Stability of Periodic Polling System with $BMAP$ arrivals

Zsolt Saffer , Miklós Telek

Department of Telecommunications, Technical University of Budapest, 1521 Budapest Hungary

Abstract

This paper considers the stability of $BMAP/GI/1$ periodic polling models with mixed service disciplines. The server attends the N stations in a repeating sequence of stages. Customers arrive to the stations according to batch Markov arrival processes ($BMAP$s). The service times of the stations are general independent and identically distributed. The characterization of global stability of the system, the order of instability of stations and the necessary and sufficient condition for the stability are given. Our stability analysis is based on the investigation of the embedded Markov chains at the polling epochs, which allows a much simpler discussion than the formerly applied approaches. This work can also be seen as a survey on stability of a quite general set of polling models, since the majority of the known results of the field is a special case of the presented ones.

Key words: queueing, stability, polling model, service discipline, BMAP.

1 Introduction

The stability of polling models has been investigated for several decades. In this paper we introduce a new framework that allows the stability analysis of a wide set of periodic polling models. In the considered set of polling models each station has an infinite queue, a $BMAP$ arrival stream and a general service time distribution (GI). The server attends the stations in repeating sequences of stages. A stage is an interval the server spends at a station. The repeating sequence of stages is referred to as cycle, in which each station is visited at least once. The time required for the server to travel from one station

⋆ This work is partially supported by the OTKA K61709 grant.
Email address: {safferzs,telek}@hit.bme.hu (Zsolt Saffer , Miklós Telek).

to the next one is the switchover time. The periodic model is one among the several polling model variants. For a summary on different polling models see the book of Takagi [1].

One of the first works on the stability of polling systems is the early paper of Kuehn in 1979 [2], who gave a heuristic sufficient stability condition for the l-limited token ring. However this condition was derived without formal proof. The proof of the stability conditions of the same polling model was given first by Georgiadis and Szpankovsky in 1992 [3].

The stability of the polling model with Poisson arrivals and with general independent service times and switchover times was studied by many authors. Altman, Konstantinopoulos and Liu [4] used Foster's criteria to derive sufficient conditions for the stability of cyclic polling systems with mixed service policies. In the work of Borovkov and Schassberger [5] the polling model with Markovian server routing and with limited gated service policy is investigated. They studied the ergodicity and stability of the model by Lyapounov functions. Fricker and Jaïbi [6] studied the stability for periodic polling model. They applied monotonicity arguments, which utilize the monotonicity property of service policies.

Stability conditions based on fluid models associated to Markov processes have been developed for more general polling models. Down [7] presented the stability condition for polling model with multiple servers and with general independent interarrival and service times. His model with server routing is a generalization of the cyclic model. Foss, Chernova and Kovalevskii [8] investigated the stability of multi-server polling models with state-independent server routing. Foss and Kovalevskii [9] introduced a generalized criterion for the stability of Markovian queueing systems and considered a polling system with two stations and two heterogeneous servers as an example.

Some stability results are also available for polling models with general stationary input. Massoulie [10] gave a sufficient but not necessary condition for the stability of polling models with Markovian server routing. In [10], the interarrival times of the customers and their service times are assumed to be jointly stationary, ergodic processes, which are independent of the switchover times and of the routings of the server. The service policy is typically gated or binomial-gated. Foss and Chernova [11] presented the sufficient and necessary stability condition for polling models with state-independent server routing allowing general assumption on service policies. The arrival times of the customers and their service times form a common general, stationary egodic input flow, where however the customers are randomly routed to the stations. The proofs of [11] are based on the monotonicity properties of the model and dominance theorems.

Recently Lillo [12] gave an ergodicity analysis for polling systems with two queues. The proofs of [12] rely on the combination of three embedded processes that were previously used in the literature. For a recent survey of the available results on polling models we refer to Vishnevskii and Semenova [13].

In this paper, we investigate the stability of the polling model, but in contrast to the above references, we consider the periodic polling models with general assumptions on service policies and with $BMAP$ arrival processes. To the best of our knowledge no stability criterions are available for these models. The intended contributions of this paper are twofold. The first one is the applied framework based on the properly chosen embedded Markov chains. This framework allows the generalization of the arrival process to BMAPs and a much simpler stability analysis than the existing ones (e.g., the one based on monotonicity properties and dominance theorems). The second contribution is the complete survey of stability results for a fairly general set of polling models, which includes:

- necessary assumptions for $BMAP$ arrival processes and for the service disciplines,
- characterization of global stability,
- order of instability of stations,
- conditions for partial stability,
- necessary and sufficient condition for the stability states of the system.

The rest of this paper is organized as follows. Section II gives the introduction to the model and the notations. In section III we characterize the global stability. The analysis of the particular stations follows in section IV. Stability relationships and the order of instability are studied in section V. In section VI we give the conditions for partial stability and for the stability of the whole system. Concluding remarks are given in section VII.

2 Model and Notation

2.1 The $BMAP/GI/1$ periodical polling model

We consider a continuous-time asymmetric polling model with N stations [1]. A single server attends the stations in a repeating sequence of V stages, defined by the polling table $t : \{1, \ldots, V\} \to \{1, \ldots, N\}$, where $t(l)$ is the station attended by the server at stage l. Each station has infinite buffer queues, which is served when the server attends that station [14]. A stage is the time interval during which the server works continuously on a single station, and the cycle is the time needed for the server to accomplish all the

$1, \ldots, V$ consecutive stages. If no customer is present at a stage, the server immediately attends the next station according to the polling table. Station i is attended by the server $V_i > 0$ times in a cycle and $\sum_{i=1}^{N} V_i = V$. We refer to the j-th stage belonging to station i as the $i(j)$-stage. An i-stage is any of these $i(j)$-stages. At each station batch of customers arrive according to a $BMAP$ process satisfying assumptions B.1, B.2 (see subsection 2.2). We call the $BMAP$ arrival process at station i as i-th $BMAP$ arrival process and λ_i denotes its stationary arrival rate. The customer who arrives to station i is called i-customer. Customers of station i are served for a general independent random service time and b_i denotes its mean. The server utilization and the overall utilization are $\rho_i = \lambda_i b_i$ and $\rho = \sum_{i=1}^{N} \rho_i$. The switchover time from stage $i(j)$ to the next stage is general independent and identically distributed with mean $r_{i(j)}$. It can be different for each i and j. Furthermore $r_i = \sum_{j=1}^{V_i} r_{i(j)}$ and $r = \sum_{i=1}^{N} r_i$.

