Stability of polling systems with exhaustive service policies and state dependent routing

Serguei Foss*and Günter Last

Novosibirsk State University and

Technical University of Braunschweig

Abstract

We consider a polling system with a finite number of stations fed by compound Poisson arrival streams of customers asking for service. A server travels through the system and upon arrival at a station the server serves all waiting customers until the queue is empty, where the service time distribution depends on the station. The choice of the station to be visited next as well as the corresponding walking time may depend on the whole current state. Examples are systems with a greedy-type routing mechanism. Under appropriate independence assumptions it is proved that the system is stable if and only if the workload is less than one.

POLLING SYSTEM; STABILITY; ERGODICITY OF MARKOV CHAINS; GREEDY SERVER

AMS 1991 SUBJECT CLASSIFICATIONS: PRIMARY 60K25, SECONDARY 60J27

1

1 Introduction

Consider a server who visits (*polls*) the stations of a queueing network. The stations are numbered 1 through K and with each of them there is associated a queue with infinite waiting capacity fed with an arrival stream of customers with intensity λ_i, $i = 1, \ldots, K$. The process of all arrival instants is assumed to be homogeneous Poisson. At a given arrival instant however, all stations may simultaneously receive a group of customers. The joint distribution of these groups should render the expected group sizes to be positive and finite. The server employs the so-called *exhaustive* service policy at each station. This means that, upon arriving at a non-empty station, he will provide service there until the moment when the station becomes empty, which includes service for all those customers who may arrive during the service of the customers present at the time of the server's arrival. The services are independent of the arrival stream and each served customer departs from the system. The service times at station i are i.i.d. and are assumed to have a finite positive mean b_i. When the server has finished the batch of services at station i or if he has found that station empty, then he walks to another station j, say, taking a walking time that might also be zero. The choice of this station and the associated walking time depend on the whole current state of the system and may even depend on future arrivals. An example is the *greedy* routing mechanism, where the server chooses a station with the maximum number of customers in a certain neighborhood waiting at the start of the walk.

The subject of the present paper is to establish a *stability* condition for the queueing network ensuring that the server can handle all the work arriving to the network. Mathematically this amounts to prove *ergodicity* of the underlying Markov chain. Under appropriate independence assumptions it will be shown that the system is stable if

and only if

$$\sum_{i=1}^{K} \lambda_i b_i < 1. \tag{1.1}$$

It is worth mentioning that our stability results can be proved under more general i.i.d. assumptions on the input using the same ideas. We have decided to work under Poisson assumption in order to avoid technicalities which could obscure the main arguments.

A special feature of the present model is that the server's decision which station to serve next depends on the actual configuration of customers in the system. In Coffman and Gilbert (1987) one can find a discussion of a polling system with a greedy server on a circle and a line as well as a conjecture about the stability condition. The only earlier paper we are aware of, however, solving the stability problem for polling systems with state-dependent routing is Schassberger (1993), where the ergodicity of a symmetric ringlike network with a limited service policy has been proved. These results are generalized in Foss and Last (1995) dealing with polling systems with a special greedy routing mechanism on a graph but with rather general service policies for each station. In this work conditions sufficient for stability and sufficient for instability are presented. These conditions coincide only in special (symmetric) cases and it is shown by examples that the determination of the exact stability region might be a difficult task. There are many other papers establishing comparison and stability results for polling systems with state-independent routing. We refer here to Levy, Sidi and Boxma (1990), Georgiadis and Szpankowski (1992), Borovkov and Schassberger (1994), Fricker and Jaibi (1994). Massoulié (1995) constructed a stationary regime for polling systems with stationary ergodic arrival process and with state-independent routing mechanism. Thereafter, for a more general model, Foss and Chernova (1995) proved the stability results using the saturation rule of Baccelli and Foss (1995). In

3

this context one should also mention the papers by Kroese and Schmidt (1992, 1994) proving the stability of a polling model on the circle and by Altman and Levy (1994) on polling systems in space. For applications of polling models we refer to Coffman and Gilbert (1987) and Takagi (1990) who gives a survey of the extensive literature.

The idea of the stability proof in this paper is to construct, by induction on the number of stations, a stopping time at which a certain linear test function satisfies a multiplicative drift condition. Ergodicity is then obtained from a general stability result on Markov chains which is a 'randomized' version of stability criteria by Malyshev and Menshikov (1982), see also Meyn and Tweedie (1993, 1994), Borovkov (1994), Dai (1995) and Fayolle, Malyshev and Men'shikov (1995).

The paper is organized as follows. Section 2 gives the extensive and exact description of the model and introduces some notation. Section 3 presents the proof of the stability result. An appendix contains some general results for discrete time Markov processes needed in the course of our paper.

2 Model description

We consider a queueing system consisting of K stations with infinite waiting capacities. With each station there is associated a queue of customers waiting for service. One server is traveling through the system. Upon arrival of the server at a station the corresponding queue is served until it is empty, including customers arriving after the server's arrival instant. Each service takes a random time and each served customer leaves the system.

We let $X_i(t)$ denote the number of customers in the ith queue at time $t \in \mathbb{R}_+$. Further we denote by $S(t)$ the number of the station which is occupied by the server while

4

serving is in progress and we let $S(t) := 0$, otherwise. Both, $X_i(t)$ and $S(t)$ are taken to be right-continuous. and we can assume that there exist the limits from the left, denoted by $X_i(t-)$ and $S(t-)$.

Let T_n, $n \in \mathbb{N}$, be the time of nth service completion and T^{n+1} the time of the beginning of the next service after time T_n. If $X_{S(T_n-)}(T_n) > 0$, then $T_n = T^{n+1}$. If $X_{S(T_n-)}(T_n) = 0$, then

$$W_{n+1} := T^{n+1} - T_n,$$

is the *walking time* taken by the server to travel from station $S(T_n-)$ to $S(T^{n+1}) = S(T_{n+1}-)$. The cases $W_{n+1} = 0$ and $S(T_n-) = S(T^{n+1})$ are not excluded. It should be kept in mind that the server may have visited several empty stations before he starts serving queue $S(T^{n+1})$. Hence in certain cases the time W_{n+1} can be regarded as a sum of walking times. If the system is empty at time T_n then a part of W_{n+1} is in fact used to wait for the next arriving customer. Hence W_{n+1} has to be positive in that case. We refer here to the model description below as well as to Examples 2.1 and 2.2.

The *initial conditions* are given by a random element $X(0) = (X_1(0), \ldots, X_K(0), S_0)$ of $\mathbb{Z}_+^K \times \{1, \ldots, K\}$. If $X_{S_0}(0) = 0$ then the server starts walking as described above. In this case $T^1 = W_1$ is defined as the time of the begin of the first service, and we set $S(0-) := S_0$ and $S(t) = 0$ for $0 \le t < T^1$. (If $T^1 = 0$ we have $S(0) = S(T^1)$). If $X_{S_0}(0) > 0$ then the server starts serving until the queue $S(0-) = S(0) := S_0$ is empty. In this case we define $W_1 = T^1 := 0$.

In order to explain how the system is operating we will now describe the arrival processes, the services and the assumptions on the routing and walking times. The underlying probability space is denoted by $(\Omega, \mathcal{F}, P)$. Define

$$\mathcal{F}_t := \sigma(X(s) : s \le t), \quad t \in \mathbb{R}_+,$$

where $X(t) := (X_1(t), \ldots, X_K(t), S(t))$. Then $\{\mathcal{F}_t : t \in \mathbb{R}_+\}$ is a right-continuous filtration describing the internal history of the process. We let $A_i(t)$ denote the number of customers who have arrived at station i by time t. Hence

$$A_i(s, t] := A_i(t) - A_i(s), \quad s \leq t,$$

is the number of customers who arrived at station i during the time interval $(s, t]$. We assume that the arrival instants are different from the times of completion of services so that

$$A_i(t) = \sum_{s : X_i(s) > X_i(s-)} \mathbf{1}\{s \leq t\}(X_i(s) - X_i(s-)), \quad t \in \mathbb{R}_+,$$

is $\mathcal{F}_t$-measurable. We also note that

$$T_{n+1} = \min\{t \geq T_n : \sum_{i=1}^K X_i(t) < \sum_{i=1}^K X_i(t-)\}, \quad n \in \mathbb{Z}_+,$$

where $T_0 := 0$, and

$$T^{n+1} = \inf\{t \geq T_n : S(t) \neq 0\}, \quad n \in \mathbb{Z}_+.$$

Hence T_n and T^n, $n \in \mathbb{N}$, are $\{\mathcal{F}_t\}$-stopping times. If T is any $\{\mathcal{F}_t\}$-stopping time then, as usual,

$$\mathcal{F}_T := \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t \in \mathbb{R}_+\},$$

denotes the history at time T, where $\mathcal{F}_\infty := \sigma(\bigcup_{s>0} \mathcal{F}_s)$.

