

Analysis of Alternating-priority Queueing Models with (Cross) Correlated Switchover Times

Robin Groenevelt, Eitan Altman

INRIA Sophia Antipolis, 06902 Sophia Antipolis, France

Email: {robin.groenevelt, eitan.altman}@sophia.inria.fr

Abstract— This paper analyzes a single server queueing system in which service is alternated between two queues and the server requires a (finite) switchover time to switch from one queue to the other. The distinction from classical results is that the sequence of switchover times from each of the queues need not be i.i.d. nor independent from each other; each sequence is merely required to form a stationary ergodic sequence. With the help of stochastic recursive equations explicit expressions are derived for a number of performance measures, most notably for the average delay of a customer and the average queue lengths under different service disciplines. With these expressions a comparison is made between the service disciplines and the influence of correlation is studied. Finally, through a number of examples it is shown that the correlation can significantly increase the mean delay and the average queue lengths indicating that the correlation between switchover times should not be ignored. This has important implications for communication systems in which a common communication channel is shared amongst various users and where the time between consecutive data transfers is correlated (for example in ad-hoc networks).

Index Terms— Stochastic processes/Queueing theory.

I. INTRODUCTION AND MOTIVATION

SO far only few explicit results have been known in queueing theory for systems whose evolution is described by general stationary ergodic processes. One line of research that allows one to handle stationary ergodic sequences is based on identifying measures that are insensitive to correlations. For example, the probability of finding a G/G/1 queue non-empty is just the ratio between the expected service time and the expected interarrival time of customers (which follows directly from Little's Law). The expected cycle duration in a polling system (under fairly general conditions) too, depends on the interarrival, service and vacation times only through their expectations under general stationary ergodic assumptions (see e.g. [1]). An example of performance measures that depend on the whole distribution of service times but is insensitive to correlations is the growth rate of the number of customers or of the sojourn time in a (discriminatory) processor sharing queue in overload [2], [3]. Other insensitivity results on bandwidth sharing in a network can be found in [4], [5].

The polling models we study do not exhibit insensitivity. Approximating correlated vacations by independent ones can result in large errors in the performance, see e.g. [6] in the context of Bluetooth. To study these systems we make use of stochastic recursive equations (SRE) introduced in [7] which

extend branching processes with migration on one hand, and linear stochastic recursive equations¹ on the other. It has already been shown in [7] that vector valued SRE can be used to describe some embedded processes appearing in polling models.²

In this paper we identify one dimensional SRE which we use in order to compute the expected waiting times and queue lengths in a system with two queues where a single server alternates between two queues and requires switch-over times (modeled as vacations) to move from one queue to the other. We consider the exhaustive service discipline where the server serves a queue until it empties before switching to the next queue as well as the gated discipline where only customers present upon the arrival of the server are served. Two systems are studied: one in which both queues are served exhaustively and one in which one queue is served exhaustively and the other according to the gated discipline. Our analytical results are then used to study numerically the impact of correlated switchover times on the performance, as well as the difference in performance due to the service discipline used.

The polling system studied in this paper, but without the correlation, has been used in the past [10] to model communication systems in which transmission between two stations can take place only in one direction at a time. The position of the server then corresponds to the direction data is traveling in. A similar situation arises in ad-hoc network; there is a common channel which needs to be shared amongst various users. The more users there are, the longer one has to wait before being able to capture the channel necessary to (re)transmit data. In particular, if one has to wait a long time before being able to transmit data, then it is very likely that there are many users around and that the next time one has to wait once again for a long period of time. For this reason the correlation of the number of users over time in an ad-hoc network inherently introduces correlation between the switchover times, and this in turn leads to an increase in the mean delay and queue lengths.

The remainder of this document is structured as follows. In section II the polling system is described in more detail, notation and some formulas are established, and the one dimensional SRE is identified. Following that, in section III a number of performance measures are derived, most notably the

¹Linear SRE have been used to study the impact of correlation of the loss process on TCP throughput [8].

²SRE have also been used recently to study the infinite server queue with correlated arrivals [9].

expected waiting times and the average queue lengths. These performance measures are then used in the examples of section IV to show the effect of correlated switchover times. Finally, conclusions are given in section V. To help the reader a list of the notation used is given in Appendix C.

II. MODEL DESCRIPTION

We examine the polling of two queues, i.e. one queue is served after which the other queue is served. No limit is specified for the length of either queue. After serving queue i ($i = 1, 2$) for the n -th time, the server requires a switchover time of duration $V_{n,i}$. Assume all $V_{n,i}$ have the same distribution as V_i ($V_{n,i} \sim V_i$), where V_i is assumed to form a general distribution with first and second moment v_i and $v_i^{(2)}$, and with variance $\delta_i^2 := v_i^{(2)} - v_i^2$, $i = 1, 2$. Let $R := v_1 + v_2$ and $\Delta^2 := \delta_1^2 + \delta_2^2$. The sequences of switchover times are assumed to be stationary ergodic instead of the usual i.i.d., and possibly dependent on each other. This implies that there can be a correlation between the switchover times of the two queues and/or within the sequence of switchover times for each queue. The arrival of customers at queue i is Poisson with rate λ_i and the service times are nonnegative, i.i.d. random variables with (finite) first and second moments for queue i given by, respectively, b_i and $b_i^{(2)}$. The load at queue i is $\rho_i := \lambda_i b_i$ and the system is stable [11, page 280] if and only if the overall load $\rho := \rho_1 + \rho_2 < 1$, which we assume throughout. Furthermore, we will continuously assume that the queues are operating under stationary regime.

Introduce the covariance functions ($i = 1, 2$)

$$\begin{aligned} c_i(n) &= \mathbb{E}[V_{0,i}V_{n,i}] - \mathbb{E}[V_{0,i}]\mathbb{E}[V_{n,i}], & n \in \mathbb{N}, \\ c_{12}(n) &= \mathbb{E}[V_{0,1}V_{n,2}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,2}], & n \in \mathbb{Z}. \end{aligned}$$

Note that $c_{12}(n)$ is defined for $n \in \mathbb{Z}$. With this convention it is not necessary to work with $c_{21}(n) := \mathbb{E}[V_{0,2}V_{n,1}] - \mathbb{E}[V_{0,2}]\mathbb{E}[V_{n,1}]$, since under stationary regime $c_{21}(n) = \mathbb{E}[V_{0,2}V_{n,1}] - \mathbb{E}[V_{0,2}]\mathbb{E}[V_{n,1}] = \mathbb{E}[V_{-n,2}V_{0,1}] - \mathbb{E}[V_{-n,2}]\mathbb{E}[V_{0,1}] = c_{12}(-n)$. In particular, if for each queue the sequence of switchover times is uncorrelated, then $c_i(0) = \delta_i^2$ and $c_i(n) = 0$, for $n \in \mathbb{N}$. If there is no correlation between the switchover times of the two queues, then $c_{12}(n) = 0$, for $n \in \mathbb{Z}$.

Because of the assumption of the queues operating under stationary regime

$$\begin{aligned} \mathbb{E}[V_{0,i}V_{n,i}] &= v_i^2 + c_i(n), & i = 1, 2, & n \in \mathbb{N}, & (1a) \\ \mathbb{E}[V_{0,1}V_{n,2}] &= v_1v_2 + c_{12}(n), & n \in \mathbb{Z}. & (1b) \end{aligned}$$

In order to establish the SRE, let $\mathcal{D}_{n,i}(N)$ be the duration of the busy period in the i^{th} queue, initiated by N customers waiting in that queue when the server arrives at that queue for the n^{th} time. Similarly, let $\mathcal{N}_{n,i}(T)$ be the number of customers arriving at queue i during a period of time T during the server's n^{th} visit to queue i .

We start by examining the exhaustive polling of two queues, i.e. one queue is served until it is empty after which the other queue is served until emptied. Consider the system at the moment the server starts serving the first queue for the n^{th} time with $L_{n,1}^*$ customers waiting in the queue. From

here on the following steps take place (see Figure 1 for a visual representation of this decomposition):

- *Exhausting the first queue.* The $L_{n,1}^*$ customers in the first queue require a busy period duration of $D_{n,1} := \mathcal{D}_{n,1}(L_{n,1}^*)$ to exhaust.
- *Switching to the second queue.* After serving the first queue the server requires a switchover time of $V_{n,1}$ units of time.
- *Exhausting the second queue.* In the time needed to switch from the second to the first queue ($V_{n-1,2}$), to exhaust the first queue ($D_{n,1}$), and to switch back to the second queue ($V_{n,1}$), there have been³ $L_{n,2}^* := \mathcal{N}_{n-1,2}(V_{n-1,2}) + \mathcal{N}_{n,2}(\mathcal{D}_{n,1}(L_{n,1}^*) + V_{n,1})$ customers arriving at the second queue. It requires $D_{n,2} := \mathcal{D}_{n,2}(L_{n,2}^*)$ units of time to empty this queue.
- *Switching back to the first queue.* After serving the second queue the server requires a switchover time of $V_{n,2}$ units of time.

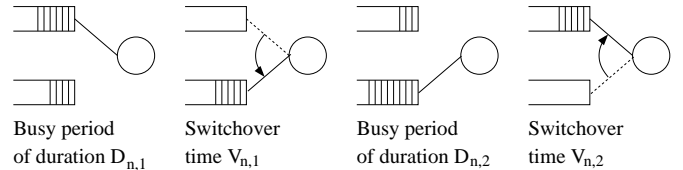


Fig. 1. Decomposition of the n^{th} cycle into busy periods ($D_{n,1}$ and $D_{n,2}$) and switchover times ($V_{n,1}$ and $V_{n,2}$).

After this the process starts over again and a new cycle begins. Hence the n^{th} cycle is made up of

$$C_n = D_{n,1} + V_{n,1} + D_{n,2} + V_{n,2}.$$

The time between the server finishing work at queue i and returning to queue i in the next cycle is the *intervisit* time $I_{n,i}$ and is given by

$$I_{n,1} = V_{n,1} + D_{n,2} + V_{n,2}, \quad (2a)$$

$$I_{n,2} = V_{n,2} + D_{n+1,1} + V_{n+1,1}. \quad (2b)$$

A SRE will be established for this quantity and we will see that it plays a central role for the derivation of the expected waiting times and queue lengths. The time $D_{n+1,i}$ spent at queue i in the $(n+1)^{\text{th}}$ cycle is related to the intervisit time according to

$$D_{n+1,1} = \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})), \quad (3a)$$

$$D_{n+1,2} = \mathcal{D}_{n+1,2}[\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(V_{n+1,1} + \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})))]. \quad (3b)$$

The expectation is the sum of the expectation of independent sub-busy periods [12, p.217] and thus

$$\mathbb{E}[D_{n,1}] = \frac{b_1 \mathbb{E}[\mathcal{N}_{n,1}(I_{n,1})]}{1 - \rho_1} = \frac{\rho_1 \mathbb{E}[I_{n,1}]}{1 - \rho_1}. \quad (4a)$$

³The star is added to $L_{n,i}^*$ to distinguish it from the average queue length $L_{n,i}$.

