Queueing Systems 11(1992)35–57

Stability, monotonicity and invariant quantities in general polling systems

Eitan Altman, Panagiotis Konstantopoulos and Zhen Liu

INRIA, Centre Sophia Antipolis, 2004 Route des Lucioles, 06560 Valbonne, France

Consider a polling system with $K \geq 1$ queues and a single server that visits the queues in a cyclic order. The polling discipline in each queue is of general gated-type or exhaustive-type. We assume that in each queue the arrival times form a Poisson process, and that the service times, the walking times, as well as the set-up times form sequences of independent and identically distributed random variables. For such a system, we provide a sufficient condition under which the vector of queue lengths is stable. We treat several criteria for stability: the ergodicity of the process, the geometric ergodicity, and the geometric rate of convergence of the first moment. The ergodicity implies the weak convergence of station times, intervisit times and cycle times. Next, we show that the queue lengths, station times, intervisit times and cycle times are stochastically increasing in arrival rates, in service times, in walking times and in set-up times. The stability conditions and the stochastic monotonicity results are extended to the polling systems with additional customer routing between the queues, as well as bulk and correlated arrivals. Finally, we prove that the mean cycle time, the mean intervisit time and the mean station times are invariant under general service disciplines and general stationary arrival and service processes.

Keywords: Polling system, stability condition, monotonicity, stochastic comparison, invariant quantities.

1. Introduction

The aim of this paper is to establish some basic properties of polling systems under general service disciplines. The polling system under consideration consists of a single server which serves customers from different queues in a cyclic order. The service discipline in each queue is of general gated-type or exhaustive-type.

In the first case, when the server arrives at a queue, the number of customers that are served is a (possibly randomized) function of the number of customers in that queue at the arrival instant of the server. This includes as special cases the (purely) gated and the gated-limited.

In the exhaustive-type discipline, service is given in a queue until the number of customers decreases by a number which is a (possibly random) function of the number of customers found by the server upon arrival. The decrementing service and the exhaustive service are special cases of this service discipline. Such general disciplines were first introduced by Levy et al. [16].

Three kinds of results are obtained under the assumption that in each queue the arrival times form a Poisson process, and that the service times, the walking times, as well as the set-up times form sequences of independent and identically distributed (i.i.d.) random variables (RVs). These assumptions are relaxed for the third result.

The first result is on the stability conditions. For the general polling system described above, we use Foster's criterion to obtain new sufficient conditions for the ergodicity of the queue lengths, as well as for the geometric ergodicity (see definition in e.g. [2, p. 144]). We also show that the ergodicity then implies the stationarity of the cycle times, and that the geometric ergodicity implies the geometric rate of convergence of the first moment of the queue lengths. In the literature, some general sufficient conditions for stability of gated and exhaustive conditions were obtained by Zhdanov and Saksonov [28], where the interarrival times are allowed to be i.i.d. random variables. Kuehn [14] and Georgiadis and Szpankowski [11] presented the stability condition for the limited service discipline. The interested reader is referred to Takagi [25] for further references.

Second, we analyze the stochastic monotonicity of the queue lengths, the cycle times, the intervisit times and the station times with respect to (w.r.t.) the system parameters such as the arrival rate, the service times, the walking times and the set-up times. We show that the queue lengths, the cycle times, the intervisit times and the station times are stochastically increasing in arrival rates, in service times, in walking times and in set-up times. This monotonicity also holds with respect to the number of queues in the system.

All these results on the stability and monotonicity are extended to the systems with customer routing and bulk and correlated arrivals. In such a case, the customers, at the end of service, may be routed to other queues in the system or rejoin the same queue for another portion of service. Polling systems with such routing mechanisms are analyzed by Sidi et al. [21] under the (purely) gated and (fully) exhaustive service disciplines. We further consider the case where the external arrivals of the customers are controlled by a single Poisson process. At an arrival epoch, a random number of customers are sent to each of the queues.

Last, we prove that the mean cycle time, the mean intervisit time and the mean station time in the steady state are invariant under general service disciplines and general stationary arrival and service processes. The basic tool used to prove this invariance is the theory of stationary point processes, such as Campbell's formula. This invariance generalizes the work of Kuehn [14] to a more general setting. Such invariant quantities were traditionally obtained using calculations that heavily depend on the discipline that is used, see e.g. Takagi [24, pp. 73–77 for the exhaustive service, p. 108 for the gated discipline and pp. 128–134 for the limited discipline (all under the assumption of Poisson arrivals)]. Zhdanov and Saksonov [28] obtained these invariant quantities for polling systems with either gated or exhaustive service disciplines, under the assumption that the interarrival times and the service times are i.i.d. random variables. Daganzo [9] relaxed the statistical assumptions by allowing the service times to be dependent.

Our paper is organized as follows. In the next section, we describe the model in detail and we define the notation used in the paper. In sections 3, 4 and 5, we prove the above three results, respectively. In section 6, we extend the stability and monotonicity results to the more general systems with customer routing and correlated external arrivals.

2. The model and notation

The polling system contains a single server and $K \geq 1$ queues with infinite capacity, numbered by $1, 2, \ldots, K$. The initial state of the polling system is described by the vector $Q_1 = (Q_1^1, \ldots, Q_1^K)$ where Q_1^k is the initial number of customers present at queue k, $1 \leq k \leq K$. The customers arrive to queue k in accordance with a Poisson process of parameter λ_k. The service times in queue k are i.i.d. with the common distribution B_k. Denote by b_k the first moment of B_k, which is assumed to be finite.

The server visits the queues in a cyclic order, viz. $1, 2, \ldots, K, 1, 2, \ldots$. Without loss of generality, we assume that the server is initially at queue 1. Thus, the nth ($n \geq 1$) queue that the server visits is queue $I(n) = (n-1) \bmod K + 1$, where $n \bmod K$ means the remainder of the division of n by K.

At the nth visit, if there are some customers in the queue and if the server decides to serve, then a set-up time (also called the switch-in time in the literature) R_n is incurred before the server starts serving customers in that queue. If the server decides not to give service to the queue, then this set-up time is not incurred (sometimes, R_n is referred to as the potential set-up time). For any given k, $1 \leq k \leq K$, the random variables in $\{R_{nK+k}\}_{n=0}^{\infty}$ are i.i.d., and their first moment, assumed finite, is denoted by r_k. Let $R = \sum_{k=1}^{K} R_k$ be the total set-up time in a cycle and denote by r the first moment of R.

The time the server takes between the departure from the nth visited queue and the arrival to the $(n+1)$st queue is called the nth walking time and is denoted by D_n. For any given k, $1 \leq k \leq K$, the random variables in $\{D_{nK+k}\}_{n=0}^{\infty}$ are i.i.d., and their first moment, assumed finite, is denoted by d_k. Let $D = \sum_{k=1}^{K} D_k$ be the total walking time in a cycle and denote by d the first moment of D.

Denote by $\rho_k = \lambda_k b_k$, $1 \leq k \leq K$, $\rho = \sum_{k=1}^{K} \rho_k$.

The system state is described by the random variables Q_n^k, $1 \leq k \leq K$, $n \geq 1$, where Q_n^k represents the number of customers in queue k when the server arrives at the nth queue that it visits. Let $Q_n = (Q_n^1, \ldots, Q_n^K)$. The state space of the process Q is $\mathbf{X} = \mathbb{N}^K$, where $\mathbb{N}$ is the set of nonnegative integers. Denote by U_n^k, $1 \leq k \leq K$, $n \geq 1$, the number of customers in queue k when the server leaves the nth visited queue.

The arrival times, the service times, the walking times and the set-up times are mutually independent, and are independent of the past and present system states.

Let $\mathcal{A}_k(T)$ denote the number of arrivals to queue k, $1 \le k \le K$, during a (possibly random) time interval of length T. Let $\mathcal{B}_n(l)$ be the random variable that represents the total service time of l customers at the nth visit of the server, $n \ge 1$. Let $\mathcal{G}_n(l)$ be the RV that represents the time spent by the server at its nth visit in order to decrease the length of the visited queue by l, $n \ge 1$. The random variable $\mathcal{G}_n(l)$ is thus distributed as the time it takes to empty an $M/G/1$ queue, the service time of which is distributed as queue $I(n)$'s, from the instant a service begins and l customers are present.

