BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
DEPARTMENT OF TELECOMMUNICATIONS

Unified Analysis of Cyclic Polling Models with BMAP

Ciklikus polling modellek egységes analízise BMAP érkezési folyamat esetén

Einheitliche Analyse von zyklischen Polling-Modellen mit BMAP

Ph.D. Thesis

by

Zsolt Saffer

Advisor:

Prof. Dr. Miklós Telek, BUTE

Budapest, Hungary

June 17, 2010

The reviews of the dissertation and the report of the thesis discussion are available at the Dean's Office of the Faculty of Electrical Engineering and Informatics of the Budapest University of Technology and Economics.

Az értekezés bírálatai és a védésen készült jegyzőkönyv elérhető a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karának Dékáni Hivatalában.

Kivonat

Polling modelleket alkalmaznak számos távközlési rendszer analízisében és optimalizálásában az 1980-as évektől fogva. A klasszikus polling modellben az egyetlen kiszolgálóegység ciklikusan látogatja az állomásokat és ennek során kiszolgálja azok sorait. A klassikus polling model Poisson érkezési folyamatának batch Markovi folyamattá (BMAP) történő kiterjesztése egy ígéretes modellt eredményez, amely a forgalom folyamat és az aktuálisan vizsgált rendszer működésének valósághűbb modellezésére képes.

A disszertáció folytonos idejű, asszimmetrikus, nullától eltérő átkapcsolási idejű ciklikus polling modellekkel foglalkozik BMAP érkezési folyamat esetén. A disszertáció tárgyalja a modell stabilításának feltételeit, a rendszerbeli igények számának stacionárius analízisét és a vizsgált polling modell legfontosabb speciális eseteit.

A stabilítás analízis módszertana a kiszolgálás-diszciplínák széles körére alkalmazható, amely magában foglalja a gated, az exhaustive és a G-limited diszciplínákat. Ezt megfelelően megválasztott beágyazott Markov láncok segítségével érjük el. Az új stabilítás eredmények csoportjában jellemezzük a globális stabilítást, áttekintő képet adunk egy adott állomás stabilítás tartományairól és megadjuk az állomások instabillá válásának sorrendjét, a részleges stabilítás feltételeit valamint a rendszer stabilítás állapotainak szükséges és elégséges feltételeit.

Két lépéses módszertant alkalmazunk a rendszerbeli igények számának stacionárius analízisében, amely a probléma szétválasztásán keresztül összességében egyszerübben kezelhető analízist eredményez. A kiszolgálás-diszciplína független részben összefüggéseket állítunk fel a kiszolgálóegység érkezésének és távozásának pillanatában értelmezett mennyiségek függvényében. A kiszolgálás-diszciplína függő részben ezeket a mennyiségeket határozzuk meg az egyes diszciplínákra egységes módszerrel. Az igények számára vonatkozó, új kiszolgálás-diszciplína független eredmények a nulla átkapcsolási idejű polling modellekre is érvényesek. A kiszolgálás-diszciplína specifikus megoldást néhány fontos kiszolgálás-diszciplína esetére vezetjük le, amely magában foglalja a legfontosabb gated és exhaustive valamint az általános G-limited és decrementing-K diszciplínákat.

A módszertannak és az eredményeknek a kiszolgálás-diszciplínák széles körére való alkalmazhatósága biztosítja az ismertetett analízis egységes jellegét.

Zusammenfassung

Polling-Modelle wurden in der Analyse und der Optimierung von zahlreichen Telekommunikationssystemen seit der 80er Jahre angewendet. In dem klassischen Polling-Modell besucht der einzige Server die Stationen in einer zyklischen Reihenfolge und die Warteschlangen der einzelnen Stationen werden während ihrer Besuche gedient. Die Erweiterung des Poisson-Ankunftsprozesses von dem klassischen Polling-Modell für den Markovschen Gruppenankunftsprozess (BMAP) ergibt sich ein vielversprechendes Modell, welches zu modellieren des Datenprozesses und des Arbeitsvorgangs von dem tatsächlichen erforschten System in realistischer Weise fähig ist.

Diese Doktorarbeit handelt von zeitkontinuierlichen asymmetrischen zyklischen Polling-Modellen mit BMAP, in denen die Umschaltungszeiten von Null verschieden sind. Die These beschäftigt sich mit den Bedingungen für Stabilität des Modells, der Analyse von der Anzahl der Kunden im System im stationären Zustand und den wichtigsten Speziallfällen von den betrachteten Polling-Modellen.

Die Methodologie verwendet zur Stabilitätsanalyse kann für die breite Klasse der Bedienungsdisziplinen einschließlich der Gated-Disziplin, der Exhaustive-Disziplin und der G-Limited-Disziplin angewenget werden. Es wird durch die Identifizierung der richtig gewählten eingebetteten Markov-Ketten erreicht. In der Gruppe der neuen Stabilitätsergebnisse werden die Charakterisierung der globalen Stabilität, die Übersicht der Stabilitätsregionen von der einzelnen Station, die Ordnung der Instabilität von den Stationen, die Bedingungen für partielle Stabilität und die notwendige und hinreichende Bedingungen für die Stabilitätszustände des Systems beschafft.

Eine Zweischrittmethode wird in der Analyse von der Anzahl der Kunden im System im stationären Zustand angewendet, die durch die Separation des Problems eine einfachere Gesamtbetrachtung der Analyse ergibt. In dem bedienungsdisziplinunabhängigen Teil werden Formeln in Abhängigkeit von den Quantitäten, die um die Abfahrtszeit und die Ankunftszeit des Servers definiert werden, aufgestellt. In dem bedienungsdisziplinabhängigen Teil werden diese Quantitäten für die einzelne Diszipline mittels eines einheitlichen Verfahrens bestimmt. Die bedienungsdisziplinspezifische Lösung wird für Polling-Modelle mit einigen wichtigen Bedienungsdisziplinen einschließlich der wichtigsten Gated-Disziplin und Exhaustive-Disziplin und der allgemeinen G-Limited-Disziplin und Decrementing-K-Disziplin abgeleitet.

Die Anwendbarkeit der Methodologie und der Ergebnisse für die breite Klasse der Bedienungsdisziplinen sichert die einheitliche Eigenschaft der vorgelegten Analyse zu.

Acknowledgements

First of all, I would like to thank my supervisor, Professor Miklós Telek for his persistent guidance and encouragement during my way leading to complete this dissertation. He helped me always on that level of problem where it was necessary. I learned a lot from him among others on the scientific art of working and on preparing scientific publications.

As next I would like to thank dr. Tibor Trón and Professor Csaba Szabó for their initial help in evolving my interest in the telecommunication networks and later on in queueing theory. I also thank my ex-wife together with her parents and my father for their support during my preliminary scientific activity. Unfortunately dr. Tibor Trón and my father are not any more among the living.

My special thank dr. Heinrich Bartosik the former technology department manager at Nuance Communications Austria GmbH., Vienna, Austria (until 10/2008 known as Philips Speech Recognition Systems GmbH.) for allowing me to change my previous full-time contract to a part-time one. Without this I could not allocate enough time for performing the work on my PhD.

My next thank goes to local government of Vienna, Austria for their grant within the framework of WAFF (Wiener ArbeitnehmerInnen Förderungsfonds) as it supported me financially during preparing this dissertation.

Further I also thank the colleges at the Department of Telecommunications, BUTE for their assistance. Especially Gábor Horváth, who was always ready to help me with his suggestions in the whole range from research discussions to administrative issues. I also thank Zsolt Pándi as the form of his dissertation served as skeleton for mine.

Last but not least, I would like to thank my mother and Andrea Gállfy for their support.

Zsolt Saffer

Contents

List of Abbreviations

AGV	automated guided vehicle
ARQ	Automatic Repeat reQuest
ATM	Asynchronous Transfer Mode
BMAP	batch Markovian arrival process
BWA	broadband wireless network
CTMC	continuous-time Markov chain
DLL	distributional Little's law
DQDB	Distributed Queue Dual Bus
DTMC	discrete-time Markov chain
FDDI	Fiber Distributed Data Interface
FIFO	First-In-First-Out
GF	probability-generating function
GG	globally gated
HDLC	High-level Data Link Control
HSLAN	High-Speed LAN
HT	heavy-traffic limit
LAN	Local Area Network
LST	Laplace-Stieljes transform
MAC	Media Access Control
MAN	Metropolitan Area Network
MAP	Markovian arrival process
MMPP	Markov-modulated Poisson process
MRP	Markov regenerative process
MTBP	multi-type branching process
MVA	mean value analysis
PH	phase type
PGF	probability-generating function
PSA	power series algorithm
QoS	Quality of Service
SDLC	Synchronous Data Link Control
TDMA	Time-Division Multiple Access
UMTS	Universal Mobile Telecommunications System

Summary of notations

Quantity types

Y, y, Γ, γ	scalars
$\mathbf{y}$, $\boldsymbol{\pi}$	vectors, hypervectors
$\mathbf{Y}$, $\boldsymbol{\Psi}$	matrices, hypermatrices
$\mathcal{F}(\mathbf{x}, \mathbf{y})$, $\mathcal{R}(\boldsymbol{\Upsilon})$	general matrix operators

Operators

Pr	the probability operator
E[ ]	the mean value operator
$[]_j$	j-th element of a vector
$[]_{j,l}$	j, l-th element of a matrix
$rank()$	the rank of a matrix
adj	the adjugate of a matrix
det	the determinant of a matrix
Tr ()	the trace of a matrix
$\mathcal{R}(\mathbf{V})$	column reduction operator
$\mathcal{M}(\mathbf{Y})$	matrix reduction operator
$\otimes$	Kronecker product
$\oplus$	Kronecker sum
$p(n_1, \ldots, n_N)$	position of 1 in the unit hypervector

Systematics

X, Y	generic continuous or nonnegative integer valued (discrete) random variables
X(t)	cumulative distribution function of the continuous random variable X
$\widetilde{X}(s)$	the LST of the continuous random variable X
x	first moment of the continuous random variable X
$x^{(k)}$	k-th moment of the continuous random variable X for $k \geq 1$
$\widehat{Y}(z)$	the PGF of the nonnegative discrete random variable Y
$y^{(0)}$	$= 1$
$y^{(k)}$	k-th factorial moment of the nonnegative discrete random variable Y for $k \geq 1$
$\widehat{\mathbf{y}}(z)$	vector GF of the nonnegative discrete random variable Y
$\mathbf{y}$	the value of $\widehat{\mathbf{y}}(z)$ at $z = 1$ i.e., $\mathbf{y} = \widehat{\mathbf{y}}(1)$
$\mathbf{y}^{(0)}$	$= \mathbf{y}$
$\mathbf{y}^{(k)}$	k-th vector factorial moment of the nonnegative discrete random variable Y for $k \geq 1$, i.e., $\mathbf{y}^{(k)} = \frac{d^k}{dz^k}\widehat{\mathbf{y}}(z)\|_{z=1}$

$\widehat{\mathbf{Y}}(z)$	matrix GF
$\mathbf{Y}$	the value of $\widehat{\mathbf{Y}}(z)$ at $z = 1$ i.e., $\mathbf{Y} = \widehat{\mathbf{Y}}(1)$
$\mathbf{Y}^{(0)}$	$= \mathbf{Y}$
$\mathbf{Y}^{(k)}$	k-th factorial moment of $\widehat{\mathbf{Y}}(z)$ for $k \geq 1$, i.e., $\mathbf{Y}^{(k)} = \frac{d^k}{dz^k}\widehat{\mathbf{Y}}(z)\|_{z=1}$
$adj\mathbf{Y}$	$= adj\widehat{\mathbf{Y}}(1)$
$[adj\mathbf{Y}]^{(0)}$	$= adj\mathbf{Y}$
$[adj\mathbf{Y}]^{(k)}$	k-th derivative of matrix $adj\widehat{\mathbf{Y}}(z)$ at $z = 1$ for $k \geq 1$, i.e., $[adj\mathbf{Y}]^{(k)} = \left.\frac{d^k\left(adj\widehat{\mathbf{Y}}(z)\right)}{dz^k}\right\|_{z=1}$
$det\mathbf{Y}$	$= det\widehat{\mathbf{Y}}(1)$
$[det\mathbf{Y}]^{(0)}$	$= det\mathbf{Y}$
$[det\mathbf{Y}]^{(k)}$	k-th derivative of matrix $det\widehat{\mathbf{Y}}(z)$ at $z = 1$ for $k \geq 1$, i.e., $[det\mathbf{Y}]^{(k)} = \left.\frac{d^k\left(det\widehat{\mathbf{Y}}(z)\right)}{dz^k}\right\|_{z=1}$

Modifiers

$\widetilde{X}_i(s)$, $\mathbf{y}_i^{(k)}$	first subscript - station index (i)
Y^f	superscript - at polling epoch
Y^m	superscript - at departure epoch
Y^s	superscript - customer service starts
Y^d	superscript - customer departures
Y^{d^-}	superscript - just before customer departures
Y^a	superscript - customer arrival
Y^*	superscript - BMAP state change epochs
Y^i	superscript - during intervisit time
Y^u	superscript - unlimited stations
Y^S	superscript - during station time
Y^I	superscript - during intervisit time
Y^e	superscript - exhaustive discipline

System variables

N	number of stations
N^l	number of limited type stations
N^u	number of instable stations
L	number of BMAP phases
$\widehat{\mathbf{D}}_i(z)$, λ_i	matrix GF and stationary arrival rate of i-th BMAP, respectively
$\widetilde{B}_i(s)$, b_i	LST and the mean of the i-customer service time, respectively
$b_i^{(k)}$	k-th moment of the i-customer service time for $k \geq 1$
$\widetilde{R}_i(s)$, r_i	LST and the mean of the switchover time after station i, respectively
$r_i^{(2)}$	the second moment of the switchover time after station i
ρ_i	$= \lambda_i b_i$ server utilization at station i
ρ	$= \sum_{i=1}^N \rho_i$ overall utilization
r	$= \sum_{i=1}^N r_i$ the sum of the mean switchover times
$r^{(2)}$	$= \sum_{i=1}^N r_i^{(2)}$ the sum of the second moments of the switchover times
v	vacation time

Random variables

g_i^{∞}	the mean number of i-customers served during an i-station time given that the number of i-customers at i-polling epoch goes to infinity
g_i^{max}	the maximum of the the mean number of i-customers which can be served during i-station time
$G_i(m)$	the number of i-customers served in the m-th polling cycle for $m \geq 1$
$A_i(m)$	the number of arriving i-customers between the m-th and $m+1$-th i-polling epoch for $m \geq 1$
$A_i^*(\ell, n)$	the number of i-BMAP state changes during the service time of the n-th i-customer in the ℓ-th polling cycle for $\ell \geq 1$ and $n = 1, \ldots, G_i(\ell) + 1$
$S_i(m)$	i-station time in the m-th polling cycle for $m \geq 1$
$I_i(m)$	i-intervisit time in the m-th polling cycle for $m \geq 1$
$C_i(m)$	polling cycle of station i between the m-th and $m+1$-th i-polling epoch
$J_{i,k}^f(m)$	the phase of k-BMAP at i-polling epoch of the m-th polling cycle for $m \geq 1$
$F_{i,k}(m)$	number of k-customers at i-polling epoch of the m-th polling cycle for $m \geq 1$
$\mathbf{Z}_i(m)$	the state of the system in the m-th i-polling epoch for $m \geq 1$
$\Lambda_i(t)$	number of i-BMAP arrivals in $(0, t]$
$J_i(t)$	the phase of i-BMAP at time t
$N_i(t)$	the number of i-customers in the system at time t

Vectors

$\boldsymbol{\pi}_i$	stationary probability vector of the phase process of i-BMAP
$\widehat{\mathbf{f}}_i(z)$	vector GF of number of i-customers at i-polling epoch
$\widehat{\mathbf{m}}_i(z)$	vector GF of number of i-customers at i-departure epoch
$\widehat{\mathbf{q}}_i^d(z)$	vector GF of number of i-customers at i-customer departure epochs
$\widehat{\mathbf{q}}_i(z)$	vector GF of number of i-customers at an arbitrary epoch
$\widehat{\mathbf{q}}_i^*(z)$	vector GF of number of i-customers at i-BMAP state change epochs
$\widehat{\mathbf{q}}_i^a(z)$	vector GF of number of i-customers at i-customer arrival epoch

Matrices

$\widehat{\mathbf{\Psi}}_i(z)$	matrix GF of the number of arriving i-customers at an i-BMAP state change
$\widehat{\mathbf{P}}_i(z,t)$	matrix GF of number of i-customers arriving in $(0,t]$
$\widehat{\mathbf{A}}_i(z)$	matrix GF of number of i-customers arriving during an i-customer service time
$\mathbf{G}_i(t)$	the time dependent first passage matrix at station i
$\widetilde{\mathbf{G}}_i(s)$	LST of the time dependent first passage matrix at station i
$\mathbf{\Delta}^l_{i+1}(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N))$	the coefficient matrix of $\mathbf{p}^m_i(n_1,\ldots,n_N)$ for the x_1-th, …, x_N-th derivatives
$\mathbf{\Delta}^r_{i+1}(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N))$	the coefficient matrix of $\mathbf{p}^m_{i+1}(n_1,\ldots,n_N)$ for the x_1-th, …, x_N-th derivatives

Hypervectors

$\mathbf{p}^f_i(n_1,\ldots,n_N)$	probability hypervector of the stationary number of customers at i-polling epochs
$\mathbf{p}^m_i(n_1,\ldots,n_N)$	probability hypervector of the stationary number of customers at i-departure epochs
$\boldsymbol{\theta}_i$	hypervector representing the unknowns of the system of linear equations

Hypermatrices

$\widehat{\mathbf{A}}_i(z_1,\ldots,z_N)$	hypermatrix GF of the number of simultaneously arriving k-customers for every $k=1,\ldots,N$ and the phase changes of every BMAP-s during an i-customer service time
$\widehat{\mathbf{H}}_i(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_N)$	hypermatrix GF of the number of simultaneously arriving k-customers for every $k=1,\ldots,i-1,i+1,\ldots N$ and the phase changes of every BMAP-s during the decrement of the number of i-customers by one
$\widehat{\mathbf{U}}_i(z_1,\ldots,z_N)$	hypermatrix GF of the number of simultaneously arriving k-customers for every $k=1,\ldots,N$ and the phase changes of every BMAP-s during the switchover time after station i
$\mathbf{R}_i(l_1,\ldots,l_N)$	transition probabilityhypermatrix of the number of simultaneously arriving k-customers for every $k=1,\ldots,N$ during the switchover time R_i when the number of simultaneously arriving $k_1,\ldots,k_N$-customers are exactly $l_1,\ldots,l_N$, respectively
$\mathbf{\Phi}_i$	hypermatrix representing the coefficients on the left side of the system of linear equations
$\mathbf{\Upsilon}_i$	hypermatrix representing the coefficients on the right side of the system of linear equations
$\mathbf{\Xi}_i$	hypermatrix representing the governing equations of the system in the relation $\boldsymbol{\theta}_i \to \boldsymbol{\theta}_i$

Composite hypervector

$\widehat{\mathbf{f}}_i(z_1,\ldots,z_{i-1},\boldsymbol{A},z_{i+1},\ldots,z_N)$	substitution of a hypermatrix A into the defining series of the hypervector GF $\widehat{\mathbf{f}}_i(z_1,\ldots,z_N)$

Miscellaneous

$\mathbf{e}$	column vector having all elements equal to one
$\mathbf{e}_j$	$=(0,\ldots,0,1,0,\ldots,0)$ the $1\times L$ vector with 1 at the j-th position
$\mathbf{e}^{L^N(X+1)^N}$	$1\times(L^N(X+1)^N)$ column vector having all elements equal to one
$\mathbf{e}_\ell^{X+1}$	$=(0,\ldots,0,1,0,\ldots,0)$ the $1\times(X+1)$ vector with 1 at the ℓ-th position
$\mathbf{I}$	identity matrix
$1_{(\mathrm{con})}$	the indicator of condition "con"
i	the imaginary unit

Imagination is more important than knowledge.
Knowledge is limited.
Imagination encircles the world.
- A. Einstein

Preface

Polling models have been applied in the analysis and optimization of numerous telecommunication systems from the beginning of 1980s. In the classical polling model the single server attends the stations in cyclic manner and serves their queues during their visits. The customers arrive according to Poisson process to each station.