Using the stationarity and the divisibility⁴ of the arrival process it can be shown that

$$\mathbb{E}[D_{n,2}] = \frac{\rho_2 \mathbb{E}[I_{n,2}]}{1 - \rho_2}. \quad (4b)$$

Since the busy periods are sums of service times, the divisibility property also holds for $\mathcal{D}_{n,i}$. This means that from (substitute 3b into 2a)

$$I_{n+1,1} = V_{n+1,1} + V_{n+1,2} + \mathcal{D}_{n+1,2} \left[\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2} \left(V_{n+1,1} + \mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1})) \right) \right],$$

$n \in \mathbb{N}$, we see a SRE (as presented and solved for stationary ergodic sequences in [7]) arising. Although the system is two dimensional (as there are two queues), the reduction to a one dimensional SRE is a key element in obtaining explicit formulas for the performance measures.

Theorem 1: (SRE for exhaustive/exhaustive system). The intervisit time of the first queue allows itself to be written as a one-dimensional SRE,

$$I_{n+1,1} = \mathcal{A}_n(I_{n,1}) + \mathcal{B}_n \quad (5)$$

with

$$\begin{aligned} \mathcal{A}_n(\cdot) &:= \mathcal{D}_{n+1,2} \left(\mathcal{N}_{n+1,2} \left(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(\cdot)) \right) \right), \\ \mathcal{B}_n &:= V_{n+1,1} + V_{n+1,2} + \mathcal{D}_{n+1,2} \left(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(V_{n+1,1}) \right). \end{aligned} \quad (6)$$

Note that from (4a) and (4b) we have $\mathbb{E}[\mathcal{A}_n(I_{n,1})] = \alpha \mathbb{E}[I_{n,1}]$ where $\alpha := \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$. \diamond

Now let the first queue be served exhaustively and the second be served in a gated manner. The time needed to serve N customers in the second queue in the n^{th} cycle is denoted by $\mathcal{S}_{n,2}(N)$. Naturally, $\mathbb{E}[\mathcal{S}_{n,2}(N)] = b_2 \mathbb{E}[N]$. The service time of the second queue, $S_{n,2}$, satisfies the following recursive relationship

$$S_{n+1,2} = \mathcal{S}_{n+1,2}(\mathcal{N}_{n,2}(S_{n,2} + V_{n,2}) + \mathcal{N}_{n+1,2}(D_{n+1,1} + V_{n+1,1})).$$

At the same time, the time the server works per cycle at the first queue, $D_{n,1}$, satisfies

$$D_{n+1,1} = \mathcal{D}_{n+1,1} \left(\mathcal{N}_{n,1}(V_{n,1} + S_{n,2} + V_{n,2}) \right).$$

By combining these two expression we obtain the following theorems.

Theorem 2: (SRE for the exhaustive/gated system). The SRE for the service time at the gated queue is given by

$$S_{n+1,2} = \mathcal{X}_n(S_{n,2}) + \mathcal{Y}_n, \quad (7)$$

⁴The divisibility property implies that $\mathcal{N}_{n,2}(a+b) = \mathcal{N}_{n,2}^{(1)}(a) + \mathcal{N}_{n,2}^{(2)}(b)$, where $\mathcal{N}_{n,2}^{(1)}(\cdot)$ and $\mathcal{N}_{n,2}^{(2)}(\cdot)$ are stochastic processes independent of each other and each with the same distribution as $\mathcal{N}_{n,2}(\cdot)$.

with

$$\begin{aligned} \mathcal{X}_n(\cdot) &:= \mathcal{S}_{n+1,2} \left(\mathcal{N}_{n,2}(\cdot) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(\cdot))) \right), \\ \mathcal{Y}_n &:= \mathcal{S}_{n+1,2} \left(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(V_{n,1} + V_{n,2})) + V_{n+1,1}) \right). \end{aligned} \quad (8)$$

Note that $\mathbb{E}[\mathcal{X}_n(S_{n,2})] = \gamma \mathbb{E}[S_{n,2}]$ where $\gamma := \frac{\rho_2}{1-\rho_1}$. \diamond

Remark: In systems with both queues served in a gated manner or with more than two queues the SRE can not be written in a one-dimensional version. Although still solvable, the analysis is more involved and will be addressed in the future.

Before presenting a number of performance measures, a number of formulas need to be established. First recall that if D_i is a random sequence with $\mathbb{E}[D_i] = d$ and $\mathbb{E}[D_i^2] = d^{(2)}$, independent of a random variable N , and

$$\tau(N) := \sum_{i=1}^N D_i, \quad (9)$$

then

$$\begin{aligned} \mathbb{E}[\tau^2(N)] &= \sum_{n=1}^{\infty} n \mathbb{E} \left[\sum_{i=1}^n D_i \cdot \sum_{i=1}^n D_i \mid N = n \right] P(N = n) \\ &= d^2 \mathbb{E}[N^2] + (d^{(2)} - d^2) \mathbb{E}[N]. \end{aligned} \quad (10)$$

Similarly,

$$\begin{aligned} \mathbb{E}[\mathcal{N}_{n,i}^2(T)] &= \int_0^{\infty} \mathbb{E}[\mathcal{N}_{n,i}^2(t) \mid T = t] dT(t) \\ &= \lambda_i^2 \mathbb{E}[T^2] + \lambda_i \mathbb{E}[T]. \end{aligned} \quad (11)$$

Next we proceed in a similar manner to obtain the second moment of the busy period generated by Y customers initially in the system. First recall that $\mathcal{D}_{n,i}(1)$ is a single busy period initiated by a single customer in an M/G/1 queue with Poisson arrivals with rate λ_i and general service time with first and second moments b_i and $b_i^{(2)}$ respectively. The first two moments of a single busy period initiated by a single customer are given by [12, equations 5.141 and 5.142]

$$d_i := \mathbb{E}[\mathcal{D}_{n,i}(1)] = \frac{b_i}{1 - \rho_i}, \quad (12a)$$

$$d_i^{(2)} := \mathbb{E}[\mathcal{D}_{n,i}^2(1)] = \frac{b_i^{(2)}}{(1 - \rho_i)^3}, \quad i = 1, 2. \quad (12b)$$

The busy period, $D_{n,i}(Y)$, generated by Y customers is the sum of Y independent single busy periods, each with distribution $D_{n,i,k} \sim \mathcal{D}_{n,i}(1)$. Hence the second moment is

$$\begin{aligned} \mathbb{E}[\mathcal{D}_{n,i}^2(Y)] &= \mathbb{E} \left[\sum_{k=1}^Y \mathcal{D}_{n,i,k} \right]^2 = \mathbb{E} \left[\sum_{k=1}^Y \mathcal{D}_{n,i}(1) \right]^2 \\ &= d_i^2 \mathbb{E}[Y^2] + (d_i^{(2)} - d_i^2) \mathbb{E}[Y]. \end{aligned} \quad (13)$$

By taking $Y = \mathcal{N}_{n,i}(T)$ we obtain ($i = 1, 2$)

$$\mathbb{E}[\mathcal{D}_{n,i}^2(\mathcal{N}_{n,i}(T))] = \lambda_i^2 d_i^2 \mathbb{E}[T^2] + d_i^{(2)} \lambda_i \mathbb{E}[T]. \quad (14)$$

III. PERFORMANCE MEASURES

Starting with a system in which both of the queues are served exhaustively, an explicit expression will be given for the first two moments of the intervisit time in the presence of correlated switchover times. Based on this a number of performance measures follow, in particular, the expected waiting times and the average queue lengths. After this the expected waiting time and average queue length for the exhaustive/gated service system will be given.

A. Exhaustive/Exhaustive Service Discipline

Central to the derivation of a number of performance measures is Theorem 1. This allows the derivation of the following theorem of which the proof is given in Appendix A. Because of the symmetry for this service discipline the results will be presented for only the first queue.

Theorem 3: (Intervisit time in exhaustive/exhaustive system). Under the stationary regime the expected intervisit time of the first queue is given by

$$E[I_{n,1}] = \frac{R(1-\rho_1)}{1-\rho}, \quad \rho := \rho_1 + \rho_2, \quad (15)$$

The second moment is given by

$$\begin{aligned} \beta \mathbb{E}[I_{n,1}^2] &= \frac{R}{1-\rho} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1-\rho_1)^2} + \lambda_2 b_2^{(2)} \right) + \delta_1^2 + 2K \\ &+ \left(1 - \frac{2\rho_2(1-\rho)}{1-\rho_1} \right) \delta_2^2 + \left(\frac{1-\rho+2\rho_1\rho_2}{1-\rho} \right) R^2 \end{aligned} \quad (16)$$

where

$$\alpha := \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}, \quad \beta := \frac{(1-\rho)(1-\rho+2\rho_1\rho_2)}{(1-\rho_1)^2}.$$

and

$$\begin{aligned} K := \sum_{j=1}^{\infty} &\left[c_1(j) + c_2(j) + \frac{(1-\rho)^2}{\rho_1(1-\rho_1)} c_2(j) \right. \\ &\left. + \frac{1-\rho_2}{\rho_1} c_{12}(-j) + \frac{1-\rho_2(1-\alpha)}{\alpha} c_{12}(j-1) \right] \alpha^j \end{aligned} \quad (17)$$

is the addition to the intervisit time due to the correlation between the switchover times. \diamond

On the basis of this theorem the first two moments of a number of performance measures quickly follow.