The service discipline of the polling system under consideration may have mixed strategies. Some queues are served according to general gated-type service disciplines and the others according to general exhaustive-type service disciplines. Suppose at the nth arrival of the server to a queue there are $Q_n^{I(n)} = x$ customers in that queue. If queue $I(n)$ is served according to a gated-type discipline, then the number of customers that will be served is given by a random integer $f_n(x)$ whose distribution depends on x and on $I(n)$. If queue $I(n)$ is of exhaustive-type service, then service is given until the number of customers decreases, with respect to $Q_n^{I(n)}$, by the random integer $f_n(x)$. The random variables $f_n(x)$, $n, x \in \mathbb{N}$, are mutually independent, and are independent of the past and present system states, the arrival times, the service times, the walking times and the set-up times. Clearly $f_n(x) \le x$.

This kind of characterization of general service disciplines was proposed by Levy et al. [16]. It includes as special cases the (purely) gated discipline (for which $f_n(x) = x$), the decrementing discipline (where $f_n(x) = \min\{x, l\}$, $l \ge 1$), the gated-limited service discipline ($f(x) = \min\{x, l\}$, $l \ge 1$) and the binomial disciplines (see Takagi [25]). Note that when there are set-up times, the exhaustive service discipline has two variants. In the first one, which we refer to as *fully* exhaustive discipline, the server serves all customers in the queue until the queue is empty. In the second one, which we refer to as *partial* exhaustive discipline, the server serves the customers in the queue until the queue length is decreased by the number of customers it found at its arrival to the queue. The two variants coincide when the set-up times are zero. The latter variant is included in our framework (it suffices to set $f_n(x) = x$), whereas the first one is not. However, it will be seen later on that all our results hold for the fully exhaustive discipline.

The performance measures of interest, besides Q_n, are:

- Station time: S_n, the time interval between the arrival times of the server to the nth queue and the $(n+1)$st queue.
- Intervisit time: $V_n = D_n + S_{n+1} + \ldots + S_{n+K-1}$, the time interval between the departure time of the server from a queue and the next arrival time to the same queue.
- Cycle time $C_n = S_n + \ldots + S_{n+K-1}$, the time interval between two successive arrivals of the server to the same queue.

3. Sufficient stability conditions

We first study the stability conditions of the above polling systems. More specifically, we provide sufficient conditions for the ergodicity and geometric ergodicity of the queue lengths, and the stability (the weak convergence to finite RVs) of the station times, intervisit times as well as the cycle times.

Define the indicator functions J_k, $1 \le k \le K$, by $J_k = 1$ if queue k is served according to the exhaustive-type discipline, and 0 otherwise; and $\bar{J}_k = 1 - J_k$. Denote by $h_k = 1/(1 - \rho_k J_k)$. For $x \in \mathbb{N}$, let $F_n(n) = E(f_n(x))$, $n = 1, 2, \ldots$.

The RVs Q_n evolve according to the following evolution equations:

$$Q_{n+1}^j =_{\text{st}} \begin{cases} Q_n^j + \mathcal{A}_j\Big(R_n \mathbf{1}_{\{f_n(Q_n^{I(n)})>0\}} + \bar{J}_{I(n)}\mathcal{B}_n(f_n(Q_n^{I(n)})) + J_{I(n)}\mathcal{G}_n(f_n(Q_n^{I(n)})) + D_n\Big), & I(n) \ne j; \\ Q_n^j - f_n(Q_n^j) + \mathcal{A}_j\Big(R_n \mathbf{1}_{\{f_n(Q_n^j)>0\}} + \bar{J}_{I(n)}\mathcal{B}_n(f_n(Q_n^j)) + D_n\Big), & I(n) = j. \end{cases} \tag{1}$$

Due to the cyclic structure of the queueing system, the Markov chain $\{Q_n\}$ has a periodic behavior with a period of K. More precisely, the Markov chain is nonhomogeneous in time since the transition probabilities at time n are a function of the queue $I(n)$. However, for every $1 \le k \le K$, $\{Q_{nK+k}\}_{n=0}^{\infty}$ is an aperiodic (homogeneous) Markov chain for which we will establish conditions of the ergodicity. In view of the following lemma, the system is said to be stable if all these K Markov chains $\{Q_{nK+k}\}_{n=0}^{\infty}$ are ergodic.

LEMMA 3.1

Assume that for all $1 \le k \le K$, and all $i \in \mathbb{N}$, $\{f_{nK+k}(i)\}_{n=0}^{\infty}$ is a sequence of i.i.d. random integers. If for all $1 \le k \le K$ the Markov chains $\{Q_{nK+k}\}_{n=0}^{\infty}$ are ergodic, then for all $1 \le k \le K$ the sequence of station times $\{S_{nK+k}\}_{n=0}^{\infty}$, intervisit times $\{V_{nK+k}\}_{n=0}^{\infty}$ and the cycle times $\{C_{nK+k}\}_{n=0}^{\infty}$ converge weakly to finite RVs.

Proof

For any k, $1 \le k \le K$, define

$$\mathcal{H}_k(i) = R_k \mathbf{1}_{\{f_k(i)>0\}} + \bar{J}_k\mathcal{B}_k(f_k(i)) + J_k\mathcal{G}_k(f_k(i)) + D_k, \quad i = 0, 1, 2, \ldots .$$

Then for all $n \ge 0$,

$$S_{nK+k} =_{\text{st}} \mathcal{H}_k\Big(Q_{nK+k}^k\Big).$$

Denote by Q^k the limit RV of the sequence $\{Q_{nK+k}^k\}_{n=0}^{\infty}$, i.e. the length of queue k when the server arrives at the queue in the steady state. Under the assumption of the lemma, such a limit RV is well defined. Let $S^k =_{\text{st}} \mathcal{H}_k(Q^k)$. Note that S^k is almost surely (a.s.) finite.

For all $1 \leq k \leq K$, $n \geq 0$, $i \geq 0$, let

$$p_n^k(i) = P\left[Q_{nK+k}^k = i\right], \quad p^k(i) = P\left[Q^k = i\right].$$

Since the sequence $\{Q_{nK+k}\}_{n=1}^{\infty}$ is ergodic, we have that for all $i \geq 0$,

$$p^k(i) = \lim_{n \to \infty} p_n^k(i).$$

It then follows that for all $x \in \mathbb{R}^+$,

$$\begin{aligned}
\lim_{n\to\infty} P[S_{nK+k} \leq x] &= \lim_{n\to\infty} P\left[\mathcal{H}_k\left(Q_{nK+k}^k\right) \leq x\right] \\
&= \lim_{n\to\infty} \sum_{i=1}^{\infty} P[\mathcal{H}_k(i) \leq x] p_n^k(i) \\
&= \sum_{i=1}^{\infty} \lim_{n\to\infty} P[\mathcal{H}_k(i) \leq x] p_n^k(i) \\
&= \sum_{i=1}^{\infty} P[\mathcal{H}_k(i) \leq x] p^k(i) \\
&= P\left[\mathcal{H}_k(Q^k) \leq x\right] = P\left[S^k \leq x\right],
\end{aligned}$$

where the change of order between the summation and the limit follows from the Portmanteau Theorem (see [3, p. 11]).

Thus, for all $1 \leq k \leq K$, the sequence of station times $\{S_{nK+k}\}_{n=0}^{\infty}$ weakly converges.

The weak convergence of the cycle times as well as the intervisit times can be shown in a similar way using the evolution eq. (1). □

Remark

It can be checked that the process $\{Q_n, S_n\}$ is in fact a "J–X" process (see e.g. [12]) for which different limit theorems exist [12,18]. For example, under the assumptions of lemma 3.1 one can establish the Strong Law of Large Numbers for the sequence $\{S_n\}$.