In the 1980s the cyclic polling model has gained much attention in the performance analysis of the token passing protocols of local area networks. In the subsequent decades it has been applied to channel access protocols in metropolitan area networks, land mobile and satellite radio communication networks and also to wireless communication networks. There was an explosive growth of research on polling systems, motivated by the increasing number of applications in the communications systems. Numerous variations and extensions of the basic polling model have been proposed and analyzed.

The extension of polling models is an ongoing activity reacting to the continuously growing demand for more precise modeling of the operation of practical systems. Parallel to this research in the last two decades also more powerful arrival processes have been introduced and analyzed. Among them one of the most general ones is the batch Markovian arrival process. Besides of its capability of modeling a wide range of real traffic sources it can also capture the dependency in traffic processes. Hence the extension of the arrival process of the classical polling system to batch Markovian arrival process results in a promising model, which is capable of a more realistic modeling of the traffic process and of the operation of the actually studied system.

This thesis deals with continuous-time asymmetric nonzero-switchover-times cyclic polling models with batch Markovian arrival process. In the first part of the thesis the stability conditions of the model are established. The main part of the thesis is devoted to the analysis of the stationary number of customers in the systems. A two-step methodology is applied, which, through the problem separation, results in a simpler overall treatment of the analysis. In the last part of the thesis the vacation models with batch Markovian arrival process and the classical cyclic polling models are discussed as the most important special cases of the considered polling model.

The methodology used for the stability analysis can be applied for a broad class of service disciplines including the gated, the exhaustive and the G-limited disciplines. This is achieved by the identification of properly chosen embedded Markov chains. The resulted stability conditions are given in terms of service discipline dependent quantities. In the group of new stability results the characterization of global stability, overview of stability regions of a particular station, order of instability of stations, conditions for partial stability and the necessary and sufficient conditions for the stability states of the system are provided. An important limitation of the stability methodology that it can not be applied to the exhaustive-type limited disciplines in its current form.

The two-step methodology used for determining the stationary number of customers in the systems is realized by separating the analysis into two parts. In the service discipline independent part expressions are established in terms of quantities at server arrival and departure epochs. In the service discipline dependent part these quantities are solved for the individual disciplines by means of a unified method. Hence the application of the service discipline independent stationary results to an individual discipline consists of applying the discipline specific solution of the above quantities in the discipline independent stationary results. The new service discipline independent results for the stationary number of customers holds also for the zero-switchover-times polling models. The service discipline specific solution is derived for polling models with several important service disciplines. This policies are divided into two groups. The first group consists of the most important gated and exhaustive disciplines, while the second one includes the general G-limited and decrementing-K disciplines.

The applicability of the methodology and the results to a broad class of service disciplines ensures the unified characteristic of the presented analysis.

Chapter 1

Introduction

A journey of a thousand miles begins with a single step.
- Confucius

1.1 Polling models

The polling model gained greater attention first in the 1970s, when it was used in the analysis of the time-sharing computer systems. In the subsequent decades it has been applied to communication networks and also to modern telecommunication networks with service facilities. There was an explosive growth of research on polling systems, motivated by the increasing number of applications in the communications systems. Numerous variations and extensions of the basic polling model have been proposed and analyzed. Hideaki Takagi, who plays a significant role in the research of polling models, manages a web site containing the full list of references on polling models until the late 1990s, which can be found at http://www.sk.tsukuba.ac.jp/ takagi/polling.html.

Research on polling models continues to the present motivated among others by applications to modern communication networks with differentiated services. Meanwhile, other lines of research have been started, including stability analysis, heavy-traffic analysis and extensions to more general arrival processes. The survey paper of Takagi [173] provide an overview of the analysis and applications of the different polling models.

1.1.1 The classical polling model

The classical polling model in its original form is a single-server continuous-time queueing system, in which the single server attends the stations in cyclic manner (see on Fig. 1.1). Each station has a queue, which is served when the server visits that station. The time required for the server to travel from one station to the next one is called *switchover time*. Usually the classical model assumes single or infinite buffer queues.

Usually the most important performance measure of the polling system is the customer *waiting time* - the time from the arrival of a randomly chosen customer to the start of its service. The mean waiting time plus the mean service time is the *mean response time*, which is important in most computer communication systems (Kleinrock [81]). Another performance measure also used in the analysis of polling systems is the *cycle time*, which is the time between two successive visits of the same station. It is also called as *polling cycle*. Another characterizing quantities of the model are the *station time*, which is the duration of the server visit to a station, and the *intervisit time* of a station, which is the duration between two consecutive server visit to that station.

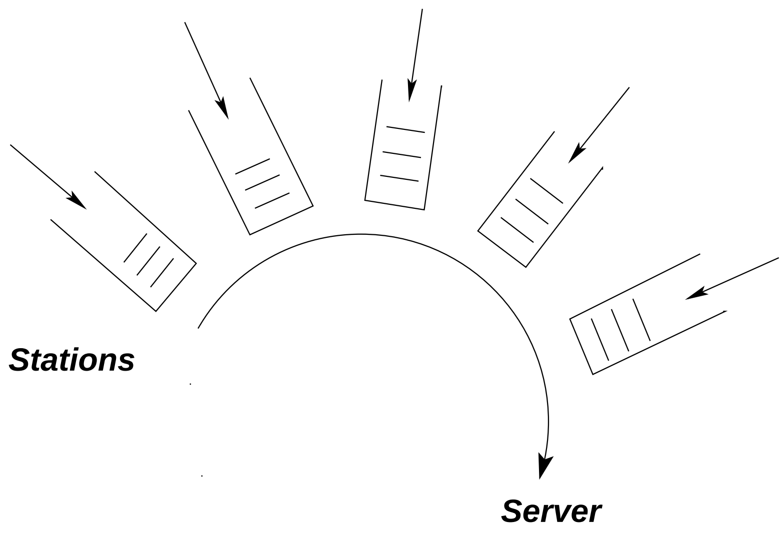

Figure 1.1: Classical polling model

The earliest paper in which the term polling was used is the paper from Hashida [67]. The term *polling* originated from the 1970s, when this model was used to study the multi-terminal computer system. In this system the central computer enquires each terminal in order to determine whether it has data to transmit or not. This explains why is the epoch, whenever the server polls a station, is called *polling epoch*. The first historical application of the basic polling systems was the investigation of a problem in the British cotton industry involving a patrolling machine repairman in the late 1950s (Mack [118]). Somewhat later in the 1960s polling models with two queues were used to analyze traffic signal control (Stidham [145]).

Various kind of early names were also used for polling models, like queues served in cyclic order (Cooper and Murray [39])), queues with periodic service (Eisenberg [47]), multiqueue systems with cyclic service (Kuehn [88]) or class of cyclic schedules (Fuhrmann [58]). Several alternative names have also appeared for polling systems in the literature, e.g. cyclic service systems (Groenendijk [63]), cyclic queueing system (Sarkar and Zangwill [146]), single-server systems with multiple customer classes (Boxma [20]).

The main characterizing stochastic processes of a polling model are determined by the stochastic characteristics of the individual stations. These are the arrival process, the customer service time distribution and the switchover time distribution. In the classical polling model at each station the arrival process of the customers is Poisson and the sequence of customer service times is general independent and identically distributed (i.i.d.). The switchover times at a station in a subsequent polling cycles are also general i.i.d. Moreover in the classical model it is also assumed that the arrival processes, the service times and the switchover times are mutually independent.

1.1.2 Service disciplines

The polling models can be differentiated according to the *service discipline* used at the stations. It is also called as service policy. In general it determines the duration of the service at a given station, i.e. the station time.

The firstly introduced and analyzed disciplines are the *exhaustive* and the *gated* ones. In the exhaustive policy the server serves the station until its queue becomes empty. This is the discipline, which is used also in the ordinary queue. In the exhaustive system with two stations this policy is also referred as *alternating priority* discipline. Such model was studied e.g. by Takács in [166].

Under gated discipline during a visit to a station the server serves only the customers that are present at the polling epoch at that station.

In the exhaustive service system customers at heavily loaded stations experience lower delays than those at lightly loaded stations. This is because at customer arrival the server is more likely to be at the heavily loaded stations. The opposite trend holds in the gated service system. Here the customers arriving to heavily loaded station experience higher delay due to their longer waiting until the next server arrival to the same station. The reason for this is again that the server is more likely at the heavily loaded station. These observations are based on numerical results (Ferguson and Aminetzah [50] and Takagi [168]). In both systems the heavily loaded stations monopolize the server, although this effect is weaker in the gated service system. This is because the customers arriving during the service of a station do not block the services of the next stations as they are served only in the next cycle.

To overcome this drawback a numerous discipline have been proposed, which provide "fairness" in service opportunity (Boxma [21]), which prevents heavily loaded stations from monopolizing the server. Among them the most commonly known is the *limited-1* policy, which is also called as *non-exhaustive* discipline. In the non-exhaustive discipline at most one customer is served at each visit to a station. If the system with non-exhaustive discipline consists only of 2 stations this policy is also called as *alternating service* discipline.

Most of the service disciplines can be classified into *gated-type* or *exhaustive-type* policies (Levy, Sidi and Boxma [104]). In case of *gated-type* disciplines whenever the server visits a station it considers only the customers to be served that are present at the polling epoch at that station. Under *exhaustive-type* policies also the customers arriving during the service of the station are considered to be served during the current visit. The exhaustive-type counterpart of *non-exhaustive* policy is the *semi-exhaustive* discipline, in which the service at a station continues until the number of customers decreases to one less than that are present at the polling epoch at that station.

In order to enable the optimization of the performance of the polling system further *general disciplines* have been introduced, from which at least one of the previously described four basic disciplines can be obtained as special case. In the *limited* policies a proper limit is put on the number of customers that can be served at a station during one visit of the server. These policies can be either *deterministic* or *probabilistic*.

Deterministic limited policies are the *G-limited*, the *E-limited* and the *decrementing-K* disciplines. In case of G-limited discipline at most K customers are served among the customers that are present at the polling epoch at that station. At a station with E-limited discipline either K customers are served or the station becomes empty before. Under decrementing-K discipline the service at a station continues until the number of customers decreases to K less than that are present at the polling epoch at that station or the station becomes empty. Note that the G-limited discipline is of gated-type and both the E-limited and the decrementing-K disciplines are of exhaustive-type. Thus the E-limited and the decrementing-K disciplines represent two different ways for putting limit on the number of customers that can be served on a basically exhaustive manner.

In the probabilistic limited policies the limit is determined randomly. One group of probabilistic limited policies consists of the *Bernoulli-exhaustive* and the *Bernoulli-gated* disciplines (Keilson and Servi [77], Servi [148] and Levy and Sidi [102]). The Bernoulli-exhaustive policy is also called as *Bernoulli schedule* or *Bernoulli scheduling*. At the station with Bernoulli-exhaustive discipline after each completed customer service the server continues to serve the station with probability p or switches to the next station with probability $1-p$. The Bernoulli-gated policy is similar except that the server can serve only the customers that are present at the polling epoch at that station.

Another kind of probabilistic limited policies are the family of binomial service policies. Under *binomial-gated* discipline (Levy [99]) every customer present at the polling epoch is served with probability p. Similarly in *binomial-exhaustive* discipline (Boxma [20]) every customer present at the polling epoch and arrived during its associated busy period is served with probability p. In the *fractional-exhaustive* discipline (Levy [98]) every customer, including also those that arrive during the actual service, is served with probability p. Again there are two different ways of applying the binomial rule resulting in exhaustive-type disciplines.

Finally in the *nonpreemtive limited-T* discipline the service time at a station during one visit of the server is limited by time T. At reaching this limit the service of the

Service discipline	**Special cases**		**Discipline-type**
G-limited	non-exhaustive ($K = 1$)	gated ($K = \infty$)	gated-type
E-limited	non-exhaustive ($K = 1$)	exhaustive ($K = \infty$)	exhaustive-type
decrementing-K	semi-exhaustive ($K = 1$)	exhaustive ($K = \infty$)	exhaustive-type
Bernoulli-gated	non-exhaustive ($p = 0$)	gated ($p = 1$)	gated-type
Bernoulli-exhaustive	non-exhaustive ($p = 0$)	exhaustive ($p = 1$)	exhaustive-type
binomial-gated		gated ($p = 1$)	gated-type
binomial-exhaustive		exhaustive ($p = 1$)	exhaustive-type
fractional-exhaustive		exhaustive ($p = 1$)	exhaustive-type
nonpreemptive limited-T		exhaustive ($T = \infty$)	exhaustive-type

Table 1.1: General service disciplines and their special cases

actual customer is still finished before the server switches to the next station. The polling system with this discipline is also referred simple as time-limited polling model. This policy brings the most "fairness" into the polling system but it makes its analysis difficult. Table 1.1 summarizes the general service disciplines.

A service discipline can be considered to be *more efficient* if the total amount of unfinished work found in the system at any time is smaller. This enables comparisons between the disciplines, which leads to dominance relations among them (Levy, Sidi and Boxma [105]).

1.1.3 Classification

Besides of the service disciplines many variations and extensions to the basic cyclic polling systems have been introduced and analyzed to provide more precise modeling of the operation of practical systems.

The variants of the polling systems can be classified by the following characteristics:

- service discipline
- *symmetric or asymmetric system*

- *single-, infinite- or finite- buffer systems*,
- *system with batch arrival process*,
- *continuous-time and discrete-time models*,
- *nonzero- or zero-switchover-times model*,
- *model with setup time*,
- *model with unreliable server and repair time*,
- *polling order*

In the symmetric system every stations have the same stochastic characteristics. On the contrary in the asymmetric system each station may have different stochastic characteristics, i.e. different arrival rate and different customer service time and switchover time distributions. The system having stations with different disciplines is called *mixed discipline system*.

In the early polling models queues with single or infinite buffers had been applied. In the single buffer model each station can have at most one outstanding customer. A new arriving customer is stored only after completing the service of the previous one, otherwise it is lost. Besides of the already mentioned paper of Mack [118] numerous paper have been published on the analysis of the single buffer polling system, among others e.g. Takagi [169], Takine [176], [177]. Finite buffers are used to model the limited capacity of the system. However such polling systems are more difficult to analyze (Tran-Gia [163], Takagi [171], Kofman [86], Jung and Un [74]).

In order to model the situation in the telecommunication systems, in which the transmission units can consists of more messages, packets, frames, etc., the polling model has been extended to have batch arrival of customers, i.e. multiple arrival (Levy and Sidi [103]).

Another motivation for a new variant of the polling systems came also from the telecommunication systems. It is from the clock-synchronized slotted operation, which lead to the introduction of the discrete-time polling models. Besides several common principles the methods and techniques used in the analysis of discrete-time polling models are quite different from their continuous-time counterparts. For methods and techniques used in discrete-time polling models we refer to the fundamental book of Bruneel and Kim [30].

In production systems often there are no switchover-times. This can be modeled by zero switchover-times model, in which at the end of a visit to a station the server immediately switches to the next station. Usually the server is off while the system is empty. In the literature of polling models, traditionally the nonzero-switchover-times and the zero-switchover-times models are considered separately. This is because some equilibrium arguments can not be directly applied to zero-switchover-times models due to server stops in the idle periods. As a consequence, a considerable research effort has been devoted to relate the zero- and nonzero-switchover-times models, see Fuhrmann [59], Cooper, Niu, and Srinivasan [40] and Srinivasan, Niu, and Cooper [154].

The polling model of production systems can incorporate also setup times before starting the service of a station. In contrast to the switchover time setup time is usually incurred only if there is waiting customer at the station. Such models have been studied by Gupta and Srinivasan [64] and Cooper, Niu and Srinivasan [41].

In order to capture failures of polling systems models with unreliable server have been introduced. The server eventually breaks down and it does not handle the customers during a random period, which is called repair time. Such model has been investigated by Nakdimon and Yechiali in [135].

The *periodic polling* is a generalization of the cyclic polling in order to model systems, in which the server visits the stations in periodic order. Such systems are e.g. an elevator in a building and the moving-arm disk device of a computer, in which the server visits the stations first in one direction and then in the reverse direction. In this model the polling is performed in a fixed order according to a *polling table*. The general analysis of periodic polling models can be found in Eisenberg [47] and Baker and Rubin [7].

The noncyclic order of polling can be either *deterministic*, like in case of periodic polling, or *probabilistic*. An example for the probabilistic polling order is the *cyclic Bernoulli polling* introduced by Altman and and Yechiali in [4]. In this polling order the server serves the actual station with probability p or switches to the next station with probability $1 - p$, i.e. the random Bernoulli decision is applied on the level of the cyclic polling order. Another probabilistic polling order is the *Markovian server routing*, in which the server visits station j after station i with given probability $p_{i,j}$. The term *random polling* is also used (see e.g. Levy [97]) for the special case in which the server selects the next station according to an i.i.d. uniform distribution. The exact analysis of several random polling systems are given by Kleinrock and Levy in [82] and by Lee and Sunjaya in [95]. In both the deterministic and the probabilistic polling orders the movement of the server do not depend on the system state. Such polling orders are called *static polling order*. Takagi has pointed out in [173] that the analysis of the basic cyclic polling models can be extended rather easily to the corresponding static polling order model. An alternative to static polling order is the *dynamic polling order*, in which the server movement depends on the system state and it is determined dynamically during the operation of the system (Levy and Sidi [102]).

A fundamental extension of queueing models is the mechanism of feedback, whose study goes back to a classical paper by Takács in 1963 [165]. In *polling models with Bernoulli feedback* the customers, whose services are completed, are fed back instantaneously to the tail of the queue of the current station with probability $1 - \sigma$ or left the system with probability σ, for $0 < sigma <= 1$. Such polling models were investigated in Takine, Takagi and Hasegawa [178], Li and Yang [106]. Another kind of feedback mechanism is the Markovian feedback. In *polling models with Markovian feedback* the customers, whose services are completed, can be fed back instantaneously to the tail of the queue of any station with a given probability or can leave the system. This model variant is also called as *polling system with probabilistic routing* (Levy and Sidi [102]). Such systems were studied by Sidi, Levy and Fuhrmann in [150] and by Hirayama in [69].