The sequence (τ_n) of all arrival instants is assumed to form a homogeneous Poisson process with intensity λ. Further there is given a sequence $(B_n) = ((B_n^1, \ldots, B_n^K))$ of independent random elements of $\mathbb{Z}_+^K$ which is independent of (τ_n) and such that $a_i := EB_1^i$ is positive and finite for all i and $\sum_i B_n^i > 0$. The arrival process $A_i(t)$ is then assumed to be given by

$$A_i(t) := \sum_{n \geq 1} \mathbf{1}\{\tau_n \leq t\} B_n^i.$$

6

It is a compound Poisson process with intensity $\lambda_i := \lambda a_i$. Note that the $\{A_i(t)\}$ are independent if and only if the components of B_1 are independent. We will assume more generally that $((\tau_n, B_n))$ is a marked $\{\mathcal{F}_t\}$-Poisson process, see e.g. Last and Brandt (1995). That is to say that

$$\sum_{n \geq 1} \mathbf{1}\{s < \tau_n \leq t, B_n \in C\}$$

is independent of $\mathcal{F}_s$ for all $s < t$ and all measurable $C \subseteq \mathbb{Z}_+^K$. In particular, $A_i(t, v]$ is independent of $\mathcal{F}_t$, for all $t < v$. This property generalizes to an $\{\mathcal{F}_t\}$-stopping time T, i.e. $A_i(T, v]$ is independent of $\mathcal{F}_T$ for all v, where $A_i((T, v]) = 0$ if $v \leq T$. Also,

$$EA_i(T) = \lambda_i ET, \quad i \in \{1, \ldots, K\},$$

for any $\{\mathcal{F}_t\}$-stopping time T. More generally, if S is another $\{\mathcal{F}_t\}$-stopping time satisfying $S \leq T$ then

$$E[A_i(S, T] | \mathcal{F}_S] = \lambda_i E[T - S | \mathcal{F}_S] \quad P - \text{a.s.}, \quad i \in \{1, \ldots, K\}. \tag{2.1}$$

We assume that service times are independent of the arrival process and that

$$P(T_n - T^n \in \cdot | \mathcal{F}_{T^n}) = G_i(\cdot) \quad P - \text{a.s. on } \{S(T^n) = i\}, \quad n \in \mathbb{N}, \tag{2.2}$$

where $G_1, \ldots, G_K$ are distributions on $(0, \infty)$ with finite and positive means $b_1, \ldots, b_K$. In particular, different service times are independent.

Assume that $S(T_n-) = i$, $n \in \mathbb{N}$, and $X_i(T_n) = 0$. Then the server stops serving station i and the station $S(T^{n+1}) = j$ to be served next satisfies $X_j(T^{n+1}) > 0$. The choice of the station j and the random variable W_{n+1} may depend on the whole actual state of the system and even on the arrivals after time T_n and the corresponding conditional distributions might be quite general, see Examples 2.1 and 2.2. However, we assume

that there is a finite constant w such that

$$E[W_1|\mathcal{F}_0] \leq w \quad P - \text{a.s. on } \{X_{S_0}(0) = 0\}, \tag{2.3}$$

as well as a constant $p > 0$ satisfying

$$P(A(W_1) = 0|\mathcal{F}_0) > p \quad P - \text{a.s. on } \{X_{S_0}(0) = 0, \sum_k X_k(0) > 0\}, \tag{2.4}$$

where

$$A(t) := \sum_i A_i(t).$$

The latter condition excludes some strange behaviour of the server and seems to be satisfied in all relevant examples. It states that given the server has just completed a batch of services and the system is non-empty, then the probability that the next batch of services starts at one of these non-empty stations is uniformly bounded from below by a positive constant.

Our assumptions imply that the T_n are almost surely finite and satisfy $\lim_{n\to\infty} T_n = \infty$ and hence we assume these properties to hold everywhere on Ω. Denote

$$\hat{S}(n) := S(T_n-), \quad n \in \mathbb{Z}_+,$$

where we recall that $T_0 = 0$, see also the discussion of the initial values above. The process

$$\hat{X}(n) := (X_1(T_n), \ldots, X_K(T_n), \hat{S}(n)), \quad \hat{\mathcal{F}}_n := \mathcal{F}_{T_n}, \quad n \in \mathbb{Z}_+, \tag{2.5}$$

is assumed to be a homogeneous Markov chain, where we dispense with making this assumption more explicit. This could be done by introducing appropriate kernels describing the conditional distribution of $(W_{n+1}, X_1(T^{n+1}), \ldots, X_K(T^{n+1}), S(T^{n+1}))$ given $\mathcal{F}_{T_n}$ and $X_{S(T_n-)}(T_n) = 0$. Because we do not need these kernels later on we prefer to illustrate our model by examples.

Example 2.1 Assume that for each $i \in \{1, \ldots, K\}$ there is a set $N(i) \subseteq \{1, \ldots, K\}$ of *neighbors* of i such that $i \in N(i)$. Assume that for all $i, j \in \{1, \ldots, K\}$ there is a sequence $k_1, \ldots, k_r \in \{1, \ldots, K\}$ such that $k_1 = i$, $k_r = j$ and $k_{m+1} \in N(k_m)$ for $1 \le m \le r - 1$. This means that the neighborhood relation equips $\{1, \ldots K\}$ with the structure of a directed and connected graph. Assume that $S(T_n-) = i$, $n \in \mathbb{N}$, and $X_i(T_n) = 0$. Then the server stops serving station i and selects another station $S(T^{n+1}) = j$, say, in such a way that $j \in N(i)$ and

$$X_j(T_n) > 0 \quad \text{if} \sum_{k \in N(i)} X_k(T_n) > 0,$$

which means that the server walks to one of the non-empty stations in his neighborhood. This a *greedy-type* routing mechanism. To describe the dynamics of this system more accurately we assume that for any $x = (x_1, \ldots, x_K, i) \in \mathbb{Z}_+^K \times \{1, \ldots, K\}$ there is a set $D_x \subseteq C_x := \{j \in N(i) : x_j > 0\}$ such that $D_x \neq \emptyset$ whenever $C_x \neq \emptyset$. If $X(0) = x$ and $C_x \neq \emptyset$, then

$$P(W_1 \in dw, X(W_1) = y_1, \ldots, X(W_K) = y_K, S(W_1) = j | \mathcal{F}_0)$$
$$= F'_{ij}(dw) P((x_1 + A_1(w), \ldots, x_K + A_K(w)) = (y_1, \ldots, y_K)) \mathbf{1}\{j \in D_x\} \frac{1}{\text{card } D_x},$$

where F'_{ij} is a distribution with a finite mean w_{ij}. Assume now that the system starts at time 0 in a state $x = (x_1, \ldots, x_K, i)$ satisfying $\sum_{k \in N(i)} x_k = 0$. Then the server chooses a station j, say, randomly among the neighbors of i. Again it takes the server a random time W^1 with distribution $F_{ij}(x)$ to walk from i to j. If $X_j(W^1) > 0$ then we put $W_1 := W^1$. If not then the server chooses another station randomly in the set $A(j)$ of all stations $k \in N(j)$ satisfying $X_k(W^1) > 0$, if there is such a station. If $\sum_{k \in N(j)} X_k(W^1) = 0$ then $A(j) := N(j)$. The server continues in this way until he arrives at a non-empty station. Assume that $\inf_x \sum_{i,j} w_{ij}(x) > 0$, where $w_{ij}(x)$ is the mean of $F_{ij}(x)$. Under

9

obvious independence assumptions it takes the server a finite random time W_1 to arrive at a non-empty station. The process satisfies (2.3) and (2.4). We conclude this example with specifying possible choices of the sets D_x.

(i) Assume that

$$D_x = \{j \in N(i) : x_j = \max\{x_k : k \in N(i)\}\}.$$

This defines the so-called *greedy* walking mechanism.