PROPOSITION 3.2

Assume that

- $\rho < 1$;
- $\forall n \geq 1$, $F_n(x)$ is a nondecreasing function of x;
- there exist a real constant ε and an integer M such that for every $n \geq 1$,

$$F_n(x) \geq \frac{\rho(d+r)+\varepsilon}{(1-\rho)h_{I(n)}b_{I(n)}} \quad \forall x > M.$$

Then for all $1 \leq k \leq K$ the Markov chain $\{Q_{nK+k}\}_{n=0}^{\infty}$ is ergodic (aperiodic and positive recurrent), and the Markov chain $\{Q_n\}_{n=0}^{\infty}$ is (periodic) positive recurrent.

Proof

Taking a conditional expectation in (1) and summing over j, we obtain:

$$E\left[\sum_{j=1}^{K} b_j Q_{n+1}^j | Q_n\right]$$

$$\leq \sum_{j=1}^{K} b_j Q_n^j + \rho d_{I(n)} + \rho r_{I(n)} + \bar{J}_{I(n)}(\rho-1)b_{I(n)}F_n(Q_n^{I(n)})$$

$$+ J_{I(n)}\left(\frac{\rho-\rho_{I(n)}}{1-\rho_{I(n)}} - 1\right)b_{I(n)}F_n(Q_n^{I(n)})$$

$$= \sum_{j=1}^{K} b_j Q_n^j + \rho d_{I(n)} + \rho r_{I(n)} + h_{I(n)}(\rho-1)b_{I(n)}F_n(Q_n^{I(n)}).$$

Similarly, we have

$$E\left[\sum_{j=1}^{K} b_j Q_{n+2}^j | Q_n\right]$$

$$= E\left\{E\left[\sum_{J=1}^{K} b_j Q_{n+2}^j | Q_{n+1}\right] | Q_n\right\}$$

$$\leq E\left\{\sum_{j=1}^{K} b_j Q_{n+1}^j + \rho d_{I(n+1)} + \rho r_{I(n+1)} + h_{I(n+1)}(\rho-1)b_{I(n+1)}F_{n+1}\left(Q_{n+1}^{I(n+1)}\right) | Q_n\right\}$$

$$\leq E\left\{\sum_{j=1}^{K} b_j Q_{n+1}^j | Q_n\right\} + \rho d_{I(n+1)} + \rho r_{I(n+1)} + h_{I(n+1)}(\rho-1)b_{I(n+1)}F_{n+1}\left(Q_n^{I(n+1)}\right)$$

$$= \sum_{j=1}^{K} b_j Q_n^j + \rho(d_{I(n)} + d_{I(n+1)}) + \rho(r_{I(n)} + r_{I(n+1)})$$

$$+ h_{I(n)}(\rho-1)b_{I(n)}F_n\left(Q_n^{I(n)}\right) + h_{I(n+1)}(\rho-1)b_{I(n+1)}F_{n+1}\left(Q_n^{I(n+1)}\right),$$

where we used the increasingness of $F_n(x)$ to obtain the last inequality. Repeating the above calculation, we obtain

$$E\left[\sum_{j=1}^{K} b_j Q_{n+K}^j | Q_n\right]$$
$$\leq \sum_{j=1}^{K} b_j Q_n^j + \rho d + \rho r + \sum_{j=0}^{K-1} h_{I(n+j)}(\rho - 1) b_{I(n+j)} F_{n+j}\left(Q_n^{I(n+j)}\right). \tag{2}$$

It then follows that if for some $0 \leq j < K$, $Q_n^{I(n+j)} > M$, then

$$E\left[\sum_{j=1}^{K} b_j \left(Q_{n+K}^j - Q_n^j\right) | Q_n\right] \leq -\varepsilon. \tag{3}$$

Moreover, for any value of Q_n, we have

$$E\left[\sum_{j=1}^{K} b_j \left(Q_{n+K}^j - Q_n^j\right) | Q_n\right] \leq \rho(d + r) < \infty. \tag{4}$$

The ergodicity of the chain Q_{nK+k} now follows from Foster's criterion (see e.g. [2, p. 18]). This completes the proof. □

Remark

The functions used in (3)–(4) for the verification of Foster's criterion, known as Lyapunov functions, are linear in the queue lengths. The weight given to a queue is the expected service time. Hence, (3) has the intuitive meaning that outside a finite set, the total expected workload should strictly decrease at each cycle.

Remark

The third assumption in proposition 3.2 has a similar interpretation. This assumption can be rewritten as

$$F_n(x) h_{I(n)} b_{I(n)} > \rho \frac{d + r}{1 - \rho}, \quad \forall x > M. \tag{5}$$

The quantity $(d + r)/(1 - \rho)$ is the expected cycle time in steady state (see (19) below) where the potential set-up times are included in the walking times. Hence, (5) states that if the length of a queue is large enough, then the total reduction of the expected workload in one cycle due to serving this queue (cf. the left-hand side of (5)) should be strictly greater than the total increase of the expected workload in one cycle in the steady state.

COROLLARY 3.3

Suppose that in the polling system, for all $1 \le k \le K$, $F_k(x)$ monotonically increases to ∞ when x goes to ∞. Then $\rho < 1$ is the necessary and sufficient condition for the stability.

COROLLARY 3.4

Suppose that the polling system consists of a set E of queues served exhaustively, a set G of queues served according to the gated discipline, and a set L of queues served according to the (gated)-limited discipline, with limits l_k, $k \in L$. If

- $\rho < 1$,
- $\forall k \in L: b_k l_k > (d + r)\rho/(1-\rho)$,

then for all $1 \le k \le K$ the Markov chain $\{Q_{nK+k}\}_{n=0}^{\infty}$ is ergodic, and the Markov chain $\{Q_n\}_{n=0}^{\infty}$ is positive recurrent.

PROPOSITION 3.5

Assume that $\forall n \ge 1$, $F_n(x)$ is a nondecreasing function of x, and that there exists some constant $0 < \alpha \le 1$ and integer M_0 such that $F_n(x) \ge \alpha x$ for all $x > M_0$, $n \ge 1$. Then the Markov chain $\{Q_{nK+k}\}_{n=0}^{\infty}$ is geometrically ergodic if and only if $\rho < 1$.

Proof

The necessary part of the proof is trivial and is omitted. For the sufficient part, it follows from (2) and the hypothesis of the proposition that

$$E\left[\sum_{j=1}^{K} b_j\left(Q_{n+K}^j - Q_n^j\right) | Q_n\right]$$

$$\le \rho d + \rho r + \alpha \sum_{j=0}^{K-1} h_{I(n+j)}(\rho - 1) b_{I(n+j)}\left(Q_n^{I(n+j)} - M_0\right)$$

$$\le \left(\rho d + \rho r + \alpha M_0 \sum_{j=1}^{K} h_j b_j + \frac{\alpha}{2} \sum_{j=0}^{K-1} h_{I(n+j)}(\rho - 1) b_{I(n+j)} Q_n^{I(n+j)}\right)$$

$$+ \frac{\alpha}{2} \sum_{j=0}^{K-1} h_{I(n+j)}(\rho - 1) b_{I(n+j)} Q_n^{I(n+j)}.$$

Let $\mathcal{K}$ be the set of states given by

$$\mathcal{K} = \left\{ x{:}\rho d + \rho r + \alpha M_0 \sum_{j=1}^{K} h_j b_j + \frac{\alpha}{2} \sum_{j=1}^{K} h_j (\rho - 1) b_j x_j > 0 \right\}. \tag{6}$$

Since $\rho < 1$, $\mathcal{K}$ is finite. Clearly, if $Q_n \notin \mathcal{K}$, then

$$E\left[\sum_{j=1}^{K} b_j \left(Q_{n+K}^j - Q_n^j \right) | Q_n \right] \le - \frac{\alpha(1-\rho)}{2} \sum_{j=1}^{K} b_j Q_n^j . \tag{7}$$