The wide variety of polling models allows more ways to realize a *priority schemes*. These realizations are also called *priority polling* (Vishnevskii and Semenova [180]). One solution to implement priority is to utilize the server monopolizing facility of the exhaustive stations. In this case the exhaustive discipline is assigned to the higher priority station and any other more fair discipline to the other stations (Srinivasan and Lee [153]). Typically gated or limited-1 polling is assigned to low priority stations. Another way to implement higher priority is assigning higher frequency of visits, thus the priorities are realized on the level of polling order (Murata, Shiomoto and Miyahara [133]). These solutions are called *queue priority* schemes. Finally priority schemes can also be realized by assigning external priorities to the arriving customers (messages), which therefore can be called *message priority* schemes (Wierman, Winands and Boxma [186], Boon and Adan [16]).

Polling systems pose several optimization problems. General service disciplines can be optimized by finding the optimal parameter of the discipline. Such optimizations are performed e.g., for the fractional-exhaustive discipline by Levy [98] and for the nonpreemptive limited-T discipline by Blanc [14]. Another possibility is to find optimal visit orders in the periodic or probabilistic polling system as it is done in the works of Boxma, Levy and Weststrate ([25] and [26]). The dynamic polling order is also suitable for optimization as in such system the server can react to the actual system state by its movement. An algorithmic optimization was provided for such zero-switchover-times polling systems by Klimov in [84] and [85]. The optimal control of server movement in nonzero-switchover-times systems is much more difficult. The only result about it is by Hofri and Ross [71]. Another dynamic control problem of periodic polling model is selecting the optimal visit order at the beginning of each cycle. Such system is referred to as polling model with *semi-dynamic server routing*. This optimization task is considered in Browne and Yechiali [29] and Yechiali [195], [194].

For further discussions on optimization issues in polling models we refer to the recent survey of Vishnevskii and Semenova [180].

1.1.4 Vacation models

The *vacation model* is a special case of the cyclic polling model, in which the number of stations is one. Thus the switchover time of the only station of the polling model becomes the *vacation period* of the vacation model.

On the other hand the vacation system can model a particular station of a polling model, where the vacation period corresponds to the intervisit time of the particular station. In this case the polling model, seen from the point of view of the particular station, is a special kind of vacation model.

Research on vacation models run parallel and frequently overlapped with the research on polling models. Vacation models are also distinguished by their service discipline that is the set of rules determining the beginning and the end of the service (vacation).

The vacation models are classified also according to the applied vacation policy. In the *multiple vacation policy* the server immediately takes the next vacation at finding the station empty upon return from vacation. In case of the *single vacation policy* if the server finds the station empty upon return from vacation then it waits for the next customer arrival. Among others both policies are included by the *generalized vacation model* defined in Fuhrmann and Cooper [60] for the $M/G/1$ queue. Note that the definition of the *generalized vacation model* implies that the length of a vacation period depends on the arrival process.

For more details on the analysis of the classical vacation models with Poisson arrival process we refer to the comprehensive survey of Doshi [43] and to the excellent book of Takagi [172].

1.1.5 Stability

Stability of polling systems were defined on more alternative ways. According to the *stability definition* of Kuehn ([88]) the system is stable if for positive service times and finite arrival rates the average number of customers at every stations are finite.

Apart from several exceptions the research on stability analysis of polling models has been started relatively late, only at the beginning of 1990s. One important exception is the above mentioned early work of Kuehn in 1979 [88], 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 [61].

Let N be the number of stations in the classical cyclic polling model. Let λ_i, b_i be the arrival rate of the Poisson arrival process at station i, the mean of the probability distribution of the customer service time at station i, for $i = 1, \ldots, N$, respectively. r_i denotes the mean of the probability distribution of the switchover time following the service of station i and we use also the notation $r = \sum_{i=1}^{N} r_i$. Furthermore $\rho_i = \lambda_i b_i$

is the server utilization at station i and $\rho = \sum_{i=1}^{N} \rho_i$ stands for the overall utilization. The necessary conditions for stability of this polling model with the most common exhaustive, gated, non-exhaustive and semi-exhaustive disciplines were given by Boxma and Groenendijk in [22] as

- exhaustive and gated: $\rho < 1$,
- non-exhaustive: $\rho + \lambda_i r < 1$, i=1,...,N,
- semi-exhaustive: $\rho + \lambda_i(1 - \rho_i)r < 1$, i=1,...,N.

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 [6] 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 [18] 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 [56] studied the stability of 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 [44] presented the stability condition for polling model with multiple servers and with general independent interarrival and service times. This model with server routing is a generalization of the cyclic model. Foss, Chernova and Kovalevskii [55] investigated the stability of multi-server polling models with state-independent server routing. The stability of the polling model with state-dependent server routing was studied by Foss and Last in [54]. Foss and Kovalevskii [53] 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 [111] gave a sufficient but not necessary condition for the stability of polling models with Markovian server routing. In [111], 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 [52] 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 ergodic input flow, where however the customers are randomly routed to the stations. The proofs of [52] are based on the monotonicity properties of the model and dominance theorems.

A recent work on stability is the one by Lillo [108], who gave an ergodicity analysis for polling systems with two queues.

1.1.6 Infinite-buffer cyclic polling models

Stationary analysis and frameworks

The focus of the analysis of polling system is usually the derivation of stationary quantities, like e.g. the probability-generating function (PGF) of the stationary number of customers at polling epoch of station i for $i = 1, \dots, N$. In most of the analysis works on polling models "limiting", "limiting average" or "stationary" quantities are used as they refer to the different assumed contexts or analysis frameworks. The different variants are illustrated on the above mentioned quantity as an example.

Let $F_i(m)$ denote the number of customers at polling epoch of station i in the m-th polling cycle for $m = 1, \dots$. The limiting PGF of the stationary number of customers at polling epoch of station i is defined as a pointwise limit and it is given as

$$\lim_{m\to\infty} E[z_i^{F_i(m)}], \quad |z_i| \le 1. \tag{1.1}$$

The limiting average PGF of the stationary number of customers at polling epoch of station i is defined as

$$\lim_{m\to\infty} \frac{\sum_{\ell=1}^{m} E[z_i^{F_i(\ell)}]}{m}, \quad |z_i| \le 1. \tag{1.2}$$

The limiting average variant of a quantity is exactly what a "random observer" sees. Given the observer arrives until the end of the m-th polling cycle, the observer's arrival is uniformly distributed on the first m polling cycles. Thus the distribution of what the observer "sees" is obtained by letting $m \to \infty$.

The stationary PGF of the stationary number of customers at polling epoch of station i can be defined e.g. in the context of an underlying *regenerative process* (for the regenerative process see Stidham [160] or Heymann and Sobel [68], Chap. 6). As the name suggests the stochastic behavior of this kind of process repeats itself from the beginning each time when a *regenerative point* is reached. In the above mentioned underlying regenerative process these regenerative points are the ones at which the polling system becomes empty. The interval between two consecutive regenerative points is called *regenerative cycle*. The above stationary quantity is the one in any polling cycle of the *stationary version of the regenerative process* and it is defined as

$$\frac{E[\sum_{\ell=1}^{C^r} z_i^{F_i(\ell)}]}{E[C^r]}, \quad |z_i| \le 1, \tag{1.3}$$

where C^r is the length of the first regenerative cycle in number of polling cycles. The stationary variant of a quantity is also called as "equilibrium" quantity. The terminology arises from the construction of stationary (equilibrium) versions of processes. Sometimes this variant is also called as "steady state" quantity.

Stidham shown in [160] that if the mean of the regenerative cycle is finite then the stationary distributions defined by means of the underlying regenerative process exist. He also shown that in that case in the queueing context (for the precise mathematical conditions see [160]) the corresponding limiting distribution also exists and these "limiting" and the "stationary" distributions are the same. Moreover the existence of the limiting distribution implies also the equality of the "limiting" and "limiting average" distributions (see Appendix C). Consequently in a "well-behaved" queueing model all of these quantities are equal and hence they can be used equally.

The stationary analysis of polling system can be based on direct definitions of the limiting or limiting average quantities by means of proper limits. This way is followed e.g. in the early work of Eisenberg [47].

Takagi applied a *regenerative process framework* for the analysis of cyclic polling systems in [167]. It is shown in [167] that the mean of the regenerative cycle is finite if and only if the mean polling cycle is finite, which follows from the stability of the system. Thus the stability of the polling system ensures the applicability of the regenerative process framework. Applying this framework it is enough to establish the average over the regenerative cycle in the definitions of the stationary quantities, as it is shown in (1.3), and hence it makes the stationary analysis simpler.

An alternative *Markov regenerative process (MRP) framework* (for the MRP see Kulkarni [89]) has been proposed for cyclic polling systems in [202]. Applying this framework the analysis becomes even simpler, since it is enough to establish the average only over the polling cycle in the definitions of the stationary quantities. We remark here that the stationary variant of a quantity, alternatively, can be also defined in the context of an underlying MRP. In this case the stationary quantity is the one in any polling cycle of the *stationary version of the MRP*. The existence of the stationary distribution is ensured directly by the finiteness of the mean polling cycle.

Stochastic decomposition

One of the most significant result of vacation models, which holds also for a station of a polling model, is the *M/G/1 decomposition property.*

Property 1.1 *(M/G/1 decomposition property.) The stationary number of customers present at a station at an arbitrary epoch can be given as the sum of two independent components, the stationary number of customers present in the corresponding standard*

M/G/1 queue of that station at an arbitrary epoch and the stationary number of customers seen at an arbitrary epoch in the intervisit time of that station.

It can be shown that similar decomposition property holds also for the waiting time under more restrictive conditions. The M/G/1 decomposition property in a more general setting was presented first by Fuhrmann and Cooper [60], who proved it under very general conditions, for the M/G/1 queue with generalized vacation.

Later, Borst and Boxma in [19] has presented an alternative derivation of the stochastic decomposition, which is valid both for the nonzero- and the zero-switchover-times polling models. Bertsimas and Mourtzinou [9] generalized the stochastic decomposition for the stationary number of customers in polling systems with arbitrary polling order. They made use of the distributional Little's law (DLL) (see Bertsimas and Nakazato [10]) and assumed First-In-First-Out (FIFO) service scheduling and that newly arriving customers do not affect the time the preceding customers spend in the system.

The application of M/G/1 decomposition property reduces the problem of finding the PGF of the stationary number of customers at an arbitrary epoch to determine the PGF of the stationary number of customers only at characteristic epoch of the system, namely at start and at end of the server visit to the considered station.

Pseudo-conservation law

An exact result for the classical cyclic polling models is the pseudo-conservation law, which is a closed-form expression for the weighted sum of the mean waiting times (Boxma and Groenendijk [22]. Let $b_i^{(2)}$ be the second moment of the probability distribution of the customer service time at station i for $i = 1, \ldots, N$. $r_i^{(2)}$ stands for the second moment of the probability distribution of the switchover time following the service of station i and $r^{(2)} = \sum_{i=1}^{N} r_i^{(2)}$. Furthermore $E\left[W_i\right]$ denotes the mean waiting time of an arbitrary customer at station i. For the stable classical cyclic polling model the general form of the pseudo-conservation law is given by

$$\sum_{i=1}^{N} \rho_i E\left[W_i\right] = \rho \frac{\sum_{i=1}^{N} \lambda_i b_i^{(2)}}{2\left(1-\rho\right)} + \rho \frac{r^{(2)}}{2r} + \frac{r}{2\left(1-\rho\right)} \left(\rho^2 - \sum_{i=1}^{N} \rho_i^2\right) + \sum_{i=1}^{N} b_i m_i,$$

where m_i stands for the number of customers at station i at a departure epoch. Thus the dependency on the given service disciplines is incorporated via the last term including m_i-s.

In [22] this law has been given for the exhaustive, gated, non-exhaustive and semi-exhaustive disciplines. The law has been extended to the Bernoulli-exhaustive discipline by Tedijanto [179] and it is also available for the binomial-gated and fractional-exhaustive disciplines (Levy [101], [98]). Recently it was also proved for the time-limited discipline with exponential timer (Katayama and Kobayashi [76]). Moreover this law holds even for non-cyclic polling models (Boxma, Groenendijk and Weststrate [24]).

Although the pseudo-conservation law does not give the exact mean waiting times in the asymmetric polling model it has several merits. It provides a qualitative insight into the behavior of the polling system, it is a measure of overall system performance, it can be used for checking the correctness of numerical computations as well as for validating simulation results. In addition the pseudo-conservation law has been extensively used to develop mean waiting time approximations (e.g. Boxma and Meister [23], Everitt [48], Groenendijk [63]).

Closed-form solutions

For symmetric system the pseudo-conservation law gives directly the exact expressions for the mean waiting time for the disciplines, for which it holds. Comparing them leads to the following relations for the mean waiting times among the exhaustive, the gated, the non-exhaustive and the semi-exhaustive disciplines (Takagi [170]):

$$E[W]_{exhaustive} \leq E[W]_{gated} \leq E[W]_{non-exhaustive}$$
$$E[W]_{exhaustive} \leq E[W]_{semi-exhaustive} \leq E[W]_{non-exhaustive}.$$

In asymmetric polling model the exact solution for the waiting time distribution is known for the cyclic Bernoulli polling, i.e. in which the cyclic polling order is randomly altered by the Bernoulli decision (Altman and Yechiali [4]).

In the cyclic polling model under asymmetric settings the only discipline for which the closed-form expression of exact mean waiting times is known is the *globally gated* (GG) discipline. Under the GG discipline during a server visit to a station only those customers are served that are present at the beginning of the cycle. Thus in contrast to the gated policy the customers arriving from the beginning of the actual cycle to the polling epoch of the station must wait until the service in the next cycle. The GG policy leads to a simpler mathematical analysis than those for polling systems with gated or any other known discipline and thus closed-form expressions can be derived also for the distribution of the stationary waiting time. This discipline has been introduced by Boxma, Levy and Yechiali in [27] for modeling a maintenance process, in which the preventive maintenance requirements are assigned to the repairman at the beginning of each cyclic order tour.

Basic methods for exhaustive and gated service systems

Infinite buffer cyclic polling systems with exhaustive and gated disciplines play a central role in the analysis of polling models, as they are the most intensively studied polling models. The main concern of the analysis of these models are the stationary waiting time, which is often derived from the mean number of customers by using Little's law (Little [109]).

Unfortunately no closed form solution exists for these polling models except special cases like the case of symmetric systems and the case of asymmetric system with GG discipline. Instead numerical approaches were developed to compute the mean waiting time. There are four basic methods for solving these polling models:

- *Buffer occupancy method*
- *Station time method*
- *Method of Hirayama et al.*
- *Mean value analysis*

The most widely used method for analyzing polling system with exhaustive and gated disciplines is the buffer occupancy method (used e.g. in Cooper and Murray [39], Cooper [38] or Eisenberg [47]). This method is based on the buffer occupancy variables $F_i(j)$ which denote the number of customers at station j at polling instant of station i, for $i, j = 1, \ldots, N$. A system of linear equations consisting of N^3 equations can be derived for the unknowns $E[F_i(j)F_i(k)]$, from which the mean number of customers can be computed. The fundamental book of Takagi [167] on the mathematical analysis of the basic cyclic polling systems applies also the buffer occupancy method for the exhaustive and gated polling systems. The book describes the analysis for both discrete-time and continuous-time models. The system of linear equations with the unknowns $E[F_i(j)F_i(k)]$ can be also iteratively solved resulting in $O(N^3 \log_\rho \epsilon)$ operations, where ρ is the overall utilization and ϵ is the required relative accuracy (Levy [100]). Two advanced techniques were also developed based on the buffer occupancy method. The first one is the *descendant set* method (Konheim, Levy Srinivasan [87]) which is based on counting the number of descendants of each customer in the system. It is an iterative technique and it can be used for computing the mean number of customers at each station independently from the other stations in $O(N \log_\rho \epsilon)$ operations. The second advanced method is the *individual station* technique (Srinivasan, Levy and Konheim [155]), which is not an iterative method. According to its name it is also suited for computation of the mean number of customers at individual station, which requires $O(N^2)$ operations.

Another basic method is the station time method (Ferguson and Aminetzah [50]) using the station time variables S_i^+, for $i = 1, \ldots, N$. S_i^+ is the duration of the server visit at station i, plus the preceding or succeeding switchover time in case of exhaustive or gated service, respectively. The mean number of customers at every stations are obtained by solving N^2 linear equations of the unknowns $E[S_i^+ S_j^+]$. This system of linear equations can be solved iteratively, for which $O(N^2 \log_\rho \epsilon)$ operations are necessary. The station time method was further developed by Sarkar and Zangwill in [146] resulting in a system of N linear equations, which is less sparse. Its solution requires $O(N^3)$ operations resulting in the mean number of customers at all stations simultaneously.

Hirayama, Hong and Krunz developed a new, third method [70] using mean waiting times conditioned on the state of the system. The method is based on a set of linear functional equations for these conditional mean waiting times, which is obtained from the analysis at polling instants. Starting from them $N(N+1)$ linear equations are derived for the unconditioned mean waiting times. Their solution requires N^6 operations and gives the mean waiting times at all stations simultaneously. In spite of the complexity of the solution this method has some similarities to the buffer occupancy method and therefore it may be possible to construct a more efficient iterative algorithm to solve this system of linear equations.

Recently Winands, Adan and Houtum established a fourth basic method [188] for computing the mean waiting times in polling system with exhaustive and gated disciplines. Their approach is based on the application of the mean value analysis (MVA) to polling system. It uses the quantities $Q_i(j)$, which denote the number of customers at station i at an arbitrary epoch within the server visit at station j for $i, j = 1, \ldots, N$. The mean waiting times at all stations can be obtained simultaneously from the solution of no more than $N(N+1)$ linear equations with the unknowns $E[Q_i(j)]$. The paper [188] left open the question whether it may be possible to construct a more efficient iterative algorithm for solving this system of linear equations or not. The main advantage of this approach is its intrinsic simplicity, since MVA enables pure probabilistic reasoning and the unknowns in the system of linear equations are all first moments.

Among the above basic methods the buffer occupancy method and the approach applying MVA can also be used in polling systems with mixed disciplines, in which some of the stations have exhaustive policy and some others have gated one. It is also known that the station time method can not be used in such mixed discipline system, while the method of Hirayama et al. was not considered for this case.

Moreover the buffer occupancy method and, as stated in [188], possibly also the MVA can be applied for a class of service discipline satisfying the *branching property* (Fuhrmann [58], Resing [143]). The following definition of this property is taken from page 2 of [143].

Property 1.2 *(Branching property.) If the server arrives at station i to find k customers there, then during the course of the server's visit each of these k customers will effectively be replaced in an i.i.d. manner by a random population having PGF $h_i(z_1, \ldots, z_N)$.*

Resing [143] has studied polling systems with disciplines satisfying this property. Besides of the exhaustive and the gated policies this class of disciplines includes also the binomial-gated, binomial-exhaustive and fractional-exhaustive ones. Unfortunately the Bernoulli-schedule policy does not satisfy the branching property. Resing introduced the *Bernoulli-type* discipline ([143]), which in the contrary satisfies the branching property. Under Bernoully-type discipline each customer that are present at the polling epoch of the station is served and its busy period is handled independently on the way that it is served according to the Bernoulli schedule with parameter p. Thus both the gated ($p = 0$) and the exhaustive ($p = 1$) disciplines are special cases of it. It is shown in [143], that for polling systems with the class of service disciplines satisfying the branching property the joint process of the number of customers at all stations can be represented by a multi-type branching process (MTBP) with immigration. The analysis of these polling models by using multi-type branching process with immigration also utilizes the stochastic decomposition property (1.1). The full and a brief description of the analysis can be found in the paper of Borst and Boxma [19] and in the paper of Adan, Boxma and Resing [1], respectively.