On the periodical polling model we impose the following assumptions:

A.1 The polling order is the same in each cycle (static polling order).

A.2 Each station can have different service discipline at each stage (mixed-discipline system).

A.3 The mean service time is positive and finite at each station, $0 < b_i < \infty$.

A.4 The mean switchover times are positive and finite after each stage, $0 < r_{i(j)} < \infty$ (nonzero-switchover-times model).

A.5 The arrival processes, the customer service times and the switchover times are mutually independent.

Definition 1 *The arrival of the server to a station and the departure of the server from a station are called* polling epoch *and* departure epoch*, respectively. We call the j-th polling epoch of station i in a cycle as $i(j)$-polling epoch. An i-polling epoch is any of the $i(j)$-polling epochs. We also use $i(j)$-departure epoch and i-departure epoch, which are defined similarly.*

Definition 2 *The sum of the duration of stages of a given station in a cycle is called* station time*.*

Definition 3 *The time the server spends away from the given station between the first server departure from that station in the actual cycle to the first server arrival to the same station in the next cycle is called* intervisit time*.*

Definition 4 *The* cycle time *of a given station is defined as the time elapsed from the first server visit to the given station in the actual cycle to the first server visit to the same station in the next cycle. It is also called as polling*

cycle time. The polling cycle of station i is called i-polling cycle.

2.2 BMAP process

At each station the arrival process is a *BMAP*, which is a generalization of the batch Poisson process, such that the arrivals are governed by a background continuous-time Markov chain (CTMC). $\Lambda(t)$ is the count of the number of arrivals in interval $(0,t]$ and $J(t)$ is the state of the background CTMC at time t, which is referred to as phase. A *BMAP* is a bivariate CTMC $\{(\Lambda(t), J(t)) ; t \geq 0\}$ on the state space $((\Lambda(t), J(t)); \Lambda(t) \in \{0,1,\ldots\}, J(t) \in \{1,2,\ldots,P\})$. It has the following infinitesimal generator:

$$\begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \ldots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \ldots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \ldots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $\{\mathbf{D}_\ell; \ell \geq 0\}$ is a set of $P \times P$ matrices.

The infinitesimal generator of the phase process is $\mathbf{D} = \sum_{\ell=0}^{\infty} \mathbf{D}_\ell$. π denotes the stationary probability vector of the phase process. If the phase process is irreducible, then $\pi\mathbf{D} = \mathbf{0}$ and $\pi e = 1$ uniquely determine π, where e denotes the column vector having all elements equal to one. λ is the stationary arrival rate of a BMAP, and $\lambda = \pi \sum_{\ell=0}^{\infty} \ell \mathbf{D}_\ell e$. For more details on BMAPs we refer to [15].

We make the following assumptions on the BMAP arrival process of the stations:

B.1 The phase process is irreducible.

B.2 The stationary arrival rate is positive and finite, $0 < \lambda_i < \infty$.

Remark 1 *The diagonal elements of the $\mathbf{D}_0$ matrix are strictly negative. Consequently the BMAP process can remain in any phase without an arrival for any finite time interval with positive probability.*

Recall that our stability analysis is based on properly chosen embedded Markov chains. The assumptions (B.1) and (B.2) together with Remark 1 are

used to show the structural properties of the state space of these embedded Markov chains.

2.3 Service discipline

Definition 5 *The* service discipline *gives the condition on the beginning and on the end of the service of a given stage.*

The most commonly known disciplines are, e.g., exhaustive, gated, binomial-exhaustive, binomial-gated, non-exhaustive, semi-exhaustive, limited-N, non-preemptive limited-T and so on [1]. In case of exhaustive discipline, the server continues serving until the station is emptied. Under gated policy only those customers are served, which are present at the polling epoch. By binomial-gated discipline every customer present at the polling epoch is served with probability pr. Similarly for binomial-exhaustive discipline every customer present at the polling epoch and arrived during its associated busy period is served with probability pr. At a station with limited-N discipline N customers are served if the station does not get empty before, in which case served leaves the station. The non-exhaustive discipline is a special case of limited-N discipline with N=1. Under semi-exhaustive policy the service continues until the number of customers becomes one less than it was at the polling epoch. In the nonpreemptive limited-T discipline either all customers are served or the station time limit, T, is reached before, in which case the server first finishes the service of the customer under service and then stops the service of that station.

We define the following notations:

$J^k_{i(j)}(m)$ - the phase of the k-th $BMAP$ arrival process at the $i(j)$-polling epoch of the m-th cycle. Furthermore $J^k_{i(j)} = \lim_{m\to\infty} J^k_{i(j)}(m)$ and $J_{i(j)} = J^i_{i(j)}$.