The geometric ergodicity then follows from Popov [19]. □

Remark

In case some of the queues have fully exhaustive service discipline, the stability conditions of propositions 3.2 and 3.5 still hold, and are obtained by the same Lyapunov functions. The evolution eq. (1) should be modified as follows:

$$Q_{n+1}^j =_{st} \begin{cases} Q_n^j + \mathcal{A}_j \Big(R_n \mathbf{1}_{\{f_n(Q_n^{I(n)})>0\}} + \bar{J}_{I(n)} \mathcal{B}_n (f_n(Q_n^{I(n)})) \\ \qquad + J_{I(n)} \mathcal{G}_n \Big(f_n(Q_n^{I(n)}) + u_n \mathcal{A}_{I(n)} (R_n \mathbf{1}_{\{f_n(Q_n^j)>0\}}) \Big) + D_n \Big), & I(n) \ne j; \\ Q_n^j - f_n(Q_n^j) + \mathcal{A}_j \Big(\bar{u}_n R_n \mathbf{1}_{\{f_n Q_n^j)>0\}} + \bar{J}_{I(n)} \mathcal{B}_n (f_n(Q_n^j)) + D_n \Big), & I(n) = j, \end{cases} \tag{8}$$

where $u_n = 1$ if queue $I(n)$ has fully exhaustive service discipline, and $u_n = 0$ otherwise, with $\bar{u}_n = 1 - u_n$. If $u_n = 1$, then it is readily checked that

$$E\left[\sum_{j=1}^{K} b_j Q_{n+1}^j | Q_n \right]$$

$$\le \sum_{j=1}^{K} b_j Q_n^j + \rho d_{I(n)} + \frac{\rho - \rho_{I(n)}}{1 - \rho_{I(n)}} r_{I(n)} + h_{I(n)} (\rho - 1) b_{I(n)} F_n \left(Q_n^{I(n)} \right)$$

$$< \sum_{j=1}^{K} b_j Q_n^j + \rho d_{I(n)} + \rho r_{I(n)} + h_{I(n)} (\rho - 1) b_{I(n)} F_n \left(Q_n^{I(n)} \right),$$

where we used the fact that $\rho < 1$ to establish the last inequality. Therefore, inequality (2) still holds even if there are queues with fully exhaustive service discipline.

Next, we establish the geometric convergence of the first moment of queue lengths.

PROPOSITION 3.6

Under the assumptions of proposition 3.5, EQ^i_{nK+k} converges geometrically fast for any $i, k = 1, \ldots, K$.

Proof

Choose any $i, k = 1, \ldots, K$. Let P be the transition matrix of the Markov chain Q^i_{nK+k}, $n = 1, 2, \ldots$ and let

$$\Pi = \lim_{n \to \infty} P^n$$

be the steady-state probability matrix corresponding to the transition matrix P. Let $\pi_x = \Pi_{yx}$ for all $x, y \in \mathbf{X}$ (note that Π has identical rows). Let $\mu : \mathbf{X} \to \mathbb{R}$ be some function. Following [22], define the μ-norm of any real vector z on $\mathbf{X}$ and of any real matrix Z on $\mathbf{X} \times \mathbf{X}$ as

$$\|z\|_\mu = \sup_{x \in \mathbf{X}} \mu_x^{-1} |z_x|$$

$$\|Z\|_\mu = \sup_{x \in \mathbf{X}} \mu_x^{-1} \sum_{y \in \mathbf{X}} |Z_{xy}| \mu_y.$$

Define the projection $r : \mathbf{X} \to \mathbb{R}$ by $r(x) = x_i$. We show below that there exists some function $\mu : \mathbf{X} \to \mathbb{R}$ with all entries greater than 1, some constants $0 < \beta < 1$ and $\gamma > 0$ such that

$$\|P^n r - \Pi r\|_\mu \le \gamma \beta^n, \quad \forall n \in \mathbb{N}, \tag{9}$$

which will then imply

$$\left| \sum_{x \in \mathbf{X}} P^n_{yx} x_i - \sum_{x \in \mathbf{X}} \pi_x x_i \right| \le \mu_y \gamma \beta^n, \quad \forall y \in \mathbf{X}, \quad \forall n \in \mathbb{N},$$

so that the assertion of the proposition holds.

In order to prove (9), we define $a = \max_{1 \le j \le K} b_j^{-1}$ and $\mu_x = a \sum_{j=1}^K b_j x_j + 1\{x = 0\} \ge 1$. Let $\mathcal{K}$ be defined as in (6), and

$$\mathcal{K}' = \{x : \mu_x \alpha (1 - \rho) < 4\}. \tag{10}$$

It then follows from (7) that for all $x \notin \mathcal{K} \cup \mathcal{K}'$,

$$\sum_{y \in \mathbf{X}} P_{xy} \mu_y \le \mu_x \left(1 - \frac{\alpha(1-\rho)}{2}\right) + 1 \le \mu_x \left(1 - \frac{\alpha(1-\rho)}{4}\right). \tag{11}$$

Relation (11) readily implies (cf. [22, Key Theorem I, p. 24]) that there exist constants $0<\beta<1$ and $\gamma'>0$ such that

$$\|P^n-\Pi\|_\mu \le \gamma'\beta^n, \quad \forall n \in \mathbb{N}.$$

Now,

$$\|P^n r-\Pi r\|_\mu \le \| P^n-\Pi\|_\mu \|r\|_\mu \le \gamma'\beta^n a^{-1} b_i^{-1}, \quad \forall n \in \mathbb{N},$$

which completes the proof by taking $\gamma=\gamma' a^{-1} b_i^{-1}$. □

4. Monotonicity of queue lengths, cycle times, intervisit times and station times

In this section, we analyze the stochastic monotonicity of the queue lengths, cycle times, intervisit times and station times with respect to the system parameters such as the arrival rates, service times, walking times and set-up times.

We first define some basic notation of stochastic ordering of random variables.

DEFINITION 4.1

Let $X, Y \in \mathbb{R}^n$ denote two random variables of dimension n. We say that X is stochastically smaller than Y, denoted by $X \le_{st} Y$, if for all nondecreasing functions $g: \mathbb{R}^n \to \mathbb{R}$, $E[g(X)] \le E(g(Y)]$, provided the expectations exist.

LEMMA 4.2 (STRASSEN [23])

Let $X, Y \in \mathbb{R}^n$ denote two random variables of dimension n. $X \le_{st} Y$ if and only if there exist two random variables $\hat{X}$ and $\hat{Y}$ defined on a common probability space such that $X =_{st} \hat{X}$, $Y =_{st} \hat{Y}$, and $\hat{X} \le \hat{Y}$ a.s., where $=_{st}$ stands for the equality in law.

Denote by Γ the polling system described in section 2. Let there be another polling system Γ' with K' queues. The initial state of the polling system is $Q_1' = (Q_1'^1, \ldots, Q_1'^{K'})$. The arrival rates, the service times, the walking times and the set-up times are described by λ_k', B_k', $\{D_n'\}_n$, and $\{R_n'\}_n$, respectively, $1 \le k \le K'$, with the assumption that these random variables are mutually independent. Let $\mathcal{A}_n'$, $\mathcal{B}_n'$, $\mathcal{G}_n'$ be the number of arrivals, the cumulative service times and the durations of busy periods corresponding to system Γ'. Denote by $Q_n'^k$ and $U_n'^k$ the state variables of the system Γ'. The station times, the intervisit times and the cycle times associated with Γ' are referred to as S_n', V_n' and C_n', respectively.

In the systems Γ and Γ', the number of customers to be served at each visit of the server is characterized by $f_n(x)$ and $f_n'(x)$, respectively. We assume that $f_{nK+k}(x) =_{st} f_{nK'+k}'(x)$ for all $n \ge 1$, $x \ge 0$, and $1 \le k \le \min(K, K')$. Throughout this section, we assume that the service discipline is "stochastically increasing" and "contractive": For all $n, x, y \in \mathbb{N}$, with $x \ge y$, there is a probability space such that

- (stochastic increasingness) $f_n(x) \geq f_n(y)$, a.s.
- and (contractiveness) $f_n(x) - f_n(y) \leq x - y$, a.s.