Number of Customers Waiting. The exhaustive nature of the server implies that the number of customers building up at the first queue is exactly the number of customers that arrived at that queue during its intervisit time. Thus,

$$L_{n+1,1}^* = \mathcal{N}_{n,1}(I_{n,1}).$$

From this we immediately obtain

$$\mathbb{E}[L_{n+1,1}^*] = \lambda_1 \mathbb{E}[I_{n,1}] = \frac{R\lambda_1(1-\rho_1)}{1-\rho}$$

as the expected length of the queue, under stationary regime, at the moment the server arrives at the first queue. The second moment follows through squaring,

$$\mathbb{E}[(L_{n+1,1}^*)^2] = \mathbb{E}[\mathcal{N}_{n,1}^2(I_{n,1})] = \lambda_1^2 \mathbb{E}[I_{n,1}^2] + \lambda_1 \mathbb{E}[I_{n,1}],$$

which leads to

$$\mathbb{E}[(L_{n,1}^*)^2] := \lambda_1^2 \mathbb{E}[I_{n,1}^2] + \frac{R\lambda_1(1-\rho_1)}{1-\rho}. \quad (18)$$

Duration of Busy Periods. The expected time per cycle, in steady state, for the server to work on the first queue is given by

$$\mathbb{E}[D_{n,1}] = \frac{R\rho_1}{1-\rho}.$$

This follows directly from $\mathbb{E}[D_{n,1}] = \mathbb{E}[\mathcal{D}_{n,1}(L_{n,1}^*)] = \frac{b_1 \mathbb{E}[L_{n,1}^*]}{1-\rho_1}$. Since $\mathbb{E}[D_{n,1}^2] = \mathbb{E}[\mathcal{D}_{n,1}^2(\mathcal{N}_{n,1}(I_{n,1}))]$, the second moment follows with (14) and is given by

$$\mathbb{E}[D_{n,1}^2] = \frac{\rho_1^2 \mathbb{E}[I_{n,1}^2]}{(1-\rho_1)^2} + \frac{R\lambda_1 b_1^{(2)}}{(1-\rho_1)^2(1-\rho)}.$$

Number Served per Cycle. To derive the first and second moments of the number of customers served per cycle, consider an $M/G/1$ queue with arrival rate λ_1 , average service time b_1 , and the second moment of the service time $b_1^{(2)}$. Then the expectation and the variance of the number of customers served in a single busy period are known to be [12, equations (5.153) and (5.154)]

$$\mathbb{E}[\Gamma_1] = \frac{1}{1-\rho_1}, \quad \text{Var}[\Gamma_1] = \frac{\rho_1(1-\rho_1) + \lambda_1^2 b_1^{(2)}}{(1-\rho_1)^3}.$$

Let $\mathcal{T}_{n,1}(N)$ be the number of customers served at the first queue during the n^{th} cycle if there are N customers in the queue at the moment of polling. Note that $\mathcal{T}_{n,1}(N)$ is the number of customers served starting with N customers, whereas Γ_1 is the number of customers served starting with just one customer. A different notation is used for these two quantities to reflect the dependency on the n -th cycle and the queue number. Since the number served is the sum of the number served during N busy periods, (9) tells us that the expected number of customers served, per cycle, at the first queue is

$$\mathbb{E}[\mathcal{T}_{n,1}] = \mathbb{E}[\mathcal{T}_{n,1}(L_{n,1}^*)] = \frac{\mathbb{E}[L_{n,1}^*]}{1-\rho_1} = \frac{\lambda_1 R}{1-\rho}. \quad (19)$$

To derive the second moment note that

$$\begin{aligned} \mathbb{E}[\mathcal{T}_{n,1}^2] &= \mathbb{E}[\mathcal{T}_{n,1}^2(L_{n,1}^*)] \\ &= (\mathbb{E}[\Gamma_1])^2 \mathbb{E}[(L_{n,1}^*)^2] + \text{Var}[\Gamma_1] \mathbb{E}[L_{n,1}^*] \\ &= \frac{\mathbb{E}[(L_{n,1}^*)^2]}{(1-\rho_1)^2} + \frac{R\lambda_1(\rho_1(1-\rho_1) + \lambda_1^2 b_1^{(2)})}{(1-\rho_1)^2(1-\rho)}. \end{aligned}$$

Using equation (18) gives

$$\mathbb{E}[I_{n,1}^2] = \frac{1}{(1-\rho_1)^2} \left(\lambda_1^2 \mathbb{E}[I_{n,1}^2] + \frac{R\lambda_1(1-\rho_1^2 + \lambda_1^2 b_1^{(2)})}{1-\rho} \right).$$

Theorem 4: (Expected waiting time and queue length for the exhaustive/exhaustive service discipline). The expected waiting time (total time in system minus service time) of a customer going through the first queue is decomposed of two parts, namely

$$\mathbb{E}[W_{q,1}] = \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{(1-\rho)\mathbb{E}[I_{n,1}^2]}{2R(1-\rho_1)}. \quad (20)$$

This gives

$$\begin{aligned} \mathbb{E}[W_{q,1}] &= \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{\lambda_1 \rho_2^2 b_1^{(2)} + \lambda_2 (1-\rho_1)^2 b_2^{(2)}}{2(1-\rho_1)(1-\rho)(1-\rho+2\rho_1\rho_2)} \\ &+ \frac{(1-\rho_1)R}{2(1-\rho)} + \left(\frac{\Delta^2}{2} - \frac{\rho_2(1-\rho)}{1-\rho_1} \delta_2^2 + K \right) \psi, \end{aligned} \quad (21)$$

where $\psi := \frac{1-\rho_1}{R(1-\rho+2\rho_1\rho_2)}$ and K (defined in (17)) is the increase in the expected waiting time due to correlated switchover times.

The average number of customers at the first queue (in service and in the queue) follows directly from Little and is

$$\mathbb{E}[L_{s,1}] = \lambda_1 \mathbb{E}[W_{q,1}] + \rho_1. \quad \diamond$$

The proof can be found in Appendix B. In the uncorrelated case ($K = 0$) this is in correspondence with [11, formula⁵ (3.12)].

B. Exhaustive/Gated Service Discipline

Now let the second queue be served with a gated discipline instead of an exhaustive one. We obtain the following theorem.

Theorem 5: (Expected waiting time and queue length for the exhaustive/gated service discipline).

The expected time a customer waits in queue i ($i = 1, 2$) until being served is given by (22) where K_1 and K_2 are the increases in the expected waiting time due to correlated switchover times.

The average queue lengths,

$$\mathbb{E}[L_{s,i}] = \lambda_i \mathbb{E}[W_{q,i}] + \rho_i, \quad i = 1, 2.$$

follow immediately because of Little. \diamond

If all of the switchover times are independent of each other, then $K_1 = 0 = K_2$ and we obtain the results given in [13, formulas (25) and (28)] or [11, formula⁶ 4.1].

Proof: The proof, although slightly more involved, runs along the same lines as the proofs of Theorems 3 and 4 and is omitted due to space constraints. The reader is referred to [14] for the full proof. \blacksquare

Although not shown here, it can be verified that for any choice of parameters $K_2 \geq K_1$. From (22) we can see that if δ_1^2 is sufficiently large, if $1 - 2(1-\rho)(1+\rho_2) > 0$, and if $K_1 = 0 = K_2$, then it may very well be possible that the

⁵The last term in [11, formula (3.12)] is missing a factor two and there is a mix up between δ_1 and δ_2 . The expression should read $\dots + \frac{[1-\rho_1-2\rho_2(1-\rho)]\delta_2^2 + (1-\rho_1)^2\delta_1^2}{2R(1-\rho+2\rho_1\rho_2)}$.

⁶The formula presented in this reference is copied incorrectly from [13]. The first term for the waiting time for customers arriving at the second queue should contain $(1+\rho_2)$ instead of $(1+\rho_1)$.

expected waiting time at the gated queue is smaller than the expected waiting time at the exhaustive queue! However, the range of parameter settings for which this is the case is fairly small. In particular, it can be shown [14] that this does not happen when the parameters for both queues are equal, if the switchover times or equal to zero, or if the system is heavily loaded.

IV. EXAMPLES

In the following paragraphs a number of examples will be considered in which the sequences of switchover times are correlated. The covariance functions will be calculated explicitly after which the impact of the correlation on the waiting times will be studied. In all of the examples the expected waiting time of a customer arriving at the first queue of an exhaustive/exhaustive served system is given by (21) whereas in the exhaustive/gated served system the expected waiting times are given by (22). The difference between each of the examples is that K , K_1 , and K_2 take on different values. The first example studies a single server queue with correlated vacations (by turning off one of the queues), whereas in the subsequent examples explicit expressions are derived for the expected waiting times and these are compared to the expected waiting times if there would be no correlation.

A. Single Server Queue with Correlated Vacations

By turning off one of the queues one obtains an M/G/1 queue with multiple correlated vacations. Let us start by turning off the second queue (by setting $\lambda_2 = 0$, $\rho_2 = 0$, $v_2 = 0$, and $v_2^{(2)} = 0$, which leads to $c_2(j) = 0$, $c_{12}(j) = 0$, and $\gamma = 0$) to end up with an exhaustively served M/G/1 queue where the expected waiting time

$$\mathbb{E}[W_{q,1}] = \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{v_1^{(2)}}{2v_1} \quad (\text{Exhaustive M/G/1})$$

is independent of the correlation between the vacations! This result was previously pointed out in [7, paragraph 3.6] which causes it to correspond to the expression for the expected waiting time but with i.i.d. vacation times [15, page 123].

On the other hand, by turning off the first queue in the exhaustive/gated system, we are left with an M/G/1 queue with a gated service discipline. After setting the appropriate parameters to zero we obtain

$$\mathbb{E}[W_{q,2}] = \frac{\lambda_2 b_2^{(2)}}{2(1-\rho_2)} + \frac{v_2^{(2)}}{2v_2} + \frac{\rho_2 v_2}{1-\rho_2} + \frac{1}{v_2} \sum_{j=1}^{\infty} c_2(j) \rho_2^j \quad (\text{Gated M/G/1})$$

as the waiting time of a customer arriving at a gated M/G/1 queue with correlated vacations. If there is no correlation then this expression is in agreement with the result previously obtained in [7, Theorem 5] and [15, equation (5.24a)].

