Polling Models with Unequal Service Rates under Limited Service Policies - Sharp Asymptotics

Douglas G. Down
downd@mcmaster.ca
Department of Computing and Software
McMaster University
1280 Main Street West
Hamilton, Ontario L8S 4L7
Canada

January 27, 2012

Abstract

We derive asymptotic expressions for the distribution of the total queue length in a polling model with two classes of customers and unequal service rates. The server employs a scheduling policy that alternately visits each queue, with the maximum number served in each visit potentially being different for each queue. We provide sufficient conditions for the behaviour to lie in one of two regimes, depending on the system parameters. The first regime, called codominant, has both queues tending to grow as the total system size grows. The single-class dominant regime has only one queue tending to grow as the total system size grows. Finally, we present numerical results that demonstrate that the developed conditions are only sufficient and comment on the implications of this observation.

1 Introduction

We consider a single server, k_i-limited polling model with two customer classes. Arrivals of class i occur according to a Poisson process of rate λ_i. The service times for a customer of class i are exponentially distributed with rate μ_i. By k_i-limited polling model, we mean that when a server visits class i, it serves k_i customers of class i (if possible) before switching to the other class. Service is non-preemptive and non-idling, so the server will not idle at an empty queue if there are customers waiting at the other. Without loss of generality, we will assume that the classes are labeled such that $\lambda_1/k_1 \geq \lambda_2/k_2$. The k_i-limited policy is a natural generalization of the round-robin policy, i.e. a policy in which the server alternately serves each queue (as long as there is a customer to serve). The round-robin policy enforces a notion of fairness. Generalizing to a different number of customers served on each visit to a queue allows a system designer to place more importance on a particular queue. Such policies have seen applications in communications [1, 5] and logistics [15].

There is a huge literature relating to polling systems, a testament to their usefulness. Two surveys, that of Takagi [16] and a more recent one by Vishnevskii and Semenova [18] are recommended. With this wealth of literature, however, the problem of determining the invariant (or steady-state) distribution for our model is one that remains open. In this work, we are interested in computing a function of the invariant distribution: the tail asymptotics of the invariant distribution of the total queue length. In other words, we are interested in exactly computing the probability that the total queue length is large. This is of interest when one is concerned with rare, potentially catastrophic events. To be precise, we would like to calculate constants c and α such that the probability that the total number in system is equal to ℓ is asymptotically equal (as ℓ goes to infinity) to $c\alpha^{-\ell}$. The *rate* (or *rough asymptotics*) for the system is given by $1/\alpha$, while if one can calculate c, the expression $c\alpha^{-\ell}$ is known as the *exact asymptotics*.

This work continues studies begun in [3, 4]. The main difference here is that [3, 4] assumed that the service rate was independent of the customer class. In that case, it is easy to see that the total queue length behaves as an M/M/1 queue, so the focus is on more detailed behaviour, such as the relative proportion of the different customers when a large total queue length is reached. Here, we remain interested in the detailed behaviour, however in this case it is also not obvious what the rate will be. So, while the problem is more difficult than that in [3], we find that in particular, the work in [2, 3, 4] makes the analysis tractable so that the techniques from McDonald [14] (elaborated on in Foley and McDonald [10]) can be applied. Over the course of our work, we find an interesting adaptation needs to be used. We in fact apply the methodology from [10, 14] twice, in order to sharpen our results. This technique may be of independent interest.

Besides the work undertaken in this research program ([3, 4]), the most relevant work is that of Delcoigne and De La Fortelle [7]. They identify the local rate function for a scaled version of the queue length process in a general polling model. The model in [7] is different than that considered here in that after each service completion, the server randomly chooses (according to some distribution) the next queue to serve. So, unfortunately, we cannot leverage their work on identifying the rate function to aid in finding the rate for the system that we consider. Other work that has been done for large deviations in polling models are in Choudhury and Whitt [6], Duffield [9], and Ioresh [12], but all of these study a different kind of service policy (exhaustive or gated types).

The organization of the paper is as follows. Section 2 constructs a Continuous Time Markov Chain model for the system. Section 3 gives the main results, while Section 4 provides the proofs of the main results. Section 5 provides numerical results that explore the issue of whether the parameter space is completely covered (it is not) and how the solution varies as particular parameters change. Section 6 provides final thoughts.

2 Continuous Time Markov Chain model

A Continuous Time Markov Chain (CTMC) for this system is given by

$$Q(t) = (Q_1(t), Q_2(t), Z(t), I(t)),$$

where $Q_i(t)$ is the number of waiting customers of class i (including the one in service, if applicable), $Z(t)$ is the class being served (we will arbitrarily set this to 1 if $Q_1(t) + Q_2(t) = 0$), and $I(t)$ is the number of service completions during the current server visit (we will assume that if this reaches $k_i - 1$ during a visit to class i and the other queue is empty, that a service completion at class i will leave $I(t)$ unchanged at $k_i - 1$). The state space for $Q(t)$ is $S = \mathbb{Z}_+ \times \mathbb{Z}_+ \times \{1, 2\} \times \{0, \ldots, \max(k_1, k_2) - 1\}$. We will look at the uniformized chain, $Q[n]$, where we assume (without loss of generality) that time has been rescaled such that $\lambda_1 + \lambda_2 + \mu_1 + \mu_2 = 1$. We will denote the transition kernel for $Q[n]$ by K, where $K(x, y)$ gives the probability of moving in one step to state y, given that the chain started in state x. For example, if $x = (i, j, 2, k)$, with $i, j > 0$, $k_2 > 1$, and $k \leq k_2 - 2$,

$$\begin{aligned}
K(x, (i+1, j, 2, k)) &= \lambda_1 \\
K(x, (i, j+1, 2, k)) &= \lambda_2 \\
K(x, x) &= \mu_1 \\
K(x, (i, j-1, 2, k+1)) &= \mu_2
\end{aligned}$$