(ii) Assume that there are positive and finite numbers $d(i,j)$ indicating a distance between $j \in N(i)$ and i and that

$$D_x = \{j \in N(i) : d(i,j) = \min\{d(i,k) : k \in N(i), x_k > 0\}\},$$

where $\min \emptyset := 0$. In this case the server tries to walk to the nearest non-empty station in his neighborhood.

(iii) $D_x = \{j \in N(i) : x_j > 0\}$.

(iv) $D_x = \{j \in N(i) : x_j = \min\{x_k : x_k > 0\}\}$.

(v) Let $0 < c < d$ and assume that $x = (x_1, \ldots, x_k, i)$ satisfies $\sum_{k \in N(i)} x_k > 0$. Let

$$D_x = \{j \in N(i) : c \le x_j \le d\}$$

if this set is non-empty and let D_x be given as in (i) otherwise.

We give now another example for a routing mechanism which was considered by Schassberger (1993) in the special case of a ringlike graph. In that case the polling system could be considered as an approximation of the greedy server on the circle.

Example 2.2 Consider the situation of Example 2.1(ii). An r-tuple $p = (k_1, \ldots, k_r) \in \{1, \ldots, K\}^r$ is called a path if $k_{m+1} \in N(k_m)$ for $1 \leq m \leq r-1$ and if all k_m are distinct. The length $d(p)$ of such a path is defined by $d(k_1, k_2) + \ldots + d(k_{m-1}, k_m)$. Assume that the system starts at time 0 in state $x = (x_1, \ldots, x_K, i)$ satisfying $\sum_k x_k > 0$. Consider the set of all paths p starting in i and ending in a station k with $x_k > 0$ and let H_x be the subset where $d(p)$ becomes minimal. The server chooses then the next station j, say, randomly in the set

$$D_x' := \{l \in N(i) : \text{there is a } p = (k_1, \ldots, k_r) \in D_x \text{ with } l = k_2\}.$$

As in Example 2.1 it takes the server a random time W^1 with some distribution F_{ij}' to walk from i to j. If $X_j(W^1) > 0$ then we put $W_1 := W^1$. If not then the server chooses another station randomly in the set $D_{X(W_1)}'$. The server continues in this way until he arrives at a non-empty station at a time W_1. If $\sum_k x_k = 0$ then the server moves as in the previous example. Once the server has completed a walk at a moment where the system is non-empty he continues walking as desribed above. It is easy to see that (2.3) and (2.4) hold. This example could be generalized by allowing the server to change his destination during a walk if there occurs an arrival at a station that is closer than that he is currently walking to. In this case the choice of the next destination depends again on future arrivals.

State-independent routings as in Fricker and Jaibi (1994) are also covered by our model. A special case is a Markovian routing as described next.

Example 2.3 Let $(r_{i,j})_{1 \leq i,j \leq K}$ be an irreducible Markovian (routing) matrix and assume that the server, upon completing a batch of services at station i chooses his next destination j, say, with probability $r_{i,j}$ independent of everything else. It takes the server a

11

random time with distribution F'_{ij} to walk from i to j. These walking times are independent of each other and independent of anything else. If at the moment of the server's arrival the jth queue is non-empty then it is served according to the exhaustive policy. Otherwise the server chooses a next destination according to the routing matrix. Assuming that the F'_{ij} have finite means w_{ij} satisfying $\sum_{ij} w_{ij} > 0$, both conditions (2.3) and (2.4) are fulfilled.

We conclude with a simple example where the only non-zero walking times are times taken to wait for the next arriving customer. Such a system is very similar to an M/GI/1–queueing system with batch arrivals. In particular, it is rather easy to see that stability is implied by inequality (1.1).

Example 2.4 Assume that after a batch of services the system is not empty. Then the server chooses one of the non-empty stations and resumes servicing instantaneously. If, on the other hand, the system is empty after a batch of services then the server waits for the next batch of arriving customers. Then he chooses one of the non-empty stations and starts serving without further delay.

3   Proof of the stability result

We consider the Markov chain $\{\hat{X}(n)\}$ defined by (2.5). All assumptions formulated in the previous section are supposed to be in force. The main result of our paper is the following theorem.

Theorem 3.1   (i) *Assume that* (1.1) *holds. Then there is a finite subset of* $\mathbb{Z}_+^K \times \{1, \ldots, K\}$ *which is positive recurrent for the Markov chain* $\{\hat{X}(n)\}$. *If the chain*

is irreducible, then it is positive recurrent and if it is in addition also aperiodic,
then it is ergodic.

(ii) *If the chain $\{\hat{X}(n)\}$ is ergodic, then* (1.1) *holds.*

Remark 3.2 In the Examples 2.1–2.4 the Markov chain $\{\hat{X}(n)\}$ is irreducible and aperiodic. In general these properties are not implied by our model assumptions and must be checked in each case.

The proof of the theorem is splitted into several lemmas. The basic idea is to prove that the number of walks can be neglected when compared with the number of services. A polling system where the server needs no time to travel to a non-empty station as in Example 2.4 can easily be proved to be stable if (1.1) is satisfied.

Unless stated otherwise we consider in this section a more general model, where the stochastic behaviour of the system is influenced by a further piecewise constant process $\{U(t) : t \geq 0\}$ taking values in some measurable space $(\mathbf{U}, \mathcal{U})$ and being right-continuous w.r.t. discrete topology. We let

$$X(t) := (X_1(t), \ldots, X_K(t), S(t), U(t))$$

and define the filtration $\{\mathcal{F}_t\}$ as before, where we now use the new process $X(t)$. We apply the assumptions on the arrival process and the services of section 2 verbatim and we also assume that

$$\hat{X}(n) := (X_1(T_n), \ldots, X_K(T_n), \hat{S}(n), U(T_n)), \tag{3.1}$$

is a homogeneous Markov process. We can assume that it fits the general setting of the Appendix. Our assumptions (2.3) and (2.4) remain unchanged, but we note that $\mathcal{F}_0$ contains also the information induced by $U(0)$.

Define $\nu_0 := 0$ and

$$\nu_{n+1} := \min\{k > \nu_n : \hat{X}_{\hat{S}(k)}(k) = 0\}, \quad n \in \mathbb{Z}_+.$$

If $\hat{X}_{\hat{S}(0)} = 0$ (resp. > 0) then ν_n, $n \in \mathbb{N}$, is the number of completed services before the server starts his $(n+1)$st (resp. nth) walk. Clearly T_{ν_n} is an $\{\mathcal{F}_t\}$-stopping time. Most of our analysis will be based on the Markov chain $(Y(n), \mathcal{G}_n)$, $n \in \mathbb{Z}_+$, which is defined by

$$Y(n) =: (Y_1(n), \ldots, Y_K(n), S_n, U_n) := \hat{X}(\nu_n) = X(T_{\nu_n}), \quad \mathcal{G}_n := \mathcal{F}_{T_{\nu_n}}.$$

Note that $(Y(0), \mathcal{G}_0) = (X(0), \mathcal{F}_0)$.

We define a function V on $\mathbb{Z}_+^K \times \{1, \ldots, K\} \times \mathbf{U}$ by

$$V(x) := b_1 x_1 + \ldots + b_K x_K, \quad x = (x_1, \ldots, x_K, i, u). \tag{3.2}$$

Further we define $|x| := x_1 + \ldots + x_K$, where x is as above.

Lemma 3.3 *Assume* (1.1). *There are constants* $d_1, d_2, d_1', d_2', \tilde{d}_1, \tilde{d}_2$ *such that*

$$E[T_{\nu_1}|\mathcal{G}_0] \leq d_1'|Y(0)| + d_2', \tag{3.3}$$

$$E[V(Y(1))|\mathcal{G}_0] \leq d_1 V(Y(0)) + d_2, \tag{3.4}$$

$$E[\nu_1|\mathcal{G}_0] \leq \tilde{d}_1|Y(0)| + \tilde{d}_2. \tag{3.5}$$

PROOF. Unless stated otherwise we assume that $\hat{X}_{\hat{S}(0)}(0) = 0$.