$F^k_{i(j)}(m)$ - the number of k-customers at the $i(j)$-polling epoch of the m-th cycle, $F^k_{i(j)} = \lim_{m\to\infty} F^k_{i(j)}(m)$, $F_{i(j)} = F^i_{i(j)}$,

$G_{i(j)}(m)$ - the number of customers served in the $i(j)$-stage in the m-th polling cycle, $G_{i(j)} = \lim_{m\to\infty} G_{i(j)}(m)$, $g_{i(j)}(m) = E\left(G_{i(j)}(m)\right)$, $g_{i(j)} = \lim_{m\to\infty} g_{i(j)}(m)$, $g_i = \sum_{j=1}^{V_i} g_{i(j)}$,

$g^\infty_{i(j)}$ - the mean number of customers served at the $i(j)$-stage given that the number of i-customers at $i(j)$-polling epoch goes to infinity: $g^\infty_{i(j)} = \sum_{q=1}^{P} P\left\{J_{i(j)} = q\right\} \lim_{n\to\infty} E\left(G_{i(j)} \mid F_{i(j)} = n, J_{i(j)} = q\right)$, $g^\infty_i = \sum_{j=1}^{V_i} g^\infty_{i(j)}$,

$g_{i(j)}^{\max}$ - the maximum of the mean number of customers, which can be served during an $i(j)$-stage: $g_{i(j)}^{\max} = \max_{n,q} E\left(G_{i(j)} \mid F_{i(j)} = n, J_{i(j)} = q\right)$, $g_i^{\max} = \sum_{j=1}^{V_i} g_{i(j)}^{\max}$.

We also use the shorthand notation $F_{i(j)} = \infty$ for $\lim_{m\to\infty} F_{i(j)}(m) = \infty$.

The set of service disciplines we allow to assign with the stages of the polling cycle is wider than the list of the most commonly known disciplines and it is limited by the following properties:

P.1 Memoryless property: In general the service discipline is independent from the history of the system.

P.2 Lack of anticipation assumption (LAA) [16]: The service discipline does not depend on the future of the system.

P.3 Work-conservation property: If the service of the actual stage begins, then it is work conserving up to the end of that stage according to the used discipline. Note, that if at least one customer is present at the polling epoch it does not necessarily mean, that the service immediately begins (e.g., binomial-gated or binomial-exhaustive disciplines).

P.4 Nonpreemptive service property: The service is nonpreemptive. Hence the server departs from a station only when the customer under service, if any, is served.

P.5 Determination property: If the service discipline of the $i(j)$-stage and the customer service time of that station are given, then the number of i-customers and the phase of the i-th $BMAP$ arrival process at the $i(j)$-polling epoch completely determines, in stochastic sense, the number of i-customers served during that stage, the length of that stage, the number of i-customers, and the phase of the i-th $BMAP$ at the $i(j)$-departure epoch. Additionally for each $k \neq i$ the number of k-customers and the number of i-customers together with the phase of the k-th and i-th $BMAP$ arrival processes at the $i(j)$-polling epoch determines also the number of k-customers, and the phase of the k-th $BMAP$ at the $i(j)$-departure epoch in stochastic sense.

P.6 Non-zero maximum property: If at least one i-customer is present at $i(j)$-polling epoch, then at least one i-customer is served with positive probability in that $i(j)$-stage. Furthermore the maximum of the mean number of customers that can be served during the $i(j)$-stage is greater than zero, $g_{i(j)}^{max} > 0$.

P.7 Maximum limit property: If the number of i-customers at the $i(j)$-polling epoch goes to infinity then the limit of mean number of i-customers served during the $i(j)$-stage equals the maximum of the mean number of customers

that can be served during that stage, $g^{\infty}_{i(j)} = g^{max}_{i(j)}$.

P.8 Mean maximum limit property: If the mean number of i-customers at the $i(j)$-polling epoch goes to infinity and $g^{\infty}_{i(j)} = \infty$ then the mean number of i-customers served during the $i(j)$-stage is also tends to infinity. That is, if $E\left(F_{i(j)}\right) = \infty$ than

$$E\left(G_{i(j)} \mid F_{i(j)}, J_{i(j)}\right) =$$
$$\sum_{\ell=0}^{\infty} P\left\{F_{i(j)} = \ell\right\} \sum_{q=1}^{P} P\left\{J_{i(j)} = q\right\} E\left(G_{i(j)} \mid F_{i(j)} = \ell, J_{i(j)} = q\right) = \infty.$$

A numerous service disciplines satisfies properties P.1-P.8. For example all the above mentioned examples fulfill these conditions. Note, that due to P.1 the service discipline is independent of the $BMAP$ arrival processes. Hence $g^{max}_{i(j)}$ and as a consequence of P.7 also $g^{\infty}_{i(j)}$ are independent of the i-th $BMAP$ arrival process. Properties P.6 and P.7 are similar to the assumptions S1 and S2 of the model of Down [7].

2.4 Stability related property of service disciplines

Definition 6 *The service discipline of the $i(j)$-stage is called* unlimited type *when $g^{\infty}_{i(j)} = \infty$.*

In this case

$$g^{\infty}_{i(j)} = g^{max}_{i(j)} = \infty, \tag{1}$$

due to P.7.

Definition 7 *The service discipline of the $i(j)$-stage is called* limited type *when $g^{\infty}_{i(j)} < \infty$.*

In this case

$$0 < g^{\infty}_{i(j)} = g^{max}_{i(j)} < \infty, \tag{2}$$

due to P.6 and P.7.

Definition 8 *A station is of* unlimited type*, if at least one stage of the station has unlimited type service discipline.*

A station is of limited type*, if all stages of the station have limited type service discipline.*

Remark 2 *If $F_{i(j)} = \infty$ at least for one stage of a limited type station i then it is also valid for every other stages of that station and it follows from (2), that $g_i = g_i^{max} < \infty$.*

The exhaustive, the gated, the binomial-gated and the binomial-exhaustive service disciplines are unlimited types. On the other hand the non-exhaustive, the semi-exhaustive, the limited-N, as well as the nonpreemptive limited-T service disciplines are limited types.

