On the stability of greedy polling systems with general service policies

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. Upon arrival at a non-empty station i, say, with $x > 0$ waiting customers, the server tries to serve there a random number B of customers if the queue length has not reached a random level $C < x$ before the server has completed the B services. The random variable B may also take the value ∞ so that the server has to provide service as long as the queue length has reached size C. The distribution $H_{i,x}$ of the pair (B, C) may depend on i and x while the service time distribution is allowed to depend on i. The station to be visited next is chosen among some neighbors according to a greedy policy. That is to say that the server always tries to walk to the fullest station in his well-defined neighborhood. Under appropriate independence assumptions two conditions are established which are sufficient for stability and sufficient for instability. Some examples will illustrate the relevance of our results.

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

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

[*]This work was done while the first author held a visiting position at Technical University of Braunschweig. Partial support was provided by the INTAS grant 93–820.

1 Introduction

Our goal in this paper is to establish conditions which are sufficient for the stability or for the instability of a Markov chain describing the evolution of a polling system with K queues, one server and rather general service policies for each station. The routing of the server is determined by the "greedy" routing mechanism. Hence, in order to choose his next destination, the server needs to know the current state of the system. It is this state-dependent routing which makes the analysis of the underlying Markov chain interesting.

This paper continues the work in [4], where it was assumed that each station is served according to the exhaustive policy. In this exceptional case the natural workload condition is sufficient and necessary for stability. This is even true for a state-dependent routing mechanism much more general than the greedy one. Here we will formulate our conditions in terms of a maximal and a minimal nominal load of the system defined in terms of the means of the underlying random variables. Example 4.1 in Section 4 shows that, unlike [4], the greedy routing seems to be an almost necessary assumption for the general stability analysis of a polling system with state-dependent routing and mixed service policies. But even for the greedy routing we are only able to derive a couple of conditions which are either sufficient or necessary for stability. These conditions coincide only in special cases and Example 4.2 explains why the determination of the exact stability region can be a very difficult task.

Similarly as in [4] our main method of proof is to establish drift conditions at certain stopping times. To avoid duplications we will refer to [4] whenever possible. However, because we allow here a much greater variety of possible service disciplines (including those which have been studied in the literature) we have to resort to additional arguments detailed in Section 3. In particular, the proofs of the key lemmas 3.4 and 3.5 are based on the specific properties of the greedy routing and fail in case of a general greedy-type routing.

There is a voluminous literature on polling systems (see Tagaki [10]) but there are only few results on systems with state-dependent routing. A special case of our results has been proved in [9] by another more direct method. In [2] the authors investigate the stability region of a special polling system with state-dependent routing and the 1-limited service strategy using the approach of [7].

To give an outline of our results we next describe more details of the model. 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. As indicated in Section 4 one might consider also more general arrival processes. Upon arrival at a non-empty station i the server decides to serve no more than a random number (possibly infinite) of customers as long as the queue is non-empty. The distribution $H'_{i,x}$ of that random number is allowed to depend on i and on the number x of customers in station i at the server's arrival time. The distribution of the service times is assumed to have a finite mean b_i. The family $\{H'_{i,x} : x > 0\}$ determines the *service policy* at station i. Different service times are independent and independent of the arrival process. If $H'_{i,x} = \delta_\infty$, then station i is served according to the *exhaustive policy*. Other examples are $H'_{i,x} = \delta_x$ (*gated* policy) or if $H'_{i,x}$ is given as the distribution of min $\{x, D_i\}$, where D_i is a a random variable with mean d_i (D_i-*limited* policy). In fact, the model described in Section 2 is more general and contains for instance also the *decrementing policy* as a special case.

The server chooses the station j, say, to be visited next in a set $N(i) \subseteq \{1, \dots, K\}$ of *neighbors* of i. It takes the server a random time with finite mean $w_{ij} \geq 0$ to *walk* from i to j. We assume here the *greedy routing mechanism*. That is to say that station j is chosen among those stations in the neighborhood with the maximum number of customers waiting at the start of the walk. It is assumed that $i \in N(i)$ and that the neighborhood relation defines a connected graph. For some slight generalizations of the greedy routing mechanism we refer to Section 4.

We shall prove that the system is stable if $\rho^+ < 1$, where

$$\begin{aligned} \rho^+ &:= \sum_{i=1}^{K} \lambda_i (b_i + w_i^+ / d_i), \qquad (1.1) \\ w_i^+ &:= \max\{w_{ij} : j \in N(i)\}, \\ d_i &:= \lim_{x \to \infty} d'_i(x) \leq \infty, \qquad (1.2) \\ d'_i(x) &:= \int y H'_{i,x}(dy), \end{aligned}$$

and the limit (1.2) is assumed to exist. On the other hand, if $d'_i(x) \leq d_i$ for all i and all x, then the system turns out to be unstable if $\rho^- > 1$, where

$$\begin{aligned} \rho^- &:= \sum_{i=1}^{K} \lambda_i (b_i + w_i^- / d_i), \qquad (1.3) \\ w_i^- &:= \min\{w_{ij} : j \in N(i)\}. \end{aligned}$$

Under weak additional assumptions the inequality $\rho^- \geq 1$ is sufficient for instability. The condition for instability are also valid for more general polling systems where the routing could depend on the current state of the system in an arbitrary manner. Note that in (1.1) and (1.3) the numbers w_i^+ and w_i^- do not matter if $d_i = \infty$. In this case we call the service policy at station i *unlimited*. Otherwise we call it *limited*, see [5] for a similar definition. If, in particular, all stations are served according to an unlimited policy, then

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

is a necessary and sufficient condition for stability.

We will show by an example that the condition $\rho^+ < 1$ is not sufficient for stability of a polling system with the more general greedy-type routing mechanism (see [4]) and with at least one station being served according to a limited policy. Another example will show that if the inequality $d_i(x) > d_i$ is allowed for some x and i with $d_i < \infty$, then, in case $\rho^- = 1$, the system can be both stable or unstable. A third example will show that in case $\rho^- < \rho^+$ the stability region may depend on the whole distribution of the underlying walking times. It seems to be a hard task to determine the region explicitly in such cases.

The paper is organized as follows. The complete model description is given in Section 2. The stability results will be proved in Section 3. The final section is devoted to examples and discussion.