Using a simple workload argument, the existence of a unique invariant distribution for $Q[n]$, π_Q, is guaranteed if the load on the system is less than one, i.e.

$$\rho := \lambda_1/\mu_1 + \lambda_2/\mu_2 < 1.$$

We will assume that this stability condition holds. Note that while the condition for stability is known, there is no known explicit expression for π_Q. (One can compute generating functions for the invariant distribution, see the work of Lee [13], for example.)

In the next section we give expressions for the probability that the total system size is large, i.e. we are interested in the asymptotic behaviour of the event $F_\ell = \{Q_1[n] + Q_2[n] = \ell\}$. One could get similar results for related events, such as the probability that a particular queue length is large.

3 Main Results

In this section, we provide the main results, sufficient conditions and expressions for the exact asymptotics under different parameter combinations. In particular, Section 3.1 identifies when the exact asymptotics are determined due to the influence of both classes of customers, while Section 3.2 discusses the case when the exact asymptotics are determined due to the influence of c;lass 1 customers only.

3.1 Codominant case

The goal of this section is to provide a sufficient condition for the case when the exact asymptotics for the total queue length are determined by state trajectories where both queues are large. Here, the exact asymptotics rely on three values: α, β and γ, which depend on the system parameters in a non-trivial fashion. Unfortunately, it is difficult to give direct intuition for these constants, other than that they are required to construct a harmonic function for $Q[n]$.

First, we need to solve the following for α. There are in general multiple solutions and it is easy to see that one of these is $\alpha = 1$.

$$1 = \left(\frac{\alpha}{\mu_1}\right)^{k_1} \left(\frac{\alpha}{\mu_2}\right)^{k_2} (1 - \lambda_1\alpha - \lambda_2\alpha - \mu_2)^{k_1} (1 - \lambda_1\alpha - \lambda_2\alpha - \mu_1)^{k_2} \tag{1}$$

Given a solution α, we also define the values:

$$\beta = \frac{\alpha}{\mu_2}(1 - \lambda_1\alpha - \lambda_2\alpha - \mu_1) \tag{2}$$

$$\gamma = \beta^{-k_2/k_1}. \tag{3}$$

Note here that it cannot be the case that there is more than one solution of (1)-(3) satisfying (4) as this would contradict the uniqueness of the invariant distribution π_Q. Of course, there may be no valid solution.

If there exist α, β, γ satisfying (1)-(3) such that $\alpha > 1$ is real-valued and

$$\lambda_2 \left(\frac{k_2}{\beta\mu_2} + \frac{k_1}{\gamma\mu_1} \right) > \frac{k_2}{\alpha^2} \tag{4}$$

then we will call the system *codominant*. In Section 4.1, we provide a more detailed discussion on solutions to (1)-(3). Here, F_ℓ is reached by both queue lengths getting large. For such systems, the asymptotic behaviour is characterized in the following theorem. For arbitrary functions f and g, the notation $f \sim g$ denotes $\lim_{\ell\to\infty} f(\ell)/g(\ell) = 1$ and $T_\ell = \min\{n \geq 0 : Q_1[n] + Q_2[n] = \ell\}$ is the first time that F_ℓ is reached.

Theorem 1 *For a codominant system*

(i)

$$E[T_\ell | Q[0] = (0,0,2,0)] \sim \alpha^\ell g^{-1}$$

where g is given in (13). Note that g may be obtained by fast simulation, i.e. we do not need to estimate events with low probability.

(ii)

$$P_{\pi_Q}\{Q_1[n] + Q_2[n] = \ell, Z[n] = 1, I[n] = k\} \sim \alpha^{-\ell}\gamma^{-k}\beta^{-k_2} \frac{\beta\mu_2}{k_2\gamma\mu_1 + k_1\beta\mu_2} f/\tilde{d}_1,$$

$$P_{\pi_Q}\{Q_1[n] + Q_2[n] = \ell, Z[n] = 2, I[n] = k\} \sim \alpha^{-\ell}\beta^{-k} \frac{\gamma\mu_1}{k_2\gamma\mu_1 + k_1\beta\mu_2} f/\tilde{d}_1,$$

where

$$\tilde{d}_1 = \frac{k_1(\alpha\lambda_1 + \alpha\lambda_2 - \gamma\mu_1/\alpha)\beta\mu_2}{k_2\gamma\mu_1 + k_1\beta\mu_2} + \frac{k_2(\alpha\lambda_1 + \alpha\lambda_2 - \beta\mu_2/\alpha)\gamma\mu_1}{k_2\gamma\mu_1 + k_1\beta\mu_2}$$

and f is given in (12) and may be obtained by fast simulation.