It is interesting to compare the difference between these two waiting times due to the server behaving differently. Assuming queues with identical parameters (by dropping the indices of

$$\mathbb{E}[W_{q,1}] = \frac{(1-\rho_1)R}{2(1-\rho)} + \frac{\rho_2(1-\rho)(1+\rho_2)\delta_1^2}{R(1-\rho_1+\rho_2)} + \frac{1-\rho_1}{2(1-\rho_1+\rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1-\rho} + \frac{\Delta^2 + 2K_1}{R} \right] \quad (22a)$$

$$\mathbb{E}[W_{q,2}] = \frac{(1+\rho_2)R}{2(1-\rho)} - \frac{\rho_1(1-\rho)(1+\rho_2)\delta_1^2}{R(1-\rho_1+\rho_2)} + \frac{1+\rho_2}{2(1-\rho_1+\rho_2)} \left[\frac{\lambda_1 b_1^{(2)} + \lambda_2 b_2^{(2)}}{1-\rho} + \frac{\Delta^2 + 2K_2}{R} \right] \quad (22b)$$

where $\gamma := \frac{\rho_2}{1-\rho_1}$ and

$$K_1 := \sum_{j=1}^{\infty} \left[c_2(j) + \rho c_{12}(j) + \left(1 + \frac{\rho_2(1-\rho)}{1-\rho_1} \right) (c_{12}(-j) + \rho c_1(j)) \right] \gamma^j + \frac{\rho_2(1-\rho_1(2-\rho))}{(1-\rho_1)^2} c_{12}(0) \quad (23a)$$

$$K_2 := \sum_{j=1}^{\infty} \left[c_2(j) + \rho c_{12}(j) + \frac{1-\rho_1(2-\rho)}{\rho_2} (c_{12}(-j) + \rho c_1(j)) \right] \gamma^j + 2\rho c_{12}(0) \quad (23b)$$

ones and twos and setting $\hat{\rho} := \rho_1 = \rho_2$) we see that

$$\begin{aligned} \mathbb{E}[W]_{\text{gated } M/G/1} - \mathbb{E}[W]_{\text{exhaustive } M/G/1} \\ = \frac{\hat{\rho}v}{1-\hat{\rho}} + \frac{1}{v} \sum_{j=1}^{\infty} c(j)\hat{\rho}^j, \end{aligned} \quad (24)$$

where the first term on the right hand side is the mean length of a service period (which is the same for the exhaustive and the gated service systems). If there is no correlation, then it is well known that the expected waiting time in an exhaustively served queue is less than that in a gated serviced queue. In the presence of correlated vacation times this difference is larger but remains a surprisingly simple expression.

B. Correlated Switchover Times

Consider a sequence of switchover times where there is no correlation between the switchover times of the two queues (this gives $c_{12}(j) = 0$, for $j \in \mathbb{Z}$). Let the individual sequence of switchover times per queue satisfy

$$V_{n+1,i} = x_i V_{n,i} + (1-x_i)\varepsilon_{n,i}, \quad i = 1, 2, \quad (25)$$

where $x_i \in [0, 1)$ is a constant and $\varepsilon_{n,i}$ are positive i.i.d. variables with finite expectation $\mathbb{E}[\varepsilon_{n,i}] =: \bar{\varepsilon}_i$ and second moment $\mathbb{E}[\varepsilon_{n,i}^2] =: \varepsilon_i^{(2)}$. The parameter x_i determines the amount of correlation in the sequence; with $x_i = 0$ the sequence is i.i.d., whereas when x_i tends to one the correlation is maximal. Notice that there exists a stationary ergodic sequence of switchover times which satisfies (25). By taking the expectation it follows that $\mathbb{E}[V_{n+1,i}] = x_i \mathbb{E}[V_{n,i}] + (1-x_i)\bar{\varepsilon}_i$. Due to the stationarity of the process $\mathbb{E}[V_{0,i}] = \mathbb{E}[V_{n,i}] = v_i$ is independent of x_i , and therefore $v_i = \bar{\varepsilon}_i$. A similar relationship can be derived for the second moments by taking the expectation over the square of (25) to give

$$\begin{aligned} \mathbb{E}[V_{n+1,i}^2] &= x_i^2 \mathbb{E}[V_{n,i}^2] + (1-x_i)^2 \mathbb{E}[\varepsilon_{n,i}^2] \\ &\quad + 2x_i(1-x_i)\mathbb{E}[\varepsilon_{n,i}]\mathbb{E}[V_{n,i}]. \end{aligned}$$

Due to the stationarity ($\mathbb{E}[V_{n+1,i}^2] = \mathbb{E}[V_{n,i}^2] = v_i^{(2)}$) this implies that

$$v_i^{(2)} = \frac{(1-x_i)\varepsilon_i^{(2)} + 2x_i\bar{\varepsilon}_i v_i}{1+x_i},$$

which gives a second relationship (since $\bar{\varepsilon}_i = v_i$),

$$\delta_i^{(2)} = \frac{1-x_i}{1+x_i} \text{Var}(\varepsilon_{n,i}).$$

Thus we see that for $x_i \in [0, 1)$ there exists a $\varepsilon_{n,i}$ such that any desired values of v_i and $\delta_i^{(2)}$ can be obtained. Now we will derive the covariance functions and the expected waiting time.

By iterating (25) a number of times it is quickly seen that

$$V_{n,i} = x_i^n V_{0,i} + (1-x_i) \sum_{k=0}^{n-1} \varepsilon_{n-1-k,i} x_i^k. \quad (26)$$

From this we obtain

$$\begin{aligned} \mathbb{E}[V_{0,i} V_{j,i}] &= x_i^j \mathbb{E}[V_{0,i}^2] + (1-x_i^j) \bar{\varepsilon}_i \mathbb{E}[V_{0,i}] \\ &= x_i^j v_i^{(2)} + (1-x_i^j) \bar{\varepsilon}_i v_i. \end{aligned}$$

This means that the covariance functions, $c_i(j) = \mathbb{E}[V_{0,i} V_{j,i}] - \mathbb{E}[V_{0,i}]\mathbb{E}[V_{j,i}]$, are given by

$$\begin{aligned} c_i(j) &= x_i^j v_i^{(2)} + (1-x_i^j) \bar{\varepsilon}_i v_i - v_i (x_i^j v_i + (1-x_i^j) \bar{\varepsilon}_i) \\ &= x_i^j (v_i^{(2)} - v_i^2) = x_i^j \delta_i^{(2)}. \end{aligned} \quad (27)$$

Since

$$\sum_{j=1}^{\infty} c_i(j) \alpha^j = \delta_i^{(2)} \sum_{j=1}^{\infty} (\alpha x_i)^j = \frac{\alpha x_i \delta_i^{(2)}}{1-\alpha x_i}, \quad (28)$$

we have from Theorem 4 that the expected waiting time in the exhaustive/exhaustive system is given by (21) with

$$K := \frac{\alpha x_1 \delta_1^{(2)}}{1-\alpha x_1} + \frac{\alpha x_2 \delta_2^{(2)}}{1-\alpha x_2} \left(1 + \frac{(1-\rho)^2}{\rho_1(1-\rho_1)} \right)$$

and $\alpha = \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$.

Equivalently, in the exhaustive/gated system we have from equation (28) and from Theorem 5 that the expected waiting times are given by (22) where $\gamma = \frac{\rho_2}{1-\rho_1}$ and

$$\begin{aligned} K_1 &= \frac{\gamma x_1 \delta_1^{(2)}}{1-\gamma x_1} \left(1 + \frac{\rho_2(1-\rho)}{1-\rho_1} \right) \rho + \frac{\gamma x_2 \delta_2^{(2)}}{1-\gamma x_2} \\ K_2 &= \frac{\gamma x_1 \delta_1^{(2)}}{1-\gamma x_1} \left(1 + \frac{\rho_1(1-\rho)^2}{\rho_2} \right) + \frac{\gamma x_2 \delta_2^{(2)}}{1-\gamma x_2}. \end{aligned}$$

Numerical examples of the influence of the correlation on the expected waiting times can be found in Figure 2. Shown

in each of the figures is the expected waiting time divided by the expected waiting time for uncorrelated sequences of switchover times. There are Poisson arrivals with $\lambda_i = 0.4$. The first two moments of the switchover times are always kept fixed (first moment for each of the switchover time distributions is fixed at $v_i = 3$) and the service times are taken to be exponential with $b_i = 0.4$ or $b_i = 1.2$.

Based on the figures and equations above the following important conclusions can be made:

- If $x_1 = x_2 = 0$ then there is no correlation between the sequences of switchover times and $K = 0$, $K_1 = 0$, and $K_2 = 0$;
- The increase in expected waiting times due to correlated switchover times can be up to several times (3.5 times in the example) the expected waiting times if there would be uncorrelated switchover times.
- The increase in the expected waiting times due to correlation grows linearly with the variance δ_i^2 of the switchover times;
- Under light traffic (α and γ are small and so) the increase in the expected waiting time is (approximately) linear in x_i .
- Under heavy traffic α and γ are close to one and, due to the factor $1 - \alpha x_i$ or $1 - \gamma x_i$ in the denominators, the increase in waiting time due to correlated switchover times can be significant. Hence the presence of correlation has the biggest impact on the waiting time if the system has a heavy load (and the switching times have a high variance). This can be seen clearly in Figure 2.
- It can be shown that, under identical parameter setting, in the exhaustive/gated system the expected waiting time at the exhaustive queue is always larger than at the gated queue. In addition to this, we see from Figure 2 that in lightly loaded systems the gated queue (Q2) suffers most from correlated switchover times whereas in heavily loaded traffic both queues are effected (relatively) equally by the correlated switchover times.