2 Model description

We consider a queueing system consisting of K stations with infinite waiting capacities. Each station receives an input of customers asking for service. We denote by $A_i(s,t]$ the number of customers who arrive at station i during the time interval $(s,t]$ and put $A_i(t) := A_i(0,t]$. We assume that $(A_1, \ldots, A_K)$ is a multivariate compound Poisson process (see [4]) defined (as all random elements in this paper) on the underlying probability space $(\Omega, \mathcal{F}, P)$. The ith *arrival intensity* λ_i is defined by the equation $EA_i(t) = \lambda_i t$. There is only one server who travels through the system. Upon arrival at a non-empty station he starts with a batch of services which may include customers arriving after the server's arrival epoch. Each served customer leaves the system immediately. The size of the batch is determined by the service policy to be described below.

We let $X_i(t)$ denote the number of customers in the ith queue at time $t \in \mathbb{R}^+$ and $S(t)$

the number of the station which is occupied by the server when servicing is in progress and we let $S(t) := 0$, otherwise. Both $X_i(t)$ and $S(t)$ are taken to be right-continuous and justified by our assumptions below we 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 completion of the nth service and T^{n+1} the time of the begin of the next service after time T_n. Note that we must always have that

$$X_{S(T^{n+1})}(T^{n+1}) > 0, \quad n \in \mathbb{Z}_+. \tag{2.1}$$

To describe the *service policies* we assume for a moment that at time T^n the server has just arrived at station $S(T^n) = i$ with $x = X_i(T^n) > 0$ waiting customers. The server then generates a pair (B_n, C_n) of random variables with values in $(\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+$ satisfying $C_n < x$ and having distribution $H_{i,x}$. This pair depends on the system history at time T^n only through (i, x). The server either serves B_n customers and thereafter departs or else departs after the first service resulting in a drop to the queue length C_n, whatever event comes first. In other words, if μ is the smallest $m > n$, for which

$$\psi_m := \min\{B_n - (m - n), \mathbf{1}\{X_i(T_m) \neq C_n\}\}$$

becomes 0 then the server serves $\mu - m$ customers. Thereafter he walks to another station j, say which is chosen according to a greedy mechanism to be described below. The walking times are allowed to depend on the whole current state of the system. The case $i = j$ is not excluded. In this case the walking time could also be an idle or a vacation time.

To formulate the assumptions on the *routing mechanism* we assume that each $i \in \{1, \ldots, K\}$ has a set $N(i) \subseteq \{1, \ldots, K\}$ of *neighbors* with $i \in N(i)$. This neighborhood relation is assumed to equip $\{1, \ldots, K\}$ with the structure of a directed and *connected* graph. If the server has just finished at time T_n a batch of services at station i then the greedy routing mechanism forces the server to choose his next destination in the set

$$\{j \in \{1, \ldots, K\} : X_j(T_n) = \max\{X_k(T_n) : k \in N(i)\}\}$$

provided that $\sum_{k \in N(i)} X_k(T_n) > 0$. Otherwise the server is allowed to choose any station.

The *initial conditions* are given by $(X_1(0), \ldots, X_K(0))$ and by a random element (S_0, B_0, C_0) of $\{1, \ldots, K\} \times (\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+$ satisfying $C_0 \leq X_{S_0}(0)$. We think of time 0 as a the moment of completion of the 0th service at station S_0 and let

$$\psi_0 := \min\{B_0 - 1, \mathbf{1}\{X_{S_0}(0) \neq C_0\}\}.$$

If $\psi_0 > 0$, then the server starts serving queue $S(0-) = S(0) := S_0$. Then $T^1 = 0$ and we let $B_1 := B_0 - 1$, $C_1 := C_0$. If $\psi_0 = 0$, then the server starts walking and the epoch T^1 of the start of the first service satisfies (2.1). We then set $S(0-) := S_0$ and $S(t) = 0$ for $0 \le t < T^1$. (If $T^1 = 0$ we have $S(0) = S(T^1)$).

It should be clear now how the system is operating and we are now going to make our stochastic assumptions precise. We are given the arrival processes $A_1, \ldots, A_K$ described above, processes $X_i(t)$, $i = 1, \ldots, K$, and $S(t)$ with values in $\mathbb{Z}_+$ and in $\{0, \ldots, K\}$, and a sequence (T^n, T_n, B_n, C_n), $n \in \mathbb{Z}_+$, where $T^0 = T_0 := 0$, $S(T^n) = S(T_n-) \in \{1, \ldots, K\}$ and B_n is the remaining number of services the server has to provide at station $S(T_n-)$ just before begin of the nth service if the queue length has not reached level C_n before. In accordance with our assumptions we assume that $\lim_{n\to\infty} T_n = \infty$ and that the times T_n, $n \in \mathbb{N}$, are different from all arrival epochs. The process $X = \{X(t) : t \ge 0\}$ containing all relevant information is

$$X(t) := (X_1(t), \ldots, X_K(t), S(t), B(t), C(t)), \quad t \in \mathbb{R}^+,$$

where

$$(B(t), C(t)) := \sum_{n\ge 0} \mathbf{1}\{T^n \le t < T^{n+1}\}(B_n, C_n).$$

By

$$\mathcal{F}_t := \sigma((S_0, B_0, C_0), X(s) : s \le t), \quad t \in \mathbb{R}^+,$$

we define a right-continuous filtration $\{\mathcal{F}_t : t \in \mathbb{R}^+\}$ describing the internal history of the process. (If $\psi_0 = 0$ and $T^1 = 0$, then $X(0)$ does not contain all information about the initial conditions.) As in [4] we assume that $A_i(t, v]$ is in fact independent of $\mathcal{F}_t$, for all $t < v$, see Section 4 for possible generalizations.

The behaviour of the system just after the completion of the nth service is determined by the random variable

$$\psi_n := \min\{B_n - 1, \mathbf{1}\{X_{S(T_n-)}(T_n) \ne C_n\}\}, \quad n \ge 0.$$

If $\psi_n > 0$, then the server continues serving queue $S(T^n)$. Then $(B_{n+1}, C_{n+1}) = (B_n - 1, C_n)$, where $\infty - 1 := \infty$. It follows that $C_n \le X_{S(T_n-)}(T_n)$ for all $n \in \mathbb{Z}_+$. In particular, $X_{S(T_n-)}(T_n) = 0$ implies $\psi_n = 0$. The service times are assumed to be independent from the arrival process and to satisfy