Numerical methods for solving polling models

The previously discussed basic methods do not provide the higher moments of the stationary number of customers or of stationary waiting time. Moreover they can not be applied on polling models with general disciplines which do not satisfy the branching property. For solving these problems several numerical techniques have been proposed. However most of these numerical techniques have relatively large computational complexity.

Choudhury and Whitt [35] developed an iterative technique based on numerical transform inversion for computing also the higher moments of both transient and steady-state performance measures of the system. They applied it for the cyclic polling model with gated service. It requires $O(N^{(1+\iota)})$ number of operations for computing the moments of a steady-state performance measure for all stations, where ι is typically in the range 0.6 to 0.8. The method can be applied also for polling models with other disciplines that satisfy the branching property.

Leung [96] developed a numerical method based on the fast Fourier transform, which can be applied to polling models with exhaustive-type disciplines. The method computes the joint PGF of the number of customers at all stations at the server visit completion instants on an iterative manner through the polling cycles until reaching the required precision. He applied it to a polling model with the newly introduced *probabilistically - limited* discipline. Under this policy the maximum number of customers served during a server visit is determined by a probability, which is independent of the system state.

Thus this is a general discipline including the exhaustive, the E-limited and the Bernoulli schedule policies as special cases.

The power series algorithm (PSA) introduced by Hooghiemstra et al. [72] is a numerical procedure which can be applied to any Markov process formally. It expresses the stationary probabilities of the process as a power series of some parameter, which is usually the load of the system in queueing models. Blanc has shown that it can be applied to polling systems with broader class of service disciplines. In a series of papers he applied the power series algorithm in polling systems with Bernoulli service, with the E-limited and the time-limited disciplines (see [12], [13] and [14]). In polling models PSA is used on balance equations, whose establishment requires that the distributions of the service times and of the switchover times have to be compositions of exponential distributions. This is a restriction, since most of the methods for polling models assumes generally distributed service times and switchover times. Moreover the convergence properties of the algorithm are unknown, and therefore the convergence of the method can not be guaranteed for an arbitrary model.

Heavy-traffic analysis of polling systems

The *heavy-traffic analysis* belongs to the relatively new research areas. In this analysis the limiting behavior of the waiting time is studied at the stability boundary of the system. It is known that all stations in the classical cyclic model with exhaustive and gated disciplines becomes unstable if $\rho \to 1$. The heavy-traffic analysis is based on the observation that the W_i, as a function of ρ has a first-order pole at $\rho = 1$. Therefore the main performance measure of interest is the *scaled delay*, which is defined as

$$\Omega_i = \lim_{\rho \to 1}(1-\rho)W_i, \quad i = 1, \ldots, N.$$

Following the fundamental work of Coffman, Puhalskii and Reiman [36] R.D. van der Mei and his co-authors provided a heavy-traffic analysis of the classical poling system with Poisson arrivals in a series of papers ([127], [128], [121], [122] and [123]). The results are closed-form expressions for the waiting-time and thus they lead to fast-to-evaluate approximations. The analysis was also extended to periodic polling models ([164]).

The rigorous proof of heavy-traffic (HT) limits requires the proof of the existence of the above mentioned first-order pole of W_i at $\rho = 1$. This has been rigorously proven for models with more than two stations driven by Poisson arrival by R.D. van der Mei in [124]. In this work he also generalize the analysis for polling models with the class of disciplines satisfying the branching property, for which he also presented a general framework ([125]).

The methodology used for heavy-traffic analysis of the classical poling system with Poisson arrivals heavily relies on the existence of non-heavy-traffic results. However such results are not available for polling models with renewal arrivals. Coffman, Puhalskii and Reiman in [37] provided an approach on heavy-traffic analysis with renewal arrivals, in which they formulated conjectures about the asymptotic waiting-time distribution. This approach was adopted by Olsen and R.D. van der Mei who generalized the results on asymptotic waiting-time distribution for polling models with exhaustive and gated services in [141]. They also provided numerical results to validate the theoretical results based on their conjectures. Later in 2007 R.D. van der Mei and Winands provided a rigorous proof of HT limits for the zero-switchover-times gated service polling system with renewal arrivals in [129].

Besides of approximations of waiting times in practical heavy-load scenarios HT limits enable also optimizing the system performance with respect to the service disciplines ([127], [128]).

1.1.7 Further model variants

Besides of the above described variants, polling models have been extended also to other directions. This includes *polling with multiple servers* analyzed by Marsan, Moraes, Donatelli and Neri [120] and by R. D. van der Mei and Borst [126], *polling systems with retrial customers* considered by Langaris in subsequent papers [92], [93].

Similarly to the closed queueing networks *closed polling systems* were also introduced. In these polling systems, a constant number of customers circulate, no customer comes to the system from outside or leaves it. Closed polling systems were studied by Altman and Yechiali in [5] and by Dror and Yechiali in [45].

Introduction of *polling systems in tandem* was motivated by manufacturing systems. Due to the complexity of such model it was analyzed under heavy-traffic conditions (Reiman and Wein [142]).

In the last years several polling models have been extended in several new directions.

In *polling model with patient server* (Boxma, Schlegel and Yechiali [28]) upon finding a station empty a server waits for customers for a specified time before moving to the next station. In *polling model with parameter regeneration* (MacPhee, Menshikov, Petritis and Popov [119]) the service time distribution and the arrival rates at a station change randomly every time when that station becomes empty. Another model variant is the *polling system with impatient customer* (Vishnevsky and Semenova [182]), in which the the service start of a new customer is allowed only in a specified time limit, otherwise the customer leaves the system unserved. In *polling model with an autonomous server* the duration of the server visit to a station according to the applied discipline is extended by an exponentially distributed period. The service of the customers arriving during this interval is allowed. Upon expiry of this interval the service of the customer under service is preempted and repeated when the server returns to the station. Such system was studied by Haan, Boucherie, and Ommeren in [65].

Recently Vishnevsky and Semenova [181] introduced the *adaptive polling order* to model the scheduling schemes of the wireless networks. In the adaptive polling order in some extent the server adapts to the traffic circumstances of the stations. The server jumps over the stations with low traffic and thus their setup times are saved. Therefore it results in higher overall server utilization and it reduces the delay.

Another challenging new model variant is the *polling model with multiple priority levels*. Such model was recently introduced and studied by Boon, Adan and Boxma [17].

Finally we mention the *fluid polling models*, which was studied recently by Yechiali and Czerniak in [42].

1.2 Application of classical polling models

1.2.1 Application areas

As already mentioned in Subsection 1.1.1 the first applications of polling models were the investigations of the patrolling machine repairman problem and of the traffic signal control problem.

After these early application examples polling models were applied in several other fields including manufacturing systems, an automated guided vehicle (AGV) system, an elevator up-and-down route, passenger transportation or circular route (Dukhovny [46]) and mail delivery systems (Nahmias and Rothkopf [134]). In [102] Levy and Sidi describe the use of cyclic polling systems for modeling robotics systems.

Polling models were intensively applied to production systems. In such cases usually non-zero-switchover times polling models used or setup time are incurred before the start of service at a station. On the application of polling models to production systems the reader is referred to Sarkar and Zangwill [146] or Karakul and Dasci [117].

The application to traffic control problems is an ongoing activity, see e.g. the recent work of Simoes, Oliveira and Costa [151] and Suzuki and Yamashita [161], who use vacation model in analyzing a problem of an alternating traffic crossing at a narrow one-lane bridge.

In the applications to computer systems the polling model was used for scheduling of moving arms in storage devices and for investigating load sharing of multiprocessor computers (Wang and Morris [185]). However the vast majority of application of polling models falls into the field of telecommunication.

1.2.2 Applications in computer communication networks

The appearance of early computer communication networks opened a new application area for the polling models. Polling and vacation models have general modeling capability and they were proven to be effective instruments in modeling and analysis of communication networks. The polling model with two queues with alternating priority discipline has been used to analyze a half-duplex transmission mode (Sykes [162]). Polling system has been used to analyze polling data link control schemes introduced for central computer - terminal communication. The application to roll-call polling and hub polling can be found in Hammond and Reilly [66] and in Schwartz [147]. Polling was implemented in numerous data link control protocol including IBM's Synchronous Data Link Control (SDLC) and CCITT's High-level Data Link Control (HDLC).

In the early 1980s, with the appearance of the Local Area Networks (LANs), the cyclic polling model has gained much attention in the performance analysis of the token passing protocols and other demand based channel access schemes. The analysis of token bus (IEEE 802.4) and token ring (IEEE 802.5) token passing schemes can be found in Bux and Truong [31] and Moraes and Rubin [132]. Levy pointed out that random polling can be applied in random-access distributed control systems [102]. He applied this model to study the reservation ALOHA access scheme (Levy [97]). For more applications to the early telecommunication networks of the 1980s we refer to the survey of Takagi [170].

In the late 1980s and 1990s this model has been applied in channel access protocols in High-Speed LANs (HSLANs) and Metropolitan Area Networks (MANs). The Fiber Distributed Data Interface (FDDI) has been also analyzed by the help of polling models. Due to the complexity of the timed token protocol of FDDI a numerical analysis has been conducted for it by using the power-series algorithm (Altman [3] and Blanc and Lenzini [11]). Another work on performance evaluation of FDDI is from LaMaire [90], who approximated the stochastic behavior of a station of the system by a vacation model. Mixed discipline polling system was used to study the first real MAN, the Distributed Queue Dual Bus (DQDB), which was standardized as IEEE 802.6 (Landry and Stavrakakis [91]).

1.2.3 Applications in modern telecommunication networks

Polling and vacation models were and are extensively used in the analysis of modern telecommunication networks including Asynchronous Transfer Mode (ATM), Mobile Radio Communications, Bluetooth networks ([152]), wireless networks and the various power saving mechanisms proposed and applied to them.

The time-limited polling model was used to evaluate the efficiency of the bandwidth allocation scheme in ATM networks (Li, Sun and Liu [107]).

In Mobile Radio Communications vacation models were used to investigate the effect of Automatic Repeat reQuest (ARQ) error control schemes (Wu, Shu, Niu and Zheng [191]) and to analyze the performance of 3rd generation (3G) and fourth generation (4G) wireless mobile systems (Srinivasan and Baras [156]).

In Bluetooth Piconets a scheduling algorithm was modeled with non-exhaustive cyclic polling by Zussman, Yechiali and Segall [197]. Another application to Bluetooth is the performance modeling with M/G/1 vacation model, which was done among others by J. Misic and V.B. Misic [131].

Polling model with adaptive order was applied to model wireless networks by Vishnevsky and Semenova in series of papers e.g. [181]. Vacation model was applied to wireless network to investigate ARQ-based error control scheme (Le, Hossain and Alfa [112]) and to study wireless scheduling (Yi, Zhangy and Chiang [196]).

Delay analysis of IEEE 802.11 ([157]) broadband wireless network (BWA) Media Access Control (MAC) was conducted by using polling model (Vishnevsky, Dudin, Klimenok, Semenova and Shpilev [184]), by applying vacation model to approximate the stochastic behavior of a station of adaptive polling model (Vishnevsky, Semenova, Dudin and Klimenok [183]) and by applying time-limited vacation model (Ghazizadeh and Fan [62]).

Vacation model was used for delay analysis of non real-time polling service (nrtPS) service flow type in IEEE 802.16 BWA ([158]) with unicast polling (Saffer and Andreev [205], Andreev, Saffer and Anisimov [206], Saffer, Andreev and Koucheryavy [210]) and with contention-based random access (Andreev, Saffer, Turlikov and Vinel [207, 209]).

Power saving mechanism was analyzed by the help of polling model for Time-Division Multiple Access (TDMA) MAC protocols by Sikdar and Yang in [192]. Vacation models were applied in analysis of power saving mechanism for 3G Universal Mobile Telecommunications System (UMTS) system (Yang [193]), for distributed multihop mesh/relay network (Fallahi, Hossain and Alfa [49]) and for IEEE 802.16e (among others Alouf, Altman and Azad [2]).

1.3 Thesis motivation

As the introduction of the numerous extensions and variations of polling models and their applications show, the driving force of the further development of the polling models is the rapid development of the modern telecommunication networks. There is a growing demand for more accurate modeling of the operation of such systems, while they become more complex. The advanced modeling makes their performance evaluation more precise, facilitates the optimization of their parameters alleviating the better tuning of these parameters to the requirements of the actual application scenario.

Further development of polling models can be achieved basically in two directions.

The first one is the introduction of new model variants and adding new extensions to the basic model according to the practical requirements of the considered systems. This became a trend in the field of polling models and continues until present. The second direction is the use of more elaborated stochastic processes. This enables more precise modeling of the stochastic characteristics of the individual stations, like first of all the arrival processes but also the service time distributions. This requires also more complex treatments and solutions of more complicated numerical procedures. Fortunately this was facilitated by the rapid development of the numerical capabilities of the computers in the 1990s.

Specifically the widespread introduction and application of data networks has played an important role in the development of advanced arrival processes. Among others the modeling of Internet traffic motivated and promoted the development, analysis and application of more powerful arrival process models. The most straightforward way of extending the capabilities of the arrival processes is the generalization of the existing ones. Theoretical objectives of this generalization are the analytical tractability and preserving the renewal character of the original process. The best candidate for this is the Poisson process with exponential interarrival times. Its early generalization, the Erlang-k arrival process, has already been used by A. K. Erlang at the beginning of the 20th century in traffic modeling of telephone calls. From practical point of view it is expected that the generalized process models the correlation among the interarrival times and it captures the internal structure of the real life traffic. These objectives resulted in the introduction of various types of new arrival processes, whose interarrival time is built up from different compositions of exponential distributions. Such useful arrival processes are e.g. the Markov-modulated Poisson process (MMPP) (see in K.S. Meier-Hellstern and W. Fischer [130]) or the phase type renewal process (PH-renewal process) introduced by M. F. Neuts [137], which are all the special cases of the Markovian arrival process (MAP) defined by D. M. Lucantoni, K. S. Meier-Hellstern and M.F. Neuts in [116]. They remained analytically tractable, they can model the correlation among the interarrival times and the MAP can capture the internal structure of the real life traffic, at least to some extent.

A further generalization of MAP to the case of batch arrivals is the batch Markovian arrival process (BMAP) introduced by Lucantoni ([114]). Many published papers have investigated the traffic modeling capability of BMAP, see e.g. Klemm, Lindemann and Lohmann [83]. It turned out that a wide range of real life traffic can be approximated by BMAP and this is valid also for traffic having non-Markovian behavior (A. Horvath [73]). However the problem of fitting the BMAP to a given real-life traffic is solved only for special classes of BMAP, e.g., for the case of two-phase MAP (Bodrog, Heindl, Horvath and Telek [15]).

Apart from queueing models with or without vacation only few results can be found on polling models, whose arrival process is any of the above-mentioned special cases of MAP. Bertsimas and Mourtzinou [9] generalized the stochastic decomposition for the case when the interarrival times have mixed generalized Erlang (MGE) distribution, which is a special case of the PH-renewal process. Lee [94] presented a polling model with more general arrival process, which enables also MMPP. In the finite buffer two queue MAP/M/1 polling model of Chakravarthy and Thiagarajan [32] the server somewhat unconventionally treats concurrently one customer from each queue if both queues have customers. A very elaborated analysis of the MAP/PH/1 polling model can be found in Frigui and Alfa [57]. They make use of treatment of each station as an individual MAP/PH/1 vacation model. All of these analysis of polling models with MAP also utilize the Markovian characteristic of the customer service time. We are unaware of any previous work on general service time polling models with MAP.

Due to the more realistic traffic modeling capability of BMAP the analysis of queueing models with BMAP attracted a great attention from the beginning of 1990s. The most of the analysis works on BMAP queueing models including vacation models are based on the standard matrix analytic-method pioneered by Neuts [138]. Relative few results has been published on BMAP vacation models with disciplines others than the exhaustive ones and on BMAP priority queueing models. However to the best of our knowledge so far none of the polling models with BMAP has been analyzed.

Summarizing all these so far, although there were need for polling models with BMAP for more precise performance evaluation of modern telecommunication networks, such polling models have not yet been investigated. This motivated us to extend the classical cyclic polling model to BMAP and investigate it.

1.4 Thesis objectives

The principal goal of this thesis is to provide a unified analysis of cyclic polling models with BMAP. Only the continuous-time nonzero-switchover-times basic model with stations having infinite buffer queues is addressed. The unified character of the analysis ensures the applicability of the applied methodology and the obtained results for a broad class of service disciplines including the most common gated and exhaustive policies and at least one general discipline, e.g. the G-limited one.

The first objective of the thesis is to establish stability results for this model for a group of service disciplines including the gated, the exhaustive and the G-limited disciplines. In the practical applications the most important stability result is the condition to ensure the stability of the system. Therefore a sufficient condition on the whole stability is targeted.

The major objective of the thesis is to present a service discipline independent analysis of the cyclic polling models with BMAP. Here the term "service discipline independent" means that the analysis results are valid for a broad class of service disciplines. The dependency on the individual disciplines are incorporated by discipline specific quantities. Such a unified analysis considerably simplifies the analysis of the system, since after having the service discipline independent results it is enough to determine only the discipline specific quantities in a second part of the analysis. Similar problem separation principles are the stochastic decomposition or PGF factorization forms (product form PGF expression), which have already been successfully applied in the analysis of queueing models (Bertsimas and Mourtzinou [9] or Chang and Takine [33]).

Thus this kind of analysis results in a two-step methodology in which the service discipline independent analysis is embedded as first part. A suitable candidate for the methodology is the possible generalization of the one used by Borst and Boxma in [19] as it has already been successfully applied in the analysis of the cyclic polling systems with Poisson arrivals.

The main target of the major objective is the set-up of a closed form expression for the vector GF of the stationary number of customers at an arbitrary instant in terms of discipline specific quantities. In the applications usually instead of PGF the mean of the stationary number of customers is the quantity of interest. Hence the second target of this objective is the determination of the vector mean of the stationary number of customers at an arbitrary instant.

The last objective of the thesis is the application of the service discipline independent results to the most common policies including the gated, the exhaustive and the G-limited disciplines. This can be achieved by the determination of the discipline specific quantities in the discipline independent results. Although this can be performed on discipline specific way, the investigation of a possibly unified method is targeted for realizing it.

Even though the analysis and the results can be extended or generalized to other variations and extensions to the basic cyclic polling models with BMAP, it is out of scope of this thesis to deal with them. Waiting time analysis of the considered model is not covered by this thesis. These topics are left for future research. Furthermore elaborating any application based on the results presented in the thesis, beyond general considerations on the application areas, is also not covered by the thesis. Instead it is a challenging future work.