C. Identical Switchover Times

Set $V_{n,2} = V_{n,1}$. This introduces cross-correlation between the two sequences of switchover times and it gives $v_2 = v_1$ and $\delta_2^2 = \delta_1^2$. In addition to this let $V_{n+1,1} = xV_{n,1} + (1-x)\varepsilon_{n,1}$ just as in the previous example. From (27) we have $c_1(j) = x^j \delta_1^2$ after which

$$c_2(j) = \mathbb{E}[V_{0,2}V_{n,2}] - \mathbb{E}[V_{0,2}]\mathbb{E}[V_{n,2}]$$

$$= \mathbb{E}[V_{0,1}V_{n,1}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,1}] = c_1(j) = x^j \delta_1^2$$

$$c_{12}(j) = \mathbb{E}[V_{0,1}V_{n,2}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,2}]$$

$$= \mathbb{E}[V_{0,1}V_{n,1}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,1}] = c_1(j) = x^j \delta_1^2$$

$$c_{12}(-j) = \mathbb{E}[V_{0,2}V_{n,1}] - \mathbb{E}[V_{0,2}]\mathbb{E}[V_{n,1}]$$

$$= \mathbb{E}[V_{0,1}V_{n,1}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,1}] = c_1(j) = x^j \delta_1^2$$

immediately follow. This means that ($i = 1, 2$)

$$\sum_{j=1}^{\infty} c_i(j)\alpha^j = \sum_{j=1}^{\infty} c_{12}(j)\alpha^j = \sum_{j=1}^{\infty} c_{12}(-j)\alpha^j = \frac{\alpha x \delta_1^2}{1 - \alpha x}$$

can all be plugged into Theorem 4 so that the expected waiting time in the exhaustive/exhaustive system is given by (21) with

$$K = \frac{\alpha x \delta_1^2}{1 - \alpha x} \left(\frac{1 - \rho_2(1 - \alpha)}{\alpha x} + 2 + \frac{(1 - \rho)^2}{\rho_1(1 - \rho_1)} + \frac{1 - \rho_2}{\rho_1} \right),$$

$$\psi := \frac{1 - \rho_1}{2v_1(1 - \rho + 2\rho_1\rho_2)}, \text{ and } \alpha = \frac{\rho_1\rho_2}{(1 - \rho_1)(1 - \rho_2)}.$$

Equivalently, the expected waiting times in the exhaustive/gated system are given by (22) with

$$K_1 = \frac{\gamma x \delta_1^2}{1 - \gamma x} \left(2 + \frac{\rho_2(1 - \rho)}{1 - \rho_1} \right) (1 - \rho) + \frac{\rho_2(1 - \rho_1(2 - \rho))}{(1 - \rho_1)^2} \delta_1^2,$$

$$K_2 = \frac{\gamma x \delta_1^2}{1 - \gamma x} \left(1 + \frac{1 - \rho_1(2 - \rho)}{\rho_2} \right) (1 - \rho) + 2\rho \delta_1^2.$$

To get a feeling of the impact of the cross correlation, the expected waiting times are plotted in Figure 3 for various switchover time distributions and traffic loads. Shown in each of the figures is the expected waiting time divided by the expected waiting time for uncorrelated sequences of switchover times. There are Poisson arrivals with $\lambda_i = 0.4$. The first two moments of the switchover times are always kept fixed (first moment for each of the switchover time distributions is fixed at $v_i = 3$) and the service times are taken to be exponential with $b_i = 0.4$ or $b_i = 1.2$.

Striking is the impact of the cross correlation on the waiting times. For example, if there is no correlation within each sequence of switchover times ($x = 0$), then there is still an increase in the expected waiting time due to the cross-correlation. For the exhaustive/exhaustive system this increase is $(1 - \rho_2(1 - \alpha))\psi\delta_1^2$ and for the exhaustive/gated system this increase is given by $\frac{\rho_2(1 - \rho_1(2 - \rho))}{(1 - \rho_1)^2}\delta_1^2$ and $2\rho\delta_1^2$ for, respectively, the exhaustive and the gated queue. For exponentially distributed switchover times this can mean an increase of tens of percents in the expected waiting time. Besides this, all of the conclusions made in the first example also hold here, with the exception that the increase in expected waiting time can up to a factor 5.

D. Stochastic Recursive Switchover Times

Consider a sequence of switchover times which satisfy the following stochastic recursive relationship

$$V_{n+1,i} = \mathcal{F}_{n,i}(V_{n,i}) + \mathcal{E}_{n,i}, \quad (29)$$

where $\mathcal{F}_{n,i}(\cdot)$ are independent, infinitely divisible stochastic processes with $\mathbb{E}[\mathcal{F}_{n,i}(T)] = x_i\mathbb{E}[T]$ and $\mathbb{E}[\mathcal{F}_{n,i}^2(T)] = x_i^{(2)}\mathbb{E}[T^2] + y_i\mathbb{E}[T]$. Here $x_i \in [0, 1)$, $y_i \geq 0$ and $x_i^{(2)} \geq x_i^2$. The sequence $\mathcal{E}_{n,i}$ is a sequence of independent variables with $\mathbb{E}[\mathcal{E}_{n,i}] = \bar{\varepsilon}_i$ and $\mathbb{E}[\mathcal{E}_{n,i}^2] = \varepsilon_i^{(2)}$. Iterating gives

$$V_{n,i} = \left(\prod_{k=0}^{n-1} \mathcal{F}_{k,i} \right) V_{0,i} + \sum_{k=0}^{n-1} \left(\prod_{l=k+1}^{n-1} \mathcal{F}_{l,i} \right) \mathcal{E}_{n-k,i},$$

and so

$$\mathbb{E}[V_{n,i}] = x_i^n v_i + \bar{\varepsilon}_i \sum_{k=0}^{n-1} x_i^k = x_i^n v_i + \frac{1 - x_i^n}{1 - x_i} \bar{\varepsilon}_i.$$

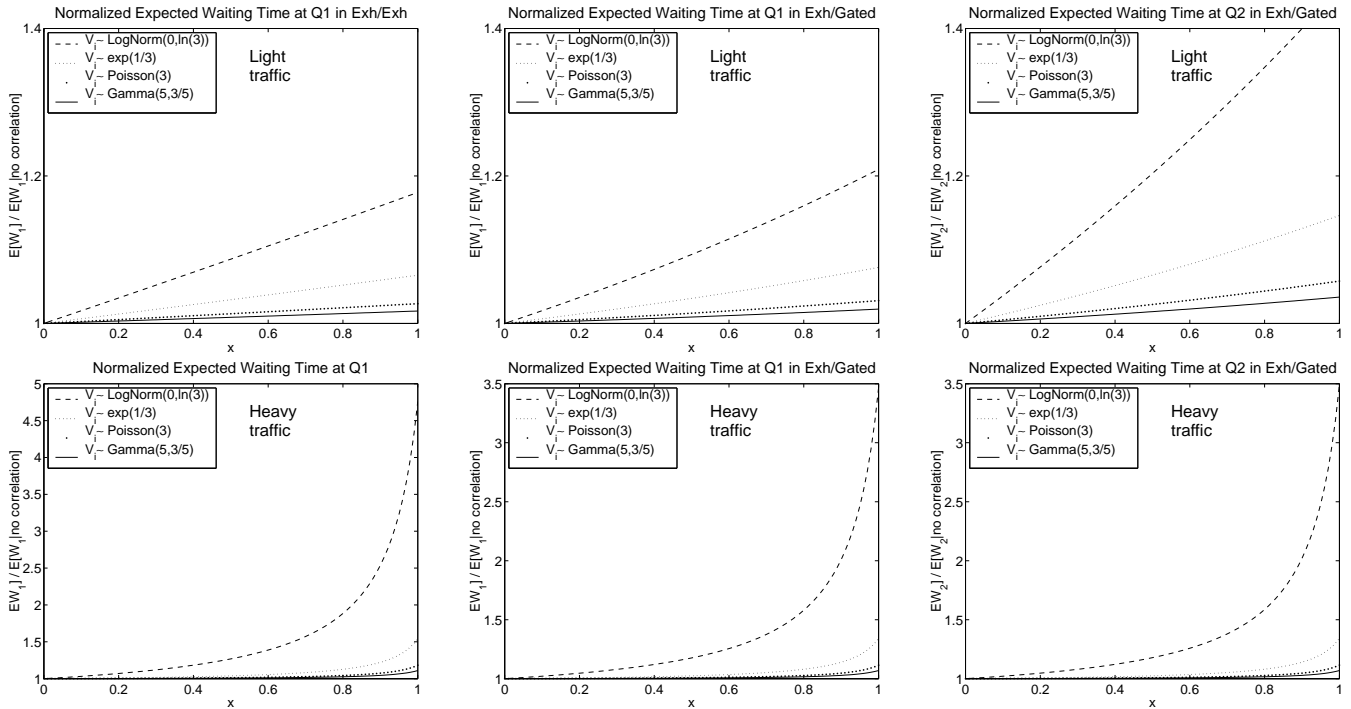


Fig. 2. Example: Correlated Switchover Times. The expected waiting time divided by the expected waiting time with uncorrelated switchover times. The different lines correspond to different switchover time distributions, all with mean $v_i = 3$. Here $x := x_1 = x_2$ determines the level of correlation, there is no cross correlation, the service times are exponential, and $\lambda_i = 0.4$. The top figures are with mean service times $b_i = 0.4$ ($\rho = 0.32$) whereas the second row of figures are under heavy traffic with $b = 1.2$ ($\rho = 0.96$). The first column of figures correspond to the exhaustive/exhaustive system, whereas the second and third column of figures show the normalized waiting times for, respectively, the exhaustive queue (Q1) and the gated queue (Q2) in the exhaustive/gated system.

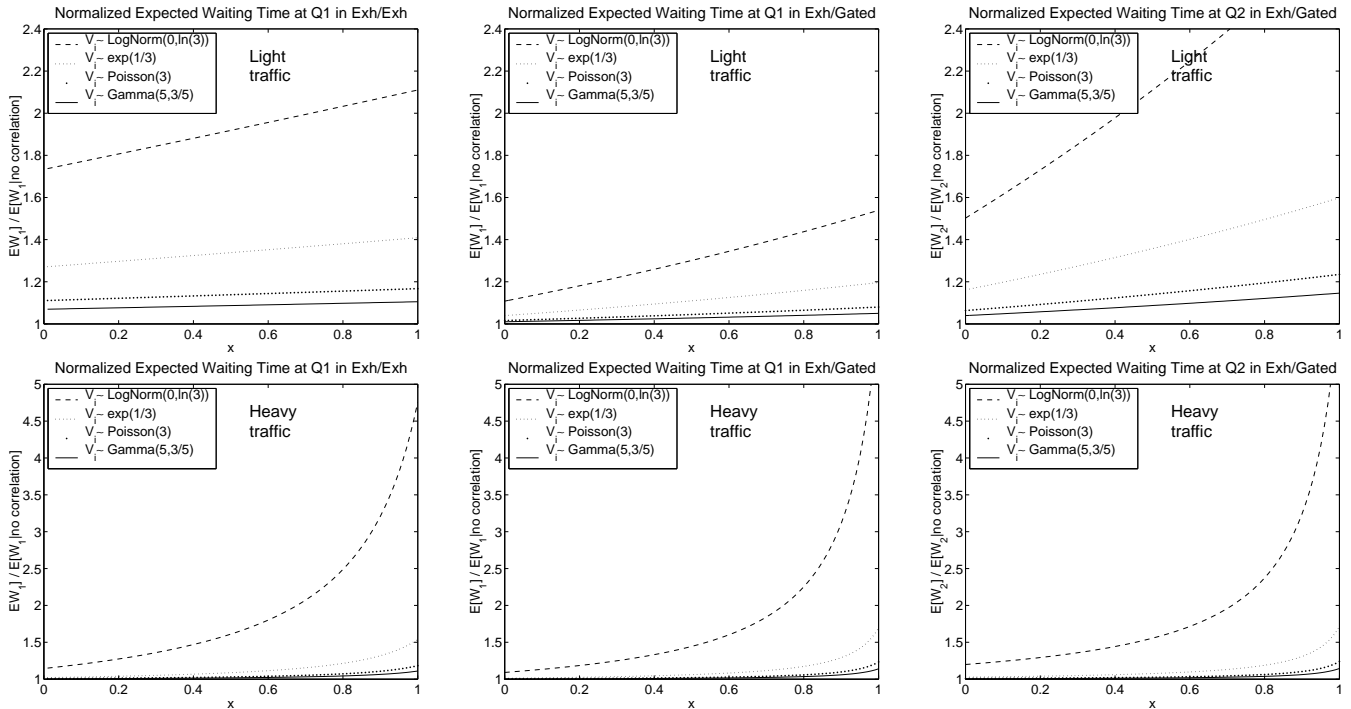


Fig. 3. Example: Identical Switchover Times. The expected waiting time divided by the expected waiting time if there would be no correlation between the switchover times. The different lines correspond to different switchover time distributions. Here x determines the level of correlation, cross correlation is introduced by setting $V_{n,2} = V_{n,1}$, the service times are exponential, and $\lambda_i = 0.4$. The figures in the first row are with mean service time $b = 0.4$ ($\rho = 0.32$) and the figures on the bottom row are under heavy traffic with $b = 1.2$ ($\rho = 0.96$). The first column of figures correspond to the exhaustive/exhaustive system, whereas the second and third column of figures show the normalized waiting times for, respectively, the exhaustive queue (Q1) and the gated queue (Q2) in the exhaustive/gated system.