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

where $G_1, \ldots, G_K$ are distributions on $(0, \infty)$ with finite means $b_1, \ldots, b_K$. If $\psi_n = 0$, then the server stops serving queue $S(T^n-)$ and travels to station $S(T^{n+1}) = S(T_{n+1}-)$

choosen according to the greedy routing mechanism, that is

$$X_{S(T^{n+1})}(T_n) = \max\{X_k(T_n) : k \in N(i)\} \tag{2.3}$$
$$P-\text{a.s. on } \left\{ S(T_n-) = i, \psi_n = 0, \sum_{k\in N(i)} X_k(T_n) > 0 \right\}$$

for all $n \in \mathbb{Z}_+$. Denote

$$W_{n+1} := T^{n+1} - T_n, \quad n \in \mathbb{Z}_+.$$

If $\psi_n = 0$ then $W_{n+1} = 0$. If $\psi_n > 0$ then W_{n+1} is the walking time taken by the server to travel from station $S(T_n-)$ to $S(T^{n+1})$. We assume that

$$E[W_{n+1}|\mathcal{F}_{T_n-}] \le w_i^+ \quad P-\text{a.s. on } \left\{ S(T_n-) = i, \psi_n = 0, \sum_{k\in N(i)} X_k(T_n) > 0 \right\}, \tag{2.4}$$

for non-negative numbers w_i^+ and

$$E[W_{n+1}|\mathcal{F}_{T_n-}] \le w \quad P-\text{a.s. on } \{\psi_n = 0\} \tag{2.5}$$

for a finite constant w. If T is a $\{\mathcal{F}_t\}$-stopping time then we define here $\mathcal{F}_{T-}$ as the σ-field generated by the initial conditions $(X_1(0), \dots, X_K(0), S_0, B_0, C_0)$, T and $\{X(t) : t < T\}$. Regarding the *service policies* we assume that

$$P((B_n, C_n) \in \cdot|\mathcal{F}_{T^n-}, S(T^n)) = H_{i,X_i(T^n)} \quad P-\text{a.s. on } \{\psi_{n-1} = 0, S(T^n) = i\}, \quad n \ge 1, \tag{2.6}$$

where the $H_{i,x}$, $(i, x) \in \{1, \dots, K\} \times \mathbb{N}$, are probability measures on $(\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+$ satisfying $H_{i,x}((\mathbb{N} \cup \{\infty\}) \times \{0, \dots, x-1\}) = 1$. Let

$$\nu_1 := \min\{k \ge 1 : \psi_k = 0\}.$$

For all $i \in \{1, \dots, K\}$ and $x \in \mathbb{Z}_+$ there is a unique number $d_i(x)$ satisfying

$$d_i(x) = E[\nu_1|\mathcal{F}_{T^1-}, S(T^1) = i] \quad P-\text{a.s. on } \{\psi_0 = 0, S(T^1) = i, X_i(T^1) = x\}.$$

This is the number of customers the server has to serve at an average at station i with x customers waiting at his arrival's epoch. We assume the existence of the limits

$$d_i := \lim_{x\to\infty} d_i(x) \le \infty. \tag{2.7}$$

If $d_i = \infty$ then we call the service policy at station i *unlimited*. If $d_i < \infty$, then we assume that there is a finite $\tilde{d}_i$ satisfying

$$E[\nu_1|\mathcal{F}_0] \leq \tilde{d}_i \quad P-\text{a.s.on } \{\psi_0 > 0, S_0 = i\}, \tag{2.8}$$

and call the policy *limited*.

We now make the natural but important *Markovian assumption* that 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 $\psi_n = 0$ is independent of n and depends only on

$$\hat{X}(n) := (X_1(T_n), \ldots, X_K(T_n), \hat{S}(n), B_n, C_n), \tag{2.9}$$

where

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

This is in accordance with the independence properties of the arrival process. Consequently, $\{\hat{X}(n)\}$ is a homogeneous Markov chain with respect to the filtration

$$\hat{\mathcal{F}}_n := \mathcal{F}_{T_n-}, \quad n \in \mathbb{Z}_+,$$

where we recall that $\hat{\mathcal{F}}_0$ is generated by the initial conditions. This chain is our main object of interest. (Notice that $\hat{X}(n)$ is indeed measurable with respect to $\hat{\mathcal{F}}_n$ and that equality $T_n = T^{n+1}$ is not excluded.) Finally we assume that there is a constant $p > 0$ satisfying

$$P(A(W_1) = 0|\mathcal{F}_0) > p \quad P-\text{a.s. on } \{\psi_0 = 0, \sum_k X_k(0) > 0\}, \tag{2.10}$$

where

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

Hence, if the server has completed a batch of services at a moment when the system is not empty, then there is a uniformly positive chance that he will reach one of these non-empty stations before the next arrival. (Note that W_1 may depend on the arrival process.)

For examples illustrating the routing mechanism and walking times as well as assumptions (2.5) and (2.10) we refer to [4]. We will add here examples of service disciplines.

Example 2.1 Let h be a measurable and non-negative function on $\{1,\dots,K\} \times \mathbb{N} \times \mathbf{Y}$, where $(\mathbf{Y}, \mathcal{Y})$ is some measurable space. Assume that

$$B_n = h(i, X_i(T^n), \eta_n^i), \quad P-\text{a.s. on } \{\psi_{n-1} = 0, S(T^n) = i\}, \quad n \geq 1,$$

and $C_n \equiv 0$, where the (η_n^i) are i.i.d. sequences of random elements of $\mathbf{Y}$ which are independent of each other and independent of the arrival process, the service times and the walking times. In this case $H_{i,x}$ can be identified with the distribution $H'_{i,x}$ discussed in the introduction and given by the distribution of $h(i, x, \eta_1^i)$. Obviously, the definitions (1.2) and (2.7) yield indeed the same values. To mention a few special cases we fix an $i \in \{1,\dots,K\}$. The choice $h(i,x,y) \equiv \infty$ determines the exhaustive policy, while $h(i,x,y) = x$ defines the gated policy. These service policies are unlimited. An example of a limited policy is obtained by taking $\mathbf{Y} = \mathbb{N}$ and $h(i,x,y) = \min\{x,y\}$, and η_1^i is assumed to have a finite mean d_i.

Our model also allows for a convenient treatment of decrementing service policies:

Example 2.2 Assume that the server upon entering station i with $x > 0$ waiting customers provides service there until the queue length after a service completion has reduced to $x - 1$. Then $H_{i,x} = \delta_{(\infty,x-1)}$ determines the decrementing policy and d_i is the mean number of steps taken by a random walk to decrease by 1. The step size of this random walk is $A_i(\eta_i) - 1$, where η_i has distribution G_i and is independent of A_i. If $\lambda_i b_i < 1$, then this policy is limited. This example can be generalized. Thinking of the customers as units of works it is quite natural to decrease the queue length not only by 1 but by a finite (possibly random) number. The random number B need not equal deterministically ∞ but could be used to model the maximal time the server is willing to stay at station i.