At time T^1 the server enters station $\hat{S}(1)$ and then the system operates (independently of $U(0)$) like a stable M/GI/1-queueing system (with batch arrivals) until station $\hat{S}(1)$ is empty. It is well-known that the *busy period* $T_{\nu_1} - T^1$ satisfies

$$E[T_{\nu_1} - T^1|\mathcal{F}_{T^1}] \leq h(X_{S(T^1)}(T^1)) \quad P - \text{a.s.},$$

where h is a linear function and where the independence assumptions have also been used. Indeed, assume that $S(T^1) = i$ and consider a random walk generated by the increment $A_i(\eta_i) - 1$, where η_i is a generic service time at station i independent of A_i. Then

$$E[\nu_1|\mathcal{F}_{T^1}] = c_i Y_i(T^1),$$

where c_i is the mean of the number of steps needed by the random walk to decrease by one. Now the desired inequality follows from Wald's identity. Using our assumption (2.3) we hence obtain that

$$
\begin{aligned}
E[T_{\nu_1}|\mathcal{G}_0] &= E[W_1|\mathcal{G}_0] + E[E[T_{\nu_1} - T^1|\mathcal{F}_{T^1}]|\mathcal{G}_0] \\
&\leq w + E[h(X_{S(T^1)}(T^1))|\mathcal{G}_0].
\end{aligned}
$$

¿From (2.1) we have

$$
\begin{aligned}
E[X_{S(T^1)}(T^1)|\mathcal{G}_0] &\leq |Y(0)| + \sum_i E[A_i(T^1)|\mathcal{G}_0] \\
&\leq |Y(0)| + w \sum_i \lambda_i
\end{aligned}
$$

and (3.3) follows. Similarly,

$$
\begin{aligned}
E[Y_i(1)|\mathcal{G}_0] &\leq Y_i(0) + E[A_i(T_{\nu_1})|\mathcal{G}_0] \\
&\leq Y_i(0) + \lambda_i E[T_{\nu_1}|\mathcal{G}_0]
\end{aligned}
$$

and (3.4) follows. The third assertion follows by similar arguments.

If $\hat{X}_{\hat{S}(0)}(0) > 0$ then the assertions follow by directly refering to the properties of an M/GI/1–system. $\qquad\square$

Lemma 3.4 *Consider a measurable set $B_\infty \subseteq \mathbb{Z}_+^K \times \{1, \ldots, K\} \times \mathbf{U}$ and assume that there is an $C \in \mathcal{F}_0$ satisfying*

$$P(T_\infty < \tau_1|\mathcal{F}_{T_n}) > p \quad P - \text{a.s. on } \{|X(T_n)| = 0\} \cap C, \quad n \in \mathbb{Z}_+, \tag{3.6}$$

15

where

$$T_\infty := \inf\{t : X(t) \in B_\infty\} \qquad (3.7)$$

and where we recall that τ_1 is the first arrival epoch. Then, for any number $N > 0$ there is a $p(N) > 0$ satisfying

$$P(T_\infty < \tau_1 | \mathcal{F}_0) \geq p(N) \quad P - \text{a.s. on } \{|X(0)| \leq N\}.$$

PROOF. Because we could always argue on the set C we will assume for simplicity that $C = \Omega$. Given $\mathcal{F}_0$ one can determine the number ξ of non-empty stations. Assume for simplicity that $X_{S_0}(0) > 0$ and put

$$T := \sum_{i=1}^{\nu_1}(T_i - T^i) + W_{\nu_1+1} + \ldots + W_{\nu_\xi-1} + \sum_{i=\nu_\xi-1+1}^{\nu_\xi}(T_i - T^i).$$

Then $\{A(T) = 0, A(T_\infty) = 0, X(T^{\nu_\xi+1}) \in B_\infty\} \subseteq \{T_\infty < \tau_1\}$ and we may use our assumptions (2.4), (3.6) and successive conditioning to obtain the estimate

$$P(T_\infty < \tau_1 | \mathcal{F}_0) \geq \tilde{p}^{|X(0)|} p^{\xi+1},$$

where

$$\tilde{p} := \min_j P(A(T_1 - T^1) = 0 | S(T^1) = j) > 0.$$

For $|X(0)| \leq N$ the right hand side of the above inequality is not smaller than $p(N) := \tilde{p}^N p^{K+1}$. □

We need to introduce some further notation. Let $\{\theta_n : n \in \mathbb{Z}_+\}$ be the flow of shift operators associated with the process $\{\hat{X}(n)\}$, see the Appendix. Then

$$\theta' := \theta_{\nu_1}$$

is the shift operator associated with $\{Y(n)\}$. The symbol O is reserved to denote a (deterministic) function $O : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\limsup_{t\to\infty} O(t)/t < \infty$. Further we

16

let

$$D(t) := \operatorname{card} \{n \geq 1 : T_n \leq t\},$$

be the number of departures by time t.

Under (1.1) the assumptions of the following lemma will be proved to hold.

Lemma 3.5 *Assume that σ is a $\{\mathcal{G}_n\}$-stopping time and $L, c_1, c_2 \in \mathbb{R}_+$, $\varepsilon, c \in (0, 1)$ are constants, satisfying*

$$E[V(Y(\sigma))|\mathcal{G}_0] \leq cV(Y(0)) \quad P - \text{a.s. on } \{V(Y(0)) > L\}, \tag{3.8}$$

$$E[\sigma|\mathcal{G}_0] \leq c_1 V(Y(0))^{1-\varepsilon} \quad P - \text{a.s. on } \{V(Y(0)) > L\}, \tag{3.9}$$

$$E[\nu_\sigma|\mathcal{G}_0] \leq c_2 V(Y(0)) \quad P - \text{a.s. on } \{V(Y(0)) > L\}. \tag{3.10}$$

Then, under the assumptions of Lemma 3.4,

$$E[\tau_\infty|\mathcal{G}_0] = O(V(Y(0))^{1-\varepsilon}) \quad P - \text{a.s. on } C, \tag{3.11}$$

$$E[D(T_\infty)|\mathcal{G}_0] = O(V(Y(0)) \quad P - \text{a.s. on } C, \tag{3.12}$$

where T_∞ is given by (3.7),

$$\tau_\infty := \min\{n \geq 1 : Y(n) \in B_\infty\},$$

and $\min \emptyset := \infty$.

PROOF. Define $\sigma_0 := 0$ and

$$\sigma_{n+1} := \sigma_n + \sigma \circ \theta'_{\sigma_n}, \quad n \in \mathbb{Z}_+,$$

$$\gamma := \min\{n \geq 1 : V(Y(\sigma_n)) \leq cV(Y(0))\}, .$$

where θ'_0 denotes the identity on Ω. By (3.8) and Lemma 4.2 we have

$$E[\gamma|\mathcal{G}_0] \leq \frac{1}{(1-c)c} \quad P - \text{a.s. on } \{V(Y(0)) > L'\}, \tag{3.13}$$

17

where $L' := L/c$. The $\{\mathcal{G}_n\}$-stopping time $\sigma' := \sigma_\gamma$ satisfies

$$V(Y(\sigma')) \leq cV(Y(0))$$

everywhere on $\{\gamma < \infty\}$. Moreover, by (3.10) we have, P-almost surely on $\{V(Y(0)) > L'\}$, that

$$
\begin{aligned}
E[\nu_{\sigma'}|\mathcal{G}_0] &= E\left[\sum_{j=0}^{\gamma-1}(\nu_{\sigma_{j+1}} - \nu_{\sigma_j})|\mathcal{G}_0\right] \\
&= E\left[\sum_{j=0}^{\infty}\mathbf{1}\{j < \gamma\}E[(\nu_{\sigma_{j+1}} - \nu_{\sigma_j})|\mathcal{G}_{\sigma_j}]|\mathcal{G}_0\right] \\
&\leq c_2 E\left[\sum_{j=0}^{\infty}\mathbf{1}\{j < \gamma\}V(Y(\sigma_j))|\mathcal{G}_0\right] \\
&\leq c_2 E\left[\sum_{j=0}^{\gamma}V(Y(\sigma_j))|\mathcal{G}_0\right].
\end{aligned}
$$

¿From Lemma 4.2 we obtain that

$$E[\nu_{\sigma'}|\mathcal{G}_0] \leq c_2'V(Y(0)) \quad P - \text{a.s. on } \{V(Y(0)) > L'\}, \tag{3.14}$$

where $c_2' := c_2/(1-c)$. In a similar way we obtain from (3.9) that

$$E[\sigma'|\mathcal{G}_0] = E\left[\sum_{j=0}^{\gamma-1}\sigma \circ \theta'_{\sigma_j}|\mathcal{G}_0\right] \leq c_1 E\left[\sum_{j=0}^{\gamma}V(Y(\sigma_j))^{1-\varepsilon}|\mathcal{G}_0\right],$$

so that, by Lemma 4.2,

$$E[\sigma'|\mathcal{G}_0] \leq c_1'V(Y(0))^{1-\varepsilon} \quad P - \text{a.s. on } \{V(Y(0)) > L'\}, \tag{3.15}$$

where $c_1' := c_1/(1-c)$. Next we iterate σ' by defining $\sigma_0' := 0$ and

$$\sigma_{n+1}' := \sigma_n' + \sigma' \circ \theta'_{\sigma_n'}, \quad n \in \mathbb{Z}_+.$$