The increasing contractive disciplines were first considered in Levy et al. [16].

PROPOSITION 4.3

Suppose that Γ and Γ' have the same number of queues: $K = K'$. If

$$\left(Q_1^1, \ldots, Q_1^K\right) \leq_{st} \left(Q_1'^1, \ldots, Q_1'^K\right),$$

$$(\lambda_1, \ldots, \lambda_K) \leq (\lambda_1', \ldots, \lambda_K'),$$

$$B_k \leq_{st} B_k', \qquad 1 \leq k \leq K,$$

$$D_n \leq_{st} D_n', \qquad n \geq 1,$$

$$R_n \leq_{st} R_n', \qquad n \geq 1,$$

then

$$\left(Q_n^1, Q_n^2, \ldots, Q_n^K\right) \leq_{st} \left(Q_n'^1, Q_n'^2, \ldots, Q_n'^K\right), \qquad n \geq 1, \tag{12}$$

$$(S_1, \ldots, S_n) \leq_{st} (S_1', \ldots, S_n'), \qquad n \geq 1, \tag{13}$$

$$(V_1, \ldots, V_n) \leq_{st} (V_1', \ldots, V_n'), \qquad n \geq 1, \tag{14}$$

$$(C_1, \ldots, C_n) \leq_{st} (C_1', \ldots, C_n'), \qquad n \geq 1. \tag{15}$$

Proof

For the sake of simplicity, we first prove by induction on n that

$$\left(Q_n^1, Q_n^2, \ldots, Q_n^K\right) \leq_{st} \left(Q_n'^1, Q_n'^2, \ldots, Q_n'^K\right), \tag{16}$$

and that relation (16) implies that

$$S_n \leq_{st} S_n', \tag{17}$$

$$\left(Q_{n+1}^1, Q_{n+1}^2, \ldots, Q_{n+1}^K\right) \leq_{st} \left(Q_{n+1}'^1, Q_{n+1}'^2, \ldots, Q_{n+1}'^K\right). \tag{18}$$

By the assumption of the proposition, relation (16) holds when $n = 1$. Assume that (16) holds for some $n \geq 1$.

According to Strassen's Theorem (cf. lemma 4.2), there is a probability space such that

$$\left(Q_n^1, Q_n^2, \ldots, Q_n^K\right) \le \left(Q_n'^1, Q_n'^2, \ldots, Q_n'^K\right) \text{ a.s.}$$

Using the fact that the service discipline is stochastically increasing and contractive, and that it is independent of the past and present system state, we can construct a common probability space such that

$$f_n(x) \ge f_n(y) \text{ a.s.} \quad \text{and} \quad f_n(x) - f_n(y) \le x - y \text{ a.s.}$$

We assume that $I(n) = 1$. The same argument goes through for the other cases.

Assume first that the discipline in queue 1 is of gated type. Since

$$B_k \le_{\mathrm{st}} B_k', \quad R_n \le_{\mathrm{st}} R_n', \qquad 1 \le k \le K, \; n \ge 1,$$

we have that

$$\mathcal{B}_n\left(f_n\left(Q_n^1\right)\right) + R_n \mathbf{1}_{\{f_n(Q_n^1)>0\}} \le_{\mathrm{st}} \mathcal{B}_1'\left(f_n\left(Q_n'^1\right)\right) + R_n' \mathbf{1}_{\{f_n(Q_n'^1)>0\}}.$$

Moreover, as

$$(\lambda_1, \ldots, \lambda_K) \le (\lambda_1', \ldots, \lambda_K'),$$

we have the relation

$$(\mathcal{A}_1(X), \mathcal{A}_2(X), \ldots, \mathcal{A}_K(X)) \le_{\mathrm{st}} (\mathcal{A}_1'(Y), \mathcal{A}_2'(Y), \ldots, \mathcal{A}_K'(Y)),$$

which holds for all $X, Y \in \mathbb{R}^+$ such that $X \le_{\mathrm{st}} Y$ (see e.g. Jean-Marie and Liu [13]). Therefore,

$$\begin{aligned}
&\left(U_n^1, U_n^2, \ldots, U_n^K\right) \\
&=_{\mathrm{st}} \Big(Q_n^1 - f_n\left(Q_n^1\right) + \mathcal{A}_1\left(\mathcal{B}_n\left(f_n\left(Q_n^1\right)\right) + R_n \mathbf{1}_{\{f_n(Q_n^1)>0\}}\right), \\
&\qquad Q_n^2 + \mathcal{A}_2\left(\mathcal{B}_n\left(f_n\left(Q_n^1\right)\right) + R_n \mathbf{1}_{\{f_n(Q_n^1)>0\}}\right), \\
&\qquad \ldots, Q_n^K + \mathcal{A}_K\left(\mathcal{B}_n\left(f_n\left(Q_n^1\right)\right) + R_n \mathbf{1}_{\{f_n(Q_n^1)>0\}}\right)\Big) \\
&\le_{\mathrm{st}} \Big(Q_n'^1 - f_n\left(Q_n'^1\right) + \mathcal{A}_1'\left(\mathcal{B}_n'\left(f_n\left(Q_n'^1\right)\right) + R_n' \mathbf{1}_{\{f_n(Q_n'^1)>0\}}\right), \\
&\qquad Q_n'^2 + \mathcal{A}_2'\left(\mathcal{B}_n'\left(f_n\left(Q_n'^1\right)\right) + R_n' \mathbf{1}_{\{f_n(Q_n'^1)>0\}}\right), \\
&\qquad \ldots, Q_n'^K + \mathcal{A}_K'\left(\mathcal{B}_n'\left(f_n\left(Q_n'^1\right)\right) + R_n' \mathbf{1}_{\{f_n(Q_n'^1)>0\}}\right)\Big) \\
&=_{\mathrm{st}} \left(U_n'^1, U_n'^2, \ldots, U_n'^K\right).
\end{aligned}$$

Using in addition that

$$D_n \leq_{\mathrm{st}} D'_n,$$

we obtain

$$\begin{aligned}
&\left(Q^1_{n+1}, Q^2_{n+1}, \ldots, Q^K_{n+1}\right) \\
&=_{\mathrm{st}} \left(U^1_n + \mathcal{A}_1(D_n), U^2_n + \mathcal{A}_2(D_n), \ldots, U^K_n + \mathcal{A}_K(D_n)\right) \\
&\leq_{\mathrm{st}} \left(U'^1_n + \mathcal{A}'_1(D'_n), U'^2_n + \mathcal{A}'_2(D'_n), \ldots, U'^K_n + \mathcal{A}'_K(D'_n)\right) \\
&=_{\mathrm{st}} \left(Q'^1_{n+1}, Q'^2_{n+1}, \ldots, Q'^K_{n+1}\right),
\end{aligned}$$

and

$$\begin{aligned}
S_n &=_{\mathrm{st}} R_n \mathbf{1}_{\{f_n(Q^1_n)>0\}} + \mathcal{B}_n\left(f_n\left(Q^1_n\right)\right) + D_n \\
&\leq_{\mathrm{st}} R'_n \mathbf{1}_{\{f_n(Q'^1_n)>0\}} + \mathcal{B}'_n\left(f_n\left(Q'^1_n\right)\right) + D'_n \\
&=_{\mathrm{st}} S'_n.
\end{aligned}$$

The case where the service discipline in queue 1 is of exhaustive type is similar. It is easy to see (using e.g. sample path analysis) that the busy period of an $M/G/1$ queue is stochastically increasing in the arrival rate. It is shown in Jean-Marie and Liu [13] that it is also stochastically increasing in the services times. Thus,

$$\mathcal{G}_n\left(f_n\left(Q^1_n\right)\right) \leq_{\mathrm{st}} \mathcal{G}'_n\left(f_n\left(Q^1_n\right)\right) \leq_{\mathrm{st}} \mathcal{G}'_n\left(f_n\left(Q'^1_n\right)\right).$$