3 Proof of the stability results

The main part of this section is devoted to the proof of a criterion for stability which is based on the maximal nominal load of the system defined by (1.1) and (2.4). For shortness we let $\mathbf{X} := \mathbb{Z}_+^K \times \{1,\dots,K\} \times (\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+$ denote the state space of the Markov chain $\{\hat{X}(n)\}$. For any $A \subseteq \mathbf{X}$ we define

$$\tau_A := \inf\{n \geq 1 : \hat{X}(n) \in A\},$$

where $\inf \emptyset := \infty$. We call A *positive recurrent* (for the Markov chain $\{\hat{X}(n)\}$) if

$$P_x(\tau_A < \infty) = 1, \quad x \in \mathbf{X},$$

and

$$\sup_{x \in A} E_x \tau_A < \infty,$$

where E_x denotes expectation with respect to P_x. Here P_x is the governing probability measure if $\hat{X}(0) = x$, which is a standard notation for Markov chains (see e.g. Meyn and Tweedie [8] or the Appendix in [4]).

Theorem 3.1 *Assume that $\rho^+ < 1$. Then:*

(i) The set $A := \{(0, \ldots, 0)\} \times \{1, \ldots, K\} \times (\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+$ is positive recurrent.

(ii) If 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 $\sum_{i=1}^K X_i(T_n) = 0$ is independent of B_n, then each essential state of $\{\hat{X}(n)\}$ is positive recurrent.

Note that the additional assumption in (ii) is very natural and could have assumed from the very beginning.

Remark 3.2 Assume that there are no batch arrivals, i.e. that $A_1, \ldots, A_K$ are independent homogeneous Poisson processes. Then there is only one absorbing class of communicating states for $\{\hat{X}(n)\}$. Indeed, taking some $j \in \{1, \ldots, K\}$ and some m with $H_{j,1}(\{m\} \times \mathbb{Z}_+) = H_{j,1}(\{(m, 0)\})$ one can use assumption (2.10) to show that state $(0, \ldots, 0, j, m, 0)$ can be reached from any other state with a positive probability.

To prove Theorem 3.1 it is now convenient to consider 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

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

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

$$\hat{X}(n) := (X_1(T_n), \ldots, X_K(T_n), \hat{S}(n), B_n, C_n, U(T_n)), \quad \hat{\mathcal{F}}_n := \mathcal{F}_{T_n}, \tag{3.1}$$

is a homogeneous Markov process fitting the general setting of the Appendix of [4].

We define a *test function* V on $\mathbf{X} \times \mathbf{U}$ by

$$V(x) := r_1 x_1 + \ldots + r_K x_K, \quad x = (x_1, \ldots, x_K, i, m, k, u), \tag{3.2}$$

where

$$r_i := b_i + \mathbf{1}\{d_i < \infty\} w_i^+/d_i.$$

If all stations are served according to an unlimited policy, then $V(x)$ is the mean work in the system with x_i customers in queue i. In general the function takes into account the walking times starting at a station with a limited service policy which could be considered as additional work the system has to cope with. Roughly speaking, if x_i is large then w_i^+/d_i is an upper bound for the mean additional work per customer. With this modified interpretation of work, function $V(x)$ provides an upper bound for the expected work load in the system. It is natural to expect that a negative drift of $V(x)$ outside bounded sets should result in a ergodic behaviour. However, we cannot establish this drift condition at deterministic times but rather at random stopping times that might depend on the whole process history. To make these ideas rigorous we formulate three lemmas the first of which is an adaption of Lemma 3.5 in [4] to the present more complicated situation. The proofs of the other lemmas however rely heavily on the properties of the greedy routing.

We need to introduce some further notation. Let $\{\hat{\theta}_n : n \in \mathbb{Z}_+\}$ be the flow of shift operators associated with the process $\{\hat{X}(n)\}$, see the Appendix in [4]. A family $Z = \{Z(n) : n \geq 0\}$ is called a *subadditive* process (w.r.t. the filtration $\{\hat{\mathcal{F}}_n\}$ and the shift $\hat{\theta}$) if $Z(n)$ is $\hat{\mathcal{F}}_n$-measurable for each $n \in \mathbb{Z}_+$ and

$$Z(m+n) \leq Z(m) + Z(n) \circ \hat{\theta}_m, \quad m, n \in \mathbb{Z}_+.$$

Define $\nu_0 := 0$ and

$$\nu_{n+1} := \min\{k > \nu_n : \psi_k = 0\}, \quad n \in \mathbb{Z}_+.$$

If $\psi_0 = 0$ (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.

Lemma 3.3 *Assume that σ is an $\{\hat{\mathcal{F}}_n\}$-stopping time and $0 \leq c < 1$, $L_0, c_1 \in \mathbb{R}_+$, are constants, satisfying*

$$E_x V(\hat{X}(\sigma)) \leq cV(x) \quad \text{if } V(x) > L_0, \tag{3.3}$$

$$E_x \sigma \leq c_1 V(x) \quad \text{if } V(x) > L_0. \tag{3.4}$$

Then

$$E_x \tau \leq a_{L_0} V(x) + c_{L_0}, \quad x \in \mathbf{X}, \tag{3.5}$$

where

$$\tau := \min\{n \geq 1 : |\hat{X}(n)| = 0\},$$

($\min\emptyset := \infty$*), and* a_{L_0}, c_{L_0} *are constants which may depend on* L_0*. Assume in addition that* $\{Z(n)\}$ *is a subadditive process satisfying*

$$E_x Z(\sigma) \le Q_L V(x) \quad \text{if } V(x) > L, \quad L \ge L_0, \tag{3.6}$$

where $\lim_{L\to\infty} Q_L = 0$*. Then, for all* $L \ge L_0$*, there are numbers* $\tilde{d}_L$ *and* $\tilde{Q}_L$ *such that*

$$E_x Z(\tau) \le \tilde{d}_L + \tilde{Q}_L V(x) \quad \text{if } V(x) > L/c \tag{3.7}$$

and $\lim_{L\to\infty} \tilde{Q}_L = 0$.

PROOF. The proof is an obvious modification of the proof of Lemma 3.5 in [4]. The role of the pair (σ, ν_σ) in the latter lemma is now played by $(Z(\sigma), \sigma)$. We omit further details. □

Let

$$N_i(m) := \text{card}\,\{0 \le n \le m : S(T^n) = i, \psi_n = 0\}$$

be the number of walks started at station i by time T_m and

$$D_i(m) := \text{card}\,\{1 \le n \le m : S(T^n) = i\}$$

be the number of departures from station i by time T_m and

$$C_i(m) := \text{card}\left\{0 \le n \le m : \psi_n = 0, \sum_{i \in N(S(T^n))} \hat{X}_i(n) = 0\right\}.$$

We would like to prove the existence of a stopping time σ satisfying the assumptions of the preceding lemma with

$$Z(n) := Z_1(n) + Z_2(n) + Z_3(n),$$

where

$$\begin{aligned} Z_1(n) &:= \sum_{i:d_i<\infty} w_i^+ \left(N_i(n) - \frac{D_i(n)}{d_i}\right), \\ Z_2(n) &:= \rho^+ w \sum_{i:d_i=\infty} N_i(n), \\ Z_3(n) &:= \rho^+ w \sum_i C_i(n), \end{aligned}$$

and w is the upper bound for the means of the walking times, see (2.5).

