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + \min_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} - (c_1 t - (1-v_1) S^{R^2}_{-t}) \} \\
&= \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + v_1 \min_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} - c_1 t + (1-v_1)[\min_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} - S^{R^2}_{-t}]\}
\end{aligned}
$$

If $v_1 < 1$, then $(1-v_1)S^{R^2}_{-t} \ge 0$. It follows from the last second equality that

$$
(40)\qquad L^1_0 \ge \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + \min_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} - c_1 t\}.
$$

Furthermore, the fact that $S^{A^2}_{-t^*,-\tau^*} = S^{R^2}_{-t^*,-\tau^*}$ and $S^{B^2}_{-\tau^*} \le S^{R^2}_{-\tau^*}$ implies $S^{A^2}_{-t^*,-\tau^*} + S^{B^2}_{-\tau^*} \le S^{R^2}_{-t^*}$. We have $\min\limits_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} \le S^{B^2}_{-t} \le S^{R^2}_{-t}$. Hence, if $v_1 \ge 1$, then $(1-v_1)[\min_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} - S^{R^2}_{-t}] \ge 0$. It follows from the last equality of (39) that

$$
(41)\qquad L^1_0 \ge \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + v_1 \min_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} - c_1 t \}.
$$

Let $\tilde{l}_i = \max\{v_i, 1\}, i = 1, 2$. Then, we can write (40) and (41) together as follows:

$$
\begin{aligned}
L^1_0 &\ge \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + \tilde{l}_1 \min_{0\le\tau\le t}\{ S^{A^2}_{-t,-\tau} + S^{B^2}_{-\tau} \} - c_1 t \} \\
&=_{st} \max_{t\in\mathbf{N}}\{ S^{A^1}_{-t} + \tilde{l}_1 \min_{0\le\tau\le t}\{ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau} \} - c_1 t \}.
\end{aligned}
$$

For any $x \ge 0$, let $t = \lceil x/\beta \rceil$, where $\beta > 0$ is an arbitrary constant. From (41), we obtain the following inequality:

$$
\begin{aligned}
(42)\qquad \liminf_{x\to\infty} \frac{1}{x} \log P\{L^1_0 > x\} &\ge \frac{1}{\beta} \liminf_{t\to\infty} \frac{1}{t} \log P\{L^1_0 > \beta t\} \\
&\ge \frac{1}{\beta} \liminf_{t\to\infty} \frac{1}{t} \log P\{ S^{A^1}_{-t} + \tilde{l}_1 \min_{0\le\tau\le t}\{ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau} \} - c_1 t > \beta t \} \\
&= \frac{1}{\beta} \liminf_{t\to\infty} \frac{1}{t} \log P\{ \frac{S^{A^1}_{-t}}{t} + \tilde{l}_1 \frac{\min_{0\le\tau\le t}\{ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau} \}}{t} > c_1 + \beta \}.
\end{aligned}
$$

*CASE*1. $v_1 \ge 1$:

(**i**) $\mathcal{A}^2 \geq \mathcal{B}^2$: Since $c_1 = q_1c_1/(q_1+q_2) + q_2c_1/(q_1+q_2) = \mathcal{B}^1 + v_1q_2c_2/(q_1+q_2) = \mathcal{B}^1 + v_1\mathcal{B}^2$,

$$
\begin{aligned}
&P\{\ \frac{S^{A^1}_{-t}}{t} + v_1\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > c_1 + \beta\ \} \\
&= P\{\ \frac{S^{A^1}_{-t}}{t} + v_1\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > \mathcal{B}^1 + v_1\mathcal{B}^2 + \beta\ \} \\
&\geq P\{\ \frac{S^{A^1}_{-t}}{t} > \mathcal{B}^1 + \beta,\quad v_1\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > v_1\mathcal{B}^2\ \} \\
&= P\{\ \frac{S^{A^1}_{-t}}{t} > \mathcal{B}^1 + \beta\ \}P\{\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > \mathcal{B}^2\ \},
\end{aligned}
$$

where, the last equality follows from the independence of $\{A^1_{-t};\ t\in\mathbf{N}\}$, $\{A^2_{-t};\ t\in\mathbf{N}\}$ and $\{B^2_{-t};\ t\in\mathbf{N}\}$. Thus,