3 Global stability

3.1 Stability

Definition 9 *The distribution of the non-negative integer valued random variable, Z, is* proper*, if $\sum_{n=0}^{\infty} P\{Z = n\} = 1$ and* degenerate *otherwise.*

The distribution of the non-negative integer variable Z is totally degenerate*, if $\sum_{n=0}^{\infty} P\{Z = n\} = 0$.*

Definition 10 *The distribution of multivariate random variable is* proper*, if the marginal distribution of each component is proper.*

The distribution of multivariate random variable is degenerate (totally degenerate)*, if the marginal distribution of at least one component is degenerate (totally degenerate).*

Definition 11 *Station i of the polling model is said to be* stable*, when the number of i-customers at each $i(j)$-polling epochs has proper limiting distribution and the limiting cycle time has a finite mean.*

Definition 12 *The polling model is said to be* stable*, when the number of customers at each $i(j)$-polling epochs have proper limiting distributions and the limiting cycle time has a finite mean.*

This stability definition of polling models is equivalent with the one of Fricker and Jaïbi [6].

P.7 implies, that an unlimited type station with proper distribution of customers at polling epoch, $F_{i(j)}$, but with infinite mean, $E\left(F_{i(j)}\right) = \infty$, results in an infinite mean number of customers served at $i(j)$-stage, $g_{i(j)} = \infty$. In

this case, the mean cycle time is also infinite because the mean service times are finite (A.3). Hence this case is excluded from the stability definition.

For the limited type stations this stability definition allows $E\left(F_{i(j)}\right) = \infty$, since in this case $g_{i(j)} < \infty$, and hence it does not lead to infinite mean cycle time. This kind of definition might be unusual. But it means, that the number of customers does not increase to infinity (does not become degenerate), instead it converges to a proper distribution, which has an infinite mean, therefore it fits to an intuitive understanding of stability. Note, that this definition is different from the stability definition of Kuehn [2] since it excludes the case $E\left(F_{i(j)}\right) = \infty$.

3.2 Global stability of the polling system

Theorem 1 *There are 3 possible stability states of the polling model:*

- *Whole stability: all stations are stable.*
- *Partial stability: 1 or more limited type stations are instable, but the rest of the stations are stable.*
- *Instability: all stations are instable and the limiting mean cycle time is infinite.*

Proof. The theorem is a straightforward consequence of the following properties:

- All unlimited type stations share the same stability state.
  When an unlimited type station becomes instable, then at least the mean number of customers to be served at its polling epochs tends to infinity. This results in an infinite mean number of customers served in the actual cycle (P.7 and P.8). Due to finite service times (A.3) the associated station time tends to infinity, and hence the other stations accumulate infinitely many customers during this time. It implies, that all stations become instable and also the mean cycle time tends to infinity, from which the first two properties follow.
- When the unlimited type stations are instable the limited type stations are instable as well, and the limiting mean cycle time is infinite.
  By the same reason as before.
- When the unlimited type stations are stable the limited type stations can be both stable and instable.
  The third property comes from the fact that for the limited type station i the maximal mean service time is finite due to $g_i \leq g_i^{max} < \infty$. Hence the other stations accumulate only a finite number of customers during this time, which might be served by the unlimited type stations. □

4 Stability of stations

4.1 State of the system

Our analysis is based on a memoryless representation of the considered polling system.

Definition 13 *The* state of the system *at an i(j)-polling epoch consists of the number of customers at the stations and the phases of the BMAP arrival processes.*

The state vector, $\underline{Y}_{i(j)}(m)$, describes the state of the system in the m-th $i(j)$-polling epoch:

$$\underline{Y}_{i(j)}(m) = (F^1_{i(j)}(m), J^1_{i(j)}(m), F^2_{i(j)}(m), J^2_{i(j)}(m), \dots F^N_{i(j)}(m), J^N_{i(j)}(m)).$$

4.2 Embedded Markov chain

Lemma 1 *For any fixed* $i \in \{1, \dots, N\}$, $j \in \{1, \dots, V_i\}$ *the* $\left\{\underline{Y}_{i(j)}(m), m > 0\right\}$ *sequence is an embedded Markov chain.*

Proof. It follows from the determination property (P.5) and the independence of switchover times (A.4). □

Theorem 2 *The* $\left\{\underline{Y}_{i(j)}(m), m > 0\right\}$ *Markov chain is homogeneous and its state space consists of one irreducible class of aperiodic recurrent states and an optional class of transient states.*

Proof. It follows from the determination property P.5, that the evolution of the system does not depend on the elapsed number of cycles. Therefore the $\left\{\underline{Y}_{i(j)}(m), m > 0\right\}$ Markov chain is homogeneous.

To study the structure of the state space, we investigate the reachability of state $(0, 1, 0, 1, \dots 0, 1)$. We show, that $(0, 1, 0, 1, \dots 0, 1)$ can be reached from any state $(l_1, p_1, l_2, p_2, \dots l_N, p_N)$, where $l_1, l_2, \dots l_N \geq 0$. The state transition from $(l_1, p_1, l_2, p_2, \dots l_N, p_N)$ to $(0, 1, 0, 1, \dots 0, 1)$ can be realized in two steps. In the first step each $BMAP$ performs the phase transition to phase 1, while customers can arrive. This occurs with positive probability in finite time since the phase processes are irreducibile (B.1) and independent (A.5). Furthermore only finite number of customers can arrive during it, because the stationary arrival rates are finite (B.2). Now the system state is