For $x = (x_1, \ldots, x_K, i, m, k, u) \in \mathbf{X} \times \mathbf{U}$ we define $|x| := x_1 + \ldots + x_K$, $s(x) := i$, and $\psi(x) := \min\{m - 1, \mathbf{1}\{x_i \ne k\}\}$.

Lemma 3.4 *For all $\varepsilon > 0$ there exists an $n \in \mathbb{N}$ such that, for all $m \geq n$,*

$$E_x[Z_1(m) + Z_2(m)] \leq \varepsilon m \quad \text{if } x_{s(x)} \geq 2m.$$

PROOF. By assumption (2.7) we find for any $\varepsilon' > 0$ and all $i \in \{1, \ldots, K\}$ with $d_i < \infty$ an $n' \in \mathbb{N}$ such that, for all $m \geq n'$,

$$|E_x\nu_1 - d_i| \leq \varepsilon' \quad \text{if } s(x) = i, \psi(x) = 0, x_i \geq m.$$

Assume that $x_{s(x)} \geq 2m$ for some $m \geq n'$. Due to the greedy walking mechanism all batches of services starting before T_m start at a station with at least m waiting customers. Hence, using also (2.8),

$$\begin{aligned}
&\sum_{i:d_i<\infty} E_x\left[N_i(m) - \frac{D_i(m)}{d_i}\right] \\
&= \sum_{i:d_i<\infty} E_x\left[\sum_{l=0}^{\infty} \mathbf{1}\{\nu_l < m \leq \nu_{l+1}\}[\mathbf{1}\{S(T^1) = i\}\frac{\nu_1}{d_i}\right. \\
&\quad \left. + \sum_{k=1}^{l-1} \mathbf{1}\{S(T^{\nu_k+1}) = i\}\left(1 - \frac{\nu_{k+1} - \nu_k}{d_i}\right) + \mathbf{1}\{S(T^{\nu_l}) = i\}\frac{(m - \nu_l)}{d_i}\right]\Bigg] \\
&\leq a_1\varepsilon' \sum_{i:d_i<\infty} E_x N_i(m) + a_2,
\end{aligned}$$

where a_1, a_2 are generic constants which do not depend on n'. Since

$$\sum_i D_i(n) = n, \tag{3.8}$$

$E_x N_i(n')$ can be bounded by a linear function of n' if $d_i < \infty$. Hence we may conclude the assertion for Z_1.

Also by assumption we find for any $C > 0$ and all $i \in \{1, \ldots, K\}$ with $d_i = \infty$ an $n' \in \mathbb{N}$ such that, for $m \geq n'$,

$$E_x\nu_1 \geq C \quad \text{if } s(x) = i, x_i \geq m.$$

Hence we get for $x_{s(x)} \geq 2m$ and $m \geq n'$ that

$$\sum_{i:d_i=\infty} E_x D_i(m)] \leq a_3 + C \sum_{i:d_i=\infty} E_x N_i(m)$$

for a constant a_3 not depending on n'. Taking into account (3.8), we get the assertion for Z_2 and the lemma is proved. □

Lemma 3.5 *Assume that $\rho^+ < 1$. Then there exists a $\{\hat{\mathcal{F}}_n\}$-stopping time σ and constants $c \in [0, 1)$, $L_0, c_1 \in \mathbb{R}_+$ such that inequalities (3.3), (3.4) and (3.6) in Lemma 3.2 are satisfied.*

PROOF. As in [4] we proceed by induction on the number of stations. For $K = 1$ we choose $\sigma := \max\{k \in \mathbb{N} : k \le |X(0)|/2\}$. Then equality (3.4) is trivial while (3.3) can be proved as in the induction step. Inequality (3.6) reduces to the corresponding inequality for Z_1 and follows from Lemma 3.4.

Now we consider a system with $K + 1$ stations. As in [4] we couple the process

$$X(t) = (X_1(t), \dots, X_{K+1}(t), S(t), B(t), C(t), U(t))$$

with another auxiliary process $\tilde{X}(t)$ describing a polling system with stations $\{1, \dots, K\}$ that behaves like the original system until the time when the server first enters station $K + 1$. Hence we define

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

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 $t \ge \tilde{T}_\infty$, where u_∞ is not in the space $\mathbf{U} \times \mathbb{Z}_+$. Consequently $\tilde{U}(t)$ takes values in the set $\mathbf{U} \times \mathbb{Z}_+ \cup \{u_\infty\}$. It is an easy technical point to define the dynamics of $\tilde{X}(t)$ after the epoch $\tilde{T}_\infty$ such that all assumptions of this section are satisfied. Therefore we can use the induction hypothesis together with a geometrical trial argument based on assumption (2.10) to conclude from Lemma 3.3 that (3.5) and (3.7) are satisfied with τ replaced by the stopping time

$$\min\{n \ge 1 : \hat{S}(n) = K + 1\},$$

if $X_{K+1}(0) > 0$, see [4]. Of course this conclusion remains true for the stopping time

$$\tilde{\sigma} := \min\{n \ge 1 : \hat{S}(n) = \xi\},$$

where ξ is an $\mathcal{F}_0$-measurable random element of $\{1, \dots, K + 1\}$ satisfying

$$X_\xi(0) \ge \frac{|X(0)|}{K + 1}.$$

We claim that

$$\sigma := \tilde{\sigma} + \max\left\{k \in \mathbb{N} : k \le \frac{|X(0)|}{2(K + 1)}\right\} \tag{3.9}$$

satisfies (3.3), (3.4) and (3.6) for suitable chosen constants. The validity of (3.4) is obvious. Using Lemma 3.4, the equation $Z_3(\sigma) = Z_3(\tilde{\sigma})$ and the definitions and properties of $\tilde{\sigma}$ and σ, we obtain for any $\varepsilon' > 0$ that

$$\begin{aligned} E_x Z(\sigma) &= E_x Z(\tilde{\sigma}) + E_x[E_x[Z(\sigma) - Z(\tilde{\sigma}) | \mathcal{F}_{T^{\tilde{\sigma}+1}}]] \\ &\leq \tilde{Q}_{L'} V(x) + \tilde{d}_{L'} + \varepsilon' \frac{|x|}{2(K+1)} \end{aligned}$$