(iii)

$$\lim_{\ell\to\infty} \left(\frac{Q_1[T_\ell]}{\ell}, \frac{Q_2[T_\ell]}{\ell} \right) = \left(\frac{\lambda_1\alpha - \frac{\gamma\mu_1\beta\mu_2}{\alpha(k_2\gamma\mu_1 + k_1\beta\mu_2)}}{\tilde{d}_1}, \frac{\lambda_2\alpha - \frac{\beta\mu_2\gamma\mu_1}{\alpha(k_2\gamma\mu_1 + k_1\beta\mu_2)}}{\tilde{d}_1} \right).$$

3.2 Class one dominant case

As in the previous section, we are interested in determining an exact asymptotic expression for the total number of customers in the system, this time for the case when the asymptotics are determined by the state trajectory where there are a large number of class one customers and a small number of class two customers. As in the codominant case, this involves three values: α, β and γ, all non-trivial functions of the system parameters. Here, a two-stage procedure is required. We also need to find additional values α', β' and γ' to define our conditions.

Step 1. Solve the following for α':

$$1 = \left(\frac{\alpha'}{\mu_2}\right)^{k_2} (\lambda_2 + \mu_2 - \lambda_2\alpha')^{k_2} \left(\frac{1}{\mu_1}\right)^{k_1} (\lambda_2 + \mu_1 - \lambda_2\alpha')^{k_1}$$

Given a solution α', define

$$\begin{aligned} \beta' &= \frac{1}{\mu_1}(\lambda_2 + \mu_1 - \lambda_2\alpha') \\ \gamma' &= \beta'^{-k_1/k_2} \end{aligned}$$

If there is a solution such that $\alpha' > 1$ is real-valued and

$$\lambda_2\alpha'\left(\frac{k_2\alpha'}{\mu_2\gamma'} + \frac{k_1}{\mu_1\beta'}\right) > k_2$$

then call the resulting α', α^*.

Step 2. Solve the following for α:

$$\alpha^3(\mu_2\lambda_1 - \lambda_1\mu_1) + \alpha^2(\mu_1 - \mu_1\mu_2 + \mu_2\lambda_2 - \mu_2 + \mu_2^2 + \lambda_1\mu_1) + \alpha(\mu_1\mu_2 - \mu_1^2 - \mu_1 + \mu_1\mu_2) + \mu_1^2 = 0 \quad (5)$$

Given a solution α, define

$$\beta = \frac{1}{\lambda_2\alpha}(1 - \lambda_1\alpha - \mu_2 - \mu_1/\alpha) \quad (6)$$

If there exists a solution to (5) and (6) with $\alpha > 1$ real-valued satisfying

$$\begin{aligned} \lambda_2\beta\left(\frac{\beta k_2}{\mu_2}+\frac{k_1}{\mu_1}\right) &< \frac{k_2}{\alpha^2} && (7) \\ \frac{\alpha^2\beta^2\lambda_2}{\mu_2}+\frac{\alpha^2\lambda_1}{\mu_1} &> 1 && (8) \\ \beta &> 1/\alpha^* \end{aligned}$$

we will call the system *class one dominant.*

Theorem 2 *For a class one dominant system*

$$P_{\pi_Q}\{Q_1[n]+Q_2[n]=\ell, Q_1[n]=j, Z[n]=k, I[n]=m\} \sim \alpha^{-\ell}\beta^{-j}\varphi(j,k,m)g/\tilde{d}_1, \qquad (9)$$

where g may be obtained by fast simulation,

$$\tilde{d}_1 = \frac{k_1\alpha}{\mu_1}\left(\frac{\alpha\lambda_1}{1-\alpha^2\beta^2\lambda_2/\mu_2}-1\right)$$

and φ is defined in Section 4.2.

Note that there can be only one solution of (5) and (6) satisfying the class one dominant conditions and also that a system cannot be both codominant and class one dominant, due to the uniqueness of the invariant distribution π_Q. Also note that in Theorem 2, we cannot provide an explicit expression for φ. However, it is useful to note that to evaluate (9), both g and φ can be estimated by a fast simulation, i.e. we do not need to estimate events with small probabilities.

Note that there is no corresponding class two dominant condition (i.e. only queue two reaches a large level). The reason is that our assumption that $\lambda_1/k_1 \geq \lambda_2/k_2$ implies that when queue 2 grows large, with high probability queue 1 gets large (as the queue length at 2 goes to infinity, this probability goes to 1, see [2]). Thus we only consider the class one dominant or codominant cases.

4 Proofs

4.1 Codominant case (Theorem 1)

We use the methodology presented in Foley and McDonald [10]. To do so, we first identify a boundary Δ that corresponds to states where the queues that we expect to be large when reaching F_ℓ are empty. We construct a series of chains using the transition kernel for $Q[n]$, K, the boundary Δ and a harmonic function for one of these constructed chains.