1.5 Methodology

1.5.1 Stability

The applied stability methodology is based on identification of properly chosen embedded Markov chains. The basic idea of this methodology incorporates several elements from the work of Fricker and Jaïbi in [56], like the limited and unlimited type of disciplines and proper embedded Markov chain. However our proposed methodology requires only relaxed conditions on the service disciplines in comparison with the monotonicity property of [56]. These conditions are rather similar to the assumptions of the model of Down [44]. The proposed methodology allows the generalization of the arrival process to BMAPs and a much simpler stability analysis than the existing ones based on monotonicity properties and dominance theorems like the [56].

Moreover it leads to a unified stability analysis, since it can be applied for a fairly general set of service disciplines.

1.5.2 Queueing models with BMAP

The vast majority of the analyzed BMAP/G/1 queueing models are based on the standard matrix analytic-method pioneered by Neuts [138] and further extended by many others (see e.g., Lucantoni [115]). It provides the vector probability-generating function (PGF or vector GF) of the number of customers at departure or at an arbitrary epoch. Determination of the waiting time distribution requires more effort, see Kasahara, Takine, Takahashi and Hasegawa [75].

The standard matrix analytic-method exploits the underlying M/G/1-type structure of the model, i.e., that the embedded Markov chain at the customer departure epochs is of M/G/1-type (Neuts [139]) in which the block size in the transition probability matrix equals to the number of phases of the BMAP. The easy identification of the exceptional boundary states makes the standard matrix analytic-method very suitable for analyzing models with exhaustive discipline. However for a broad class of BMAP/G/1 queuing models the number of customers and the phase of BMAP at customer departure epochs does not form a kind of M/G/1-type Markov chain, in which the block size in the transition probability matrix equals to the number of phases of the BMAP ([203]). As a consequence of it the proper definition of the system states, which results in an M/G/1-type structure, makes the state space very complicated and hence the analysis becomes cumbersome. This holds for vacation models with e.g. gated and G-limited disciplines.

As a consequence of it very few work is available in the literature on such BMAP/G/1 vacation models and they apply different methods. Ferrandiz [51] used Palm-martingale calculus to analyze a flexible vacation scheme. Shin and Pearce [149] studied queue-length dependent vacation schedules by using the semi-Markov process technique. Recently Banik, Gupta and Pathak [8] studied the BMAP/G/1/N queue with vacations and E-limited service discipline. They applied supplementary variable technique to get the queue length distributions and several system performance measures.

Unfortunately the M/G/1 decomposition property can not be generalized completely for queuing models with BMAP. Instead factorization forms can be derived for numerous queuing models in PGF domain. Kim, Chae and Chaudhry [79] shown the invariance relation, which relates the stationary number of customers in the system and the stationary number of customers in the queue on a very simple way. It can be applied to get factorization forms also in BMAP/G/1 queueing models. Chang, Takine, Chae and Lee [34] established a factorization property for the BMAP/G/1 queues with generalized vacations. The vector GF of the stationary queue length is factored into two PGFs of proper random variables. One of them is the vector GF of the conditional stationary queue length given that the server is on vacation (or idle). Chang and Takine [33] applied this factorization property together with matrix analytic-method to get analytical results for several fundamental $BMAP/G^B/1$ queueing models with generalized vacations and with exhaustive discipline. In the factorization forms usually some matrix generating functions (matrix GFs), which occur in the forms, has no stochastic interpretation. In spite of this drawback the factorization forms make the analysis of vacation models with BMAP considerably easier due to their closed-form.

Relative few results can be found in the literature on BMAP queuing models with multiple queues. One of such models is the nonpreemptive priority queue, which is slightly different from the zero-swithover times polling model with exhaustive service, since that model has common underlying arrival process forming a finite state continuous-time Markov chain (CTMC). Nishimura [140] gave a spectral method for the nonpreemptive BMAP/G/1 priority queue with two priority classes. Takine [174] has analyzed that nonpreemptive priority queue model with MAP arrivals. In his work he pointed out, that the matrix analytic-method allows only one random variable having countable infinite space to describe the system dynamics. However the priority queueing model, exactly like the cyclic polling model, requires mutually dependent random variables, each of which is defined in the countable infinite space, i.e. the number of customers in each priority class. Consequently the matrix analytic-method alone is not enough to analyze such a model. Hence for the analysis of this model he also utilized a relation among the stationary PGF of the queue length at the embedded customer departure epochs and of the time-average queue length (Takine and Takahashi [175]).

As a summary it can be stated that the main methods used in the analysis of queueing models with BMAP are the standard matrix analytic-method, semi-Markov process technique, supplementary variable technique and establishment of factorization forms.

1.5.3 The unified analysis approach

A behavior of a station during a server visit can be modeled by a vacation system in which the vacation period corresponds to the intervisit time of the considered station. Thus the BMAP/G/1 cyclic polling models can be seen as the generalization of both the corresponding vacation and priority queueing models. Hence they inherit the analytical difficulties of both models:

BMAP/G/1 vacation models: for these models with disciplines other than exhaustive the proper definition of the system states, which results in an $M/G/1$-type structure, would make the state space very complicated

BMAP priority queueing models: in contrast to the matrix analytic-method, which allows only one random variable having countable infinite space to describe the system dynamics, these models require mutually dependent random variables, each of which is defined in the countable infinite space, i.e. the number of customers in each priority class

To overcome these drawbacks we separate the analysis into two parts based on quantities at server arrival and departure epochs treating only one of the above mentioned difficulties in both of them. This results in simplification in the overall analysis. In the first part dealing only with the vacation model part of the problem enables to establish relations in terms of quantities at server arrival and departure epochs. The closed-form of these factorization results makes the analysis considerably easier. In the second part due to the problem separation the above quantities are to be determined at server arrival and departure epochs, which requires the description of the system dynamics only at server arrival and departure epochs. This is achieved by the help of relating the joint PGFs of the stationary number of customers and the phases of the BMAPs at that epochs by means of a unified method. This results in a simpler mathematical structure compared to the possible descriptions at other system epochs (like e.g. at customer departure times) or to the application of other methods (like e.g. the supplementary variable technique).

The results in the first part are valid for a group of disciplines since the dependency on the given discipline is incorporated by the quantities at server arrival and departure epochs. Hence this part is called as service discipline independent part of the analysis. In the second part the required quantities at server arrival and departure epochs and their determination are discipline specific. This part is called as service discipline dependent part of the analysis.

Summarizing this two-step methodology, the results provided in the service discipline independent part are valid for a broad class of disciplines. Furthermore they can be applied to the individual disciplines by determining the discipline specific quantities in the discipline dependent part of the analysis by means of a unified method. Hence this methodology provides a unified approach for analysis of polling models with BMAP.

More specifically the two parts of the analysis can be described as follows

Service discipline independent part: In this part we establish the relation of the vector GF of the stationary number of customers at a particular station ($\widehat{\mathbf{q}}_i(z)$) in terms of the vector GFs of the stationary number of customers at server arrival ($\widehat{\mathbf{f}}_i(z)$) and departure epochs ($\widehat{\mathbf{m}}_i(z)$), i.e., $\widehat{\mathbf{q}}_i(z) = \mathcal{F}\left(\widehat{\mathbf{f}}_i(z), \widehat{\mathbf{m}}_i(z)\right)$. In its derivation we rely on the stationary relationship from Takine and Takahashi [175] ($\widehat{\mathbf{q}}_i(z) = \mathcal{H}\left(\widehat{\mathbf{q}}_i^d(z)\right)$), where $\widehat{\mathbf{q}}_i^d(z)$ is the vector GF of the stationary number of customers at customer departure epochs at a particular station. To relate $\widehat{\mathbf{q}}_i^d(z)$ to $\widehat{\mathbf{f}}_i(z)$ and $\widehat{\mathbf{m}}_i(z)$ we utilize only the evolution of the system from server arrival to server departure epochs at a particular station of the same cycle. During this period the service process is completely independent of the other stations. Therefore in this part we consider only the number of customers at the particular station without representing the state of the other stations in the notation. This way the inter-dependency of stations is not needed and hence it is not visible in this part of the thesis (chapter 4).

Service discipline dependent part: The description of the system dynamics for a given discipline by mutually dependent discrete random variables (number of customers and the phases of the BMAPs at the stations) is necessary only at server arrival and departure epochs due to the form of the service discipline independent results. We setup the governing equations of the system at these epochs in terms of the joint PGFs of the stationary number of customers at every stations and the phases of every BMAPs at server arrival ($\widehat{\mathbf{f}}_i(z_1, \ldots, z_N)$) and departure ($\widehat{\mathbf{m}}_i(z_1, \ldots, z_N)$) epochs, i.e., $\widehat{\mathbf{f}}_i(z_1, \ldots, z_N) \to \widehat{\mathbf{m}}_i(z_1, \ldots, z_N)$ and $\widehat{\mathbf{m}}_i(z_1, \ldots, z_N) \to \widehat{\mathbf{f}}_{i+1}(z_1, \ldots, z_N)$. The required quantities at server arrival and departure epochs are computed from the solution of these governing equations. In these parts of the thesis (chapters 5 and 6) the inter-dependencies of stations are also captured due to the treatment of the mutually dependent discrete random variables.

This two-steps methodology can be seen as a generalization of the one, which has also been used for analyzing cyclic polling models with Poisson arrival by Borst and Boxma in [19]. In the service discipline independent part the derivation of the factorization form of $\widehat{\mathbf{q}}_i(z)$ utilizes also the generalization of an argument from an early work of Eisenberg [47]. Furthermore in the service discipline dependent part the description of the system by mutually dependent random variables at server arrival and departure epochs is realized by the generalization of the buffer occupancy method (Cooper [38]), which was also used in the analysis of the classical cyclic polling model. Consequently besides of incorporating several elements of the matrix analytic-methods the whole methodology can be seen rather as the natural generalization of several methods used for the analysis of the classical cyclic polling model.

1.6 Thesis structure

Chapter 2 gives a detailed description of the cyclic polling model with BMAP. The limitations on the model and the service disciplines, which determine the validity scope of the major statements of the thesis, are also specified.

Chapter 3 deals with the stability of the model. The new stability results are provided, which include the characterization of global stability, overview of stability regions of a particular station, order of instability of stations, and the necessary and sufficient conditions for the stability states of the system. At the end of the chapter also the discipline specific forms of the condition for the whole stability of the system are summarized for the gated, for the exhaustive and for the G-limited disciplines.

Chapter 4 presents the discipline independent stationary relations. The content of this chapter realizes the first part of the applied two-step methodology. The vector GF of the stationary number of customers at an arbitrary instant and a new formula for its vector factorial moments are derived in terms of the vector GFs of the stationary number of customers at server arrival and departure epochs and their vector factorial moments, respectively. The expressions of the vector GF of the stationary number of customers at customer departure and arrival epochs and their vector factorial moments are also provided.

Chapter 5 discusses the determination of the discipline specific terms in the polling model with gated and exhaustive disciplines, which are needed for the application of the discipline independent results. This requires the solution of the discipline specific governing equations of the system, which relates the joint PGFs of the stationary number of customers and the phases of BMAPs at server arrival and departure epochs. The detailed discussion of the numerical solutions for the case of both the gated and the exhaustive disciplines is also provided.

Chapter 6 gives the discipline specific solutions for the polling model with G-limited and decrementing-K disciplines. Following the same line of description as in chapter 5 the set-up and the solution of the governing equations of the system are discussed. The brief discussion of the numerical solutions for the case of both the G-limited and the decrementing-K disciplines closes the chapter.

Chapter 7 provides specialized results for BMAP vacation models and for classical cyclic models as the most important special cases of the BMAP cyclic polling models considered in the thesis.

Chapter 8 gives a brief summary of the contributions of this thesis and discusses the applications of the results presented in the thesis. Finally several future research directions are identified, which are related to the cyclic polling model with BMAP.

At the end of this thesis several appendices are also placed. Appendix A summarizes the most familiar special cases of BMAP. Appendix B discusses new properties of model specific key matrices needed for the derivation of the new factorial moments formula. Appendix C shows the equivalence of different definitions of several quantities. Appendix D closes the thesis with the proof of Lemma 4.1.

Chapter 2

Model description

2.1 BMAP process

BMAP is a natural generalization of the (batch) Poisson arrival process. It characterizes a very general arrival process and hence it includes many familiar arrival processes as special cases (see Appendix A).

In this Section we give a summary only on those BMAP related definitions and notations, which we use in the thesis. For a more detailed description on BMAP see Lucantoni [114].

In the BMAP the arrivals are governed by a background CTMC. It is referred to as phase process and its state is called phase. $\Lambda(t)$ denotes the count of the number of arrivals in $(0, t]$ and $J(t)$ is the phase at time t. The BMAP batch arrival process is characterized by the $\{(\Lambda(t), J(t))\,; t \geq 0\}$ bivariate CTMC on the state space $\{0, 1, \ldots\} \times \{1, 2, \ldots, L\}$ where $(\Lambda(t) \in \{0, 1, \ldots\}, J(t) \in \{1, 2, \ldots, L\})$. Its infinitesimal generator is given as:

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

where $\mathbf{0}$ and $\mathbf{D}_k (k \geq 0)$ are $L \times L$ matrices.

$\mathbf{D}_0$ governs the transitions without any arrival. Similarly $\mathbf{D}_k (k \geq 1)$ governs the transitions with batch arrivals, in which the batch size is k.

Remark 2.1 *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.*

Remark 2.1 is used in the stability analysis to show the structural properties of the state space of properly chosen embedded Markov chains.

The irreducible infinitesimal generator of the phase process is $\mathbf{D} = \sum_{k=0}^{\infty} \mathbf{D}_k$. Let $\boldsymbol{\pi}$ be the stationary probability vector of the phase process. Then $\boldsymbol{\pi}\mathbf{D} = \mathbf{0}$ and $\boldsymbol{\pi}\mathbf{e} = 1$

uniquely determine $\boldsymbol{\pi}$, where $\mathbf{e}$ is the column vector having all elements equal to one. It implies that

$$rank\,(\mathbf{D}) = L - 1. \tag{2.1}$$

The matrix generating function (matrix GF) of $\mathbf{D}_k$, $\widehat{\mathbf{D}}(z)$ is defined as

$$\widehat{\mathbf{D}}(z) = \sum_{k=0}^{\infty} \mathbf{D}_k z^k, \quad |z| \leq 1. \tag{2.2}$$

The stationary arrival rate of a BMAP is

$$\lambda = \boldsymbol{\pi} \left. \frac{d\widehat{\mathbf{D}}(z)}{dz} \right|_{z=1} \mathbf{e} = \boldsymbol{\pi} \sum_{k=0}^{\infty} k\mathbf{D}_k \mathbf{e}. \tag{2.3}$$

$[\mathbf{P}(k,t)]_{j,\ell}$ denotes the probability $Pr\left\{\Lambda(t) = k, J(t) = \ell | \Lambda(0) = 0, J(0) = j\right\}$ and it is the (j,ℓ)-th element of an $L \times L$ matrix $\mathbf{P}(k,t)$. It is shown in Lucantoni [114] that

$$\widehat{\mathbf{P}}(z,t) = \sum_{k=0}^{\infty} \mathbf{P}(k,t) z^k = e^{\widehat{\mathbf{D}}(z)t}. \tag{2.4}$$

We define the embedded epochs of batch arrivals and/or phase changes of the BMAP process as *BMAP state changes*.

In the following we introduce the description of the number of i-customers arriving at BMAP state changes. Let $\mathbf{D}_0^d$ be a diagonal matrix, which is composed by the diagonal elements of $\mathbf{D}_0$. We define the $L \times L$ matrix $\widehat{\boldsymbol{\Psi}}(z)$ as

$$\widehat{\boldsymbol{\Psi}}(z) = \left(-\mathbf{D}_0^d\right)^{-1} \left(\widehat{\mathbf{D}}(z) - \mathbf{D}_0^d\right) = \left(-\mathbf{D}_0^d\right)^{-1} \widehat{\mathbf{D}}(z) + \mathbf{I}, \tag{2.5}$$

where $\mathbf{I}$ denotes the identity matrix. $\boldsymbol{\Psi} = \widehat{\boldsymbol{\Psi}}(1)$ is stochastic, since $\boldsymbol{\Psi}\mathbf{e} = \mathbf{e}$ and its elements are non-negative. From (2.5) matrix $\boldsymbol{\Psi}$ can be expressed as

$$\boldsymbol{\Psi} = \left(-\mathbf{D}_0^d\right)^{-1} \left(\mathbf{D} - \mathbf{D}_0^d\right). \tag{2.6}$$

Matrix $\widehat{\boldsymbol{\Psi}}(z)$ is the matrix GF of the number of arriving customers at BMAP state changes. Thus matrix $\boldsymbol{\Psi}$ is interpreted as the transition probability matrix at BMAP state changes. Applying (2.6) $\mathbf{D}$ can be expressed in terms of $\boldsymbol{\Psi}$ by

$$\mathbf{D} = \left(\mathbf{D}_0^d\right)(\mathbf{I} - \mathbf{\Psi}). \tag{2.7}$$

For the sake of clarity so far we used the conventional BMAP notations, i.e. the station index i was suppressed throughout this Section. However from now on the first subscript stands for the station index, i.e. $\mathbf{D}_i$ denotes matrix $\mathbf{D}$ of the BMAP at station i. If also an index of the conventional BMAP notation is necessary then it follows the station index, i.e. $\mathbf{D}_{i,0}$ denotes matrix $\mathbf{D}_0$ of the $BMAP$ at station i.

2.2 The BMAP/G/1 cyclic polling model

We consider a continuous-time asymmetric polling model with N stations. A single server attends the stations in a cyclic manner. Each station has an infinite buffer queue, which is served when the server attends that station. If no customer is present at a station at server arrival, the server immediately attends the next station. In this model, each station can have different service discipline (mixed-discipline system). At each station batch of customers arrive according to BMAP process. We call the BMAP at station i as i-th BMAP and λ_i denotes its stationary arrival rate. Similarly the BMAP state change epochs at station i are called as i-th BMAP state change epochs. The customer who arrives to station i is called i-customer. The customer service times at station i are general independent and identically distributed. B_i stands for the customer service time at station i and $\widetilde{B}_i(s)$, $B_i(t)$, $b_i^{(k)}$ denote its Laplace-Stieljes transform (LST), its cumulated distribution function and its k-th moments, for $k \geq 1$, respectively. The mean of the customer service time is also denoted by b_i, i.e. $b_i = b_i^{(1)}$. The model enables only nonzero-switchover-times. R_i denotes the switchover time from station i to the next one. The R_i switchover times of the consecutive cycles are general independent and identically distributed. $\widetilde{R}_i(s)$, $R_i(t)$ and r_i are its LST, cumulated distribution function and its mean, respectively. On the BMAP/G/1 cyclic polling model we impose the following assumptions:

A.1 At each station the phase process of the BMAP is irreducible and the stationary arrival rate is positive and finite, $0 < \lambda_i < \infty$.

A.2 The mean customer service time and the mean switchover time are positive and finite at each station, $0 < b_i < \infty$, $0 < r_i < \infty$.

A.3 The arrival processes, the service times and the switchover times are mutually independent.