This leads to the condition $v_i = \frac{\bar{\varepsilon}_i}{1-x_i}$. The second moment of the switchover times is found by taking the expectation over the square of (29). Doing this produces

$$\begin{aligned}\mathbb{E}[V_{n+1,i}^2] &= \mathbb{E}[\mathcal{F}_{n,i}^2(V_{n,i})] + \mathbb{E}[\mathcal{E}_{n,i}^2] + 2\mathbb{E}[\mathcal{F}_{n,i}(V_{n,i}) \cdot \mathcal{E}_{n,i}] \\ &= x_i^{(2)} v_i^{(2)} + y_i v_i + \varepsilon_i^{(2)} + 2x_i v_i \bar{\varepsilon}_i.\end{aligned}$$

Since, by definition, $\mathbb{E}[V_{n+1,i}^2]$ also equals $v_i^{(2)}$, we have

$$v_i^{(2)} = \frac{\varepsilon_i^{(2)} + (y_i + 2x_i \bar{\varepsilon}_i) v_i}{1 - x_i^{(2)}}.$$

Thus we see that there exists a $\mathcal{E}_{n,i}$ which generates a sequence $V_{n,i}$ with arbitrary correlation and any desired values of v_i and $v_i^{(2)}$. Similarly, it is easy to show that

$$\mathbb{E}[V_{0,i} V_{n,i}] = x_i^n v_i^{(2)} + \frac{1 - x_i^n}{1 - x_i} \bar{\varepsilon}_i v_i.$$

This gives the covariance functions

$$\begin{aligned}c_i(j) &= \mathbb{E}[V_{0,i} V_{j,i}] - \mathbb{E}[V_{0,i}] \mathbb{E}[V_{j,i}] \\ &= x_i^j v_i^{(2)} + \frac{1 - x_i^j}{1 - x_i} \bar{\varepsilon}_i v_i - v_i \left(x_i^j v_i + \frac{1 - x_i^j}{1 - x_i} \bar{\varepsilon}_i \right) = x_i^j \delta_i^2\end{aligned}$$

which are completely identical to (27)! This means that if the first two moments of the switchover times are kept fixed while varying $x \in [0, 1)$, that the expected waiting times are once again given by (21) and (22) with K , K_1 , and K_2 as given in the first example. Furthermore, the conclusions of the first example also hold here.

As a special case of (29) we can take $V_{n+1,i} = x_i V_n + \varepsilon_{n,i}$ where $x_i \in [0, 1)$ is a constant and $\varepsilon_{n,i}$ is a positive sequence of i.i.d. variables.

V. CONCLUSIONS

We have studied the performance of alternating-priority queues with very weak assumptions on the switchover time sequences; all we assume is that these sequences are stationary ergodic. In spite of this generality we were able to derive explicit expressions for the expected waiting times and number of customers in each queue. The expressions obtained involve the weighted sum of all correlations where the weights decrease exponentially fast to zero. With the help of our explicit expressions, we studied numerically the role of correlation and gave examples where they add up to 400% to the expected waiting times. This has important implications for (ad-hoc) networks where a common communication channel is shared amongst a number of users and the number of users between consecutive data transfers are correlated.

REFERENCES

- [1] Eitan Altman, Panagiotis Konstantopoulos, and Zhen Liu, "Stability, monotonicity and invariant quantities in general polling systems," *Queueing Systems*, vol. 11, pp. 35–57, 1992, Special issue on 'Polling Systems'.
- [2] Eitan Altman, Tanja Jimenez, and Daniel Kofman, "DPS queues with stationary ergodic service times and the performance of TCP in overload," in *Proceedings of IEEE Infocom*, Hong-Kong, March 2004.

- [3] Alain Jean-Marie and Philippe Robert, "On the transient behaviour of the processor sharing queue," *QUESTA*, vol. 17, no. 1-2, pp. 129–136, 1994.
- [4] Thomas Bonald, A. Proutière, G. Régnié, and J.W. Roberts, "Insensitivity results in statistical bandwidth sharing," in *ITC*, Salvador, Brazil, 2001, number 17.
- [5] Thomas Bonald, M. Jonckheere, and A. Proutière, "Insensitive load balancing," in *ACM Sigmetrics Performance Evaluation Review*, June 2004, vol. 32, pp. 367–377.
- [6] Gil Zussman, Uri Yechiali, and Adrian Segall, "Exact probabilistic analysis of the limited scheduling algorithm for symmetrical bluetooth piconets," in *Personal Wireless Communications (PWC)*, Venice, Italy, Sept. 23-25 2003.
- [7] Eitan Altman, "Stochastic recursive equations with applications to queues with dependent vacations," *Annals of Operations Research*, vol. 112, no. 1, pp. 43–61, 2002.
- [8] Eitan Altman, Chadi Barakat, and Konstantin Avratchenkov, "A stochastic model of TCP/IP with stationary ergodic random losses," in *ACM-Sigcomm*, Aug. 28 - Sept. 1 2000, See also INRIA Research Report RR-3824.
- [9] Eitan Altman, "On stochastic recursive equations and infinite server queues," in *Proc. of IEEE Infocom 2005*, Miami, FL, March 13-17 2005.
- [10] Jack S. Sykes, "Simplified analysis of an alternating-priority queueing model with setup times," *Operations Research*, vol. 18, no. 6, pp. 1182–1192, November-December 1970.
- [11] Hideaki Takagi, "Queueing analysis of polling models: An update," *Stochastic Analysis of Computer and Communication Systems*, pp. 267–318, 1990.
- [12] Leonard Kleinrock, *Queueing Systems, Volume I: Theory*, vol. I, John Wiley and Sons, 1976.
- [13] T. Ozawa, "An analysis for multi-queueing systems with cyclic-service discipline - models with exhaustive and gated services," Tech. Rep. 4, The Institute of Electronics, Information and Communication Engineers (IEICE) (in Japanese), 1987.
- [14] Robin Groenevelt and Eitan Altman, "Analysis of alternating-priority queueing models with (cross) correlated switchover times," Tech. Rep. RR-5368, INRIA, Sophia-Antipolis, November 2004.
- [15] Hideaki Takagi, *Queueing Analysis, Vacations and Priority Systems, Part 1*, vol. 1, Elsevier Science Publishers B.V., The Netherlands, 1991.
- [16] David M. Lucantoni, "New results on the single server queue with a batch markovian arrival process," *Commun. Statist. - Stoch. Models*, vol. 7, pp. 1–46, 1991.
- [17] David M. Lucantoni, Kathleen S. Meier-Hellstern, and Marcel F. Neuts, "A single server queue with server vacations and a class of non-renewal arrival processes," *Adv. Appl. Prob.*, vol. 22, pp. 676–705, 1990.
- [18] Martin Eisenberg, "Queues with periodic service and changeover time," *Operations Research*, vol. 20, no. 2, pp. 440–451, March-April 1972.
- [19] J.D.C.Little, "A proof for the queueing formula: $L = \lambda W$," *Operations Research*, vol. 9, pp. 383–387, May-June 1961.
- [20] F. Baccelli and P. Brémaud, *Elements of Queueing Theory*, Springer-Verlag, 1994.
- [21] Robert B. Cooper, Shun-Chen Niu, and Mandyam M Srinivasan, "A decomposition theorem for polling models: The switchover times are effectively additive," *Operations Research*, vol. 44, no. 4, pp. 629–633, Jul-Aug 1996.
- [22] S. W. Fuhrmann and Robert B. Cooper, "Stochastic decompositions in the M/G/1 queue with generalized vacations," *Operations Research*, vol. 33, no. 5, pp. 1117–1129, 1985.
- [23] Hideaki Takagi, *Analysis of Polling Systems*, MIT Press, Cambridge, Massachusetts, 1986.