The first step is to construct a chain W from Q such that its first component is at least ℓ on F_ℓ. Here, we use $W[n] = (Q_1[n] + Q_2[n], Q_1[n], Z[n], I[n])$. At this point, we construct the free chain, $W^\infty[n] = (\tilde{W}^\infty[n], \hat{W}^\infty[n])$, from W and Δ by extending the transition structure of W on S/Δ to the state space $S^\infty = \mathbb{Z} \times \mathbb{Z} \times \{1, 2\} \times \{0, \ldots, \max(k_1, k_2) - 1\}$ such that if we define $\tilde{W}^\infty[n] = (Q_1^\infty[n] + Q_2^\infty[n], Q_1^\infty[n])$ and $\hat{W}^\infty[n] = (Z^\infty[n], I^\infty[n])$, then $W[n]$ is Markov additive, with $\hat{W}^\infty[n]$ the Markovian part and $\tilde{W}^\infty[n]$ the additive part. Here, the transition kernel for W^∞, K_W^∞ is simply one where the server serves exactly k_i customers at each visit to class i, with a negative number of customers allowed at each class. We need to construct two more chains, the so-called *twisted free chain* and a *twisted basic chain*. The first is standard from [10], the second was introduced in Chang and Down [4] (although the terminology *twisted basic chain* was not introduced in [4]). The idea is that we modify the underlying distribution such that the rare events are common. We can then use the Markov additive structure to derive appropriate expressions for the asymptotic behaviour (this is all in [10]).

The appropriate twist requires finding a harmonic function of the form

$$h(x) = \alpha^{\tilde{x}_1} \hat{a}(z, k)$$

for the W^∞ chain and where the state is $x = (\tilde{x}_1, \tilde{x}_2, z, k)$. If we let $\hat{a}(1, k) = \gamma^k \beta^{k_2}$ for $0 \le k \le k_1 - 1$

and $\hat{a}(2,k) = \beta^k$ for $0 \leq k \leq k_2 - 1$, then by enumerating $K^\infty h(x) = h(x)$ over all possible states, we get that (α, β, γ) must satisfy

$$\begin{aligned} 1 &= \left(\frac{\alpha}{\mu_1}\right)^{k_1}\left(\frac{\alpha}{\mu_2}\right)^{k_2}(1-\lambda_1\alpha-\lambda_2\alpha-\mu_2)^{k_1}(1-\lambda_1\alpha-\lambda_2\alpha-\mu_1)^{k_2} \qquad (10) \\ \beta &= \frac{\alpha}{\mu_2}(1-\lambda_1\alpha-\lambda_2\alpha-\mu_1) \\ \gamma &= \beta^{-k_2/k_1} \end{aligned}$$

Characterizing solutions to this seems a difficult task. For example, the number of solutions to (10) with real-valued $\alpha > 1$ can vary. For example, $\lambda_1 = 0.1692$, $\lambda_2 = 0.0758$, $\mu_1 = 0.6217$, $\mu_2 = 0.1332$, $k_1 = 8$, $k_2 = 7$ yields no real-valued roots $\alpha > 1$, while $\lambda_1 = 0.0926$, $\lambda_2 = 0.0573$, $\mu_1 = 0.6396$, $\mu_2 = 0.2105$, $k_1 = 3$, $k_2 = 2$ yields three real-valued roots $\alpha > 1$. The twisted free chain, $\mathcal{W}^\infty[n]$ is described by its kernel, $\mathcal{K}^\infty(x,y)$, given by

$$\mathcal{K}^\infty(x,y) = K^\infty(x,y)h(y)/h(x). \qquad (11)$$

We will postpone the description of the twisted basic chain, but at this point, we note that the twisted free chain can be thought of as taking the free chain and multiplying the arrival rates λ_1 and λ_2 by α, the service rate μ_2 by β/α and the service rate μ_1 by γ/α.

We are now ready to show Conditions C.1-C.12 in [10] (in the interest of space, we do not repeat all of these conditions). Conditions C.1-C.5 have been satisfied by the construction given above.

For Condition C.6, if we let $\varphi(j,k)$ be the invariant probability for $\hat{\mathcal{W}}^\infty[n]$, we see that if it exists it satisfies the balance equations

$$\begin{aligned}
\gamma\varphi(1,k) &= \gamma\varphi(1,k-1), \quad k=1,\ldots,k_1-1 \\
\gamma\mu_1\varphi(1,0) &= \beta\mu_2\varphi(2,k_2-1) \\
\beta\varphi(2,k) &= \beta\varphi(2,k-1), \quad k=1,\ldots,k_2-1 \\
\beta\mu_2\varphi(2,0) &= \gamma\mu_1\varphi(1,k_1-1) \\
\sum_{j,k}\varphi(j,k) &= 1
\end{aligned}$$

The unique solution to this set of equations is

$$\begin{aligned}
\varphi(1,k) &= \frac{\beta\mu_2}{k_2\gamma\mu_1+k_1\beta\mu_2}, \quad k=0,\ldots,k_1-1 \\
\varphi(2,k) &= \frac{\gamma\mu_1}{k_2\gamma\mu_1+k_1\beta\mu_2}, \quad k=0,\ldots,k_2-1.
\end{aligned}$$

We next move to Condition C.8. This is equivalent to showing that with positive probability, both queues for the twisted free process go to infinity. Using Theorem 3 and Corollary 4 of Chang and Lam [2] exactly as in Lemma 1(ii) of [4], we see that as (4) is satisfied for $i=1,2$, both queues in the twisted free process go to infinity with probability one. This also immediately yields C.7, which requires that

$$\begin{aligned}
\tilde{d}_1 &= \sum_{\hat{x}\in\hat{S}}\varphi(\hat{x})E[\tilde{\mathcal{W}}^\infty[1]|\hat{\mathcal{W}}^\infty[0]=\hat{x}] \\
&= (\tilde{\lambda}_1+\tilde{\lambda}_2-\tilde{\mu}_1)k_1\varphi(1,0)+(\tilde{\lambda}_1+\tilde{\lambda}_2-\tilde{\mu}_2)k_2\varphi(2,0)
\end{aligned}$$

is finite and strictly positive. It is trivially finite and if it were not positive, it would contradict C.8. Condition C.9 is trivially satisfied as $\hat{S}$ is finite.