The server utilization at station i and the overall utilization are $\rho_i = \lambda_i b_i^{(1)}$ and $\rho = \sum_{i=1}^{N} \rho_i$, respectively. We assume that all stations of the polling system are stable.

Definition 2.1 *Polling epoch, departure epoch: 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 polling epoch of station i as i-polling epoch. Similarly the departure epoch of station i is an i-departure epoch.*

F_i and M_i stand for the generic random variable describing the number of i-customers at i-polling epoch and at i-departure epoch, respectively.

Definition 2.2 *Station time: The* station time *of a given station is defined as the time elapsed from the arrival of the server to station i until its next departure. The station time of station i is called i-station time.*

Definition 2.3 *Intervisit time: The* intervisit time *of a given station is defined as the time elapsed from the departure of the server from station i until its next arrival to the same station. The intervisit time of station i is called i-intervisit time.*

S_i and I_i denote the generic random variable for i-station time and for i-intervisit time, respectively.

Definition 2.4 *Cycle time: The* cycle time *of a given station is defined as the time elapsed from the server visit to station i in the actual cycle to the server visit to the same station in the next cycle. It is also called as polling cycle. The polling cycle of station i is called i-polling cycle.*

C_i denotes the generic random variable of the duration of an i-polling cycle.

The basic terms and their notations are illustrated on Fig. 2.1.

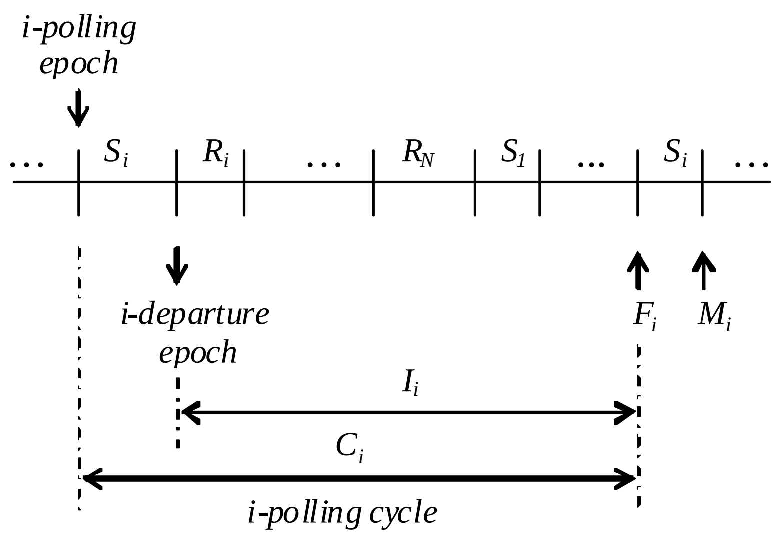

Figure 2.1: Classical polling model

$adj\mathbf{Y}$, $det\mathbf{Y}$ and $Tr(\mathbf{Y})$ denote the adjugate, the determinant and the trace of matrix $\mathbf{Y}$, respectively. Furthermore $[\mathbf{Y}]_{j,l}$ stands for the j,l-th element of matrix $\mathbf{Y}$. Similarly $[\mathbf{y}]_j$ denotes the j-th element of vector $\mathbf{y}$.

When $\widehat{\mathbf{Y}}(z)$, $|z| \leq 1$ is a matrix GF, $\mathbf{Y}^{(k)}$ denotes its k-th ($k \geq 1$) factorial moment, i.e., $\mathbf{Y}^{(k)} = \frac{d^k}{dz^k}\widehat{\mathbf{Y}}(z)|_{z=1}$ and $\mathbf{Y}$ as well as $\mathbf{Y}^{(0)}$ denote its value at $z = 1$, i.e., $\mathbf{Y} = \mathbf{Y}^{(0)} = \widehat{\mathbf{Y}}(1)$. Similarly when $\widehat{\mathbf{y}}(z)$, $|z| \leq 1$ is a vector GF, $\mathbf{y}^{(k)}$ denotes its k-th ($k \geq 1$) factorial moment, i.e., $\mathbf{y}^{(k)} = \frac{d^k}{dz^k}\widehat{\mathbf{y}}(z)|_{z=1}$ and $\mathbf{y}$ as well as $\mathbf{y}^{(0)}$ denote its value at $z = 1$, i.e., $\mathbf{y} = \mathbf{y}^{(0)} = \widehat{\mathbf{y}}(1)$. We also use notation $y^{(k)} = \mathbf{y}^{(k)}\mathbf{e}$.

2.3 The number of arrivals during a customer service time

Let $\Lambda_i(t)$ and $J_i(t)$ be the number of arrivals of the i-th $BMAP$ in interval $(0, t]$ and the phase of the i-th $BMAP$ at time t, respectively. $\Lambda_i(t)$ and $J_i(t)$ are right continuous, and thus every trajectories of every quantities based on them are also right continuous. t_i^s and t_i^{d-} denote the instants at the start of the service and just before the departure of any i-customer, respectively.

We define matrix $\mathbf{A}_i(k)$, whose (j,l)-th element, for $k \geq 0$, $1 \leq j,l \leq L$, is given by

$$[\mathbf{A}_i(k)]_{j,l} = P\left\{\Lambda_i(t_i^{d-}) = k, J_i(t_i^{d-}) = l|\Lambda_i(t_i^s) = 0, J_i(t_i^s) = j\right\},$$

and it is interpreted as the conditional probability that during an i-customer service time the number of i-th BMAP arrivals is k and the final phase of the i-th $BMAP$ is l given that the initial phase of the i-th BMAP is j. Matrix GF $\widehat{\mathbf{A}}_i(z)$ is defined as

$$\widehat{\mathbf{A}}_i(z) = \sum_{k=0}^{\infty} \mathbf{A}_i(k)z^k, \quad |z| \leq 1. \tag{2.8}$$

By applying the definition of $\mathbf{P}_i(k,t)$ in (2.8)and using (2.4) $\widehat{\mathbf{A}}_i(z)$ can be explicitly expressed by

$$\widehat{\mathbf{A}}_i(z) = \int_{t=0}^{\infty} e^{\widehat{\mathbf{D}}_i(z)t} dB_i(t) = E[e^{\widehat{\mathbf{D}}_i(z)B_i}], \tag{2.9}$$

see Lucantoni [114].

The elements of $\mathbf{A}_i$ are interpreted as the probabilities of the phase transitions of the i-th BMAP from the start to the end of an i-customer service time, and therefore $\mathbf{A}_i$ is stochastic. It follows that the structures of matrix $(\mathbf{I} - \mathbf{A}_i)$ and matrix $\mathbf{D}_i$ (see (2.7)) are similar. Utilizing the non-singularity of the diagonal matrix in (2.7) and applying (2.1) yields

$$rank\,(\mathbf{I} - \mathbf{A}_i) = L - 1. \tag{2.10}$$

2.4 Service discipline

Definition 2.5 *The* service discipline *gives the condition on the beginning and on the end of the service at a given station.*

The most commonly known disciplines are, e.g., the exhaustive, the gated, the binomial-exhaustive, the binomial-gated, the non-exhaustive, the semi-exhaustive, the G-limited and the decrementing-K. For more details on disciplines see Subsection 1.1.2.

2.4.1 Relevant class of service disciplines

The allowed service disciplines have the following properties:

P.1 Work-conserving property: If the service begins at the actual station, then it is work conserving up to the end of the service at that station according to the used discipline.

P.2 Nonpreemptive service property: The service is nonpreemptive. Hence the server first finishes the service of the customer under service before leaving the station.

P.3 Determination property: If the service discipline, the arrival process and the customer service time of station i are given, then the number of i-customers and the phase of i-th BMAP at the i-polling epoch completely determine the number of i-customers served at that station, the duration of that service as well as the number of i-customers and the phase of the i-th BMAP at any time during the service of station i (including the i-departure epoch) in stochastic sense (in distribution). Additionally for each $k \neq i$ the number of k-customers, the number of i-customers and the phase of the k-th and i-th BMAPs at the i-polling epoch determine the number of k-customers and the phase of the k-th BMAP at any time during the service of station i (including the i-departure epoch) in stochastic sense. It follows that number of k-customers and the phases of k-th BMAPs for every $k = 1, \ldots, N$ at the i-polling epoch completely determine the evolution of the system including the number of k-customers and the phases of k-th BMAPs for every $k = 1, \ldots, N$ at any time during the service of station i (including the i-departure epoch) in stochastic sense.

P.4 Memoryless property: In general the service discipline is independent of the history of the system. A numerous service disciplines satisfy the properties **P.1**-**P.4**. For example all the above mentioned examples fulfill these properties.

2.4.2 Stability related limitations of service disciplines

Definition 2.6 *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* $\sum_{n=0}^{\infty} P\{Z=n\}=0$.

Let $Z(m)$ be a sequence of non-negative integer valued random variables for $m \geq 1$. Let $\lim_{m \rightrightarrows \infty} Z(m)$ denote the convergence of $Z(m)$ in distribution, i.e. the convergence of $\lim_{m\to\infty} P\{Z(m)=n\}$, for $n \geq 0$. This convergence is interpreted in generalized sense, i.e. $\sum_{n=0}^{\infty} \lim_{m\to\infty} P\{Z(m)=n\} \leq 1$. Thus also the case of degenerate limiting distribution is included. Note that in different contexts, the case of totally degenerate limiting distribution is interpreted as $Z(m)$ does not converge.

We define the following notations:

$J_{i,k}^{f}(m)$ - the phase of the k-th BMAP arrival process at the i-polling epoch of the m-th cycle, $J_{i,k}^{f}=\lim_{m \rightrightarrows \infty} J_{i,k}^{f}(m)$ and $J_i^f=J_{i,i}^f$.

$F_{i,k}(m)$ - the number of k-customers at the i-polling epoch of the m-th cycle, $F_{i,k}=\lim_{m \rightrightarrows \infty} F_{i,k}(m)$, $F_i=F_{i,i}$,

$G_i(m)$ - the number of customers served in the i-station time in the m-th polling cycle, $G_i=\lim_{m \rightrightarrows \infty} G_i(m)$, $g_i(m)=E\left(G_i(m)\right)$, $g_i=\lim_{m\to\infty} g_i(m)$,

g_i^{∞} - the mean number of customers served at the i-station time given that the number of i-customers at i-polling epoch goes to infinity:
$g_i^{\infty}=\sum_{j=1}^{L} P\left\{J_i^f=j\right\} \lim_{n\to\infty} E\left(G_i \mid F_i=n, J_i^f=j\right)$,

$g_i^{\max}$ - the maximum of the mean number of customers, which can be served during an i-station time: $g_i^{\max}=\max_{n,j} E\left(G_i \mid F_i=n, J_i^f=j\right)$,.

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

The disciplines allowed to the stability analysis are further limited by the following properties:

P.5 Non-zero maximum property: If at least one i-customer is present at i-polling epoch, then at least one i-customer is served with positive probability in that i-station time. Furthermore the maximum of the mean number of customers that can be served during the i-station time is greater than zero, $g_i^{max}>0$.

P.6 Maximum limit property: If the number of i-customers at the i-polling epoch goes to infinity then the limit of mean number of i-customers served during the i-station time equals the maximum of the mean number of customers that can be served during that stage, $g_i^{\infty}=g_i^{max}$.

P.7 Mean limit property: If $g_i^{\infty}=\infty$, and the mean number of i-customers at the i-polling epoch goes to infinity, then the mean number of i-customers served during the i-station time also tends to infinity. That is, if the service discipline satisfies $g_i^{\infty}=\infty$, then $E\left(F_i\right)=\infty$ implies

$$E(G_i) = \sum_{\ell=0}^{\infty} \sum_{j=1}^{L} P\left\{F_i = \ell, J_i^f = j\right\} E\left(G_i \mid F_i = \ell, J_i^f = j\right) = \infty.$$

A numerous service disciplines satisfies properties **P.5 - P.7**. For example from the above mentioned examples the exhaustive, the gated, the binomial-exhaustive, the binomial-gated, the non-exhaustive and the G-limited disciplines fulfill these conditions. Properties **P.5** and **P.6** are similar to the assumptions S1 and S2 of the model of Down [44].

2.4.3 Stability related properties of service disciplines

Definition 2.7 *The service discipline at station i is called* unlimited type *when* $g_i^\infty = \infty$. *Due to* **P.6**

$$g_i^\infty = g_i^{max} = \infty. \tag{2.11}$$

The service discipline at station i is called limited type *when* $g_i^\infty < \infty$. *Due to* **P.5** *and* **P.6**

$$0 < g_i^\infty = g_i^{max} < \infty. \tag{2.12}$$

Definition 2.8 *A station is of* unlimited type*, if that station has unlimited type service discipline. A station is of* limited type*, if that station has limited type service discipline.*

Remark 2.2 *If $F_i = \infty$ at the limited type station i then it follows from (2.12), 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 G-limited, as well as the decrementing-K service disciplines are limited types.

Chapter 3

Stability

In this chapter the stability of the polling model is investigated. The applied methodology is based on identification of properly chosen embedded Markov chains at the polling epochs. Several properties of service disciplines (**P.3** and **P.5 - P.7**) play crucial role in the completion of the stability proofs. Properties **P.5 - P.7** are relaxed conditions on the service disciplines in comparison with the monotonicity property of Fricker and Jaïbi in [56], and some of them is rather similar to the assumptions of the model of Down [44]. The new stability results include the characterization of global stability, overview of stability regions of a particular station, order of instability of stations, conditions for partial stability and the necessary and sufficient conditions for the stability states of the system. As the condition for the whole stability of the system is given in terms of service discipline dependent quantities, its discipline specific form is summarized at the end of the chapter for the gated, for the exhaustive and for the G-limited disciplines.

We note that the stability analysis presented in this chapter is the special case of the one for the periodic polling system with BMAP, which is published in the journal paper [198].

3.1 Global stability

3.1.1 Stability

Definition 3.1 *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 3.2 *Station i of the polling model is said to be* stable*, when the number of i-customers at i-polling epoch has proper limiting distribution and the limiting cycle time has a finite mean as the number of polling cycles goes to infinity.*

Definition 3.3 *The polling model is said to be* stable*, when the number of customers at every stations at i-polling epoch have proper limiting distributions and the limiting cycle time has a finite mean as the number of polling cycles goes to infinity.*

This stability definition of polling models is equivalent with the one of Fricker and Jaïbi [56].

P.7 implies, that an unlimited type station with proper distribution of customers at polling epoch, F_i, but with infinite mean, $E(F_i) = \infty$, results in an infinite mean number of customers served at station i, $g_i = \infty$. In this case, the mean cycle time is also infinite because the mean service times are positive (**A.2**). Hence this case is excluded from the stability definition.

For the limited type stations this stability definition allows $E(F_i) = \infty$, since in this case $g_i < \infty$, 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 [88], since it excludes the case $E(F_i) = \infty$.

3.1.2 Global stability of the polling system

Theorem 3.1 *(Global stability states of the system.) 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 the unlimited type stations are instable the limited type stations are instable as well, and the limiting mean cycle time is infinite.
- When the unlimited type stations are stable the limited type stations can be both stable and instable.

When an unlimited type station becomes instable, then the mean number of customers to be served at its polling epoch tends to infinity. This results in an infinite mean number of customers served in the actual cycle (**P.7**). Due to positive service times (**A.2**) 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.

The third property comes from the fact that for the limited type station i the maximal mean service time is finite due to $g_i \leq g_i^{max} < \infty$ (**P.6**). Hence the other stations accumulate only a finite number of customers during this time, which might be served by the unlimited type stations. □

3.2 Stability of stations

3.2.1 State of the system

Definition 3.4 *The* state of the system *at an i-polling epoch consists of the number of customers at the stations and the phases of the BMAP arrival processes.*

The state vector, $\mathbf{Z}_i(m)$, describes the state of the system in the m-th i-polling epoch:

$$\mathbf{Z}_i(m) = (F_{i,1}(m), J_{i,1}^f(m), F_{i,2}(m), J_{i,2}^f(m), \ldots F_{i,N}(m), J_{i,N}^f(m)).$$

3.2.2 Embedded Markov chain

The stability analysis is based on properly chosen embedded Markov chains. In this Subsection the structural properties of the state space of these embedded Markov chains are shown.

Lemma 3.1 *(Identification of the proper embedded Markov chain.) For any fixed $i \in \{1, \ldots, N\}$ the $\{\mathbf{Z}_i(m), m > 0\}$ sequence is an embedded Markov chain.*

Proof. It follows from the determination property (**P.3**) and the independence of switchover times (**A.3**). □

Theorem 3.2 *(Properties of the embedded Markov chain.) The $\{\mathbf{Z}_i(m), m > 0\}$ 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.3**, that the evolution of the system does not depend on the elapsed number of cycles. Therefore the $\{\mathbf{Z}_i(m), m > 0\}$ Markov chain is homogeneous.

To study the structure of the state space, we investigate the reachability of state $(0, 1, 0, 1, \ldots 0, 1)$. We show, that $(0, 1, 0, 1, \ldots 0, 1)$ can be reached from any state $(l_1, p_1, l_2, p_2, \ldots l_N, p_N)$, where $l_1, l_2, \ldots l_N \geq 0$. The state transition from $(l_1, p_1, l_2, p_2, \ldots l_N, p_N)$ to $(0, 1, 0, 1, \ldots 0, 1)$ 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 irreducible (**A.1**) and independent (**A.3**). Furthermore only finite number of customers can arrive during it, because the stationary arrival rates are finite (**A.1**). 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 (**A.1**), finite mean switchover times and finite mean service times (**A.2**) the duration of state transitions $\mathbf{Z}_i(m) \to \mathbf{Z}_i(m+1)$ is finite for every finite $m > 0$. It follows, that the first step 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-polling epoch is reached, while the phases of the BMAPs 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.5**) 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 BMAPs can remain unchanged without any arrival for this finite time with positive probability (Remark 2.1). 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 $\{\mathbf{Z}_i(m), m > 0\}$ 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 $\mathbf{Z}_i$ be the following limit: $\mathbf{Z}_i = \lim_{m \rightrightarrows \infty} \mathbf{Z}_i(m)$.

Theorem 3.3 *(Property of the distribution of $\mathbf{Z}_i$.) The distribution of $\mathbf{Z}_i$ is either proper or totally degenerate.*

Proof. Due to Theorem 3.2 the $\{\mathbf{Z}_i(m), m > 0\}$ 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 $\mathbf{Z}_i$ 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. Hence the statement holds for both cases. □

We define the following quantities:
$\mathbf{J}_i^f(m) = \left(J_{i,1}^f(m), J_{i,2}^f(m), \dots, J_{i,N}^f(m)\right)$, $\mathbf{J}_i^f = \lim_{m \rightrightarrows \infty} \mathbf{J}_i(m)$,
$\mathbf{F}_i(m) = (F_{i,1}(m), F_{i,2}(m), \dots, F_{i,N}(m))$, $\mathbf{F}_i = \lim_{m \rightrightarrows \infty} \mathbf{F}_i(m)$.