Since $Y(\sigma'_{n+1}) = Y(\sigma') \circ \theta'_{\sigma'_n}$, we have

$$V(Y(\sigma'_n)) \leq c^n V(Y(0)). \tag{3.16}$$

Denote

$$\tau' := \min\{n \geq 1 : V(Y(n)) \leq L'\}.$$

Then we can bound $\tau' \leq \sigma'_{\gamma'}$, where

$$\gamma' := \min\{n \geq 1 : V(Y(\sigma'_n)) \leq L'\}.$$

Arguing as above we get from (3.15) and (3.16), P-almost surely on $\{V(Y(0)) > L'\}$, that

$$
\begin{aligned}
E[\tau'|\mathcal{G}_0] &\leq E\left[\sum_{j=0}^{\gamma'-1}(\sigma'_{j+1} - \sigma'_j)|\mathcal{G}_0\right] \\
&= \sum_{j=0}^{\infty} E[\mathbf{1}\{j < \gamma'\}E[\sigma' \circ \theta'_{\sigma'_j}|\mathcal{G}_{\sigma'_j}]|\mathcal{G}_0] \\
&\leq c'_1 \sum_{j=0}^{\infty} E[\mathbf{1}\{j < \gamma'\}V(Y(\sigma'_j))^{1-\varepsilon}|\mathcal{G}_0] \\
&\leq c'_1 V(Y(0))^{1-\varepsilon} \sum_{j=0}^{\infty} c^{(1-\varepsilon)j}
\end{aligned}
$$

and hence

$$E[\tau'|\mathcal{G}_0] \leq \tilde{c}_1 V(Y(0))^{1-\varepsilon} \quad P - \text{a.s. on } \{V(Y(0)) > L'\}, \tag{3.17}$$

where $\tilde{c}_1 := c'_1/(1 - c^{(1-\varepsilon)})$. Similarly we obtain from (3.14) that

$$E[\nu_{\tau'}|\mathcal{G}_0] \leq \tilde{c}_2 V(Y(0)) \quad P - \text{a.s. on } \{V(Y(0)) > L'\}, \tag{3.18}$$

where $\tilde{c}_2 := c'_2/(1 - c)$.

19

To prove (3.11) and (3.12) we define inductively for $n \in \mathbb{Z}_+$

$$T'_{n+1} \quad := \quad \inf\{t \geq T^*_n : V(X(t)) \leq L'\},$$

$$T^*_{n+1} \quad := \quad \inf\{t > T'_{n+1} : A(t) > A(t-)\},$$

where $T^*_0 := 0$. Note that $T'_{n+1} = T^*_n$ if $V(X(T^*_n)) \leq L'$. Let

$$B(t) := \operatorname{card} \{n \geq 1 : T_{\nu_n} \leq t\},$$

be the number of completed batches of services by time t. We have the following inequalities

$$T'_1 \leq T_{\nu_{\tau'}}, \quad D(T'_1) \leq \nu_{\tau'}, \quad B(T'_1) \leq \tau'. \tag{3.19}$$

Let

$$\sigma^* := \min\{n \geq 1 : X(t) \in B_\infty \text{ for some } t \text{ with } T'_n \leq t \leq T^*_n\}.$$

On the set $\{V(X(0)) > L'\}$ we have by (3.18) and (3.19)

$$
\begin{aligned}
&E[D(T^*_{\sigma^*})|\mathcal{F}_0] \\
= \quad & E\left[\sum_{j=0}^\infty \mathbf{1}\{j < \sigma^*\}(D(T^*_{j+1}) - D(T^*_j))|\mathcal{F}_0\right] \\
= \quad & E\left[\sum_{j=0}^\infty \mathbf{1}\{j < \sigma^*\}(D(T^*_{j+1}) - D(T'_{j+1}))|\mathcal{F}_0\right] \\
+ \quad & E\left[\sum_{j=0}^\infty \mathbf{1}\{j < \sigma^*\}E[D(T'_{j+1}) - D(T^*_j)|\mathcal{F}_{T^*_j}]|\mathcal{F}_0\right] \\
\leq \quad & L''E[\sigma^*|\mathcal{F}_0] + \tilde{c}_2 E\left[\sum_{j=0}^\infty \mathbf{1}\{j < \sigma^*\}\mathbf{1}\{V(X(T^*_j)) > L'\}V(X(T^*_j))|\mathcal{F}_0\right] \\
\leq \quad & (L'' + \tilde{c}_2(L' + 1))E[\sigma^*|\mathcal{F}_0] + \tilde{c}_2 V(X(0)),
\end{aligned}
$$

where L'' is a finite constant satisfying $|x| \leq L''$ whenever $V(x) \leq L'$. Here we have also used, for $n \geq 1$, the easy to prove inequality

$$E[V(X(T^*_n))|\mathcal{F}_{T'_n}] \leq L' + \max_i \lambda_i b_i,$$

20

where we assume for notational simplicity that $\lambda_i b_i \leq 1$ for all i. Similarly we obtain from (3.17) and (3.19) that

$$E[B(T^*_{\sigma^*})|\mathcal{F}_0] \leq (K + \tilde{c}_1(L'+1)^{1-\varepsilon})E[\sigma^*|\mathcal{F}_0] + \tilde{c}_2 V(X(0))^{1-\varepsilon}.$$

By standard arguments it follows that $\tau'' := B(T^*_{\sigma^*})$ is a $\{\mathcal{G}_n\}$-stopping time. Since clearly, $\nu_{\tau''} \leq D(T^*_{\sigma^*})$ and $\tau_\infty \leq \tau''$ we could use our estimates to conclude the assertions (3.11) and (3.12) if we knew that $E[\sigma^*|\mathcal{F}_0]$ is bounded on $\{V(X(0)) > L'\}$. To this end we note that Lemma 3.4 implies, P-a.s. on C,

$$P(X(t) \in B_\infty \text{ for some } t \text{ with } T'_n \leq t \leq T^*_n\}|\mathcal{F}_{T'_n}) \geq p(L''),$$

so that a geometrical trial argument completes the proof. □

Lemma 3.6 *Assume* (1.1)*. Then there exists a $\{\mathcal{G}_n\}$-stopping time σ and constants $L, c_1, c_2 \in \mathbb{R}_+$, $\varepsilon, c \in (0,1)$, such that equations* (3.8)–(3.10) *in Lemma 3.5 are satisfied.*

PROOF. We proceed by induction on the number of stations. For $K = 1$ we can take $\sigma = 1$. Then (3.8) and (3.9) are trivial while (3.10) is a well-known property of a stable M/GI/1-queueing system, see Lemma 3.3.

Now we suppose that the assertion is true for all polling systems with K stations satisfying the our general assumptions and (1.1). We consider a polling system with $K+1$ stations described by the process

$$X(t) = (X_1(t), \ldots, X_{K+1}(t), S(t), U(t)).$$

To exploit our induction hypothesis we couple $X(t)$ with another auxiliary polling system with stations $\{1, \ldots, K\}$ described by the process $\tilde{X}(t)$. This auxiliary process

should behave like the original process until the time when the server first enters station $K + 1$. Hence we define

$$(\tilde{X}_1(t), \ldots, \tilde{X}_K(t), \tilde{S}(t)) := (X_1(t), \ldots, X_K(t), S(t)), \quad t < \tilde{T}_\infty,$$

where

$$\tilde{T}_\infty := \inf\{t : S(t) = K + 1\}.$$

Further we let $\tilde{U}(t) := (U(t), X_{K+1}(0) + A_{k+1}(t))$ for $t < \tilde{T}_\infty$ and $\tilde{U}(t) = u_\infty$ for $t \geq \tilde{T}_\infty$, where u_∞ is not in the space $\mathbf{U} \times \mathbb{Z}_+$. Hence the process $\tilde{U}(t)$ takes values in the set $\mathbf{U} \times \mathbb{Z}_+ \cup \{u_\infty\}$. Assume that $\tilde{T}_\infty = T^{n+1}$ for $n \in \mathbb{Z}_+$. In particular $X_{S(T_n-)}(T_n) = 0$ and we recall that $T^{n+1} = T_n$ is possible. We distinguish two cases. First we assume that $\sum_{i \neq K+1} X_i(T_n) > 0$. Then we let $\tilde{T}^{n+1} := T^{n+1}$ be the moment of the beginning of the $(n + 1)$th service for the process $\{\tilde{X}(t)\}$ and assume that

$$\tilde{X}(T^{n+1}) = (X_1(T^{n+1}), \ldots, X_K(T^{n+1}), \tilde{S}(T^{n+1}), u_\infty),$$

where $\tilde{S}(T^{n+1})$ is assumed to be in the set $\{j \neq K + 1 : X_j(T^{n+1}) > 0\}$. The second case is $\sum_{i \neq K+1} X_i(T^{n+1}) = 0$. In that case we put

$$\tilde{T}^{n+1} := \inf\{t > T^{n+1} : \sum_{i \neq K+1} A_i(t) > \sum_{i \neq K+1} A(t-)\}$$

and

$$\tilde{X}_i(t) = X_i(T_n) + A_i(T_n, t], \quad T_n \leq t \leq \tilde{T}^{n+1}.$$