$(l_1+n_1,1,l_2+n_2,1,\dots l_N+n_N,1)$, where $n_1,n_2,\dots n_N \geq 0$ are the numbers of newly arrived customers. Due to finite stationary arrival rates (B.2), finite mean switchover times (A.4) and finite mean service times (A.3) the duration of state transitions $\underline{Y}_{i(j)}(m) \to \underline{Y}_{i(j)}(m+1)$ is finite for every finite $m>0$. It follows, that the $(l_1,p_1,l_2,p_2,\dots l_N,p_N) \to (l_1+n_1,1,l_2+n_2,1,\dots l_N+n_N,1)$ transition occurs in finite number of state transitions with positive probability. In the second step the system is let to become empty, that is no arrival occurs until all customers in the system are served and the next $i(j)$-stage is reached, while the phases of the $BMAP$s remain unchanged. The numbers of customers $l_1+n_1,l_2+n_2,\dots,l_N+n_N$ are finite, thus due to non-zero maximum property (P.6) the system becomes empty in finite number of state transitions. Due to finite duration of state transitions this happens in finite time. Also the phases of the $BMAP$s can remain unchanged without any arrival for this finite time with positive probability (Remark 1). All these together ensures, that also the second step occurs in finite number of state transitions with positive probability. Therefore the chain has at least one state, which can be reached from all states in finite number of state transitions. This implies, that this state is recurrent. In general the states belonging to different irreducible classes of recurrent states of a Markov chain can not reach each other. It follows, that the state space of the $\left\{\underline{Y}_{i(j)}(m), m>0\right\}$ Markov chain has only one irreducible class of recurrent states, which includes the $(0,1,0,1,\dots 0,1)$ state. This state can be reached also from itself, so this state is aperiodic. As a consequence all states of the irreducible class are aperiodic. □

Let $\underline{Y}_{i(j)}$ be the following limit: $\underline{Y}_{i(j)} = \lim_{m\to\infty} \underline{Y}_{i(j)}(m)$.

Theorem 3 *The distribution of $\underline{Y}_{i(j)}$ is either proper or totally degenerate.*

Proof. Due to Theorem 2 the $\left\{\underline{Y}_{i(j)}(m), m>0\right\}$ Markov chain is homogeneous, and it has one irreducible class of aperiodic recurrent states. Assuming that there are no additional transient states, it follows, that the chain is either positive recurrent or null recurrent. In the positive recurrent case the distribution of $\underline{Y}_{i(j)}$ is proper, while in the null recurrent case it is totally degenerate. Assuming that there are also transient states in the Markov chain the probability that the chain is in the transient class tends to zero as the number of polling cycles goes to infinity. From which the theorem follows. □

We define the following quantities:

$$\underline{J}_{i(j)}(m) = \left(J^1_{i(j)}(m), J^2_{i(j)}(m), \dots, J^N_{i(j)}(m)\right),\ \underline{J}_{i(j)} = \lim_{m\to\infty} \underline{J}_{i(j)}(m),$$

$$\underline{F}_{i(j)}(m) = \left(F^1_{i(j)}(m), F^2_{i(j)}(m), \dots, F^N_{i(j)}(m)\right),\ \underline{F}_{i(j)} = \lim_{m\to\infty} \underline{F}_{i(j)}(m).$$

Corollary 1 *Both $\underline{Y}_{i(j)}$ and $\underline{F}_{i(j)}$ have either proper or totally degenerate distributions.*

Proof. $\underline{J}_{i(j)}$ has a proper distribution, since each BMAP process has finite number of phases. Since $\underline{Y}_{i(j)}$ is the union of $\underline{F}_{i(j)}$ and $\underline{J}_{i(j)}$ the statement follows from Theorem 3. □

4.3 System description in partial stability

Let $U \leq N$ is the number of the stable stations and $k_1, \ldots, k_U$ are their indexes. In addition $k_{U+1}, \ldots, k_N$ denote the indexes of the instable stations. To evaluate the system properties in partial stability, we introduce $\underline{Y}^*_{i(j)}(m)$ similar to $\underline{Y}_{i(j)}(m)$. Supposing that for the instable stations $F^k_{i(j)}(m) = \infty$, $\forall k \in \{k_{U+1}, \ldots, k_N\}, m \geq 0$, we define $\underline{Y}^*_{i(j)}(m)$ as the state of the stable stations, i.e., $\underline{Y}^*_{i(j)}(m) = (F^{k_1}_{i(j)}(m), J^{k_1}_{i(j)}(m), \ldots F^{k_U}_{i(j)}(m), J^{k_U}_{i(j)}(m))$.

Lemma 2 *$\underline{Y}^*_{i(j)}(m)$ is a discrete time Markov chain with a proper limiting distribution and* $\lim_{m\to\infty} P\left(\underline{Y}^*_{i(j)}(m) = (0, 1, \ldots, 0, 1)\right) > 0$.

Proof. $\underline{Y}_{i(j)}(m)$ is a Markov chain. All the instable stations are of limited type (Theorem 1) with $g_i = g_i^{max} < \infty$, and thus their mean station times are independent of their number of customers and of phases of their arrival processes at their polling epochs. Therefore in partial stability the number of customers and the phases of the arrival processes of instable stations do not play any role in the evolution of the state of stable stations.

It follows from the stability of stations $k_1, \ldots, k_U$, that all of the one dimensional marginal distributions of $\underline{Y}^*_{i(j)}(m)$ are proper. The topology of $\underline{Y}^*_{i(j)}(m)$ is similar to the one of $\underline{Y}_{i(j)}(m)$, i.e., the (0,1,...,0,1) state is reachable from each state in finite time with positive probability. Consequently (0,1,...,0,1) state is positive recurrent and hence $\lim_{m\to\infty} P\left(\underline{Y}^*_{i(j)}(m) = (0, 1, \ldots, 0, 1)\right) > 0$. □

4.4 Stability of a particular station

Theorem 4 *The necessary and sufficient condition of the stability of station i is*

$$g_i < g_i^{max}. \tag{3}$$

Proof. We show that the stability of station i implies $g_i < g_i^{max}$ and its instability implies $g_i = g_i^{max}$.