At this point, we construct the twisted basic chain $\mathcal{W}$. This is done by using the same twist as in (11) on the original chain W. Note that the resulting kernel may not define a Markov chain (as the function h is not harmonic for W), so we adjust the kernel by adding self-loops to give a probability kernel. If these adjustments are made on a finite number of states, then we can use the methodology introduced in [4]. Here, it is easy to see that the only state at which h is not harmonic for W is $(0,0,2,0)$. Condition C.10 (typically the most difficult condition to check) follows from Theorem 3 in [4], Condition C.11 follows as in [4] and C.12 follows from the fact that $\hat{S}$ is finite.

In order to complete the proof, we define the constants f and g used in Theorem 1. Here,

$$\begin{aligned} f &= \sum_{x\in\Delta} \pi(x)h(x)H(x) && (12)\\ g &= f\sum_{z\in\hat{S}} \mu(0,z)/h(z) && (13)\end{aligned}$$

where $H(x)$ is the probability that the twisted free chain never hits Δ, starting at x, μ is the stationary distribution of $(\tilde{\mathcal{W}}_1^\infty[\mathcal{T}_\ell^\infty]-\ell, \hat{\mathcal{W}}^\infty[\mathcal{T}_\ell^\infty])$, and $\mathcal{T}_\ell^\infty$ is the first time that the first component of the twisted free chain is ℓ. Note that f and g can be evaluated by a fast simulation.

Part (i) of Theorem 1 follows from Theorem 6 of [10], (ii) from Theorem 5 of [10] and the expression for $\varphi(j,k)$, and (iii) from Corollary 2 of [10]. ♢

4.2 Class one dominant case (Theorem 2)

We perform a similar construction as in the previous section, with the main difference being that the boundary Δ, and hence the subsequent chains, are changed. In this case, Δ corresponds to removing the constraint that queue 1 is empty, i.e. for the free chain, we will allow exactly k_1 customers to be served during each visit to class 1.

Here, we have $W[n] = (Q_1[n]+Q_2[n], Q_2[n], Z[n], I[n])$ and $W^\infty[n] = (\tilde{W}^\infty[n], \hat{W}^\infty[n])$ where $\tilde{W}^\infty[n] = (Q_1^\infty[n]+Q_2^\infty[n])$, $\hat{W}^\infty[n] = (Q_2^\infty[n], Z^\infty[n], I^\infty[n])$ and W^∞ evolves on $S^\infty = \mathbb{Z}\times\mathbb{Z}_+\times\{1,2\}\times\{0,\dots,\max(k_1,k_2)-1\}$ (note the changes from the previous section: as we expect F_ℓ to be reached through queue 1 only, we include the queue length at class 2 in $\hat{W}^\infty$, with a corresponding adjustment to the state space).

To construct the twisted free chain, we search for a harmonic function of the form

$$h(x) = \alpha^{\tilde{x}_1}\hat{a}(j,z,k)$$

for the W^∞ chain. Letting $\hat{a}(j,z,k) = \beta^j$, by evaluating $K^\infty h(x) = h(x)$ over all possible states, we see that (α, β) must satisfy

$$\begin{aligned} 0 &= \alpha^3(\mu_2\lambda_1 - \lambda_1\mu_1) + \alpha^2(\mu_1 - \mu_2\mu_1 + \mu_2\lambda_2 - \mu_2 + \mu_2^2 + \lambda_1\mu_1) \\ &\quad +\alpha(\mu_2\mu_1 - \mu_1^2 - \mu_1 + \mu_2\mu_1) + \mu_1^2 \\ \beta &= \frac{1}{\lambda_2\alpha}(1 - \lambda_1\alpha - \mu_2 - \mu_1/\alpha) \end{aligned}$$

As in the previous section, the twisted free chain, $\mathcal{W}^\infty[n]$ has transition kernel

$$\mathcal{K}^\infty(x,y) = K^\infty(x,y)h(y)/h(x),$$

which is equivalent to multiplying λ_2 by $\alpha\beta$, λ_1 by α, and dividing μ_1 by α and μ_2 by $\alpha\beta$. Let these modified rates be given by $\tilde{\lambda}_2$, $\tilde{\lambda}_1$, $\tilde{\mu}_1$, and $\tilde{\mu}_2$, respectively.

Once again Conditions C.1-C.5 in [10] have been satisfied by the above construction.

Condition C.6 is equivalent to showing the stability of queue 2 in the twisted free chain. Using Theorem 3 and Corollary 4 of [2], we have that queue 2 is stable (and thus a stationary probability φ for $\hat{\mathcal{W}}[n]$ exists) if

$$\tilde{\lambda}_2\left(\frac{k_1}{\tilde{\mu}_1} + \frac{k_2}{\tilde{\mu}_2}\right) < k_2.$$

(This has the physical interpretation that the maximum expected number of arrivals in the longest server cycle is less than the number of services in that cycle.) This is equivalent to (7). Using the same results from [2], (8) yields instability of class 1, which implies C.7 and C.8. In addition $\tilde{d}_1$ is simply the expected change in the total queue length in a cycle, which is

$$\tilde{d}_1 = \frac{k_1}{\tilde{\mu}_1}\left(\frac{\tilde{\lambda}_1}{1-\tilde{\lambda}_2/\tilde{\mu}_2} - 1\right)$$

which must be positive as queue 1 is unstable for the twisted free process.