Corollary 3.1 *(Common property of the distributions of* $\mathbf{Z}_i$ *and* $\mathbf{F}_i$*.) Both* $\mathbf{Z}_i$ *and* $\mathbf{F}_i$ *have either proper or totally degenerate distributions.*

Proof. $\mathbf{J}_i^f$ has a proper distribution, since each BMAP process has finite number of phases. Since $\mathbf{Z}_i$ is the union of $\mathbf{F}_i$ and $\mathbf{J}_i^f$ the statement follows from Theorem 3.3. □

3.2.3 System description in partial stability

Let N^l denote the number of limited type stations. In addition let N^u be the number of instable stations, and due to partial stability $1 \leq N^u \leq N^l$ (Theorem 3.1). Without the loss of generality we assume, that the instable stations are the first N^u stations, i.e., $1, \dots, N^u$ are their indexes. To evaluate the system properties in partial stability, we introduce $\mathbf{Z}_i^*(m)$ similar to $\mathbf{Z}_i(m)$. Supposing that for the instable stations $F_{i,k}(m) = \infty$, $\forall k \in \{1, \dots, N^u\}, m \geq 0$, we define $\mathbf{Z}_i^*(m)$ as the state of the stable stations, i.e., $\mathbf{Z}_i^*(m) = (F_{i,N^u+1}(m), J_{i,N^u+1}^f(m), \dots F_{i,N}(m), J_{i,N}^f(m))$.

Lemma 3.2 *(Properties of the embedded Markov chain in partial stability.)* $\mathbf{Z}_i^*(m)$ *is a discrete time Markov chain with a proper limiting distribution and* $\lim_{m\to\infty} P\left(\mathbf{Z}_i^*(m) = (0, 1, \dots, 0, 1)\right) > 0$.

Proof. $\mathbf{Z}_i(m)$ is a Markov chain. All the instable stations are of limited type (Theorem 3.1) with $g_i = g_i^{max} < \infty$, and thus their 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 affect the evolution of the state of the stable stations. It follows, that $\mathbf{Z}_i^*(m)$ is a Markov chain.

The stability of stations $N^u + 1, \dots, N$ implies, that all of the one dimensional marginal distributions of $\mathbf{Z}_i^*(m)$ are proper. The topology of $\mathbf{Z}_i^*(m)$ is similar to the one of $\mathbf{Z}_i(m)$, 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 $\lim_{m\to\infty} P\left(\mathbf{Z}_i^*(m) = (0, 1, \dots, 0, 1)\right) > 0$. □

3.2.4 Stability of a particular station

Theorem 3.4 *(Stability of a particular station - 1.) The necessary and sufficient condition of the stability of station i is*

$$g_i < g_i^{max}. \tag{3.1}$$

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 has a proper distribution. The limiting cycle time has finite mean and the mean service times are positive (A.2), hence $g_i < \infty$. According to property **P.5** $g_i^{max} > 0$. If no customer is present at an i-polling epoch, then no service occurs at that station, i.e., $E(G_i|F_i = 0) = 0$. However Lemma 3.2 implies that $P(F_i = 0) > 0$. It follows

$$\begin{aligned} g_i &= P(F_i = 0)\, E(G_i|F_i = 0) + \Big(1 - P(F_i = 0)\Big) E(G_i|F_i > 0) \\ &\leq P(F_i = 0) \cdot 0 + \Big(1 - P(F_i = 0)\Big) g_i^{max} < g_i^{max}\ . \end{aligned}$$

Therefore $g_i < g_i^{max}$.

Now we consider the case when station i is instable. In this case the distribution of F_i is not proper 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 3.1 implies that the not proper distribution of F_i must be totally degenerate. Or the mean limiting cycle time is infinite, in which case infinitely many customers accumulate during it. This implies that the distribution of F_i is totally degenerate. Hence in both cases $F_i = \infty$, and according to Remark 2.2 $g_i = g_i^{max}$. 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.6**, **P.7** and $g_i^{max} = \infty$. □

3.3 Stability relationships

3.3.1 Stability regions of a particular station

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

Lemma 3.3 *(Stability implies equilibrium.) If station i is stable, then the limiting mean number of arriving and served i-customers are the same,*

$$a_i = g_i. \tag{3.2}$$

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

Lemma 3.4 *(Possible relations of of a_i and g_i.) The following relation holds for station i:*

- if station i is of limited type, then

$$a_i \geq g_i, \tag{3.3}$$

- if station i is of unlimited type, then

$$a_i = g_i. \tag{3.4}$$

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. It follows, that $a_i \geq g_i$. In fact $a_i > g_i$ holds above the stability boundary $a_i = g_i = g_i^{max}$. For an unlimited type station g_i is not bounded by the service discipline ($g_i^{max} = \infty$) hence $a_i = g_i$. □

Proposition 3.1 *(Stability of a particular station - 2.) Station i is stable if and only if:*

$$a_i < g_i^{max}. \tag{3.5}$$

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 (3.5) follows from Lemma 3.3 and Theorem 3.4.

If station i is instable then it follows from Theorem 3.4, 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 (3.3) and (3.4) we have $a_i \geq g_i^{max}$. □

Table 3.1 gives an overview of stability regions of station i.

<table>
<tr><th>Stability</th><th>Unlimited type station i</th><th>Limited type station i</th></tr>
<tr><td>Stable
$g_i < g_i^{max}$</td><td colspan="2">$a_i = g_i$; $a_i < g_i^{max}$</td></tr>
<tr><td rowspan="2">Instable
$g_i = g_i^{max}$</td><td colspan="2">Stability boundary
$a_i = g_i$; $a_i = g_i^{max}$</td></tr>
<tr><td></td><td>Above stability boundary
$a_i > g_i$; $a_i > g_i^{max}$</td></tr>
</table>

Table 3.1: Stability regions of a particular station

3.3.2 Order of instability of stations

Let $C_i(m)$ be the polling cycle time of station i from the m-th i 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 $c = c_i$.

Lemma 3.5 *(Expression of a_i.) 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:*

$$a_i = \lambda_i c. \tag{3.6}$$

Proof. If the system is stable or partially stable, then it can be shown by applying the MRP framework (see in [202]), that 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 3.1), and thus $a_i = \infty$. Hence (3.6) holds for this case as well. □

Corollary 3.2 *(Stability of a particular station - 3.) Station i is stable if and only if:*

$$\lambda_i c < g_i^{max}. \tag{3.7}$$

Proof. The statement follows from Proposition 3.1 by applying (3.6). □

Let $\mathbf{D}_{i,\ell}$ 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 ξ, such that $\mathbf{D}_{i,\ell}(\xi) = \xi \mathbf{D}_{i,\ell}$, $l \geq 0$. This way $\lambda_i(\xi) = \xi\lambda_i$, and thus the relative ratios of station arrival rates remain fixed.

Theorem 3.5 *(Order of instability of stations.) Scaling the traffic intensity from 0 to ∞ the stations gets instable in order $i_1, i_2, \ldots, i_N$, where*

$$\frac{\lambda_{i_1}}{g_{i_1}^{max}} \geq \frac{\lambda_{i_2}}{g_{i_2}^{max}} \geq \ldots \geq \frac{\lambda_{i_N}}{g_{i_N}^{max}}. \tag{3.8}$$

Proof. It follows from Corollary 3.2, that station k gets instable, when $c\lambda_k(\xi) = g_k^{max}$. 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 $\frac{\lambda_1}{g_1^{max}} \geq \frac{\lambda_2}{g_2^{max}} \geq \ldots \geq \frac{\lambda_N}{g_N^{max}}$. If there are N^l limited type stations, it follows from Theorem 3.1, that the first N^l indexes identify the limited type stations.

3.3.3 Mean cycle time

Let $S_i(m)$ be the station time at station i in the m-th polling cycle for every $i \in \{1, \ldots, N\}$. Additionally we define $s_i(m) = E(S_i(m))$, $s_i = \lim_{m\to\infty} s_i(m)$.

Theorem 3.6 *(Mean cycle time in partial stability.) If the first N^u limited type stations ($1 \leq N^u \leq N^l$) are out of stability and the remaining $N - N^u$ stations are stable, then the mean cycle time is*

$$c = \frac{r + \sum_{k=1}^{N^u} g_k^{\max} b_k}{1 - \sum_{k=N^u+1}^{N} \rho_k}. \tag{3.9}$$

Proof. If $N^u \leq N^l$ 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 ($1 \leq k \leq N^u$) Corollary 3.1 implies, that the not proper distribution of F_k is totally degenerate. Hence $F_k = \infty$ and it follows from Remark 2.2, that $g_k = g_k^{\max}$. For the stable stations it follows from (3.6) and (3.2), that $g_k = c\lambda_k$. Substituting them we get:

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

Solving it for c results in (3.9). □

If $N^u = 0$, that is the whole system is stable, then (3.9) is reduced to the well-known form:

$$c = \frac{r}{1-\rho}. \tag{3.10}$$

3.4 Stability conditions

3.4.1 Partial stability

Theorem 3.7 *(Stability of limited type station.) Station i of limited type ($i \leq N^l$) is stable if and only if*

$$\sum_{k=i}^{N} \rho_k + \frac{\lambda_i}{g_i^{max}} \left(r + \sum_{k=1}^{i-1} g_k^{max} b_k \right) < 1. \tag{3.11}$$

Proof. For $\xi \geq 0$, $1 \leq i \leq N^l$, we define $c^{(i-1)}(\xi)$ as

$$c^{(i-1)}(\xi) = \frac{r + \sum_{k=1}^{i-1} g_k^{max} b_k}{1 - \sum_{k=i}^{N} \lambda_k(\xi) b_k}.$$

According to Theorem 3.6, $c^{(i-1)}(\xi)$ is the mean cycle time when the first $i-1$ stations are instable and the rests are stable and the value of the scaling parameter is ξ.

Indicating the dependency of λ_i on ξ, (3.11) can be rearranged as

$$\lambda_i(\xi) c^{i-1}(\xi) < g_i^{max}. \tag{3.12}$$

Therefore it is enough to show, that (3.12) holds if and only if station i is stable.

Let $g_i(\xi)$ and $c(\xi)$ stand for g_i and c as functions of ξ, respectively. In addition let $\xi^{(i)}$ be the value of the scaling parameter at the stability boundary of station i ($1 \leq i \leq N^l$) and $\xi^{(0)} = 0$.

Starting from $\xi \in [\xi^{(i-1)}, \xi^{(i)})$, $g_i(\xi) \to g_i^{max}$ as $\xi \to \xi^{(i)}$, since station i is of limited type. In this range station i is still stable and $c(\xi) = c^{(i-1)}(\xi)$. Using them and applying Lemma 3.3 and 3.5 we get:

$$\lim_{\xi \to \xi^{(i)}} g_i(\xi) = \lim_{\xi \to \xi^{(i)}} \lambda_i(\xi) c(\xi) = \lim_{\xi \to \xi^{(i)}} \lambda_i(\xi) c^{(i-1)}(\xi) = g_i^{max}.$$

Both $\lambda_i(\xi)$ and $c^{(i-1)}(\xi)$ are continuous functions of ξ, hence $\lambda_i(\xi_i) c^{(i-1)}(\xi_i) = g_i^{max}$. It follows from the monotonicity of $\lambda_i(\xi)$ and $c^{(i-1)}(\xi)$, that (3.12) holds if and only if $\xi < \xi^{(i)}$, i.e. if and only if station i is stable. □

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

$$\rho^u = \sum_{k=N^l+1}^{N} \rho_k. \tag{3.13}$$

Theorem 3.8 *(Stability of unlimited type station.) Station i of unlimited type ($i > N^l$) is stable if and only if*

$$\rho^u < 1. \tag{3.14}$$

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 3.6, 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 ($1 \leq j \leq N^l$) are the instable limited type ones. Applying $c < \infty$ and utilizing $c > 0$ implies $\sum_{k=j+1}^{N} \rho_k < 1$. Furthermore using $\sum_{k=N^l+1}^{N} \rho_k \leq \sum_{k=j+1}^{N} \rho_k$ and notation (3.13) results in $\rho^u < 1$.

To study the case when station i is instable, we start increasing the traffic load of the system to the boundary situation, when all stations of limited type are already instable, but the stations of unlimited type are still stable. We get the mean cycle time for this case by setting $j = N^l$ in the expression of the mean cycle time: $c = \frac{r + \sum_{k=1}^{N^l} g_k^{max} b_k}{1 - \sum_{k=N^l+1}^{N} \rho_k} = \frac{r + \sum_{k=1}^{N^l} g_k^{max} b_k}{1 - \rho^u}$. Further increase in the traffic intensity does not change the expression of c, but it leads to the instability of the system causing infinite cycle time. Hence in this case $\rho^u = 1$. It follows also, that ρ^u can not be greater than 1, which completes the proof. □

3.4.2 Stability of the system

Theorem 3.9 *(Stability conditions for the stability states of the system.) The system is in*

- *whole stability if and only if*

$$\rho + \left(\frac{\lambda_1}{g_1^{max}}\right) r < 1, \tag{3.15}$$

- *partial stability if and only if*

$$\rho + \left(\frac{\lambda_1}{g_1^{max}}\right) r \geq 1 \text{ and } \rho^u < 1, \tag{3.16}$$

- *instability if and only if*

$$\rho^u \geq 1. \tag{3.17}$$

Proof. The first statement comes from Theorem 3.7 by setting $i = 1$. The second and the third statements are consequences of Theorem 3.1, Theorem 3.8 and the first statement. □

3.5 Extensions

The stability analysis presented in this chapter can be extended for other polling models, like e.g. polling model with Markovian server routing.

The work-conservation property (**P.1**) and the nonpreemptive service property (**P.2**) 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 polling model.

3.6 Application to given service disciplines

The stability condition for a given discipline can be established by applying the service discipline specific $g_1^{\max}$ to the result on the whole stability (3.15).

Service discipline	$g_1^{\max}$	**Stability condition**
gated	∞	$\rho < 1$
exhaustive	∞	$\rho < 1$
G-limited	K_1	$\rho + \left(\frac{\lambda_1}{K_1}\right) r < 1$

Table 3.2: Stability condition for given service disciplines

Table 3.2 summarizes the condition of the whole stability for the gated, the exhaustive and the G-limited disciplines. In case of the G-limited discipline K_1 stands for the discipline limit at station 1.

Remark 3.1 *The stability condition shown in the table 3.2 for the gated discipline holds also for the binomial-gated and the binomial-exhaustive disciplines, since all unlimited type service disciplines have the same stability condition.*

Remark 3.2 *The formula of the whole stability (3.15) can be applied also to the Bernoulli-gated discipline, as it satisfies all the required discipline properties, but it is out of scope of this thesis.*

Chapter 4

Service discipline independent relations

In this chapter the service discipline independent part of the analysis is presented. First a general statement - called as fundamental relationship - is derived by the help of the generalization of an argument from an early work of Eisenberg [47]. Using it the expression of the vector GF of the stationary number of i-customers at i-customer departure epochs is given in terms of the vector GFs of the stationary number of i-customers at i-polling and i-departure epochs. Applying it in the stationary relationship between the vector GF of the stationary number of i-customers at an arbitrary instant and at i-customer departure epochs (see Takine and Takahashi [175]) yields to the expression of the vector GF of the stationary number of i-customers at an arbitrary instant. This is the most significant result of this chapter. Based on this expression the new formula for the vector factorial moments of the stationary number of i-customers at an arbitrary instant is derived. For this derivation also newly established properties of model specific key matrices (see Appendix B) are utilized. The relation of the vector factorial moments of the stationary number of i-customers at i-customer departure epochs to the set of the vector factorial moments of the stationary number of i-customers at an arbitrary instant is also given.

Another line of arguments is used to get the vector GF of the stationary number of i-customers at i-customer arrival epochs. The BMAP state change epochs plays a central role in this part of the analysis. This is because the stationary number of i-customers seen at these epochs can be related to the stationary number of i-customers both at i-customer arrival epochs and at an arbitrary instant. A relationship is also established between the vector factorial moments of the stationary number of i-customers at i-customer arrival epochs and the proper set of the vector factorial moments of the stationary number of i-customers at an arbitrary instant.

The major part of the analysis and results presented in this chapter have been published in the journal paper [199].

4.1 Fundamental relationship

Let $N_i(t)$ be the right continuous number of i-customers in the system at time t. Furthermore, $t_i^f(\ell)$ and $t_i^m(\ell)$ denote the i-polling epoch and the i-departure epoch in the ℓ-th cycle, respectively.

We define the $1 \times L$ vector $\boldsymbol{\nu}_i^f(m,n)$, whose j-th element represents the number of i-polling epochs in the first m polling cycle, at which the number of i-customers is n, and the phase of the i-th BMAP is j. That is, for $m \geq 1$, $n \geq 0$ and $1 \leq j \leq L$,

$$[\boldsymbol{\nu}_i^f(m,n)]_j = \sum_{\ell=1}^{m} 1_{\left(N_i(t_i^f(\ell))=n\right)} 1_{\left(J_i(t_i^f(\ell))=j\right)},$$

where $1_{(\text{con})}$ denotes the indicator of condition "con".

Let $\mathbf{e}_j = (0,\ldots,0,1,0,\ldots,0)$ denote the $1 \times L$ vector with 1 at the j-th position. Using it the vector $\boldsymbol{\nu}_i^f(m,n)$ can be expressed as

$$\boldsymbol{\nu}_i^f(m,n) = \sum_{\ell=1}^{m} 1_{\left(N_i(t_i^f(\ell))=n\right)} \sum_{j=1}^{L} 1_{\left(J_i(t_i^f(\ell))=j\right)} \mathbf{e}_j.$$

We define the $1 \times L$ indicator vector $\mathbf{1}_{(J_i(t))}$ as

$$\mathbf{1}_{(J_i(t))} = \sum_{j=1}^{L} 1_{(J_i(t)=j)} \mathbf{e}_j.$$

By the help of this notation the vector $\boldsymbol{\nu}_i^f(m,n)$ can be defined as

$$\boldsymbol{\nu}_i^f(m,n) = \sum_{\ell=1}^{m} 1_{\left(N_i(t_i^f(\ell))=n\right)} \mathbf{1}_{\left(J_i(t_i^f(\ell))\right)}.$$

The corresponding stationary vector GF, $\widehat{\mathbf{f}}_i(z)$ is defined as

$$\widehat{\mathbf{f}}_i(z) = \sum_{n=0}^{\infty} \lim_{m\to\infty} \frac{E[\boldsymbol{\nu}_i^f(m,n)]}{m} z^n. \quad |z| \leq 1. \tag{4.1}$$

Note that the limit and infinite sum in (4.1) is interchangeable, because $\lim_{m\to\infty} \frac{E[\boldsymbol{\nu}_i^f(m,n)]}{m}$ converges uniformly in n. This follows from the existence of the limiting distribution of the embedded Markov chain $\{(N_1(t_i^f(m)),\ldots,N_N(t_i^f(m)), J_1(t_i^f(m)),\ldots,J_N(t_i^f(m))); m \geq 1\}$, which holds due to the stability of the polling model.