if $|x| > L'$ and L' has been chosen large enough. For any $\varepsilon > 0$ this expression can be made smaller than $\frac{\varepsilon}{2} V(x) + \tilde{d}_{L''}$ if $V(x) > L''$ and L'' is chosen large enough. Hence there is an $L > 0$ with

$$E_x Z(\sigma) \leq \varepsilon V(x) \quad \text{if } V(x) > L,$$

proving (3.6). It remains to check the drift condition (3.3). Using a similar calculation as in [4] we obtain that

$$\begin{aligned} E_x T_\sigma &= E_x \Bigg[\sum_{m=0}^{\infty} \mathbf{1}\{\nu_m < \sigma \leq \nu_{m+1}\} \\ &\quad \Bigg(\Bigg(\sum_{k=0}^{m-1} W_{\nu_k+1} + \sum_{n=\nu_k+1}^{\nu_{k+1}} (T_n - T^n) \Bigg) + \Bigg(W_{\nu_m+1} + \sum_{n=\nu_m+1}^{\sigma} (T_n - T^n) \Bigg) \Bigg) \Bigg] \\ &\leq \sum_{i: d_i < \infty} w_i^+ E_x N_i(\sigma) + w \sum_{i: d_i = \infty} E_x N_i(\sigma) \\ &\quad + \sum_i b_i E_x D_i(\sigma) + w \sum_i E_x C_i(\sigma), \end{aligned}$$

where we recall the definition $W_1 := 0$ if $\psi_0 > 0$. From

$$\hat{X}_i(\sigma) = \hat{X}_i(0) + A_i(T_\sigma) - D_i(\sigma),$$

the assumptions on the arrival process, definition (3.2) of test function V, and definition (1.1) of ρ^+ we have

$$\begin{aligned} E_x V(\hat{X}(\sigma)) &= V(x) + \sum_i r_i \lambda_i E_x T_\sigma - \sum_i r_i E_x D_i(\sigma) \\ &\leq V(x) - (1 - \rho^+) E_x \kappa(\sigma) + E_x Z(\sigma), \end{aligned}$$

where

$$\kappa(n) := \sum_{i: d_i < \infty} w_i^+ N_i(n) + \sum_i b_i D_i(n). \tag{3.10}$$

By definition (3.9) of σ, $E_x\kappa(\sigma) \geq \tilde{C}V(x)$ for a certain constant $\tilde{C}$ and hence

$$E_x V(\hat{X}(\sigma)) \leq c' V(x) + E_x Z(\sigma),$$

where $c' < 1$. Choosing a c with $c' < c < 1$ and using (3.6) we conclude (3.3) for large enough L_0. □

In the remainder we return to the more specific setting of Section 2, where process $U(t)$ is constant.

Proof of Theorem 3.1: The first assertion follows directly from (3.5). To prove the second we now restrict $\{\hat{X}(n)\}$ to an absorbing class B of essential states. Let $\tau^{(n)}$, $n \in \mathbb{N}$, be the time of the nth visit of the set A defined in the theorem. (We must have $B \cap A \neq \emptyset$.) By assumption, $Z(n) := \hat{S}(\tau^{(n)})$, $n \in \mathbb{N}$, is a homogeneous Markov chain and we denote the set of its essential states by **S**. By (3.5) this set is not empty because, starting with any initial conditions in B, the chain $\hat{X}(n)$ must hit at least one of the sets

$$A_i := \{(0, \ldots, 0, i)\} \times (\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+$$

$i = 1, \ldots, K$ infinitely often. By irreducibility of $\{\hat{X}(n)\}$ on B, all states of **S** must communicate and since **S** is finite, all states in **S** are positive recurrent for the chain $\{Z(n)\}$. Using the strong Markov property and positive recurrence of A it is now easy to prove that A_i is positive recurrent for $\{\hat{X}(n)\}$ whenever $i \in$ **S**. Now the assertion can be deduced from standard splitting techniques (see e.g. [8], pp. 102, pp. 422) upon noticing that the conditional distribution of $\{X(t) : t \geq T_n\}$ given $\hat{X}_n$ does not depend on (B_n, C_n) provided that $\hat{X}(n) \in A_i$.

Now we prove conditions sufficient for instability. We use the notation introduced before Lemma 3.4.

Theorem 3.6 *For all $i \in \{1, \ldots, K\}$ with $d_i < \infty$ let w_i^- be a number satisfying*

$$E_x W_1 \geq w_i^- \quad \text{if } s(x) = i, \psi(x) = 0,$$

and let $w_i^- := 0$ if $d_i = \infty$. Assume that $d_i(x) \leq d_i$ for all i with $d_i < \infty$ and all x. If the chain $\{\hat{X}(n)\}$ is ergodic (i.e. aperiodic with only one absorbing class of positive recurrent states), then $\rho^- \leq 1$, where ρ^- is given by (1.3).

PROOF. For all $n \in \mathbb{N}$ we have as in [4]

$$E_x X_i(T_{\nu_n}) = x_i + \lambda_i E_x(W(n) + \zeta_n) - E_x D_i(n), \tag{3.11}$$

where

$$\zeta_n := \sum_{i=1}^{K} b_i D_i(n),$$
$$W(n) := \sum_{m=0}^{n-1} W_{\nu_m+1}.$$

Assume that $\{\hat{X}(n)\}$ is ergodic. Then the chain $Y(n) := \hat{X}(\nu_n)$, $n \in \mathbb{N}$, is also ergodic and we denote its equilibrium distribution by π. By our Markovian assumption we have a stochastic kernel K from $\mathbf{X}$ to $\mathbb{Z}_+^K \times \{1, \ldots, K\}$ satisfying

$$P_y(Z(n) \in \cdot | Y(n)) = K(Y(n), \cdot) \quad P_y - \text{a.s.}$$

for all $y \in \mathbf{X}$, where $Z(n) := (X_1(T^{\nu_n+1}), \ldots, X_K(T^{\nu_n+1}), S(T^{\nu_n+1}))$. Obviously,

$$\pi'(\cdot) := \sum_{x \in \mathbf{X}} K(x, \cdot)\pi(x)$$

is a equilibrium distribution for the chain $\{Z(n) : n \in \mathbb{N}\}$. As in [4] we may conclude that

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

satisfies $\rho \neq 1$ and that

$$\bar{w} := \sum_{y} w(y)\pi(y),$$

is a positive number, where

$$w(y) := E_y W_1 \quad \text{if } \psi(y) = 0,$$

is the mean walking time if y is the state of the system just after completion of a batch of services.