$$
\begin{aligned}
&\liminf_{x\to\infty}\frac{1}{x}\log P\{L^1_0 > x\} \\
&\geq \frac{1}{\beta}\left(\liminf_{t\to\infty}\frac{1}{t}\log P\{\ \frac{S^{A^1}_{-t}}{t} > \mathcal{B}^1 + \beta\ \} + \liminf_{t\to\infty}\frac{1}{t}\log P\{\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > \mathcal{B}^2\ \}\right) \\
&\geq -\frac{1}{\beta}\inf_{\alpha\geq\mathcal{B}^1+\beta}\Lambda^*_{A^1}(\alpha) - \frac{1}{\beta}\inf_{\alpha\geq\mathcal{B}^2}\Lambda^*_{E^2}(\alpha) = -\frac{1}{\beta}\inf_{\alpha\geq\mathcal{B}^1+\beta}\Lambda^*_{A^1}(\alpha),
\end{aligned}
$$

where, the last equality follows from the definition (48) of $\Lambda^*_{E^2}(\alpha)$ that $\inf_{\alpha\geq\mathcal{B}^2}\Lambda^*_{E^2}(\alpha) = \Lambda^*_{E^2}(\mathcal{B}^2) = 0$ if $\mathcal{A}^2 \geq \mathcal{B}^2$. As β is arbitrary we have that

$$
\begin{aligned}
\liminf_{x\to\infty}\frac{1}{x}\log P\{L^1_0 > x\} &\geq -\inf_{\beta>0}\inf_{\alpha>\mathcal{B}^1+\beta}\left\{\frac{\Lambda^*_{A^1}(\alpha)}{\beta}\right\} = -\inf_{\alpha>\mathcal{B}^1}\inf_{\alpha-\mathcal{B}^1>\beta}\left\{\frac{\Lambda^*_{A^1}(\alpha)}{\beta}\right\} \\
&= -\inf_{\alpha>\mathcal{B}^1}\left\{\frac{\Lambda^*_{A^1}(\alpha)}{\alpha-\mathcal{B}^1}\right\} = -\theta^*_{12}(v_1),
\end{aligned}
$$

where, the last second equality follows from the fact that $1/x$ is a continuous decreasing function for $x > 0$, and the last equality follows from Lemma 4.2. Here $\theta^*_{12}(v_1)$ is the unique positive solution of the equation: $\Lambda_{A^1}(\theta) = \mathcal{B}^1\theta$. However, by the definition (21), we have $\Lambda_{E^2}(\theta) = \mathcal{B}^2\theta$ in the case $\mathcal{A}^2 \geq \mathcal{B}^2$. Thus, $c_1\theta - \Lambda_{E^2}(v_1\theta) = c_1\theta - \mathcal{B}^2v_1\theta = c_1\theta - v_1q_2c_2/(q_1+q_2)\theta = q_1c_1/(q_1+q_2)\theta = \mathcal{B}^1\theta$. So $\theta^*_{12}(v_1)$ is in fact the unique solution of the equation $\Lambda_{A^1}(\theta) + \Lambda_{E^2}(v_1\theta) = c_1\theta$, which is identical to the equation (25) because $l_1 = v_1$ in this case.

(**ii**) $\mathcal{A}^2 < \mathcal{B}^2$: Let $\alpha_i \geq \mathcal{A}^i, i = 1,2$ and $\alpha_1 + \alpha_2 > c_1 + \beta$. Then,

$$
\begin{aligned}
&P\{\ \frac{S^{A^1}_{-t}}{t} + v_1\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > c_1 + \beta\ \} \\
&\geq P\{\ \frac{S^{A^1}_{-t}}{t} > \alpha_1,\quad v_1\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > \alpha_2\ \} \\
&= P\{\ \frac{S^{A^1}_{-t}}{t} > \alpha_1\ \}P\{v_1\frac{\min_{0\le\tau\le t}\{\ S^{A^2}_{-\tau} + S^{B^2}_{-t,-\tau}\ \}}{t} > \alpha_2\ \}.
\end{aligned}
$$