Further we set $S(t) = 0$ for $T_n \leq t < \tilde{T}^{n+1}$ and we assume that $\tilde{S}(T^{n+1})$ is an element of of $\{i \neq K + 1 : \tilde{X}_i(\tilde{T}^{n+1}) > 0\}$. In both cases we assume that the process $(\tilde{X}_1(t), \ldots, \tilde{X}_K(t), \tilde{S}(t))$ evolves after time $\tilde{T}^{n+1}$ as in Example 2.4.

So far we have given a pathwise description of the process $\tilde{X}(t)$. We skip here the obvious technical details describing the joint distribution of $\{\tilde{X}(t)\}$ and $\{X(t)\}$ and

22

ensuring that $\{\tilde{X}(t)\}$ will indeed become a process satisfying the general assumptions of this section. Assumption (2.4) also implies (3.6) for $B_\infty := \mathbb{Z}_+^K \times \{1,\ldots,K\} \times \{u_\infty\}$ and $C = \{X_{K+1}(0) > 0\}$, a set which is $\sigma\{\tilde{X}(0)\}$-measurable, by definition. Hence we can apply the induction hypothesis to obtain (3.11) and (3.12) with τ_∞ replaced by

$$\tilde{\tau}_\infty := \min\{n \geq 1 : S_n = K + 1\}$$

and T_∞ replaced by $\tilde{T}_\infty$ This argument works just as well for any station other than $K + 1$.

There is a $\mathcal{G}_0$-measurable random element ξ of $\{1,\ldots,K+1\}$ with the property

$$Y_\xi(0) \geq \frac{|Y(0)|}{(K+1)}. \tag{3.20}$$

We claim that (3.8)–(3.10) are satisfied with

$$\sigma := \begin{cases} \min\{n \geq 1 : \xi = S_n\} & \text{if } S_0 \neq \xi, \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

Since ξ is $\mathcal{G}_0$-measurable we can use the above argument to conclude that

$$E[\sigma|\mathcal{G}_0] = O(V(Y(0))^{1-\varepsilon}), \tag{3.22}$$

$$E[D(\tilde{T})|\mathcal{G}_0] = O(V(Y(0))), \tag{3.23}$$

where

$$\tilde{T} := \inf\{t \geq 0 : S(t) = \xi\}.$$

Next we are going to show that σ does satisfy the drift condition (3.8). We have

$$Y_i(\sigma) = Y_i(0) + A_i(T_{\nu_\sigma}) - \nu_\sigma^i, \tag{3.24}$$

where

$$\nu_n^i := \text{card}\,\{k \in \mathbb{N} : k \leq \nu_n : \hat{S}(k) = i\}, \quad n \in \mathbb{N}. \tag{3.25}$$

23

Using the very definitions of the polling system as well as (2.2) and (2.3) we can calculate

$$E[T_{\nu_\sigma}|\mathcal{F}_0] = E\left[\sum_{m=0}^{\infty}\mathbf{1}\{m+1 \le \sigma\}(W_{\nu_m+1} + \sum_{n=\nu_m+1}^{\nu_{m+1}}(T_n - T^n))|\mathcal{F}_0\right]$$

$$= E\left[\sum_{m=0}^{\infty}\mathbf{1}\{m+1 \le \sigma\}E[W_{\nu_m+1}|\mathcal{F}_{T_{\nu_m}}]|\mathcal{F}_0\right]$$

$$+E\left[\sum_{m=0}^{\infty}\mathbf{1}\{m+1 \le \sigma\}\sum_{i=1}^{K+1}\mathbf{1}\{S(T^{\nu_m+1}) = i\}\right.$$

$$\left.\sum_{n=1}^{\infty}\mathbf{1}\{\nu_m+1 \le n \le \nu_{m+1}\}E[T_n - T^n|\mathcal{F}_{T^n}]|\mathcal{F}_0\right]$$

$$\le wE\left[\sum_{m=0}^{\infty}\mathbf{1}\{m+1 \le \sigma\}|\mathcal{F}_0\right]$$

$$+E\left[\sum_{m=0}^{\infty}\mathbf{1}\{m+1 \le \sigma\}\sum_{i=1}^{K+1}\mathbf{1}\{S(T^{\nu_m+1}) = i\}b_i(\nu_{m+1} - \nu_m)|\mathcal{F}_0\right],$$

where we have taken advantage of the stopping time properties of σ and ν_m which imply, for example, that

$$\{m+1 \le \sigma, S(T^{\nu_m+1}) = i, \nu_m + 1 \le n \le \nu_{m+1}\} \in \mathcal{F}_{T^n}.$$

Hence

$$E[T_{\nu_\sigma}|\mathcal{F}_0] \le wE[\sigma|\mathcal{F}_0] + E[\zeta|\mathcal{F}_0],$$

where

$$\zeta := \sum_{i=1}^{K+1}b_i\nu_\sigma^i.$$

Using

$$E[A_i(T_{\nu_\sigma})|\mathcal{F}_0] = \lambda_i E[T_{\nu_\sigma}|\mathcal{F}_0]$$

we obtain from (3.22) and (3.24) that

$$E[V(Y(\sigma))|\mathcal{F}_0] \le V(Y(0)) - (1-\rho)E[\zeta|\mathcal{F}_0] + O(V(Y(0))^{1-\tilde{\varepsilon}}, \qquad (3.26)$$

24

where the workload

$$\rho := \sum_{i=1}^{K+1} \lambda_i b_i$$

is srictly less than one by assumption. From the definitions (3.20) and (3.21) we have

$$\nu_\sigma = \sum_{i=1}^{K+1} \nu_\sigma^i \geq \frac{|Y(0)|}{K+1},$$

yielding (all b_i are positive) that

$$E[\zeta|\mathcal{F}_0] \geq \tilde{c} V(Y(0)),$$

where

$$\tilde{c} = \frac{\min b_i}{(K+1)\max b_i}.$$

Inserting this into (3.26) we obtain (3.8) for a suitable chosen constant L.

Next we want to prove (3.10) making use of the equality

$$E[\nu_\sigma|\mathcal{F}_0] = E[D(\tilde{T})|\mathcal{F}_0] + E[E[D(T_{\nu_\sigma}) - D(\tilde{T})|\mathcal{F}_{\tilde{T}}|\mathcal{F}_0].$$

The first summand can be bounded according to (3.23). For the second we can use (3.5) and

$$E[V(X(\tilde{T}))|\mathcal{F}_0] = O(V(X(0)),$$

which can be proved as (3.26). □

In the remainder of the section we return to the setting of Section 2 which is obtained by letting $U(t) \equiv 0$.

PROOF OF THEOREM 3.1 (I).