Let us start with the case when station i is stable. In this case $F_{i(j)}$ ($\forall j \in \{1 \dots V_i\}$) has a proper distribution and the limiting cycle time has finite mean. The mean service times are finite (A.3), and hence $g_{i(j)} < \infty$. Property P.6 implies, that $g_{i(j)}^{max} > 0$. If no customer is present at an $i(j)$-polling epoch, then no service occurs at that stage, i.e., $E\left(G_{i(j)}|F_{i(j)}=0\right) = 0$. However from Lemma 2 we have $P\left(F_{i(j)}=0\right) > 0$, from which, $\forall j \in \{1 \dots V_i\}$

$$\begin{aligned} g_{i(j)} &= E\left(G_{i(j)}\right) \\ &= P\left(F_{i(j)}=0\right) E\left(G_{i(j)}|F_{i(j)}=0\right) \\ &\quad + \left(1 - P\left(F_{i(j)}=0\right)\right) E\left(G_{i(j)}|F_{i(j)}>0\right) \\ &\leq P\left(F_{i(j)}=0\right) \cdot 0 + \left(1 - P\left(F_{i(j)}=0\right)\right) g_{i(j)}^{max} \\ &< g_{i(j)}^{max} \ . \end{aligned} \tag{4}$$

Now we consider the case when station i is instable. In this case the distribution of $F_{i(j)}$ is not proper for at least one $j \in \{1 \dots V_i\}$ or the limiting cycle time has infinite mean. Now we distinguish limited and unlimited type stations. First we consider the case, when station i is of limited type. Either the mean limiting cycle time is finite and Corollary 1 implies that the not proper distribution of $F_{i(j)}$ must be totally degenerate. Or the mean limiting cycle time is infinite, in which case infinitely many customers accumulate during it. In both cases the distribution of $F_{i(j)}$ is totally degenerate. Hence $F_{i(j)} = \infty$, and according to Remark 2 it holds for every $j \in \{1 \dots V_i\}$ and $g_i = g_i^{max}$. If station i of unlimited type is instable, then at least the mean number of i-customers to be served at its polling epochs tends to infinity. In this case $g_i = g_i^{max}$ follows from P.7 and P.8. □

5 Stability relationships

5.1 Stability of stations

Let $A_{i(j)}(m)$ be the number of arriving i-customers between the mth and the m+1th $i(j)$-polling epoch for every $i \in \{1, \dots, N\}, j \in \{1, \dots, V_i\}$. In addition we define $a_{i(j)}(m) = E(A_{i(j)}(m))$, $a_{i(j)} = \lim_{m\to\infty} a_{i(j)}(m)$ and $a_i = a_{i(1)}$.

Lemma 3 *If station i is stable, then the limiting mean number of arriving and served i-customers are the same,*

$$a_i = g_i. \tag{5}$$

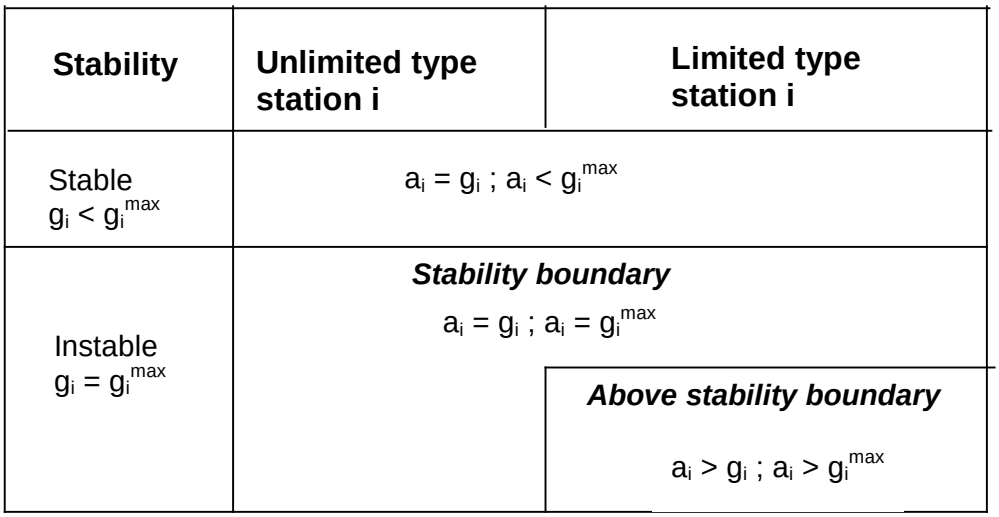

Fig. 1. Stability regions of the stations

Proof. From

$$F_{i(1)}(m+1) = F_{i(1)}(m) - \sum_{j=1}^{V_i} G_{i(j)}(m) + A_{i(1)}(m)$$

we have

$$E\left(F_{i(1)}(m+1)\right) = E\left(F_{i(1)}(m)\right) - \sum_{j=1}^{V_i} E\left(G_{i(j)}(m)\right) + E\left(A_{i(1)}(m)\right).$$

If station i has a proper limiting distribution, then $\lim_{m\to\infty} E\left(F_{i(1)}(m+1) - F_{i(1)}(m)\right) = 0$, which gives the lemma. □

Lemma 4 *The following relation holds for station i:*

- if station i is of limited type then

$$a_i \geq g_i, \tag{6}$$

- if station i is of unlimited type then

$$a_i = g_i. \tag{7}$$

Proof. The system cannot serve more customers than arrive, hence $a_i \geq g_i$ holds for both station types. For an instable limited type station g_i is bounded by the service discipline ($g_i \leq g_i^{max} < \infty$) while a_i can be any large value. Thus a_i can be greater than g_i It follows, that $a_i \geq g_i$. For an unlimited type station g_i is not bounded by the service discipline ($g_i^{max} = \infty$) hence $a_i = g_i$. □

Proposition 1 *Station i is stable if and only if:*

$$a_i < g_i^{max}. \tag{8}$$

Proof. We show, that the stability of station i implies $a_i < g_i^{max}$, and its instability implies $a_i \geq g_i^{max}$. If station i is stable then (8) follows from Lemma 3 and Theorem 4.