Therefore,

$$\begin{aligned}
&\left(U^1_n, U^2_n, \ldots, U^K_n\right) \\
&=_{\mathrm{st}} \Big(Q^1_n - f_n\left(Q^1_n\right) + \mathcal{A}_1\left(R_n \mathbf{1}_{\{f_n(Q^1_n)>0\}}\right), Q^2_n + \mathcal{A}_2\left(\mathcal{G}_n\left(f_n\left(Q^1_n\right)\right) + R_n \mathbf{1}_{\{f_n(Q^1_n)>0\}}\right), \\
&\quad \ldots, Q^K_n + \mathcal{A}_K\left(\mathcal{G}_n\left(f_n\left(Q^1_n\right)\right) + R_n \mathbf{1}_{\{f_n(Q^1_n)>0\}}\right)\Big) \\
&\leq_{\mathrm{st}} \Big(Q'^1_n - f_n\left(Q'^1_n\right) + \mathcal{A}'_1\left(R'_n \mathbf{1}_{\{f_n(Q'^1_n)>0\}}\right), Q'^2_n + \mathcal{A}'_2\left(\mathcal{G}'_n\left(f_n\left(Q'^1_n\right)\right) + R'_n \mathbf{1}_{\{f_n(Q'^1_n)>0\}}\right), \\
&\quad \ldots, Q'^K_n + \mathcal{A}'_K\left(\mathcal{G}'_n\left(f_n\left(Q'^1_n\right)\right) + R'_n \mathbf{1}_{\{f_n(Q'^1_n)>0\}}\right)\Big) \\
&=_{\mathrm{st}} \left(U'^1_n, U'^2_n, \ldots, U'^K_n\right),
\end{aligned}$$

which implies that

$$
\begin{aligned}
&\left(Q_{n+1}^1, Q_{n+1}^2, \ldots, Q_{n+1}^K\right) \\
&\quad =_{\mathrm{st}} \left(U_n^1 + \mathcal{A}_1(D_n), U_n^2 + \mathcal{A}_2(D_n), \ldots, U_n^K + \mathcal{A}_K(D_n)\right) \\
&\quad \leq_{\mathrm{st}} \left(U_n'^1 + \mathcal{A}_1'(D_n'), U_n'^2 + \mathcal{A}_2'(D_n'), \ldots, U_n'^K + \mathcal{A}_K'(D_n')\right) \\
&\quad =_{\mathrm{st}} \left(Q_{n+1}'^1, Q_{n+1}'^2, \ldots, Q_{n+1}'^K\right).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
S_n &=_{\mathrm{st}} R_n \mathbf{1}_{\{f_n(Q_n^1)>0\}} + \mathcal{G}_n\left(f_n\left(Q_n^1\right)\right) + D_n \\
&\leq_{\mathrm{st}} R_n' \mathbf{1}_{\{f_n(Q_n'^1)>0\}} + \mathcal{G}_n'\left(f_n\left(Q_n'^1\right)\right) + D_n' \\
&=_{\mathrm{st}} S_n'.
\end{aligned}
$$

Hence, by induction the relations (16)–(18) hold for all n.

Denote by $S_n(x_1, \ldots, x_{n-1})$ the random variable whose distribution is given by

$$P[S_n(x_1, \ldots, x_{n-1}) \leq x] = P[S_n \leq x \mid S_1 = x_1, \ldots, S_{n-1} = x_{n-1}].$$

Using the same argument, one can show by induction on n that

$$
\begin{aligned}
&S_n(x_1, \ldots, x_{n-1}) \leq_{\mathrm{st}} S_n(y_1, \ldots, y_{n-1}), && \forall (x_1, \ldots, x_{n-1}) \leq (y_1, \ldots, y_{n-1}), \\
&S_n(x_1, \ldots, x_{n-1}) \leq_{\mathrm{st}} S_n'(x_1, \ldots, x_{n-1}), && \forall (x_1, \ldots, x_{n-1}).
\end{aligned}
$$

Using theorem 4 of Veinott [27] (see also lemma 2 of Arjas and Lehtonen [1]) readily yields (13), which further implies relations (14) and (15). □

PROPOSITION 4.4

Suppose that there are more queues in Γ' than in Γ: $K' \geq K$. If

$$
\begin{aligned}
&\left(Q_1^1, \ldots, Q_1^K\right) \leq_{\mathrm{st}} \left(Q_1'^1, \ldots, Q_1'^K\right), \\
&(\lambda_1, \ldots, \lambda_K) \leq (\lambda_1', \ldots, \lambda_K'), \\
&B_k \leq_{\mathrm{st}} B_k', && 1 \leq k \leq K, \\
&D_{nK+k} \leq_{\mathrm{st}} D_{nK'+k}', && 1 \leq k \leq K, \quad n \geq 1, \\
&R_{nK+k} \leq_{\mathrm{st}} R_{nK'+k}', && 1 \leq k \leq K, \quad n \geq 1,
\end{aligned}
$$

then

$$\{Q_n^k,\ 1 \le k \le K\} \le_{\rm st} \{Q_n'^k,\ 1 \le k \le K\}, \qquad n \ge 1,$$

$$\{S_{iK+k},\ 1 \le i \le n,\ 1 \le k \le K\} \le_{\rm st} \{S'_{iK'+k},\ 1 \le i \le n,\ 1 \le k \le K\}, \qquad n \ge 1,$$

$$\{V_{iK+k},\ 1 \le i \le n,\ 1 \le k \le K\} \le_{\rm st} \{V'_{iK'+k},\ 1 \le i \le n,\ 1 \le k \le K\}, \qquad n \ge 1,$$

$$\{C_{iK+k},\ 1 \le i \le n,\ 1 \le k \le K\} \le_{\rm st} \{C'_{iK'+k},\ 1 \le i \le n,\ 1 \le k \le K\}, \qquad n \ge 1.$$

Proof

The proof is analogous to the one above, and is omitted. □

Remark

In case some of the queues have fully exhaustive service discipline, the stochastic monotonicity of propositions 4.3 and 4.4 still holds in view of eq. (8).

Remark

The stochastic monotonicity of the queue lengths, the cycle times, the intervisit times and the station times still holds in the steady state whenever there is a weak convergence of these quantities.

Remark

The assumption that the sequences of walking times and set-up times are i.i.d. random variables can be relaxed. In fact, the results in this section hold when these random variables are mutually independent.

5. Invariant quantities

In this section, we establish some invariant quantities in the steady state under fairly mild assumptions. In particular, we will not restrict ourselves to the model described in section 2. More specifically, we assume that the cycle times, the workload, as well as the interarrival times and the service times at every queue are stationary. The service disciplines are assumed to admit a steady state, but are otherwise arbitrary.

All the stochastic processes defined below are assumed to be stationary and ergodic, defined on the same probability space $(\Omega, \mathcal{F}, P, \theta_t)$, where θ_t is the shift operator (see e.g. Baccelli and Brémaud [4] for such a formalism). Let λ_k and μ_k be the arrival rate and the service rate of the customers in queue k, respectively, $1 \le k \le K$.

Let $N(t)$ be the point process that counts the number of arrivals of the *server* to queue 1 until time t. Let $\ldots < T_{-1} < T_0 \le 0 < T_1 < T_2 < \ldots$ be the corresponding time instants of the arrivals of the server to queue 1. Define the cycle times $C_n^1 = T_{n+1} - T_n$. Let P^0 be the Palm probability related to N and let E^0 be the expectation w.r.t. that measure. Denote by ν the intensity of $N(t)$. Note that $\nu^{-1} = E^0 C_0$. It is assumed that $\nu > 0$ and $EC_0 < \infty$.

It is immediate (using e.g. Neveu's exchange formula) to see that the expected number of arrivals to queue k during a cycle (w.r.t. P^0) is given by λ_k/v, $1 \le k \le K$.