Assume (1.1). By Lemma 3.6 we can use Theorem 4.3 with $(\hat{X}(n), \nu_\sigma)$ in place of $(X(n), \sigma)$. The assumption (4.6) of that theorem is clearly satisfied and the set $\{x : V(x) \leq L\}$ is finite for any $L > 0$. The other assertions are standard. □

25

In case $\{Y(n)\}$ is ergodic we let π denote its stationary initial distribution and define

$$\bar{w} := \sum_x w(x)\pi(x),$$

where $w(x)$, $x \in \mathbb{Z}_+^K \times \{1,\ldots,K\}$, satisfies

$$w(x) = E[W_1|\mathcal{F}_0] \quad P - \text{a.s. on } \{Y(0) = x\}.$$

Hence $\bar{w}$ is the average time taken by a walk after a batch of services. Since $\pi((0,\ldots,0,i))$ > 0 for all $i \in \{1,\ldots,K\}$ it is clear from the definition of W_1 that $\bar{w}$ has to be positive. Consider now the Markov chain $Y'(n)$ denoting the system states at the consecutive polling instants, i.e.

$$Y'(n) := X(T^{\nu_n+1}), \quad n \in \mathbb{Z}_+,$$

and let π' denote its equlibrium distribution if the chain is ergodic. Note that π' has to be concentrated on the set of all those $x = (x_1,\ldots,x_K,i)$ satisfying $x_i > 0$. On this set we define a function $B(\cdot)$ by

$$B(x) = E[\nu_1|\mathcal{F}_0] \quad P - \text{a.s. on } \{Y(0) = x\}.$$

Let

$$\pi_i := \pi'(\mathbb{Z}_+^K \times \{i\}) = \pi(\mathbb{Z}_+^K \times \{i\}).$$

Then

$$\bar{B}_i := \pi_i^{-1} \sum_{x \in \mathbb{Z}_+^K} B((x,i))\pi'(\{(x,i)\}),$$

is the stationary average batch size given that the server is at station i.

Next we turn to the proof of the necessity part of Theorem 3.1. This proof will also yield an equilibrium equation for the process $\{\hat{X}(n)\}$. A shorter proof could be obtained by comparing the system with a polling system as in Example 2.4. The latter system can easily be proved to be unstable if (1.1) does not hold.

26

PROOF OF NECESSITY OF (1.1).

Let $n \in \mathbb{N}$. We can use the calculations which led to (3.25) (with n in place of σ) to obtain that

$$E[Y_i(n)|\mathcal{F}_0] = Y_i(0) + \lambda_i E[(W(n) + \zeta_n)|\mathcal{F}_0] - E[\nu_n^i|\mathcal{F}_0], \tag{3.27}$$

where

$$\zeta_n := \sum_{i=1}^{K} b_i \nu_n^i,$$

$$W(n) := \sum_{m=0}^{n-1} W_{\nu_m+1}$$

and ν_i^n is given by (3.25).

In the remainder of the proof we follow Borovkov and Schassberger (1994). Unless stated otherwise we assume that $\rho \neq 1$, where

$$\rho := \sum_{i=1}^{K} \lambda_i b_i.$$

Let $f : \mathbb{R}^K \to \mathbb{R}^K$ be the linear transformation given by

$$f(e_i) := e_i + \frac{b_i}{1-\rho}\vec{\lambda}, \quad i \in \{1, \ldots, K\},$$

where e_i is the ith unit vector and the ith component of $\vec{\lambda}$ is λ_i. We apply this transformation to obtain from (3.27) by a straightforward calculation that

$$E[f_j((Y_1(n), \ldots, Y_K(n)) - (Y_1(0), \ldots, Y_K(0)))|\mathcal{F}_0] \tag{3.28}$$
$$= \frac{\lambda_j}{1-\rho} E[W(n)|\mathcal{F}_0] - E[\nu_n^j|\mathcal{F}_0], \quad j \in \{1, \ldots, K\},$$

where f_j is the jth component of f.

Assume now that $\{\hat{X}(n)\}$ and hence also $\{Y(n)\}$ and $\{Y'(n)\}$ are ergodic. Dividing both sides of (3.28) by n and letting $n \to \infty$ the left-hand side tends to 0 and we obtain

that

$$\frac{\lambda_j \bar{w}}{1 - \rho} = \bar{B}_j \pi_j, \quad j \in \{1, \ldots, K\}. \tag{3.29}$$

In particular we have $\rho < 1$. Assume now that $\rho = 1$. Dividing (3.27) by n and letting $n \to \infty$ we obtain

$$\bar{B}_i \pi_i = \lambda_i \bar{w} + \lambda_i \sum_{j=1}^{K} b_j \bar{B}_j \pi_j.$$

Multiplying with b_i and summing up yields an equation contradicting $\bar{w} > 0$. $\qquad \square$

Corollary 3.7 *If $\{\hat{X}(n)\}$ is ergodic then* (3.29) *holds.*

Remark 3.8 Consider Example 2.1 or Example 2.2 and let $R(t)$ denote the residual walking time at time $t \in \mathbb{R}_+$ if $S(t) = 0$ and let $R(t)$ be the residual service time, otherwise. Then $\{(X(t), R(t)) : t \in \mathbb{R}_+\}$ is an homogeneous Markov process. Using Theorem 3.1 and the properties of the Poisson process, it is not hard to prove that $\{(X(t), R(t))\}$ is Harris ergodic if (1.1) holds. Under natural additional assumptions this conclusion could also be made in the general case.

4 Appendix

This Appendix contains some results for discrete time Markov processes and Markov chains. We consider a Markov process $X = \{X(n) : n \in \mathbb{Z}_+\}$ taking values in the measurable space $(\mathbf{X}, \mathcal{X})$. The process X is assumed to fulfill the following (standard) properties. The random elements $X(n)$ are defined on $(\Omega, \mathcal{F})$ and are measurable with respect to $\mathcal{F}_n$, where $\{\mathcal{F}_n : n \in \mathbb{Z}_+\}$ is a increasing sequence of σ-fields. For each $n \in \mathbb{Z}_+$ there is a measurable mapping $\theta_n : \Omega \to \Omega$ such that $\theta_{m+n} = \theta_m \circ \theta_n$. Hence θ_n is the nth iteration of the *shift operator* $\theta := \theta_1$. We assume that

$$X \circ \theta_m(n) = X(m+n), \quad m, n \in \mathbb{Z}_+,$$

where the process $X \circ \theta_m$ is given by $X \circ \theta_m(n)(\omega) := X(n)(\theta_m \omega)$. We call θ_n, $n \in \mathbb{Z}_+$, the family of shift operators associated with X. Further there is a family P_x, $x \in \mathbf{X}$, of probability measures on $(\Omega, \mathcal{F})$ such that $P_x(X(0) = x) = 1$ and the mapping $x \to P_x(A)$ is measurable for all $A \in \mathcal{F}$ and such that for any $\{\mathcal{F}_n\}$-stopping time T

$$P_x(X \circ \theta_T \in \cdot | \mathcal{F}_T) = P_{X(T)}(X \in \cdot) \quad P_x - \text{a.s. on } \{T < \infty\},$$

where we interpret all processes as random elements of the corresponding function space equipped with Kolmogorov's product σ-field. This is the strong Markov property of X. In particular,

$$P_y(X \in \cdot | \mathcal{F}_0) = P_x \quad P_y - \text{a.s. on } \{X(0) = x\}, \quad y \in \mathbf{X}.$$

Assume that σ is a finite $\{\mathcal{F}_n\}$-stopping time and define

$$\sigma_{n+1} := \sigma_n + \sigma \circ \theta_{\sigma_n}, \quad n \in \mathbb{Z}_+, \tag{4.1}$$

where $\sigma_0 := 0$ and θ_0 is the identity on Ω. Then the σ_n are stopping times and $Y(n) := X(\sigma_n)$, $n \in \mathbb{Z}_+$ is again a Markov process in discrete time. One can take the same family P_x, $x \in \mathbf{X}$, of probability measures while the shift operator is given by θ_σ and the filtration by $\{\mathcal{F}_{\sigma_n} : n \in \mathbb{Z}_+\}$.

In the following P denotes a probability measure on $(\Omega, \mathcal{F})$ of the form

$$P(\cdot) = \int P_x(\cdot) \mu(dx),$$

where μ is a probability measure on $(\mathbf{X}, \mathcal{X})$.

Under additional technical assumptions the conditions of the next lemma are known to imply the so-called geometric ergodicity. The proofs of this as well as of the next result are not difficult and can follow along the lines of Meyn and Tweedie (1993, 1994) for example.