We have that

$$
\liminf_{x\to\infty}\frac{1}{x}\log P\{L^1_0 > x\} \tag{43}
$$

$$\geq \frac{1}{\beta}\left(\liminf_{t\to\infty}\frac{1}{t}\log P\{\ \frac{S^{A^1}_{-t}}{t}>\alpha_1\ \}+\liminf_{t\to\infty}\frac{1}{t}\log P\{\frac{\min_{0\leq\tau\leq t}\{\ S^{A^2}_{-\tau}+S^{B^2}_{-t,-\tau}\ \}}{t}>v_2\alpha_2\ \}\right)$$
$$\geq -\frac{1}{\beta}\left(\inf_{x\geq\alpha^1}\Lambda^*_{A^1}(x)+\inf_{x\geq v_2\alpha^2}\Lambda^*_{E^2}(x)\right)=-\frac{1}{\beta}(\Lambda^*_{A^1}(\alpha_1)+\Lambda^*_{E^2}(v_2\alpha_2)),$$

where, the last equality follows from the increasing properties of $\Lambda^*_{A^2}(\cdot)$ and $\Lambda^*_{E^2}(\cdot)$, and $v_2=1/v_1$. As β is arbitrary we have that

$$\liminf_{x\to\infty}\frac{1}{x}\log P\{L^1_0>x\}\geq -\inf_{\beta>0}\frac{1}{\beta}\inf_{\{\alpha_i\in\mathbf{R},\alpha_i\geq\mathcal{A}^i,i=1,2;\alpha_1+\alpha_2>c_1+\beta\}}\{\Lambda^*_{A^1}(\alpha_1)+\Lambda^*_{E^2}(v_2\alpha_2)\}$$
$$= -\inf_{\{\alpha_i\in\mathbf{R},\alpha_i\geq\mathcal{A}^i,i=1,2;\alpha_1+\alpha_2>c_1\}}\inf_{\alpha_1+\alpha_2-c_1>\beta}\left\{\frac{\Lambda^*_{A^1}(\alpha_1)+\Lambda^*_{E^2}(v_2\alpha_2)}{\beta}\right\}\tag{44}$$
$$= -\inf_{\{\alpha_i\in\mathbf{R},\alpha_i\geq\mathcal{A}^i,i=1,2;\alpha_1+\alpha_2>c_1\}}\left\{\frac{\Lambda^*_{A^1}(\alpha_1)+\Lambda^*_{E^2}(v_2\alpha_2)}{\alpha_1+\alpha_2-c_1}\right\}$$
$$= -\inf_{\alpha>c_1}\left\{\frac{I^*(\alpha)}{\alpha-c_1}\right\}=-\theta^*_{12}(v_1),$$

where,

$$I^*(\alpha)\equiv -\inf_{\{\alpha_i\in\mathbf{R},\alpha_i\geq\mathcal{A}^i,i=1,2;\alpha_1+\alpha_2>c_1\}}\{\Lambda^*_{A^1}(\alpha_1)+\Lambda^*_{E^2}(v_2\alpha_2)\},$$

and $\theta^*_{12}(v_1)$ is the unique solution of the equation $\Lambda_{A^1}(\theta)+\Lambda_{E^2}(v_1\theta)=c_1\theta$. Let $I(\theta)=\sup_{\alpha\in\mathbf{R}}\{\theta\alpha-I^*(\alpha)\}$. By Lemma 4.2, if we can prove that $I(\theta)=\Lambda_{A^2}(\theta)+\Lambda_{E^2}(v_1\theta)$ and $I'(0)<c_1$, then the last equality in (44) is obtained. First, we have that

$$I(\theta)=\sup_{\alpha\in\mathbf{R}}\{\ \theta\alpha-\inf_{\{\alpha_i\in\mathbf{R},\alpha_i\geq\mathcal{A}^i,i=1,2;\alpha_1+\alpha_2>c_1\}}\{\Lambda^*_{A^1}(\alpha_1)+\Lambda^*_{E^2}(v_2\alpha_2)\}\ \}$$
$$=\sup_{\alpha\in\mathbf{R}}\sup_{\{\alpha_i\in\mathbf{R},\alpha_i\geq\mathcal{A}^i,i=1,2;\alpha_1+\alpha_2>c_1\}}\{\ \theta\alpha-\Lambda^*_{A^1}(\alpha_1)-\Lambda^*_{E^2}(v_2\alpha_2)\ \}$$
$$=\sup_{\alpha_1\in\mathbf{R},\alpha_2\in\mathbf{R}}\{\ \theta\alpha_1+\theta\alpha_2-\Lambda^*_{A^1}(\alpha_1)-\Lambda^*_{E^2}(v_2\alpha_2)\ \}$$
$$=\sup_{\alpha_1\in\mathbf{R},\alpha_2\in\mathbf{R}}\{\ (\theta\alpha_1-\Lambda^*_{A^1}(\alpha_1))+(v_1\theta v_2\alpha_2-\Lambda^*_{E^2}(v_2\alpha_2))\ \}$$
$$=\sup_{\alpha_1\in\mathbf{R}}\{\ \theta\alpha_1-\Lambda^*_{A^1}(\alpha_1)\ \}+\sup_{\alpha_2\in\mathbf{R}}\{\ v_1\theta v_2\alpha_2-\Lambda^*_{E^2}(v_2\alpha_2)\ \}=\Lambda_{A^2}(\theta)+\Lambda_{E^2}(v_1\theta).$$