We now turn to Condition C.9. We must show that

$$\sum_{\hat{x}\in\hat{S}} \varphi(\hat{x})/\hat{a}(\hat{x}) < \infty.$$

It is not hard to see that proving this is related to the asymptotics of $\hat{W}^\infty$. To do this, we apply the methodology of [10] a second time, with a goal of identifying the rough asymptotics of the $\hat{W}^\infty$ chain. The $\hat{W}^\infty$ chain describes a system in which a server serves a queue until either k_2 services have been performed or the queue becomes empty. This is then followed by the server going on vacation for an Erlang distributed period of time, with k_1 stages, each with mean $1/\mu_1$. The dynamics for the free chain (derived from $\hat{W}^\infty$) allow queue 2 to be negative, in other words k_2 customers are always served. Let K_2^∞ be the kernel for the free chain derived from $\hat{W}^\infty$. We give a couple of elements of K_2^∞ here, in the interest of space we do not give it in its entirety. For example,

$$\begin{aligned}
K_2^\infty((j,2,k),(j-1,2,k+1)) &= \mu_2, \quad -\infty < j < \infty,\ 0 \le k < k_2 \\
K_2^\infty((j,1,k),(j,1,k+1)) &= \mu_1, \quad -\infty < j < \infty,\ 0 \le k < k_1 \\
K_2^\infty((j,i,k),(j+1,i,k)) &= \lambda_2, \quad -\infty < j < \infty,\ i = 1,2,\ 0 \le k < k_i
\end{aligned}$$

Now, we need to find a harmonic function for K_2^∞ of the form $h_2(j,i,k) = \alpha'^j \hat{a}(i,k)$. By enumerating over all possible values of j, i, k, and letting $\hat{a}(1,k) = \gamma'^k \beta'^{k_2}$ and $\hat{a}(2,k) = \beta'^k$, we find that

α', β', γ' must satisfy

$$\begin{aligned} 1 &= \left(\frac{\alpha'}{\mu_2}\right)^{k_2} (\lambda_2 + \mu_2 - \lambda_2\alpha')^{k_1} \left(\frac{1}{\mu_1}\right)^{k_1} (\lambda_2 + \mu_1 - \lambda_2\alpha')^{k_1} \\ \beta' &= \frac{1}{\mu_1}(\lambda_2 + \mu_1 - \lambda_2\alpha') \\ \gamma' &= \beta'^{-k_1/k_2} \end{aligned}$$

We form the twisted chain as before, using the transition function $\mathcal{K}_2^\infty(x, y) = K_2^\infty(x, y)h_2(y)/h_2(x)$. All of the technical conditions except for C.7 are satisfied either by construction or due to the fact that the additive part is simply a single queue length, or the finiteness of the state for the Markovian part. It is easy to see that in this case, C.7 corresponds to

$$\lambda_2\alpha'\left(\frac{k_2\alpha'}{\mu_2\gamma'} + \frac{k_1}{\mu_1\beta'}\right) > k_2.$$

If there exists an α' which satisfies this, we will call it α^*. If this holds, then we have from Theorem 5 of [10], that for some $0 < c < \infty$, $P\{Q_1[n] = \ell\} \sim c(\alpha^*)^\ell$. Note that α^* would be unique if it exists, as the invariant distribution is unique. This vacation model could be of independent interest (vacation models are a well studied area, see the surveys by Doshi [8] and Takagi [17]). It may be worthwhile to derive sharp asymptotics for these, but at this point, we only need the rate. Returning to the original problem, given the rate α^*, it is easy to see that $\beta > 1/\alpha^*$ allows C.9 to be satisfied for the W^∞ chain.

The use of the methodology twice appears to be a novel approach to demonstrating this condition. While we do not see any immediate applications of this technique, it may be useful for future problems where the harmonic function cannot be explicitly computed, or for systems (such as polling models) where one queue remains stable while others become unstable.

Finally, for the W^∞ chain, conditions C.10 and C.11 follow in the same manner as the codominant case.

The main result in Theorem 2 then follows from Theorem 5 of [10]. The form of g is given in (13), with the set Δ changed to match the class one dominant case. Note that as compared to the codominant case, we are unable to compute φ. However, we only require its value around the origin, so it can be computed by fast simulation. $\diamondsuit$

5 Numerical Results

5.1 Parameter space coverage

At this point, one key question is: Do Theorems 1 and 2 cover the entire parameter space? In other words, given arbitrary system parameters, is the resulting behaviour either codominant or class one dominant?