Let $\{\sigma_n^k\}_{n=0}^\infty$ be the sequence of the times spent by the server at the nth visit to queue k, $1 \le k \le K$. Let $\{\delta_n^k\}_{n=0}^\infty$ be the sequence of the walking times spent by the server between the end of the nth service of queue k until it arrives to the next queue, $1 \le k \le K$. The sequence $\{\sigma_n^1, \delta_n^1, \sigma_n^2, \delta_n^2, \ldots, \sigma_n^K, \delta_n^K\}_{n=0}^\infty$ can thus be considered as a sequence of marks associated with the process $N(t)$. There are no set-up times. Nevertheless, the walking times can be dependent on the system state. Denote $d = E^0 \sum_{i=1}^N \delta_0^i$, $s_1 = E^0 \sigma_0^1$, $d_1 = E^0 \delta_0^1$.

Let $W_k(t)$ be the workload in queue k at time t. Let $\pi(t)$ refer to the position of the server at time t, i.e. if it is serving queue k at time t, then $\pi(t) = k$, otherwise $\pi(t) = 0$.

LEMMA 5.1

For all k, $1 \le k \le K$,

$$P(\pi(0) = k) = \rho_k.$$

Proof

We have

$$-\frac{\mathrm{d}^+}{\mathrm{d}t} W_k(t) = \begin{cases} 1, & \pi(t) = k, \\ 0, & \text{otherwise}, \end{cases}$$

where $\mathrm{d}^+/\mathrm{d}t$ stands for the right derivative. Hence,

$$P(\pi(0) = k) = P(\pi(t) = k) = -\frac{\mathrm{d}^+}{\mathrm{d}t} W_k(t).$$

By the Rate Conservation Principle [17], we have:

$$-E \frac{\mathrm{d}^+}{\mathrm{d}t} W_k(t) = \frac{\lambda_k}{\mu_k},$$

which completes the proof. □

The main result of this section is the following invariant expressions for the expected cycle times, the expected intervisit times and the expected station times.

PROPOSITION 5.2

In the steady state

(i) The expected cycle time C_0 (w.r.t. P^0) of queue 1 is given by

$$E^0C_0 = \frac{d}{1-\rho}; \tag{19}$$

(ii) the expected intervisit time V_0 (w.r.t. P^0) to queue 1 is given by

$$E^0V_0 = \frac{(1-\rho_1)d}{1-\rho}; \tag{20}$$

(iii) the expected station time S_0 (w.r.t. P^0) of queue 1 is given by

$$E^0S_0 = E^0(\sigma_0^1 + \delta_0^1) = \frac{\rho_1 d}{1-\rho} + d_1. \tag{21}$$

Proof

Define

$$H_k(t) = \sum_{n=-\infty}^{\infty} \sigma_n^k \mathbf{1}_{\{0 \le T_n < t\}}.$$

By Campbell's formula, we have:

$$E\,H_k(t) = \nu t E^0 \sigma_0^k. \tag{22}$$

Since

$$\int_0^t \mathbf{1}_{\{\pi(s)=k\}}\,\mathrm{d}s - \sigma_0^k \le H_k(t) \le \int_0^t \mathbf{1}_{\{\pi(s)=k\}}\,\mathrm{d}s + \sigma_0^k \circ \theta_t\,,$$

and due to the stationarity, we have

$$|E\,H_k(t) - P(\pi(0)=k)t| < E\left[\sigma_0^k\right] \le EC_0. \tag{23}$$

It follows from lemma 5.1 and from (22)–(23) that

$$|\nu E^0 \sigma_0^k - \rho_k| < \frac{EC_0}{t}.$$

Using the fact that EC_0 is finite and that the last inequality holds for all t, we obtain

$$E^0\sigma_0^k = \rho_k/\nu, \quad k = 1, \dots, K. \tag{24}$$

On the other hand, $\sum_{k=1}^{K}(\sigma_0^k + \delta_0^k) = C_0$ and hence it follows that

$$d = (1-\rho)E^0C_0.$$

The result (ii) follows from (19) and (24) by noting that the intervisit time at queue 1 is given by

$$V_0 = \sum_{k=1}^{K}\left(\sigma_0^k + \delta_o^k\right) - \sigma_0^1 = C_0 - \sigma_0^1.$$

The result (iii) is an immediate consequence of (19) and (24). □

6. Extensions

The results obtained in sections 3 and 4 on the stability conditions and stochastic monotonicities can readily be extended to the case where the customers, at the end of service, may be routed to other queues in the system or rejoin the same queue for another portion of service. Polling systems with such routing mechanisms are analyzed by Sidi et al. [21] under the (purely) gated and (fully) exhaustive service disciplines. In our general system, we further assume that the external arrivals of the customers are controlled by a single Poisson process. At an arrival epoch, a random number of customers are sent to each of the queues.

In such a system, customers arrive to the system in bulk according to a Poisson process with parameter λ. At the nth arrival epoch, η_n^k customers are sent to queue k, $1 \le k \le K$, $n = 1, 2, \ldots$. For all $1 \le k \le K$, the sequence $\{\eta_n^k\}_n$ is composed of non-negative integer i.i.d. random variables with mean q_k. Let η^k denote the generic version of η_n^k, $n = 1, 2, \ldots$. Note that our initial polling model corresponds to the case where $\sum_{k=1}^{K} \eta_n^k = 1$ for all $n = 1, 2, \ldots$.

When the service of a customer is finished at queue k, $1 \le k \le K$, it joins queue j with probability $p_{k,j} \ge 0$, $1 \le j \le K$, and leaves the system with probability $p_{k,0} = 1 - \sum_{j=1}^{K} p_{k,j} \ge 0$. Assume that for all k, $1 \le k \le K$, there are some indices $1 \le k_1, k_2, \ldots, k_i \le K$, such that

$$p_{k,k_1} p_{k_1,k_2} \cdots p_{k_{i-1},k_i} p_{k_i,0} > 0.$$

In other words, all customers will eventually leave the system.

The service times at queue k are i.i.d. random variables with mean b_k, $1 \le k \le K$. Let $\bar{b}_k$ be the mean length of total service times (before leaving the system) of a customer arriving at queue k:

$$\bar{b}_k = b_k + \sum_{j=1}^{K} p_{k,j} \bar{b}_j. \tag{25}$$

Denote by $\lambda_k = \lambda q_k$, $\rho_k = \lambda_k b_k$, $1 \le k \le K$, $\rho = \sum_{k=1}^{K} \lambda_k \bar{b}_k$. Note that $\rho \ge \sum_{k=1}^{K} \rho_k$. Let $h_k = 1/(1 - \rho_k J_k)$.

Let $\mathcal{L}_{j,k}(l)$ be the random variable of the number of customers routed to queue k out of l customers leaving queue j, $1 \le j, k \le K$. Let $\mathcal{M}_n(l)$ be the random variable of the number of customers served by the server at its nth visit in order to decrease the length of the visited queue by l, $n \geq 1$. Note that $\mathcal{M}_n(l)$ is distributed as the number of customers served in order to empty an $M^X/G/1$ queue from the instant a service begins and l customers are present, with the service time $B_{I(n)}$, the arrival rate λ and the bulk size η^k.

As in the original polling model, let $\mathcal{A}_k(T)$ denote the number of customers arrived at queue k, $1 \leq k \leq K$, during a (possibly random) time interval of length T. Let $\mathcal{G}_n(l)$ be the time spent by the server at its nth visit in order to decrease the length of the visited queue by l, $n \geq 1$. Thus, $\mathcal{G}_n(l)$ is distributed as the time it takes to empty an $M^X/G/1$ queue from the instant a service begins and l customers are present, with the service time $B_{I(n)}$, the arrival rate λ and the bulk size η^k.