Secondly, since $\mathcal{A}^2<\mathcal{B}^2$ and the stability condition (3), we have

$$I'(\theta)|_{\theta=0}=(\Lambda'_{A^1}(\theta)+v_1\Lambda'_{E^2}(v_1\theta))|_{\theta=0}=\mathcal{A}^1+v_1\mathcal{A}^2<\mathcal{B}^1+v_1\mathcal{B}^2=c_1.$$

Hence, we obtain the lower bound (24) in the case $\mathcal{A}^2<\mathcal{B}^2$.

*CASE*2. $v_1<1$:

(**i**) $\mathcal{A}^2\geq\mathcal{B}^2$: Note that $v_1\mathcal{B}^2<\mathcal{B}^2$ in this case. We still have that $\inf_{\alpha\geq v_1\mathcal{B}^2}\Lambda^*_{E^2}(\alpha)=0$. By the same procedure used in *CASE*1(**i**), we have

$$\liminf_{x\to\infty}\frac{1}{x}\log P\{L^1_0>x\}\geq=-\inf_{\alpha>\mathcal{B}^1}\left\{\frac{\Lambda^*_{A^1}(\alpha)}{\alpha-\mathcal{B}^1}\right\}=-\theta^*_{12}(v_2)$$

where, the last equality follows from Lemma 4.2, namely, $\theta^*_{12}(v_2)$ is the unique positive solution of the equation: $\Lambda_{A^1}(\theta)=\mathcal{B}^1\theta$. Again, we have that $\Lambda_{E^2}(\theta)=\mathcal{B}^2\theta$ in this case.

Then, $c_1\theta - \Lambda_{E^2}(v_1\theta) = c_1\theta - \mathcal{B}^2 v_1\theta = \mathcal{B}^1\theta$. Hence, $\theta_{12}(v_2)$ is actually the unique solution of the equation $\Lambda_{A^1}(\theta) + \Lambda_{E^2}(\theta) = c_1\theta$, which is identical to the equation (25) because $l_1 = v_1$ in the case $\mathcal{A}^2 \geq \mathcal{B}^2$.

(ii) $\mathcal{A}^2 < \mathcal{B}^2$: The lower bound can be obtained by replacing v_1 by 1 in the proof of $CASE1$(**ii**). We omit it here. □

Remark 1. From the relation (32), we obtain that $\Lambda_{B^i}(-\theta) = -c_i\theta + \Lambda_{B^j}(v_i\theta), i \neq j$. Substituting it into the equation $\Lambda_{A^i}(\theta) + \Lambda_{B^i}(-\theta) = 0$ yields that δ_i^* is the largest solution of the equation $\Lambda_{A^i}(\theta) + \Lambda_{B^j}(v_i\theta) = c_i\theta$, i.e. $\alpha_{A^i}(\theta) + v_i\alpha_{B^j}(v_i\theta) = c_i$. As $\Lambda_{E^i}(\theta) \leq \Lambda_{D^i}(\theta) \leq \Lambda_{B^i}(\theta)$ for $\theta \geq 0$, we have that $\theta_{ij}^*(v_i) \geq \Theta_{ij}^*(v_i) \geq \delta_i^*$, which means that the roots of the equations (23) and (25) can not be obtained before δ_i^*.

5. Conclusion

In this paper, we have analyzed a discrete-time polling system under the Bernoulli service schedule and presented the large deviations upper and lower bounds of the buffer overflow probabilities. These results can be used in traffic management of high-speed communication networks such as call admission control and bandwidth allocation problems. For instance, utilizing the relations obtained between the large deviations bounds and the parameters p_i, q_i, we can guarantee the different QoS requirements for the two queues via controlling the values of p_i, q_i.

As have been seen, the large deviations upper and lower bounds here do not match exactly. The reason is that we used the effective bandwidths of the stationary departure process in deducing the upper bound and the effective bandwidths of the transient departure process in deducing the lower bound. When the server allocates its service capacity to a queueing system randomly, a large deviations in the departure from the stationary queue may be encouraged. This results in the difference between the two rate functions of the large deviations of the stationary departure process and the transient departure processes. This phenomenon has been observed by Chang and Zajic [11] and it does not occur when the service capacity is deterministic, e.g. GPS service policy in [27]. Therefore, developing a method to give matched large deviations upper and lower bounds for the polling system still is an open problem. We will take this as the further investigation subject.