If station i is instable then it follows from Theorem 4, that $g_i \geq g_i^{max}$, but by its definition g_i^{max} can not be less than g_i, hence $g_i = g_i^{max}$. Combining it with (6) and (7) we have $a_i \geq g_i^{max}$. □

Table 1 summarizes the stability relevant relationships of particular station i.

5.2 Order of instability of stations

Let $C_i(m)$ be the polling cycle time of station i from the m-th $i(1)$ polling epoch to the m+1-th $i(1)$ polling epoch for every $i \in \{1, \ldots, N\}$. Additionally we define $c_i(m) = E\left(C_i(m)\right)$, $c_i = \lim_{m\to\infty} c_i(m)$, as well as $c = c_1$.

Lemma 5 *The mean number of arriving i-customers during a cycle equals to the mean cycle time multiplied by the stationary arrival rate to station i:*

$$a_i = \lambda_i c. \tag{9}$$

Proof. If the system is stable or partially stable, then we can compute the limiting arrival rate of station i as the mean number of arriving i-customers during the polling cycle divided by the mean length of that cycle, $\lambda_i = \frac{a_i}{c}$.

If the whole system is instable, λ_i is finite, $c = \infty$ (Theorem 1), and thus $a_i = \infty$. Hence (9) holds for this case as well. □

Corollary 2 *Station i is stable if and only if:*

$$\lambda_i c < g_i^{max}. \tag{10}$$

Proof. Taking into account Lemma 5 the statement follows from Proposition 1 by applying (9). □

Let $\mathbf{D}_\ell^{(i)}$ denote the $\mathbf{D}_\ell$ matrix of BMAP process of station i ($\ell \geq 0, i \in \{1, \ldots, N\}$). Let us control the traffic intensity by applying scaling parameter α, such that $\mathbf{D}_l^{(i)}(\alpha) = \alpha \mathbf{D}_l^{(i)}$, $l \geq 0$. This way $\lambda_i(\alpha) = \alpha\lambda_i$, and thus the relative ratios of station arrival rates remain fixed.

Theorem 5 *Scaling the traffic intensity from 0 to ∞ the stations gets instable in order $i_1, i_2, \ldots, i_N$, where*

$$\frac{\lambda_{i_1}}{g_{i_1}^{max}} \geq \frac{\lambda_{i_2}}{g_{i_2}^{max}} \geq \ldots \geq \frac{\lambda_{i_N}}{g_{i_N}^{max}}. \tag{11}$$

Proof. It follows from Corollary 2, that station k gets instable, when $c\lambda_k(\alpha) = g_k^{max}$. Since c is common to all stations the λ_k/g_k^{max} ratio determines the order of instability. □

From now on, we assume that the stations are indexed such that $\frac{\lambda_1}{g_1^{max}} \geq \frac{\lambda_2}{g_2^{max}} \geq \ldots \geq \frac{\lambda_N}{g_N^{max}}$. According to Theorem 1, if there are L limited type stations, it follows, that the first L indexes identify the limited type stations.

5.3 Mean cycle time

Let $S_{i(j)}(m)$ be the station time of the $i(j)$-stage in the mth polling cycle for every $i \in \{1, \ldots, N\}, j \in \{1, \ldots, V_i\}$. Additionally we define $s_{i(j)}(m) = E(S_{i(j)}(m))$, $s_{i(j)} = \lim_{m\to\infty} s_{i(j)}(m)$ and $s_i = \sum_{j=1}^{V_i} s_{i(j)}$.

Theorem 6 *If the first U limited type stations ($1 \leq U \leq L$) are out of stability and the remaining $N - U$ stations are stable, then the mean cycle time is*

$$c = \frac{r + \sum_{k=1}^{U} g_k^{\max} b_k}{1 - \sum_{k=U+1}^{N} \rho_k}. \tag{12}$$

Proof. If $U \leq L$ limited type stations are out of stability and the remaining stations are stable, then the mean cycle time, c, is finite. We express c as the sum of switchover times and station times:

$$c = r + \sum_{k=1}^{N} s_k = r + \sum_{k=1}^{N} g_k b_k.$$

For the instable limited type station k ($1 \leq k \leq U$) Corollary 1 implies, that the not proper distribution of $F_{k(j)}$ is totally degenerate. Hence $F_{k(j)} = \infty$ and it follows from Remark 2, that $g_k = g_k^{\max}$. For the stable stations it follows from (9) and (5), that $g_k = c\lambda_k$. Substituting them we get:

$$c = r + \sum_{k=1}^{U} g_k^{\max} b_k + \sum_{k=U+1}^{N} c\lambda_k b_k.$$

Solving it for c results in (12). □

If $U = 0$, that is the whole system is stable, then (12) is reduced to the well-known form:

$$c = \frac{r}{1-\rho}. \tag{13}$$

6 Stability conditions

6.1 Partial stability

Theorem 7 *If the first $i-1$ stations are instable then station i of limited type ($i \leq L$) is stable if and only if*

$$\sum_{k=i}^{N} \rho_k + \frac{\lambda_i}{g_i^{max}} \left(r + \sum_{k=1}^{i-1} g_k^{max} b_k \right) < 1. \tag{14}$$

Proof. First we show that the stability of station i results in (14). Since station i is stable $\lambda_i c < g_i^{max}$ (Corollary 2) and applying (12), $\lambda_i \frac{r+\sum_{k=1}^{i-1} g_k^{max} b_k}{1-\sum_{k=i}^{N} \rho_k} < g_i^{max}$, results the inequality.

Starting from (14)

==========

MASIK IRANY HIANYZIK !!!!