We denote

$$\pi_i := \sum_{x \in \mathbf{X}} \mathbf{1}\{s(x) = i\}\pi(x) = \sum_{x \in \mathbb{Z}_+^K} \pi'((x, i)).$$

As in [4] we get

$$\frac{\lambda_j \bar{w}}{1 - \rho} = \bar{B}_j \pi_j, \quad j \in \{1, \ldots, K\}, \tag{3.12}$$

where

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

is the stationary average size of a batch of services given that the server is at station i. If $w_i^- = 0$ for all i, then $\rho^- = \rho < 1$. Therefore we will assume now that

$$w^- := \sum_{i: d_i < \infty} w_i^- \pi_i$$

is positive. By assumption, $\bar{w} \geq w^-$ and $\bar{B}_i \leq d_i$. Hence

$$\frac{\lambda_j w^-}{(1-\rho) d_j} \leq \pi_j. \tag{3.13}$$

Multiplying this equation with w_j^- and summing up over all j satisfying $d_j < \infty$ we obtain that

$$\sum_{j: d_j < \infty} \frac{\lambda_j w_j^-}{(1-\rho) d_j} \leq 1$$

which implies $\rho^- \leq 1$. □

We note in passing that the preceding proof did not use any specific assumptions on the routing mechanism.

Corollary 3.7 *If the chain $\hat{X}(n)$ is ergodic, then (3.12) holds.*

In most cases the assumptions of Theorem 3.6 imply the strict inequality $\rho^- < 1$:

Corollary 3.8 *Let the assumptions of Theorem 3.6 be satisfied and let π and π' be as introduced in the proof above. Any of the following conditions implies that $\rho^- < 1$.*

(i) $d_i = \infty$ *for some* i *with* $\pi(\{(0,i)\} \times (\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+) > 0$.

(ii) $w(x) > w_i^-$ *for some* $i \in \{1, \ldots, K\}$ *and* $x \in \mathbf{X}$ *with* $s(x) = i$, $d_i < \infty$ *and* $\pi(\{x\} \times (\mathbb{N} \cup \{\infty\}) \times \mathbb{Z}_+) > 0$.

(iii) $d_i(x) < d_i$ *for some* $i \in \{1, \ldots, K\}$ *and* $x \in \mathbb{N}$ *with* $d_i < \infty$ *and* $\pi'((x,i)) > 0$.

PROOF. In the first case we have $\bar{w} > w^-$ because $w(x) > 0$ if $|x| = 0$. Hence inequality (3.13) is in fact strict implying that $\rho^- < 1$. The other two cases follow similarly because we then have $\bar{w} > w^-$ or $\bar{B}_i < d_i$, respectively. □

Remark 3.9 If $\rho^- = 1$, then the chain $\hat{X}(n)$ might be null-recurrent. However, if $\rho^- > 1$, then it is not difficult to see that

$$\lim_{n\to\infty} \sum_k \hat{X}_k(n) = \infty \quad P-\text{a.s.}$$

4 Examples and discussion

The proof of the stability result in the last section relies essentially on the greedy walking mechanism. Our first example shows that if we allowed for a greedy-type mechanism as in [4], then the system might become unstable even if $\rho^+ < 1$.

Example 4.1 Consider a polling system with two stations, where the first station is served according to the exhaustive policy and the second according to $H'_{2,x} = \delta_{\min\{x,1\}}$, see Example 2.1. After the server has completed a batch of services at station 1 or one service at station 2 he chooses a non-empty station to be served next, where each station gets an equal chance if both are non-empty. If the system is empty, then the server waits till the next customer will have arrived. It takes the server a fixed time w to get from station 1 to 2. All other walking times are supposed to be zero. Customers arrive at each station according to a Poisson process with intensity λ and service times are supposed to equal 1. This model fits the setting of the present paper except for the walking mechanism which is greedy-type, see [4].

Let us assume that $\rho^+ := 2\lambda < 1$. Assume that at time 0 the server is at station 1 which is empty and let Z_0 be the corresponding queue length at station 2. Let Z_1 denote the number of customers in station 2 after the completion of the first batch of services at station 1. A routine calculation shows that

$$E[Z_1|Z_0] \geq Z_0 + \lambda w + c \quad \text{if } Z_0 \geq L_0,$$

where L_0 and c are constants and c might be negative. Hence the underlying Markov chain cannot be ergodic if w is large enough. The reason is clearly that lots of customers may arrive during the walk from station 1 to station 2. Because the walking mechanism is not greedy but only greedy-type the server leaves station 2 after a geometrically distributed random time as soon as station 1 became non-empty no matter what the queue size at station 2 is.

Our next example will show that in case $\rho^- < \rho^+$ the stability region can depend on the whole distributions of the underlying walking times. Without additional assumptions it seems to be a difficult task to determine that region explicitly.

Example 4.2 Consider a system of two stations each of which is fed by a Poisson process with intensity λ. Each queue is served according to the policy $H'_{1,x} = H'_{2,x} = \delta_1$ and service times are deterministically equal one. After completion of a service the server chooses the station with the longest queue for his next visit, where each station gets an equal chance if the queue lengths coincide. If the system is empty, then the server stays at the station he is at and waits for the next arriving customer. It takes the server no time to travel from station i to i, $i = 1, 2$, while the walking times from one station to the other follow a distribution W with finite mean w. Note that $\rho^- = \lambda$ and $\rho^+ = \lambda(1 + w)$. The transition probabilities of the chain $\hat{X}(n)$ are homogeneous on $\mathbb{N} \times \mathbb{N} \times \{1, 2\}$ and can hence be used to define an (auxiliary) chain $X'(n) = (X_1'(n), X_2'(n), S'(n))$ with state space $\mathbb{Z} \times \mathbb{Z} \times \{1, 2\}$ which behaves on $\mathbb{N} \times \mathbb{N} \times \{1, 2\}$ like $\hat{X}(n)$. The process $M(n) := (X_1'(n) - X_2'(n), S'(n))$ is again an irreducible aperiodic Markov chain and satisfies, for $i = 1, 2$,

$$E[|M_1||M_0 = (m, i)] = |m| - 1, \quad |m| \geq 1,$$

because the arrival intensities coincide. Foster's criterion shows that $M(n)$ is ergodic. Let p be the probability of the event $S'(0) \neq S'(1)$ under the equilibrium distribution of $M(n)$. Let us run the chain $X'(n)$. After a random time β satisfying $E[\beta|X'(0)] = c_1|X(0)|$, where $|X(n)| := X_1'(n) + X_2'(n)$ and c_1 is a constant, the chain $M(n)$ couples with its stationary version. Thereafter, on an average, $\lambda + \lambda wp$ customers arrive between two service completions. Therefore, letting $T := c_1(1 + C)|X'(0)|$, for some (large) $C > 0$, one can prove that

$$E[|X'(n)| - |X'(0)||X'(0)] = c_1 C(\lambda + \lambda wp - 1)|X'(0)| + O(|X'(0)|).$$

Hence the chain $\hat{X}(n)$ is stable if $\lambda + \lambda wp < 1$. If $\lambda + \lambda wp > 1$, then the chain is transient. The stability region is hence determined by the value of the constant p. Since the transition probabilities of the chain $M(n)$ are determined by the distribution W, the constant p may, in general, depend on the whole distribution W and, in particular, on any finite number of moments of W.