Unfortunately, the answer appears to be no. Consider a system with $\lambda_1 = 0.0944$, $\lambda_2 = 0.0238$, $\mu_1 = 0.8456$, $\mu_2 = 0.0363$, $k_1 = 3$, $k_2 = 1$. Figure 1 shows a typical sample path for reaching a total queue length of 200. These parameters do not satisfy either the codominant nor the class one dominant conditions. The reason appears to be that the trajectory is such that queue 2, while staying "small", spends no time on the boundary (empty). This appears to be the so-called "bridge" phenomenon described in Foley and McDonald [11]. (The term "bridge" was coined to denote a trajectory that begins and ends on an axis, but in between never or rarely touches the axis.) One might hope that one could apply their methodology here, however the fact that we do not have an explicit form for the harmonic function makes it difficult to see how one could do so. We leave this as a topic for future work.

Explicitly identifying the range of parameters for which either codominant or class one dominant behaviour holds appears to be quite difficult, due to the complexity of the expressions that define them. To try to get a better idea of how well the parameter space is covered, we generated 300 systems at random in the following manner. Here, the values of k_i may be up to an order of magnitude different, and the arrival and service rates can range over the entire stability region.

1. k_i were chosen independently and uniformly from $\{1, \dots, 10\}$
2. μ_i were chosen independently and randomly according to a uniform distribution on $(0, 1)$
3. λ_1 was chosen by multiplying μ_1 by a sample from a uniform distribution on $(0, 1)$
4. λ_2 was chosen by multiplying $(1 - \lambda_1/\mu_1)\mu_2$ by a sample from a uniform distribution on $(0, 1)$
5. types 1 and 2 were interchanged if $\lambda_2/k_2 > \lambda_1/k_1$
6. the arrival and service rates were normalized so that their sum was one

The results were quite encouraging in that our conditions covered all but 4 percent of the generated systems. For every one of those remaining 4 percent, it appears that the behaviour is of the "bridge" variety discussed earlier. At this point, we conjecture that our conditions cover all cases not giving rise to the "bridge" phenomenon.

5.2 Additional insights

We present several additional results in this section. Our goal here is to give an idea of how the asymptotics depend on the parameters of the system. Our first results are in Table 1, where we fix λ_2, μ_1 and μ_2, while letting λ_1 vary. The parameters are chosen as follows: let $\bar{\lambda}_2 = 0.7$, $\bar{\mu}_1 = 10$, $\bar{\mu}_2 = 7$, and let $\bar{\lambda}_1$ vary as in Table 1. Then λ_i, μ_i are chosen by taking the corresponding value of $\bar{\lambda}_i$ or $\bar{\mu}_i$ and dividing by $\bar{\lambda}_1 + \bar{\lambda}_2 + \bar{\mu}_1 + \bar{\mu}_2$. Here, $k_1 = k_2 = 1$. We see as λ_1 increases, there is a transition from the codominant regime, through the bridge regime and finally the class one dominant regime. Note that $1/\alpha$, the tail decay rate, is always larger than the load on the system

(ρ), but appears to approach ρ as the load approaches one. In Table 2, the only change is to set $k_2 = 10$. Here, the transition to the class one dominant regime occurs at lower values of λ_1, which is not surprising, as the larger k_2, the more "priority" is placed on the second queue. Note that in Tables 1 and 2, when both systems are class one dominant, the tail decay rate does not depend on k_1 and k_2 (as expected). These observations are reinforced in Table 3, which takes the same system, but with $k_1 = 2$, $k_2 = 1$. Then, class 1 is visited longer when queue 1 is nonempty, so the effect seen in Table 2 is reversed.

λ_1	Regime	ρ	$1/\alpha$
1.5	codominant	0.25	0.2771
2.0	codominant	0.30	0.3373
2.5	codominant	0.35	0.3974
3.0	bridge	0.40	
3.5	bridge	0.45	
4.0	bridge	0.50	
4.5	class one dominant	0.55	0.5919
5.0	class one dominant	0.60	0.6331
5.5	class one dominant	0.65	0.6759
6.0	class one dominant	0.70	0.7200
6.5	class one dominant	0.75	0.7652
7.0	class one dominant	0.80	0.8111
7.5	class one dominant	0.85	0.8576
8.0	class one dominant	0.90	0.9047
8.5	class one dominant	0.95	0.9522

Table 1: $k_1 = k_2 = 1$

6 Conclusion

We have presented results that give sufficient conditions for two types of behaviour in a two queue polling model. In some sense, this work appears to be at the limit of what can be accomplished using the techniques in [10, 14], due to the implicit nature of the constructed harmonic function. More work could be done on trying to explicitly characterize the region of the parameter space

$\bar{\lambda}_1$	Regime	ρ	$1/\alpha$
1.5	codominant	0.25	0.3077
2.0	bridge	0.30	
2.5	bridge	0.35	
3.0	bridge	0.40	
3.5	bridge	0.45	
4.0	class one dominant	0.50	0.5530
4.5	class one dominant	0.55	0.5919
5.0	class one dominant	0.60	0.6331
5.5	class one dominant	0.65	0.6759
6.0	class one dominant	0.70	0.7200
6.5	class one dominant	0.75	0.7652
7.0	class one dominant	0.80	0.8111
7.5	class one dominant	0.85	0.8576
8.0	class one dominant	0.90	0.9047
8.5	class one dominant	0.95	0.9522

Table 2: $k_1 = 1$, $k_2 = 10$

covered by the conditions in this paper. For that part of the parameter space that is not covered, the use of the methodology in [11] could be explored. Both of these seem extremely challenging. Finally, extending the results to more than two classes appears to not be conceptually more difficult, although it would certainly would be much more complex algebraically.