□

Let ρ^u denote the utilization of the unlimited stations:

$$\rho^u = \sum_{k=L+1}^{N} \rho_k. \tag{15}$$

Theorem 8 *Station i of unlimited type ($i > L$) is stable if and only if*

$$\rho^u < 1. \tag{16}$$

Proof. We show, that the stability of station i implies $\rho^u < 1$, while from its instability $\rho^u \geq 1$ follows.

Let us start with the case, when station i is stable. Applying Theorem 6, we get $c = \frac{r+\sum_{k=1}^{j} g_k^{max} b_k}{1-\sum_{k=j+1}^{N} \rho_k}$, where the first j stations ($1 \leq j \leq L$) are the instable limited type ones. Applying $c < \infty$ and utilizing $c > 0$ implies $\sum_{k=j+1}^{N} \rho_k < 1$. Furthermore using $\sum_{k=L+1}^{N} \rho_k \leq \sum_{k=j+1}^{N} \rho_k$ and applying notation (15) results in $\rho^u < 1$.

To study the case when station i is instable, we start increasing the traffic load of the system from $\alpha = 0$ and increase α to the boundary situation, when all stations of limited type are already instable, but the stations of unlimited type are still stable. By setting $j = L$ in the expression of the mean cycle time we get

$$c = \frac{r + \sum_{k=1}^{L} g_k^{max} b_k}{1 - \sum_{k=L+1}^{N} \rho_k} = \frac{r + \sum_{k=1}^{L} g_k^{max} b_k}{1 - \rho^u}$$

for this case. Further increase in the traffic intensity does not change the expression of c, but it leads to the instability of the system due to infinite cycle time, when $\rho^u = 1$. □

6.2 Stability of the system

Theorem 9 *The system is in*

- *whole stability if and only if*

$$\rho + \left(\frac{\lambda_1}{g_1^{max}}\right) r < 1, \tag{17}$$

- *partial stability if and only if*

$$\rho + \left(\frac{\lambda_1}{g_1^{max}}\right) r \geq 1 \text{ and } \rho^u < 1, \tag{18}$$

- *instability if and only if*

$$\rho^u \geq 1. \tag{19}$$

Proof. The first statement comes from Theorem 7 by setting $i = 1$. The second and the third statements are consequences of Theorem 1, Theorem 8 and the first statement. □

7 Concluding remarks

We have analyzed the stability of $BMAP/GI/1$ periodical polling model. We have got the following results for this model:

- necessary assumptions for $BMAP$ arrival processes (Assumptions B.1 and B.2) and for the service disciplines (Properties P.1 - P.7),
- characterization of global stability (Theorem 1),
- order of instability of stations (Theorem 5),
- conditions for partial stability (Theorems 7 and 8),
- necessary and sufficient condition for the stability of the system (Theorem 9).

Several properties of service disciplines play crucial role in the completion of the stability proofs. These key properties are:

- Property P.5 used in Theorem 2,
- Property P.6 used in Theorem 2 and Theorem 4,
- and Properties P.7 and P.8 used in Theorem 1 and Theorem 4.

The properties P.6, P.7 and P.8 are relaxed condition on the service disciplines in comparison with the monotonicity property of Fricker and Jaïbi in [6].

The work-conservation property (P.3) and the nonpreemptive service property (P.4) can be relaxed. In this case other quantities can be handled in the evolution of the system, like e.g., set-up time or repair time. However these quantities may depend only on the state of the system at the polling epochs prior to the start of set-up time, repair time, respectively. Hence state-dependent set-up times, repair times can be allowed in the model.

We believe that it is straightforward to extend the technique used in this paper for other polling models. The possible extensions are:

- general independent batch customer service times ($GI^{[y]}$),
- arrival process dependent customer service times,
- polling model with Markovian server routing,
- dependence of the service disciplines, and the customer service time on the past through finite valued discrete variables.

References

[1] H. Takagi, Analysis of Polling Systems (MIT Press, 1986).

[2] P.J. Kuehn, Multiqueue Systems with Nonexhaustive Cyclic Service, The Bell System Technical Journal 58 (1979) 671-698.

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

[4] E. Altman, P. Konstantopoulos, and Z. Liu, Stability, monotonicity and invariant Quantities in General Polling Systems, Queueing Systems 11 (1992) 35-57.

[5] A. A. Borovkov and R. Schassberger, Ergodicity of polling network, Stochastic Processes and their Applications 50 (1994) 253-262.

[6] C. Fricker and M.R.Jaïbi, Monotonicity and stability of periodic polling models, Queueing Systems 15 (1994) 211-238.

[7] G. Down, On the stability of polling models with multiple servers, Tech. Report BS-R9605, CWI, 1996.

[8] S. Foss, N. Chernova, and A. Kovalevskii, Stability of polling systems with state-independent routing, in Proceedings of 34-th Annual Allerton Conference in Communication, Control and Computing, Monticello, Illionis, oct 1996, pp. 220-227.

[9] S. Foss and A. Kovalevskii, A stability criterion via fluid limits and its application to a polling system, Queueing Systems 32 (1999) 131-168.

[10] L.Massoulie, Stability of non-Markovian polling systems, Queueing Systems 21 (1995) 67-96.

[11] S. Foss, and N. Chernova, Dominance theorems and ergodic properties of polling systems, translated from Problemy Predachi Informatsii, Vol. 32, No. 4, pp. 46-71, October-December, 1996.

[12] R. E. Lillo, Ergodicity and analysis of the process describing the system state in polling systems with two queues, European Journal of Operational Research 167 Issue 1 (2005) 144-162.

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

[14] M. Eisenberg, Queues with Periodic Service and Changeover Time, Operations Research 20 (1972) 440-451.

[15] D. M. Lucantoni, New results on the single-server queue with a batch Markovian arrival process, Stochastic Models 7 (1991) 1-46.

[16] R. W. Wolff, Poisson Arrivals See Times Averages, Operations Research 30 (1982) 223-231.