Theorem 3.5 uses the assumption $d_i(x) \leq d_i$ for all i and x. If this assumption fails, then we do not know whether inequality $\rho^- > 1$ implies the transience. For this conclusion we need additional assumptions on the service disciplines. Assume for example that the service policies are of the type described in Example 2.1. If for all stations i with a limited policy the function $h(i, \cdot, y)$ is monotone decreasing for all y and $\lim_{x \to \infty} h(i, x, \eta_1^i)$ is integrable, then it is possible to prove that $\rho^- > 1$ is sufficient for transience.

If the assumptions of Theorem 3.5 are violated, then in case $\rho^- = 1$ both is possible, stability or instability of the system:

Example 4.3 Consider a model with only one station, i.e. a single server with *vacations*. Assume that service times and walking (or vacation) times are identically 1 and

w, respectivly. The service discipline is defined as in Example 2.1 with $\mathbf{Y} = \mathbb{N}$ and

$$h(x, y) := \min\{x, 1 + \mathbf{1}\{y > x\}\},$$

i.e., there is an i.i.d. sequence (η_n) such that, with obvious notation,

$$B_n = h(X(T^n), \eta_n) \quad P-\text{a.s. on } \{\psi_{n-1} = 0\}, \quad n \geq 1.$$

Then, for $x \geq 2$, $d(x) = 1 + a_x$, where $a_x := P(\eta_1 > x)$ and $d := \lim_{x\to\infty} d(x) = 1$. In the following we assume that $\rho^+ = \rho^- = \lambda(w + 1) = 1$. A direct computation yields, for $x \geq 2$,

$$E[\hat{X}(1) - x|\hat{X}(0) = x] = -(1 - \lambda)a_x.$$

This term is negative but at the moment nothing can be said on recurrence of the chain $\hat{X}(n)$. Taking a quadratic test function one obtains, using straightforward computations, that

$$E[(\hat{X}(1))^2 - x^2|\hat{X}(0) = x] = 1 - 2(1 - \lambda)xa_x + a_x - \lambda a_x + \lambda^2 a_x.$$

Assume that $\lim_{x\to\infty} xa_x = m$ for some positive constant m. In particular, $E\eta_1 = \infty$. Then the above expression tends to $1-2(1-\lambda)m$ and if this number is negative positive recurrence follows. If $2(1 - \lambda)m < 1$, then one can use the approach of Lamperti [6] and Fayolle [1] to prove transience of the chain $\hat{X}(n)$.

In the following we skip the Poisson assumptions on the input and discuss the possibility to extend our results to more general arrival processes of the form

$$A_i(t) = \sum_{n\geq 1} \mathbf{1}\{\tau_n \leq t\}B_n^i, \quad t \geq 0, i = 1, \ldots K,$$

where $\tau_1 < \tau_2 \ldots$ are the arrival epochs of batches of customers and $B_n = (B_n^1, \ldots, B_n^K)$, $n \geq 1$, are random elements of $\mathbb{Z}_+^K$ satisfying $\sum_{i=1}^K B_n^i > 0$. A general reasonable assumption is that, at any time $t \geq 0$, the conditional distribution of the future of the arrival process $A := \{(A_1(t), \ldots, A_K(t)) : t \geq 0\}$ given the complete history $\mathcal{F}_t$, depends only on the corresponding history of A. In order to conclude the results of this paper one needs further regenerative-type assumptions. We give two examples.

Example 4.4 (i) The $(\tau_n - \tau_{n-1}, B_n)$, $(\tau_0 := 0)$ are independent and have the same distribution for $n \geq 2$.

(ii) There are K independent sequences $(\tau_n^i, \tilde{B}_n^i)$ of random elements of $\mathbb{R}_+ \times \mathbb{N}$ such that

$$A_i(t) = \sum_{n\geq 1} \mathbf{1}\{\tau_n^i \leq t\}\tilde{B}_n^i, \quad t \geq 0, i = 1, \ldots K.$$

The $(\tau_n^i - \tau_{n-1}, \tilde{B}_n^i)$, $(\tau_0^i := 0)$ are independent and have the same distribution for $n \geq 2$.

Taking the arrival processes in the above examples, the results of this paper as well as those of [4] can still be proved under additional smoothness assumptions on the underlying interarrival distributions. This remains even true for a more general class of marked point processes with an inherent regenerative structure, see [3] for the definition. However, a proof would require a lot of technical effort without yielding more insight into the basic ideas. Similar generalizations apply to service and walking times.

Our final remark concerns possible modifications of the greedy routing mechanism.

Remark 4.5 Consider the following generalization of the greedy routing mechanism. Assume that the server is at station i and chooses his next customer randomly in the set

$$\{j : g_j(x_j) \geq \max\{x_k : k \in N(i)\}\}$$

where the $g_j : \mathbb{Z}_+ \to \mathbb{R}_+$ are functions satisfying

$$\liminf_{m\to\infty} g_j(m)/m > 0.$$

It is easy to see from the proof that inequality $\rho^+ < 1$ is still sufficient for ergodicity.

Acknowledgment

We would like to thank Professor Rolf Schassberger for many fruitful and stimulating discussions on the subject of this paper.

References

[1] Fayolle, G. (1989) On random walks arising in queueing systems: ergodicity and transience via quadratic forms as Lyapunov functions-Part I. *Queueing Systems* **5**, 167–184

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

[3] Foss, S. (1991) Ergodicity of queueing networks. *Sib. Math. J.* **32**, 184–203

[4] Foss, S., Last, G. (1994) Stability of polling systems with exhaustive service policies and state dependent routing. *Ann. Appl. Probab.* **6**, 116-137

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

[6] Lamperti, J. (1960) Criteria for the recurrence or transience of stochastic processes. *I. J. Math. Anal. Appl. 1*, 314–330

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

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

[9] Schassberger, R. (1995) Stability of polling networks with state-dependent server routing. *Probab. Eng. Inform. Sc.* **9**, 539-550

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