29

Lemma 4.1 *Assume that there are numbers $c \in (0, 1)$ and $L \in \mathbb{R}_+$ as well as a measurable function $V : \mathbf{X} \to \mathbb{R}_+$ such that*

$$E[V(X(1))|\mathcal{F}_0] \leq cV(X(0)) \quad P - \text{a.s. on } \{V(X(0)) > L\}.$$

Define

$$\tau_L := \min\{n \geq 1 : V(X(n)) \leq L\}, \tag{4.2}$$

where $\min \emptyset := \infty$. Then, for any $0 < \delta \leq 1$,

$$E\left[\sum_{n=0}^{\tau_L} V(X(n))^\delta | \mathcal{F}_0\right] \leq \frac{V(X(0))^\delta}{1 - c^\delta} \quad P - \text{a.s. on } \{V(X(0)) > L\}$$

and

$$E[\tau_L|\mathcal{F}_0] < \infty \quad P - \text{a.s. on } \{V(X(0)) > L\}.$$

Lemma 4.2 *Assume that the assumptions of Lemma 4.1 hold and denote*

$$\gamma := \min\{n \geq 1 : V(X(n)) \leq cV(X(0))\}.$$

Then, for any $0 < \delta \leq 1$,

$$E\left[\sum_{n=0}^{\gamma} V(X(n))^\delta | \mathcal{F}_0\right] \leq \frac{V(X(0))^\delta}{1 - c^\delta} \quad P - \text{a.s. on } \{V(X(0)) > L/c\}. \tag{4.3}$$

Moreover,

$$E[\gamma|\mathcal{F}_0] \leq \frac{1}{(1-c)c} \quad P - \text{a.s. on } \{V(X(0)) > L/c\}.$$

The next result is similar to Theorem 2.2 in Meyn and Tweedie (1994). For the reader's convenience we will present the proof.

Theorem 4.3 *Assume that there are numbers $c \in (0, 1)$ and $L, c_2, d > 0$, a measurable function $V : \mathbf{X} \to \mathbb{R}_+$ and a finite $\{\mathcal{F}_n\}$-stopping time $\sigma \geq 1$ such that*

$$E[V(X(\sigma))|\mathcal{F}_0] \leq cV(X(0)) \quad P - \text{a.s. on } \{V(X(0)) > L\}, \tag{4.4}$$

$$E[\sigma|\mathcal{F}_0] \leq c_2 V(X(0)) \quad P - \text{a.s. on } \{V(X(0)) > L\}, \tag{4.5}$$

$$E[V(X(1))|\mathcal{F}_0] \leq d \quad P - \text{a.s. on } \{V(X(0)) \leq L\}. \tag{4.6}$$

Then

$$E[\tau_L|\mathcal{F}_0] \leq c' \max\{V(X(0)), L\} \quad P - \text{a.s.}$$

for some finite c', where τ_L is given by (4.2)*. In particular, the set $\{x \in \mathbf{X} : V(x) \leq L\}$ is positive recurrent, in the sense that*

$$E[\tau_L|\mathcal{F}_0] < c'L \quad P - \text{a.s. on } \{V(X(0)) \leq L\},$$

$$\tau_L < \infty \quad P - \text{a.s. on } \{V(X(0)) > L\}.$$

PROOF. Let $\sigma_0 := 0$ and, recursively, $\sigma_{n+1} := \sigma_n + \sigma \circ \theta_{\sigma_n}$, $n \geq 1$. Denoting

$$\tau := \min\{n \geq 1 : V(X(\sigma_n)) \leq L\},$$

we obtain from (4.5) that

$$
\begin{aligned}
E[\sigma_\tau|\mathcal{F}_0] &= E\left[\sum_{i=0}^{\tau-1} \sigma \circ \theta_{\sigma_i}|\mathcal{F}_0\right] \\
&= E\left[\sum_{i=0}^{\infty} \mathbf{1}\{i < \tau\} E[\sigma \circ \theta_{\sigma_i}|\mathcal{F}_{\sigma_i}]|\mathcal{F}_0\right] \\
&\leq c_2 E\left[\sum_{i=0}^{\infty} \mathbf{1}\{i < \tau\} V(X(\sigma_i))|\mathcal{F}_0\right].
\end{aligned}
$$

By (4.4) we can apply Lemma 4.1. Hence,

$$E[\sigma_\tau|\mathcal{F}_0] \leq \frac{c_2}{1-c} V(X(0)) \quad P - \text{a.s. on } \{V(X(0)) > L\}. \tag{4.7}$$

31

Since $\tau_L \leq \sigma_\tau$, this yields the assertion for $V(X(0)) > L$. To prove the other case we can use inequality $\tau_L \leq 1 + \sigma_\tau \circ \theta_1$ for $V(X(1)) > L$ to obtain from (4.7)

$$
\begin{aligned}
E[\tau_L|\mathcal{F}_0] &= E[\mathbf{1}\{V(X(1)) \leq L\}\tau_L|\mathcal{F}_0] + E[\mathbf{1}\{V(X(1)) > L\}\tau_L|\mathcal{F}_0] \\
&\leq P(V(X(1)) \leq L|\mathcal{F}_0) + E\left[\mathbf{1}\{V(X(1)) > L\}E[\sigma_\tau \circ \theta_1 + 1|\mathcal{F}_1]|\mathcal{F}_0\right] \\
&\leq 1 + \frac{c_2}{1-c}E[\mathbf{1}\{V(X(1)) > L\}V(X(1))|\mathcal{F}_0].
\end{aligned}
$$

By assumption (4.6) the proof is complete. □

Acknowledgment

This research has been motivated by the work of Rolf Schassberger. We would like to thank him for fruitful discussions. We also like to thank Frank Kelly and Richard Tweedie for interesting comments and for drawing our attention to a class of greedy-type polling systems during a conference on Applied Probability in Oberwolfach.

References

Altman, E., Levy, H. (1994) Queueing in space. *Adv. Appl. Prob.* **26**, 1095-1116

Baccelli, F., Foss, S. (1995) On the saturation rule for the stability of queues. *J. Appl. Prob.* **32**, 494-507

Borovkov, A.A. (1994) *Ergodicity and Stability of Stochastic Processes.* TVP Publishers, Moscow, (to appear)

Borovkov, A.A., Schassberger, R. (1994) Ergodicity of a polling network. *Stochastic Process. Appl.* **50**, 253-262

Coffman Jr., E.G., Gilbert, E.N. (1987) Polling and greedy servers on a line. *Queuing Systems* **2**, 115-145

Dai, J.G. (1995) On the positive Harris recurrence for multiclass queueing networks: A unified approach via fluid limit models. *Ann. Appl. Probab.* (to appear)

Fayolle, G., Lasgouttes, J. (1994) A state-dependent polling model with Markovian routing. INRIA-Report 2279

Fayolle, G., Malyshev, V.A., Menshikov, M.V. (1995) *Topics in the Constructive Theory of Countable Markov Chains.* Cambridge University Press

Foss, S., Chernova, N. (1994) Ergodic properties of polling systems. *Problems of Information Transmission*, (to appear)

Foss, S., Last, G. (1995) On the stability of greedy polling systems with general service policies. Preprint.

Fricker, C., Jaibi, M.R. (1994) Monotonicity and stability of periodic polling systems. *Queueing Systems* **15**, 211-238

Georgiadis, L., Szpankowski, W. (1992) Stability of token passing rings, *Queueing Systems* **11**, 7-34

Kroese, D.P., Schmidt, V. (1992) A continuous polling system with general service times. *Ann. Appl. Probab.* **2**, 906-927

Kroese, D.P., Schmidt, V. (1994) Single-server queues with spatially distributed arrivals. *Queueing Systems*, **17**, 317-345

Last, G., Brandt, A. (1995) *Marked Point Processes on the Real Line: The Dynamic Approach*. Springer. New York.

Levy, H., Sidi, M. and Boxma, O.J. (1990) Dominance relations in polling systems. *Queueing Systems* **6**, 155-171

Malyshev, V.A., Men'shikov, M.V. (1982) Ergodicity, continuity and analyticity of countable Markov chains. *Trans. Moscow Math. Soc.* **1**, 1-48

Massoulié, L. (1995) Stability of non-Markovian polling systems. *Queueing Systems*, (to appear)

Meyn, S.P., Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability.* Springer-Verlag, London

Meyn, S.P., Tweedie, R.L. (1994) State-dependent criteria for convergence of Markov chains. *Ann. Appl. Probab.* **4**, 149-168

Schassberger, R. (1993) Stability of polling networks with state-dependent server routing. Technical report 12/93, Technical University Braunschweig, (to appear in *Probability in the Engineering and Informational Sciences*)

Takagi, H. (1990) Queueing analysis of polling systems. in: *Stochastic Analysis of Computer and Communication systems.* ed. H. Takagi, North-Holland, Amsterdam, 267-318