APPENDIX

A. Proof of Theorem 3:

Taking the expectation on both sides of equation (5) gives

$$\mathbb{E}[I_{n+1,1}] = R + \frac{\rho_2}{1 - \rho_2} \left(R + \frac{\rho_1 \mathbb{E}[I_{n,1}]}{1 - \rho_1} \right).$$

Under the stationary regime $\mathbb{E}[I_{n+1,1}] = \mathbb{E}[I_{n,1}]$ which immediately leads to the first moment

$$E[I_{n,1}] = \frac{R(1 - \rho_1)}{1 - \rho}. \quad (30)$$

To obtain the second moment we take the expectation of the square of (5) to give

$$\mathbb{E}[I_{n+1,1}^2] = \mathbb{E}[\mathcal{A}_n(I_{n,1})]^2 + \mathbb{E}[\mathcal{B}_n^2] + 2\mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n]. \quad (31)$$

This expression is central to proof in this section and each of the terms will be derived piece by piece. First of all, $\mathbb{E}[\mathcal{A}_n^2(I_{n,1})] =$

$$\begin{aligned} &= \mathbb{E} \left[\mathcal{D}_{n+1,2}^2 \left(\mathcal{N}_{n+1,2} \left(\mathcal{D}_{n+1,1} \left(\mathcal{N}_{n,1}(I_{n,1}) \right) \right) \right) \right] \\ &= \lambda_2^2 d_2^2 \mathbb{E}[\mathcal{D}_{n+1,1}^2(\mathcal{N}_{n,1}(I_{n,1}))] \\ &\quad + \lambda_2 d_2^{(2)} \mathbb{E}[\mathcal{D}_{n+1,1}(\mathcal{N}_{n,1}(I_{n,1}))] \\ &= \lambda_2^2 d_2^2 \left(\lambda_1^2 d_1^2 \mathbb{E}[I_{n,1}^2] \right. \\ &\quad \left. + \lambda_1 d_1^{(2)} \mathbb{E}[I_{n,1}] \right) + \lambda_1 d_1 \lambda_2 d_2^{(2)} \mathbb{E}[I_{n,1}]. \end{aligned}$$

Plugging equation (12) and (15) into this results in

$$\begin{aligned} \mathbb{E}[\mathcal{A}_n^2(I_{n,1})] &= \frac{\rho_1^2 \rho_2^2 \mathbb{E}[I_{n,1}^2]}{(1-\rho_1)^2(1-\rho_2)^2} \\ &+ \frac{R}{(1-\rho_2)^2(1-\rho)} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1-\rho_1)^2} + \lambda_2 \rho_1 b_2^{(2)} \right). \quad (32) \end{aligned}$$

Next we proceed with the second unknown of expression (31), $\mathbb{E}[\mathcal{B}_n^2]$, where we recall that \mathcal{B}_n is defined in equation (6). We have

$$\begin{aligned} \mathbb{E}[\mathcal{B}_n^2] &= \mathbb{E} \left[V_{n+1,1} + V_{n+1,2} \right. \\ &\quad \left. + \mathcal{D}_{n+1,2} \left(\mathcal{N}_{n,2}(V_{n,2}) + \mathcal{N}_{n+1,2}(V_{n+1,1}) \right) \right]^2 \\ &= v_1^{(2)} + v_2^{(2)} + 2v_1 v_2 + 2c_{12}(0) \\ &\quad + 2\lambda_2 d_2 \left(v_1^{(2)} + v_2^{(2)} + 2v_1 v_2 + c_2(1) + c_{12}(-1) + c_{12}(0) \right) \\ &\quad + \lambda_2^2 d_2^2 \left(v_1^{(2)} + v_2^{(2)} + 2v_1 v_2 + 2c_{12}(-1) \right) + \lambda_2 d_2^{(2)} R \\ &= \frac{\Delta^2 + R^2}{(1-\rho_2)^2} - \frac{2\rho_2 \delta_2^2}{1-\rho_2} + \frac{R\lambda_2 b_2^{(2)}}{(1-\rho_2)^3} \\ &\quad + \frac{2\rho_2 c_2(1) + 2c_{12}(0)}{1-\rho_2} + \frac{2\rho_2 c_{12}(-1)}{(1-\rho_2)^2}. \quad (33) \end{aligned}$$

To solve the last part first notice that the processes $\mathcal{N}_{n,1}(\cdot)$, $\mathcal{N}_{n,2}(\cdot)$, $\mathcal{N}_{n+1,1}(\cdot)$, $\mathcal{N}_{n+1,2}(\cdot)$, $\mathcal{D}_{n+1,1}(\cdot)$, and $\mathcal{D}_{n+1,2}(\cdot)$ are all independent of each other, and each of them is independent of $I_{n,1}$, $V_{n,2}$, $V_{n+1,1}$, and $V_{n+1,2}$. This means that

$$\mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n] = \alpha \mathbb{E}[I_{n,1}\mathcal{B}_n], \quad (34)$$

with once again $\alpha := \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$. The last piece of the puzzle can be derived with the help of Theorem 2 in [7] which states that

$$I_{n,1} = \sum_{j=0}^{\infty} \left(\prod_{i=n-j}^{n-1} \mathcal{A}_i^{(n-j)} \right) (\mathcal{B}_{n-j-1}), \quad n \in \mathbb{Z},$$

where for each integer i , $\{\mathcal{A}_i^{(-j)}\}_j$ are independent of each other and have the same distribution as $\mathcal{A}_i(\cdot)$. To apply the theorem it is sufficient to have $\alpha < 1$, which turns out to be

equivalent to $\rho < 1$, and that $\mathbb{E}[\mathcal{B}_n] < \infty$ (see Lemma 1 in [7]). The latter indeed holds as $\mathbb{E}[\mathcal{B}_n] = \frac{R}{1-\rho_2}$.

Applying the theorem gives $\mathbb{E}[I_{n,1}\mathcal{B}_n] =$

$$\begin{aligned} &= \sum_{j=0}^{\infty} \mathbb{E} \left[\prod_{i=n-j}^{n-1} \left(\mathcal{D}_{i+1,2} \left(\mathcal{N}_{i+1,2} \left(\mathcal{D}_{i+1,1} \left(\mathcal{N}_{i,1}(\mathcal{B}_{n-j-1}) \right) \right) \right) \right) \mathcal{B}_n \right] \\ &= \sum_{j=0}^{\infty} \alpha^j \mathbb{E}[\mathcal{B}_{n-j-1}\mathcal{B}_n] = \sum_{j=0}^{\infty} \alpha^j \mathbb{E}[\mathcal{B}_0\mathcal{B}_{j+1}]. \quad (35) \end{aligned}$$

because of the independence of the processes $\mathcal{D}_{i,1}(\cdot)$, $\mathcal{D}_{i,2}(\cdot)$, $\mathcal{N}_{i,1}(\cdot)$, and $\mathcal{N}_{i,2}(\cdot)$, for all $i \in \mathbb{Z}$. Writing out the last term yields $\mathbb{E}[\mathcal{B}_0\mathcal{B}_{j+1}] =$

$$\begin{aligned} &= \mathbb{E} \left[\left(V_{1,1} + V_{1,2} + \mathcal{D}_{1,2} \left(\mathcal{N}_{0,2}(V_{0,2}) + \mathcal{N}_{1,2}(V_{1,1}) \right) \right) \right. \\ &\quad \cdot \left(V_{j+2,1} + V_{j+2,2} \right. \\ &\quad \left. \left. + \mathcal{D}_{j+2,2} \left(\mathcal{N}_{j+1,2}(V_{j+1,2}) + \mathcal{N}_{j+2,2}(V_{j+2,1}) \right) \right) \right] \\ &= \frac{R^2}{(1-\rho_2)^2} + \frac{c_1(j+1) + c_2(j+1)}{(1-\rho_2)^2} \\ &\quad + \frac{\rho_2(c_2(j) - 2c_2(j+1) + c_2(j+2))}{1-\rho_2} \\ &\quad + \frac{c_{12}(-j-1) + c_{12}(j+1)}{1-\rho_2} + \frac{\rho_2(c_{12}(-j-2) + c_{12}(j))}{(1-\rho_2)^2}. \quad (36) \end{aligned}$$

Putting (34)-(36) together and re-indexing the summation (for example, $\sum_{j=0}^{\infty} \alpha^{j+1} c_2(j) = \sum_{j=1}^{\infty} \alpha^j c_2(j-1) = \alpha c_2(0) + \sum_{j=1}^{\infty} \alpha^j \alpha c_2(j)$) produces $\mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n] =$

$$\begin{aligned} &= \sum_{j=1}^{\infty} \left[\frac{R^2 + c_1(j) + c_2(j)}{(1-\rho_2)^2} + \frac{\rho_2 \left(\alpha - 2 + \frac{1}{\alpha} \right) c_2(j)}{1-\rho_2} \right. \\ &\quad \left. + \frac{c_{12}(-j)}{1-\rho_2} \left(1 + \frac{\rho_2}{\alpha(1-\rho_2)} \right) + \frac{c_{12}(j)}{1-\rho_2} \left(1 + \frac{\alpha \rho_2}{1-\rho_2} \right) \right] \alpha^j \\ &\quad + \frac{\rho_2(\alpha c_2(0) - c_2(1))}{1-\rho_2} + \frac{\rho_2(-c_{12}(-1) + \alpha c_{12}(0))}{(1-\rho_2)^2}. \end{aligned}$$

All of the terms with v_1 and v_2 can be pulled out of the summation and under stationary regime $c_2(0) = \delta_2^2$. This gives

$$\begin{aligned} \mathbb{E}[\mathcal{A}_n(I_{n,1})\mathcal{B}_n] &= \frac{\rho_1 \rho_2 R^2}{(1-\rho_2)^2(1-\rho)} + \frac{\rho_1 \rho_2^2 \delta_2^2}{(1-\rho_1)(1-\rho_2)^2} \\ &\quad - \frac{\rho_2 c_2(1)}{1-\rho_2} + \frac{\rho_2(-c_{12}(-1) + \alpha c_{12}(0))}{(1-\rho_2)^2} \\ &\quad + \frac{1-\rho_2(1-\alpha)}{(1-\rho_2)^2} \sum_{j=1}^{\infty} c_{12}(j) \alpha^j + \frac{1}{(1-\rho_2)^2} \sum_{j=1}^{\infty} \left[c_1(j) \right. \\ &\quad \left. + c_2(j) + \frac{(1-\rho)^2 c_2(j)}{\rho_1(1-\rho_1)} + \frac{(1-\rho_2)c_{12}(-j)}{\rho_1} \right] \alpha^j \quad (37) \end{aligned}$$

Putting equations (32),(33), and (37) into (31) and collecting

terms gives

$$\begin{aligned} \mathbb{E}[I_{n+1,1}^2] &= \frac{\rho_1^2 \rho_2^2 \mathbb{E}[I_{n,1}^2]}{(1-\rho_1)^2(1-\rho_2)^2} \\ &+ \frac{R}{(1-\rho_2)^2(1-\rho)} \left(\frac{\lambda_1 \rho_2^2 b_1^{(2)}}{(1-\rho_1)^2} + \lambda_2 b_2^{(2)} \right) \\ &+ \frac{1}{(1-\rho_2)^2} \left(\Delta^2 - \frac{2\rho_2(1-\rho)\delta_2^2}{1-\rho_1} + \left(\frac{1-\rho+2\rho_1\rho_2}{1-\rho} \right) R^2 \right) \\ &+ \frac{2(1-\rho_2(1-\alpha))}{(1-\rho_2)^2} \sum_{j=0}^{\infty} c_{12}(j) \alpha^j + \frac{2}{(1-\rho_2)^2} \sum_{j=1}^{\infty} \left[c_1(j) \right. \\ &\left. + c_2(j) + \frac{(1-\rho)^2 c_2(j)}{\rho_1(1-\rho_1)} + \frac{(1-\rho_2)c_{12}(-j)}{\rho_1} \right] \alpha^j. \end{aligned}$$