References

[1] D. Bertsimas, I.C. Paschalidis and J.N. Tsitsiklis, “Asymptotic buffer overflow probabilities in Multiclass multiplexers: An optimal control approach,” IEEE Trans. Automatic Control 43(1998)315-335.

[2] D. Bertsimas, I.C. Paschalidis and J.N. Tsitsiklis, “Large deviations analysis of the generalized processor sharing policy,” Queueing Systems 32(1999)319-349.

[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(1995)293-320.

[4] O.J. Boxma and W.P. Groenendijk, “Two queues with alternating service and switching times,” in: *Queueing Theory and its Applications,* ed. O.J. Boxma, and R. Syski, North-Holland, Amsterdam, (1988)261-281.

[5] H. Bruneel, “Queueing behavior of statistical multiplexers with correlated inputs,” IEEE Trans. Commu. 36(1988)1339-1341.

[6] H. Bruneel and B. Steyaert, “Storage requirements in ATM switching elements with co-

rrelated arrivals and independent uniform routing," Performance Evaluation 25(1996)193 -209.

[7] J.A. Buchlew, *Large-Deviation Techniques in Decision, Simulation and Estimation*. Wiley, New York, 1990.

[8] C.S. Chang, "Stability, queue length and delay of deterministic and stochastic queueing networks," IEEE Trans. Automatic Control 39(1994)913-931.

[9] C.S. Chang, "Sample path large deviations and intree network," Queueing Systems 20 (1995)45-66.

[10] C.S. Chang and J.A. Thomas, "Effective bandwidth in high-speed digital networks," IEEE J. Select. Arears Commun. 13(1995)1091-1101.

[11] 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.

[12] F. Delcoigne and A. de La Fortelle, Large deviations rate function for polling systems, Queueing Systems, 41 (2002) 13-44.

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

[14] N.G. Duffield, "Exponential bounds for queues with Markovian arrivals," Queueing Systems 17(1994)413-430.

[15] A.I. Elwalid and D. Mitra, "Effective bandwidth of general Markovian traffic sources and admission control of high speed networks," IEEE/ACM Trans. Networking 1(1993) 329-343.

[16] M. Eisenberg, "Two queues with changeover times," Oper. Res. 19(1971)386-401.

[17] M. Eisenberg, "Two queues with alternating service," SIAM J. Appl. Math. 36(1979)287 -303.

[18] W. Feng, M. Kowada and K. Adachi, "A two-queue model with Bernoulli service schedule and switching times," Queueing Systems 30(1998)405-434.

[19] P.W. Glynn and W. Whitt, "Logarithmic asymptotics for steady-state tail probabilities in a single-server queue," J. Appl. Probab. 31A(1994)131-156.

[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] T. Katayama and Y. Takahashi, "Analysis of a two-class priority queue with Bernoulli schedules," J. Oper. Res. Jap. 35(1992)236-249.

[22] D.S. Lee, "Analysis of a cyclic server queue with Bernoulli schedules," J. Appl. Prob. 34(1997)176-191.

[23] H. Takagi, *Analysis of Polling Systems*, MIT Press Cambridge, Massachusetts, 1986.

[24] H. Takagi, "Queuing analysis of polling models," ACM Comput. Surveys 20(1988)5-28.

[25] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," IEEE Trans. Inform. Theo. 39(1993)466-478.

[26] Tedijanto, "Exact results for the cyclic-service queue with a Bernoulli schedule," Performance Evaluation 11(1990)107-115.

[27] A. Weiss, "An introduction to large deviations for communication networks," IEEE J. Selec. Areas Commun. 13(1993)938-385.

[28] Z.-L. Zhang, "Large deviations and the generalized processor sharing schedule for a two-queue system," Queueing Systems 26(1997)229-254.

[29] Z.-L. Zhang, "Large deviations and the generalized processor sharing schedule for a multiple-queue system," Queueing Systems 28(1998)349-276.

[30] W. Whitt, "Tail probabilities with statistical multiplexing and effective bandwidths in multiclass queue," Telecommun. Syst. 2(1993)71-107.

‡ Department of Engineering physics, Electronics and Mechanics, Graduate School of Engineering, Nagoya Institute of Technology

† The Faculty of Business Administration of Information Science, Chubu University