It is easy to check that the following evolution equation holds:

$$Q_{n+1}^j =_{st} \begin{cases} Q_n^j + \mathcal{A}_j\Big(R_n 1_{\{f_n(Q_n^{I(n)})>0\}} + \bar{J}_{I(n)}\mathcal{B}_n\big(f_n(Q_n^{I(n)})\big) + J_{I(n)}\mathcal{G}_n\big(f_n(Q_n^{I(n)})\big) + D_n\Big) \\ \quad + \mathcal{L}_{I(n),j}\Big(\bar{J}_{I(n)}f_n\big(Q_n^{I(n)}\big) + J_{I(n)}\mathcal{M}_n\big(f_n\big(Q_n^{I(n)}\big)\big)\Big), & I(n) \neq j; \\ Q_n^j - f_n\big(Q_n^j\big) + \mathcal{A}_j\Big(R_n 1_{\{f_n(Q_n^j)>0\}} + \bar{J}_{I(n)}\mathcal{B}_n\big(f_n\big(Q_n^j\big)\big) + D_n\Big) \\ \quad + \mathcal{L}_{j,j}\Big(\bar{J}_j f_n\big(Q_n^j\big) + J_j\mathcal{M}_n\big(f_n\big(Q_n^j\big)\big)\Big), & I(n) = j. \end{cases} \tag{26}$$

Taking a conditional expectation in (26) and summing over j, we obtain:

$$\begin{aligned} E\left[\sum_{j=1}^K \bar{b}_j Q_{n+1}^j | Q_n\right] &\leq \sum_{j=1}^K \bar{b}_j Q_n^j + \rho d_{I(n)} + \rho r_{I(n)} + \bar{J}_{I(n)}(\rho - 1) b_{I(n)} F_n\big(Q_n^{I(n)}\big) \\ &\quad + J_{I(n)}\left(\frac{\rho - \rho_{I(n)}}{1 - \rho_{I(n)}} - 1\right) b_{I(n)} F_n\big(Q_n^{I(n)}\big) \\ &= \sum_{j=1}^K \bar{b}_j Q_n^j + \rho d_{I(n)} + \rho r_{I(n)} + h_{I(n)}(\rho - 1) b_{I(n)} F_n\big(Q_n^{I(n)}\big), \end{aligned}$$

where we used relation (25) and the fact that the expected time to decrease the length of queue $I(n)$ by 1 from the instant a service starts is $b_{I(n)}/(1 - \rho_{I(n)})$, and that the expected number of service completions during this period is $1/(1 - \rho_{I(n)})$.

Therefore, as in the proof of proposition 3.2, we obtain

$$E\left[\sum_{j=1}^{K}\bar{b}_j Q_{n+K}^j \,|\, Q_n\right]$$
$$\leq \sum_{j=1}^{K}\bar{b}_j Q_n^j + \rho d + \rho r + \sum_{j=0}^{K-1} h_{I(n+j)}(\rho - 1) b_{I(n+j)} F_{n+j}\left(Q_n^{I(n+j)}\right). \tag{27}$$

Consequently, all the results in section 3 hold and can be proved by the Lyapunov function $\bar{L}(x) = \sum_{j=1}^{K}\bar{b}_j x_j$ in place of $L(x) = \sum_{j=1}^{K} b_j x_j$.

Consider now the stochastic monotonicity. Denote by Γ the polling system described above. Let there be another polling system Γ' with K' queues. In the systems Γ and Γ', the number of customers to be served at each visit of the server is characterized by $f_n(x)$ and $f_n'(x)$, respectively. We assume that $f_{nK+k}(x) =_{st} f'_{nK'+k}(x)$ for all $n \geq 1, x \geq 0$, and $1 \leq k \leq \min(K, K')$. We also assume that the service discipline is stochastically increasing and contractive.

Analogously to the proof of proposition 4.3, we can readily show, by using the evolution eq. (26), that the stochastic comparison results of propositions 4.3 and 4.4 still hold in our generalized polling system if

$$p_{j,k} \leq p'_{j,k}, \quad 1 \leq j,k \leq K,$$

and if the condition

$$(\lambda_1, \ldots, \lambda_K) \leq (\lambda_1', \ldots, \lambda_K')$$

is replaced by

$$\eta^k \leq_{st} \eta'^k, \quad 1 \leq k \leq K.$$

Acknowledgements

The authors are grateful to Dr. Flora Spieksma for her constructive comments on the stability conditions presented in section 3, and to Professor Moshe Sidi who suggested the way to generalize the stability conditions to polling systems with routing as discussed in section 6.

References

[1] E. Arjas and T. Lehtonen, Approximating many queues by means of single server queues, Math. Oper. Res. 3(1978)205–223.

[2] S. Asmussen, *Applied Probability and Queues* (Wiley, 1987).

[3] P. Billingsley, *Convergence of Probability Measures* (Wiley, 1968).

[4] F. Baccelli and P. Brémaud, *Palm Probabilities and Stationary Queues*, Lecture Notes in Statistics (Springer, 1980).

[5] O.J. Boxma, Workloads and waiting times in single-server systems with multiple customer classes, Queueing Systems 5(1989)185–214.
[6] O.J. Boxma and W.P. Groenendijk, Pseudo-conservation laws in cyclic-service systems, J. Appl. Prob. 24(1987)949–964.
[7] O.J. Boxma, H. Levy and U. Yechiali, Cyclic reservation schemes for efficient operation of multiple-queue single-server systems, Ann. Oper. Res. 35(1992)187–208.
[8] G.L. Choudhury and H. Takagi, Comments on "Exact results for nonsymmetric token ring systems", IEEE Trans. Commun. COM-38(1990)1125–1127.
[9] C.F. Daganzo, Some properties of polling systems, Queueing Systems 6(1990)137–154.
[10] M. Eisenberg, Queues with periodic service and changeover time, Oper. Res. 20(1972)440–451.
[11] L. Georgiadis and W. Szpankowski, Stability of token passing rings, Queueing Systems 11(1991), this issue.
[12] S. Grigorescu and G. Oprişan, Limit theorems for J–X processes with general state space, Z. Wahrscheinlichkeitstheorie verw. Gebiete 35(1976)65–73.
[13] A. Jean-Marie and Z. Liu, Stochastic orderings on partial sums of random variables and on counting measures of random intervals, INRIA Research Report No. 1422 (April, 1991), Adv. Appl. Prob., to appear.
[14] P. Kuehn, Multiqueue systems with nonexhaustive cyclic service, Bell. Syst. Tech. J. 58(1979) 671–698.
[15] H. Levy, Analysis of binomial-gated serice, *Proc. Performance of Distributed and Parallel Systems*, ed. T. Hasegawa, H. Takagi and Y. Takahashi (North-Holland, 1989), pp. 127–139.
[16] H. Levy, M. Sidi and O.J. Boxma, Dominance relations in polling systems, Queueing Systems 6(1990)155–172.
[17] M. Miyazawa, The derivation of invariance relations in complex queueing systems with stationary inputs, Adv. Appl. Prob. 22(1983)874–885.
[18] G.I. O'Brien, Limit theorems for sums of chain-dependent processes, J. Appl. Prob. 11(1974) 582–587.
[19] N.N. Popov, Conditions for geometric ergodicity of countable Markov chains, Sov. Math. Dokl. 18(1977)676–679.
[20] L.D. Servi, Average delay approximation of the *M/G/*1 cyclic service queues, IEEE Trans. Sel. Areas Commun. SAC-4(1986)813–822. Correction in SAC-5(1987)547.
[21] M. Sidi, H. Levy and S.W. Fuhrmann, A queueing network with a single cyclically roving server, Queueing Systems 11(1992), this issue.
[22] F.M. Spieksma, Geometrically ergodic Markov chains and the optimal control of queue, Ph.D. Thesis, University of Leiden.
[23] V. Strassen, The existence of probability measures with given marginals, Ann. Math. Statist. 36(1965) 423–439.
[24] H. Takagi, *Analysis of Polling Systems* (MIT Press, 1986).
[25] H. Takagi, Queueing analysis of polling models: An update, in: *Stochastic Analysis of Computer and Communications Systems*, ed. H. Takagi (Elsevier, 1990), pp. 267–318.
[26] Tedijanto, Exact results for the cyclic service queue with a Bernoulli schedule, Report, SRC, University of Maryland (1988).
[27] A.F. Veinott, Jr., Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes, Oper. Res. 13(1965)761–778.
[28] V.S. Zhdanov and E.A. Saksonov, Conditions of existence of steady-state modes in cyclic queueing systems, Autom. Remote Control 40(1979)176–184.