Acknowledgment. The author's research is supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Borst, S.C., Boxma, O.J., and Levy, H. (1995). The use of service limits for efficient operation of multistation single-medium communication systems. *IEEE/ACM Transactions on Networking,* 3:602-612.

[2] Chang, R.K.C. and Lam, S. (2000). A novel approach to queue stability analysis of polling

λ_1	Regime	ρ	$1/\alpha$
1.5	codominant	0.25	0.2639
2.0	codominant	0.30	0.3200
2.5	codominant	0.35	0.3761
3.0	codominant	0.40	0.4322
3.5	codominant	0.45	0.4884
4.0	bridge	0.50	
4.5	bridge	0.55	
5.0	class one dominant	0.60	0.6331
5.5	class one dominant	0.65	0.6759
6.0	class one dominant	0.70	0.7200
6.5	class one dominant	0.75	0.7652
7.0	class one dominant	0.80	0.8111
7.5	class one dominant	0.85	0.8576
8.0	class one dominant	0.90	0.9047
8.5	class one dominant	0.95	0.9522

Table 3: $k_1 = 2$, $k_2 = 1$

models. *Performance Evaluation*, 40, 27-46.

[3] Chang, W. and Down, D.G. (2002). Exact asymptotics for k_i-limited exponential polling models. *Queueing Systems,* 42, 401-419.

[4] Chang, W. and Down, D.G. (2007). Polling models under limited service policies: sharp asymptotics. *Stochastic Models*, 23, 129-147.

[5] Charzinski, J., Renger, T., and Tangemann, M. (1994). Simulative comparison of the waiting time distributions in cyclic polling systems with different service strategies. *Proceedings of the 14th International Teletraffic Congress,* 719-728.

[6] Choudhury, G.L. and Whitt, W. (1996). Computing distributions and moments in polling models by numerical transform inversion. *Performance Evaluation,* 25, 267–292.

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

[8] Doshi, H.T. (1986). Queueing systems with vacations - a survey. *Queueing Systems*, 1, 29-66.

[9] Duffield, N.G. (1997). Exponents for the tails of distributions in some polling models. *Queueing Systems*, 26, 105–119.

[10] Foley, R.D. and McDonald, D.R. (2001). Join the shortest queue: stability and exact asymptotics. *Annals of Applied Probability,* 11, 569-607.

[11] Foley, R.D. and McDonald, D.R. (2005). Bridges and networks: exact asymptotics. *Annals of Applied Probability,* 15, 542–586.

[12] Ioresh, M. (2005). Large deviations for a polling system with exhaustive service. M.Sc. Thesis, Technion.

[13] Lee, D.-S. (1996) A two-queue model with exhaustive and limited service disciplines. *Stochastic Models,* 12, 285-305.

[14] McDonald, D. (1999). Asymptotics of first passage times for random walk in a quadrant. *Annals of Applied Probability,* 9, 110–145.

[15] Sox, C.R., Jackson, P.L., Bowman, A., and Muckstadt, J.A. (1999). A review of the stochastic lot scheduling problem. *International Journal of Production Economics,* 62, 181-200.

[16] Takagi, H. (1988) Queueing analysis of polling models. *ACM Computing Surveys,* 20, 5-28.

[17] Takagi, H. (1991) *Queueing analysis: a foundation of performance evaluation, volume 1: vacation and priority systems.* North-Holland, Amsterdam, 1991.

[18] Vishnevskii, V.M. and Semenova, O.V. (2006). Mathematical methods to study the polling systems. *Automation and Remote Control,* 67, 173-220.

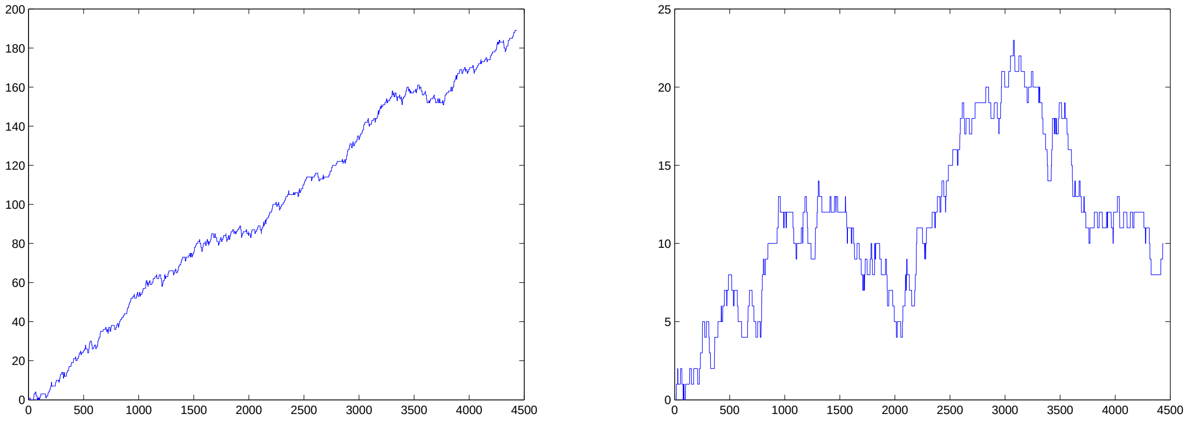

Figure 1: Queue length at queue 1 (left), queue 2 (right)