Under stationary regime $\mathbb{E}[I_{n+1,1}^2] = \mathbb{E}[I_{n,1}^2]$. The Theorem follows by putting these terms on the same side and by making use of the identity

$$1 - \left(\frac{\rho_1}{1-\rho_1} \right)^2 \left(\frac{\rho_2}{1-\rho_2} \right)^2 = \frac{(1-\rho)(1-\rho+2\rho_1\rho_2)}{(1-\rho_1)^2(1-\rho_2)^2}.$$

B. Proof of Theorem 4:

For a customer arriving at the first queue the system behaves as an M/G/1 queue where the server goes on vacation as soon as the queue is empty. The random variable for the n^{th} "vacation" from the first queue is exactly $I_{n,1}$. Conditioning the waiting time in the queue on whether or not a customer arrives when the server is busy or on vacation produces

$$\mathbb{E}[W_{q,1}] = \frac{\mathbb{E}[I_{n,1}]}{\mathbb{E}[C_{n,1}]} \mathbb{E}[W_{q,1}|vac] + \frac{\mathbb{E}[D_{n,1}]}{\mathbb{E}[C_{n,1}]} \mathbb{E}[W_{q,1}|busy]. \quad (38)$$

A tagged customer that arrives during a vacation has to wait for the vacation to finish plus the time needed to serve the customers that arrived before him/her in the vacation. The expected remaining vacation time is $\mathbb{E}[I_{n,1}^2]/2\mathbb{E}[I_{n,1}]$ and the expected number of customer that arrived before the tagged customer is $\lambda_1 \mathbb{E}[I_{n,1}^2]/2\mathbb{E}[I_{n,1}]$. This means that

$$\mathbb{E}[W_{q,1}|vac] = \frac{\mathbb{E}[I_{n,1}^2]}{2\mathbb{E}[I_{n,1}]} (1 + \lambda_1 b_1). \quad (39)$$

A tagged customer that arrives when the server is busy has to wait for the current customer in service to finish plus the expected time needed to serve the $L_{q,1}$ customers that arrived at (and still are in) the queue before the tagged customer did. This gives

$$\mathbb{E}[W_{q,1}|busy] = \frac{b_1^{(2)}}{2b_1} + b_1 \mathbb{E}[L_{q,1}|busy]. \quad (40)$$

To obtain the number of customers in the queue, first realize that the expected waiting time of a customer in the system is $\mathbb{E}[W_{s,1}] = b_1 + \mathbb{E}[W_{q,1}]$. Little [19] tells us that the expected number of customers, $L_{s,1}$, at the first queue (in service and in the queue) is

$$\mathbb{E}[L_{s,1}] := \lambda_1 \mathbb{E}[W_{s,1}] = \rho_1 + \lambda_1 \mathbb{E}[W_{q,1}].$$

On the other hand,

$$\begin{aligned} \mathbb{E}[L_{s,1}] &= \frac{\mathbb{E}[D_{n,1}]}{\mathbb{E}[C_{n,1}]} \mathbb{E}[L_{s,1}|busy] + \frac{\mathbb{E}[I_{n,1}]}{\mathbb{E}[C_{n,1}]} \mathbb{E}[L_{s,1}|vac] \\ &= \rho_1 (1 + \mathbb{E}[L_{q,1}|busy]) + (1-\rho_1) \frac{\lambda_1 \mathbb{E}[I_{n,1}^2]}{2\mathbb{E}[I_{n,1}]}. \end{aligned}$$

Combining these last two equations gives

$$\mathbb{E}[L_{q,1}|busy] = \frac{1}{b_1} \left(\mathbb{E}[W_{q,1}] - \frac{(1-\rho_1)\mathbb{E}[I_{n,1}^2]}{2\mathbb{E}[I_{n,1}]} \right). \quad (41)$$

Putting together equations (38)-(41) gives

$$\begin{aligned} \mathbb{E}[W_{q,1}] &= (1-\rho_1) \frac{\mathbb{E}[I_{n,1}^2]}{2\mathbb{E}[I_{n,1}]} (1 + \lambda_1 b_1) \\ &+ \rho_1 \left(\frac{b_1^{(2)}}{2b_i} + \mathbb{E}[W_{q,1}] - \frac{(1-\rho_1)\mathbb{E}[I_{n,1}^2]}{2\mathbb{E}[I_{n,1}]} \right) \\ &= \frac{\lambda_1 b_1^{(2)}}{2(1-\rho_1)} + \frac{\mathbb{E}[I_{n,1}^2]}{2\mathbb{E}[I_{n,1}]}. \end{aligned} \quad (42)$$

- The theorem follows by plugging in the values of $\mathbb{E}[I_{n,1}]$ and $\mathbb{E}[I_{n,1}^2]$. Notice that not once have we assumed the switchover times and the busy periods to be uncorrelated! ■

C. List of Notations

$\mathcal{A}_n(\cdot)$ = A nested combination of stochastic processes defined in (6).

\mathcal{B}_n = A nested combination of stochastic processes defined in (6).

$\mathcal{D}_{n,i}(N)$ = Total busy period generated by N customers in queue i with arrival rate λ_i and first and second moment of the service time b_i and $b_i^{(2)}$, respectively.

$\mathcal{D}_{n,i,k}$ = Single busy period generated by the k -th customer in queue i with arrival rate λ_i and first and second moment of the service time b_i and $b_i^{(2)}$, respectively.

$\mathcal{S}_{n,i}(N)$ = Service time of N customers at queue i with first and second moment of the service time b_i and $b_i^{(2)}$, respectively.

$\mathcal{N}_{n,i}(T)$ = Number of arrivals at queue i in time T in the n^{th} cycle (cycle starting from the polling instant of the first queue).

$\mathcal{T}_{n,i}(N)$ = The number of customers served at queue i during the n^{th} cycle if there are N customers in the queue at the moment of polling.

$\mathcal{X}_n(\cdot)$ = A nested combination of stochastic processes defined in (8).

\mathcal{Y}_n = A nested combination of stochastic processes defined in (8).

- $C_{n,i}$ = Duration of the n^{th} cycle starting from the polling instant of the i^{th} queue.
- $D_{n,i}$ = Duration of the busy period at queue i in the n^{th} cycle.
- $I_{n,i}$ = Intervisit of the i^{th} queue in the n^{th} cycle. This is the time between the server switching away from queue i until the time that the server comes back to queue i . ($I_{n,1} = V_{n,1} + D_{n,2} + V_{n,2}$ for exhaustive/exhaustive queues and $I_{n,1} = V_{n,1} + S_{n,2} + V_{n,2}$ for exhaustive/gated queues).
- K = Some constant used in the expressions for the intervisit time and the waiting times. Sometimes K_1 or K_2 is used if there is a difference between the two queues.
- $L_{n,i}^*$ = Number of customers in queue i in the n^{th} cycle at the moment the queue is polled.
- $L_{q,i}$ = Average number of customers in queue i . This is also the number of customers that arrived at queue i during a vacation, and are still in the queue, before a tag customer arrived in that same vacation.
- $L_{s,i}$ = Average number of customers at queue i (including the customer in service). This is also the number of customers that arrived at queue i during a vacation (and are in the queue or in service) before a tag customer arrived in that same vacation.
- $S_{n,i}$ = Service time at queue i in the n^{th} cycle (=similar to the duration $D_{n,i}$ of the busy period but then with no arrivals).
- $T_{n,i}$ = The number of customers served, per cycle, at queue i .
- $V_{n,i}$ = Switching time from queue i to the other queue in the n^{th} cycle.
- $W_{q,i}$ = Random variable for the waiting time of a customer in queue i (not including service).
- $W_{s,i}$ = Random variable for the total sojourn time of a customer at queue i (waiting time plus service time).
- Γ_i = The number of customers served during a busy period, where the arrival rate is λ_i , average service time is b_i , and the second moment of the service time is $b_i^{(2)}$.

$\alpha = \frac{\rho_1 \rho_2}{(1-\rho_1)(1-\rho_2)}$ = Central quantity in the exhaustive/exhaustive queueing system. Comes forth from $\mathbb{E}[\mathcal{A}_n(I)] = \alpha \mathbb{E}[I]$, with $\mathcal{A}_n(\cdot)$ defined in (6).

$b_i = \mathbb{E}[B_i]$ = Expected service time at queue i .

$b_i^{(2)} = \mathbb{E}[B_i]^2$ = Second moment of the service time at queue i .

$c_i(n) = \mathbb{E}[V_{0,i}V_{n,i}] - \mathbb{E}[V_{0,i}]\mathbb{E}[V_{n,i}]$ = Covariance function for the vacation sequences at queue i .

$c_{12}(n) = \mathbb{E}[V_{0,1}V_{n,2}] - \mathbb{E}[V_{0,1}]\mathbb{E}[V_{n,2}]$ = Covariance function for the vacation sequences between the two queues.

$d_i = \rho_i/(1-\rho_i)$ = The expected duration of a single busy period, where the arrival rate is λ_i and the average service time is b_i .

$d_i^{(2)} = b_i^{(2)}/(1-\rho_i)^3$ = Second moment of the duration of a single busy period, where the arrival rate is λ_i , the average service time is b_i , and the second moment of the service time is $b_i^{(2)}$.

$\delta_i^2 = v_i^{(2)} - v_i^2$ = Variance of the switching time from queue i to the other queue.

$\gamma = \frac{\rho_2}{1-\rho_1}$ = Central quantity in the exhaustive/gated queueing system. Comes forth from $\mathbb{E}[\mathcal{X}_n(I)] = \gamma \mathbb{E}[I]$, with $\mathcal{X}_n(\cdot)$ defined in (8).

λ_i = (Poisson) arrival rate at queue i .

$\rho_i = \lambda_i d_i$ = load at queue i .

$\rho = \rho_1 + \rho_2$ = load of the system. The assumption is made throughout that $\rho < 1/2$.

$\hat{\rho} = \rho_1 = \rho_2$ if the parameter settings for the two queues are equal.

$R = v_1 + v_2$.

$v_i = \mathbb{E}[V_{n,i}]$ = Expected switching time from queue i to the other queue.

$v_i^{(2)} = \mathbb{E}[V_{n,i}^2]$ = Expected second moment of the switching time from queue i to the other queue.