Similarly to (4.1), we define the $1 \times L$ vector $\boldsymbol{\nu}_i^m(m,n)$, whose j-th element represents the number of i-departure epochs in the first m polling cycle, at which the number of

i-customers is n, and the phase of the i-th BMAP equals j. That is, for $m \geq 1$ and $n \geq 0$,

$$\boldsymbol{\nu}_i^m(m,n) = \sum_{\ell=1}^{m} 1_{\left(N_i(t_i^m(\ell))=n\right)} \mathbf{1}_{\left(J_i(t_i^m(\ell))\right)},$$

and the corresponding stationary vector GF, $\widehat{\mathbf{m}}_i(z)$ is defined as

$$\widehat{\mathbf{m}}_i(z) = \sum_{n=0}^{\infty} \lim_{m\to\infty} \frac{E[\boldsymbol{\nu}_i^m(m,n)]}{m} z^n, \quad |z| \leq 1. \tag{4.2}$$

Additionally $t_i^s(\ell,k)$ and $t_i^d(\ell,k)$ denote the instants at the service start and at the departure of the k-th i-customer in the ℓ-th cycle, for $\ell \geq 1$ and $1 \leq k \leq G_i(\ell)$, respectively.

We define the $1 \times L$ vector $\boldsymbol{\nu}_i^s(m,n)$, whose j-th element represents the number of i-customer service starts in the first m polling cycle, at which the number of i-customers is n, and the phase of the i-th BMAP equals j. That is, for $m \geq 1$ and $n \geq 0$,

$$\boldsymbol{\nu}_i^s(m,n) = \sum_{\ell=1}^{m} \sum_{k=1}^{G_i(\ell)} 1_{\left(N_i(t_i^s(\ell,k))=n\right)} \mathbf{1}_{\left(J_i(t_i^s(\ell,k))\right)},$$

and the corresponding stationary vector GF, $\widehat{\mathbf{q}}_i^s(z)$ is defined as

$$\widehat{\mathbf{q}}_i^s(z) = \sum_{n=0}^{\infty} \lim_{m\to\infty} \frac{E[\boldsymbol{\nu}_i^s(m,n)]}{E[\sum_{\ell=1}^{m} G_i(\ell)]} z^n, \quad |z| \leq 1.$$

We also define the $1 \times L$ vector $\boldsymbol{\nu}_i^d(m,n)$, whose j-th element represents the number of i-customer departures in the first m polling cycle, at which the number of i-customers is n, and the phase of the i-th BMAP equals j. That is, for $m \geq 1$ and $n \geq 0$,

$$\boldsymbol{\nu}_i^d(m,n) = \sum_{\ell=1}^{m} \sum_{k=1}^{G_i(\ell)} 1_{\left(N_i(t_i^d(\ell,k))=n\right)} \mathbf{1}_{\left(J_i(t_i^d(\ell,k))\right)},$$

and the corresponding stationary vector GF, $\widehat{\mathbf{q}}_i^d(z)$ is defined as

$$\widehat{\mathbf{q}}_i^d(z) = \sum_{n=0}^{\infty} \lim_{m\to\infty} \frac{E[\boldsymbol{\nu}_i^d(m,n)]}{E[\sum_{\ell=1}^{m} G_i(\ell)]} z^n, \quad |z| \leq 1.$$

Theorem 4.1 *(Fundamental relationship.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds among the vector GFs of the stationary number of i-customers at station i, at different instants:*

$$\widehat{\mathbf{q}}_i^s(z) - \widehat{\mathbf{q}}_i^d(z) = \frac{1}{g_i}\left(\widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z)\right). \tag{4.3}$$

Proof. The proof utilizes an observation of Eisenberg [47]. We adopt and slightly modify his argument, generalize it to vector quantities and apply it for our polling model. The crucial observation is that each time an i-polling epoch or i-customer departure occurs, this coincides with either an i-customer service start or an i-departure epoch (see Fig. 4.1).

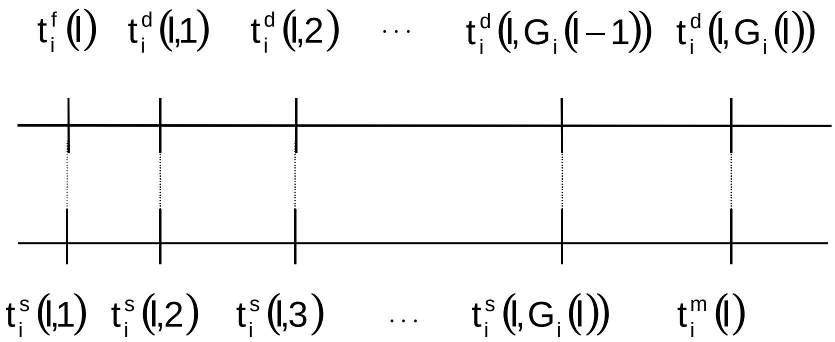

Figure 4.1: Coincidence of characteristic epochs at station i in the l-th polling cycle

Therefore the following relation holds for the ℓ-th cycle, for $\ell \geq 1$:

$$\begin{aligned}
&1_{\left(N_i\left(t_i^f(\ell)\right)=n\right)}\mathbf{1}_{\left(J_i\left(t_i^f(\ell)\right)\right)} + \sum_{k=1}^{G_i(\ell)} 1_{\left(N_i\left(t_i^d(\ell,k)\right)=n\right)}\mathbf{1}_{\left(J_i\left(t_i^d(\ell,k)\right)\right)} \\
&= \sum_{k=1}^{G_i(\ell)} 1_{\left(N_i\left(t_i^s(\ell,k)\right)=n\right)}\mathbf{1}_{\left(J_i\left(t_i^s(\ell,k)\right)\right)} + 1_{\left(N_i\left(t_i^m(\ell)\right)=n\right)}\mathbf{1}_{\left(J_i\left(t_i^m(\ell)\right)\right)}.
\end{aligned} \tag{4.4}$$

Summing from $\ell = 1$ to m on both sides of (4.4) and applying the definitions of $\boldsymbol{\nu}_i^f(k,n)$, $\boldsymbol{\nu}_i^d(k,n)$, $\boldsymbol{\nu}_i^s(k,n)$ and $\boldsymbol{\nu}_i^m(k,n)$ leads to

$$\boldsymbol{\nu}_i^f(k,n) + \boldsymbol{\nu}_i^d(k,n) = \boldsymbol{\nu}_i^s(k,n) + \boldsymbol{\nu}_i^m(k,n). \tag{4.5}$$

We take the expectations of all four terms in (4.5), divide them by the expectation of the total number of i-customer departures in the first m polling cycle ($E[\sum_{\ell=1}^m G_i(\ell)]$) and take the limit for $m \to \infty$. Thus we get a relation among the four stationary probabilities for each j phase of the i-th BMAP. In terms of vector generating functions, this yields

$$\gamma_i\widehat{\mathbf{f}}_i(z) + \widehat{\mathbf{q}}_i^d(z) = \widehat{\mathbf{q}}_i^s(z) + \gamma_i\widehat{\mathbf{m}}_i(z). \tag{4.6}$$

Here $\gamma_i = \lim_{m\to\infty} \frac{m}{E[\sum_{\ell=1}^{m} G_i(\ell)]}$ is the long-term ratio of the number of i-polling epochs to the number of i-customer departures, which equals $\frac{1}{g_i}$. Substituting it into (4.6) and rearranging it results in the statement. □

Now we relate the mean of the i-station time in the ℓ-th polling cycle ($E[S_i(\ell)]$) to the mean of the number of i-customer services in the ℓ-th polling cycle ($E[G_i(\ell)]$) for $\ell \geq 1$. $S_i(\ell) = \sum_{k=1}^{G_i(\ell)} B_i$ and $G_i(\ell)$ is stopping time, since it does not depend on the i-th customer service times after the i-departure epoch in the ℓ-th polling cycle. Hence Wald's equation can be applied, which yields

$$E[S_i(\ell)] = E[G_i(\ell)]b_i. \tag{4.7}$$

The number of arriving i-customers during $S_i(\ell)$ is given by $\Lambda_i(t_i^m(\ell)) - \Lambda_i(t_i^f(\ell))$. The stationary arrival rate of the i-th BMAP during the i-station time, λ_i^S is defined as

$$\lambda_i^S = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}(\Lambda_i(t_i^m(\ell)) - \Lambda_i(t_i^f(\ell)))\right]}{E\left[\sum_{\ell=1}^{m} S_i(\ell)\right]}. \tag{4.8}$$

Furthermore we define

$$\rho_i^S = \lambda_i^S b_i. \tag{4.9}$$

Corollary 4.1 *(Mean equilibrium relationship.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds for the mean stationary quantities:*

$$f_i^{(1)} - m_i^{(1)} = \left(1 - \rho_i^S\right) g_i. \tag{4.10}$$

Proof. Let $[\mathbf{q}_i^s]^{(1)}$ and $[\mathbf{q}_i^d]^{(1)}$ denote the first derivatives of $\widehat{\mathbf{q}}_i^s(z)$ and $\widehat{\mathbf{q}}_i^d(z)$ at $z = 1$, respectively, i.e., $[\mathbf{q}_i^s]^{(1)} = \left.\frac{d\left(\widehat{\mathbf{q}}_i^s(z)\right)}{dz}\right|_{z=1}$ and $[\mathbf{q}_i^d]^{(1)} = \left.\frac{d\left(\widehat{\mathbf{q}}_i^d(z)\right)}{dz}\right|_{z=1}$. Taking the first derivative of (4.3) at $z = 1$ and post-multiplying it by $\mathbf{e}$ yields

$$[\mathbf{q}_i^s]^{(1)}\mathbf{e} - [\mathbf{q}_i^d]^{(1)}\mathbf{e} = \frac{1}{g_i}\left(f_i^{(1)} - m_i^{(1)}\right). \tag{4.11}$$

The number of i-customers at i-customer departure equals the number of i-customers at the previous i-customer service start plus those who arrived in between minus the one who left the system at the current i-customer departure. Using it and also (4.7) we have

$$
\begin{aligned}
[\mathbf{q}_i^d]^{(1)}\mathbf{e} &= \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{G_i(\ell)} N_i(t_i^d(\ell,k))\right]}{E\left[\sum_{\ell=1}^{m} G_i(\ell)\right]} = \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{G_i(\ell)} N_i(t_i^s(\ell,k))\right]}{E\left[\sum_{\ell=1}^{m} G_i(\ell)\right]}\\
&+ \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{G_i(\ell)} (\Lambda_i(t_i^d(\ell,k)) - \Lambda_i(t_i^s(\ell,k)))\right] b_i}{E\left[\sum_{\ell=1}^{m} G_i(\ell)\right] b_i} - \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{G_i(\ell)} 1\right]}{E\left[\sum_{\ell=1}^{m} G_i(\ell)\right]}\\
&= [\mathbf{q}_i^s]^{(1)}\mathbf{e} + \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}(\Lambda_i(t_i^m(\ell)) - \Lambda_i(t_i^f(\ell)))\right]}{E\left[\sum_{\ell=1}^{m} S_i(\ell)\right]} b_i - 1.
\end{aligned}
\tag{4.12}
$$

Applying (4.8) in (4.12) leads to

$$
[\mathbf{q}_i^d]^{(1)}\mathbf{e} = [\mathbf{q}_i^s]^{(1)}\mathbf{e} + \lambda_i^S b_i - 1. \tag{4.13}
$$

Substituting (4.13) into (4.11) results in

$$
1 - \lambda_i^S b_i = \frac{1}{g_i}\left(f_i^{(1)} - m_i^{(1)}\right). \tag{4.14}
$$

Applying (4.9) in (4.14) and rearranging gives the statement. □

Remark 4.1 *(Interpretation of the mean equilibrium relationship.) The number of arriving i-customers during the i-intervisit time is $f_i^{(1)} - m_i^{(1)}$. The number of arriving i-customers during the i-station time is $\lambda_i^S b_i g_i$. Adding it to both sides of (4.10) the relation leads to the mean equilibrium relationship: the mean stationary number of arriving and served i-customers are the same during an i-polling cycle.*

4.2 Vector GF relations for the stationary number of i-customers

4.2.1 Stationary number of i-customers at i-customer departure epochs

Theorem 4.2 *(Expression of $\widehat{\mathbf{q}}_i^d(z)$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds for vector GF of the stationary number of i-customers at i-customer departure epochs:*

$$
\widehat{\mathbf{q}}_i^d(z)\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right) = (1-\rho_i^S)\frac{\widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z)}{f_i^{(1)} - m_i^{(1)}}\widehat{\mathbf{A}}_i(z). \tag{4.15}
$$

Proof. The vector GF of the number of i-customers just before the departure of the i-customer is $z\widehat{\mathbf{q}}_i^d(z)$. For this vector GF the following BMAP specific relation holds:

$$z\widehat{\mathbf{q}}_i^d(z) = \widehat{\mathbf{q}}_i^s(z)\,\widehat{\mathbf{A}}_i(z). \tag{4.16}$$

Post-multiplying the fundamental relationship (4.3) by $\widehat{\mathbf{A}}_i(z)$, applying (4.16) yields

$$\widehat{\mathbf{q}}_i^d(z)\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right) = \frac{\left(\widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z)\right)}{g_i}\widehat{\mathbf{A}}_i(z). \tag{4.17}$$

The theorem can be obtained by expressing g_i in (4.17) from the mean equilibrium relationship (4.10). □

Note that the contribution of the applied service discipline in (4.15) is represented by $\widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z)$. The term $\frac{(1-\rho_i^S)}{\left(f_i^{(1)}-m_i^{(1)}\right)}$ can be expressed from the mean equilibrium relationship (4.10) as $\frac{1}{g_i}$. Using (3.6), (3.2) and (3.10) implies that in the nonzero-switchover-times polling model also $g_i = \lambda_i c = \lambda_i \frac{\sum_{i=1}^N r_i}{1-\rho}$ holds. Therefore, for this model, $\frac{(1-\rho_i^S)}{\left(f_i^{(1)}-m_i^{(1)}\right)}$ can be computed in a discipline independent way as:

$$\frac{\left(1-\rho_i^S\right)}{\left(f_i^{(1)} - m_i^{(1)}\right)} = \frac{1-\rho}{\lambda_i \sum_{i=1}^N r_i}. \tag{4.18}$$

4.2.2 Stationary number of i-customers at an arbitrary instant

We define vector GF of the stationary number of i-customers $\widehat{\mathbf{q}}_i(z)$ as

$$\widehat{\mathbf{q}}_i(z) = \lim_{t\to\infty}\sum_{n=0}^{\infty} E[1_{(N_i(t)=n)}\mathbf{1}_{(J_i(t))}]z^n, \quad |z| \le 1.$$

Theorem 4.3 *(Expression of $\widehat{\mathbf{q}}_i(z)$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** *-* **A.3** *and properties* **P.1** *-* **P.4** *the following relation holds for vector GF of the stationary number of i-customers:*

$$\widehat{\mathbf{q}}_i(z)\widehat{\mathbf{D}}_i(z)\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right) = \lambda_i(1-\rho_i^S)(z-1)\frac{\widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z)}{f_i^{(1)} - m_i^{(1)}}\widehat{\mathbf{A}}_i(z). \tag{4.19}$$

Proof. Multiplying (4.15) by $\lambda_i(z-1)$ leads to

$$\lambda_i(z-1)\widehat{\mathbf{q}}_i^d(z)\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right) = \lambda_i(1-\rho_i^S)(z-1)\frac{\widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z)}{f_i^{(1)} - m_i^{(1)}}\widehat{\mathbf{A}}_i(z). \tag{4.20}$$

Takine and Takahashi proved the following stationary relationship between $\widehat{\mathbf{q}}_i(z)$ and $\widehat{\mathbf{q}}_i^d(z)$ under general setting in [175]:

$$\widehat{\mathbf{q}}_i(z)\widehat{\mathbf{D}}_i(z) = \lambda_i(z-1)\widehat{\mathbf{q}}_i^d(z). \tag{4.21}$$

Applying (4.21) in (4.20) gives the theorem. □

The dependency on the applied service discipline in (4.19) is expressed by $\widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z)$.

4.3 Factorial moments of the stationary number of i-customers

In this Section we provide formulas for the vector factorial moments of the stationary number of i-customers at an arbitrary instant and at i-customer departure epochs. They enable the applications of the expressions for $\widehat{\mathbf{q}}_i(z)$ and $\widehat{\mathbf{q}}_i^d(z)$ ((4.19 and (4.15)).

4.3.1 Factorial moments at an arbitrary instant

Let $\mathbf{q}_i^{(n)}$ denote the n-th ($n \geq 1$) factorial moment of $\widehat{\mathbf{q}}_i(z)$, i.e., $\mathbf{q}_i^{(n)} = \frac{d^n}{dz^n}\widehat{\mathbf{q}}_i(z)|_{z=1}$ and $\mathbf{q}_i$ and $\mathbf{q}_i^{(0)}$ denote its value at $z=1$, i.e., $\mathbf{q}_i = \mathbf{q}_i^{(0)} = \widehat{\mathbf{q}}_i(1)$.

Matrix $\widehat{\mathbf{T}}_i(z)$ is defined as:

$$\widehat{\mathbf{T}}_i(z) = \widehat{\mathbf{D}}_i(z)\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right).$$

Let $[det\mathbf{T}_i]^{(k)}$ denote the k-th ($k \geq 1$) derivative of $det\widehat{\mathbf{T}}_i(z)$ at $z=1$, i.e., $[det\mathbf{T}_i]^{(k)} = \left.\frac{d^k\left(det\widehat{\mathbf{T}}_i(z)\right)}{dz^k}\right|_{z=1}$. Furthermore $det\mathbf{T}_i$ denotes $det\widehat{\mathbf{T}}_i(1)$ and by definition $[det\mathbf{T}_i]^{(0)} = det\mathbf{T}_i$. Similarly $[adj\mathbf{T}_i]^{(k)}$ denotes the k-th ($k \geq 1$) derivative of matrix $adj\widehat{\mathbf{T}}_i(z)$ at $z=1$, i.e., $[adj\mathbf{T}_i]^{(k)} = \left.\frac{d^k\left(adj\widehat{\mathbf{T}}_i(z)\right)}{dz^k}\right|_{z=1}$. Furthermore $adj\mathbf{T}_i$ denotes $adj\widehat{\mathbf{T}}_i(1)$ and by definition $[adj\mathbf{T}_i]^{(0)} = adj\mathbf{T}_i$.

Let $\binom{n}{k_1,\ldots,k_m}$ denote the multinomial coefficient for $n \geq 0$, $m \geq 2$ and $k_1,\ldots,k_m = 0,\ldots,n$, i.e. $\binom{n}{k_1,\ldots,k_m} = \frac{n!}{k_1!\ldots k_m!}$.

Theorem 4.4 *(The vector factorial moment formula for $\mathbf{q}_i^{(n)}$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the recursive formula for computing the factorial moments of the stationary number of i-customers at an arbitrary instant is given by:*

$$\mathbf{q}_i^{(n)} = \frac{\lambda_i(1-\rho_i^S)}{f_i^{(1)}-m_i^{(1)}} \sum_{l=0}^{n+1}\sum_{k=0}^{n+1-l} \binom{n+2}{1,l,n+1-k-l,k} \left(\mathbf{f}_i^{(l)}-\mathbf{m}_i^{(l)}\right) \mathbf{A}_i^{(n+1-k-l)}$$
$$\frac{[adj\mathbf{T}_i]^{(k)}}{(1+2n+1_{(n\geq 2)}\binom{n}{2})[det\mathbf{T}_i]^{(2)}} - \boldsymbol{\pi}_i \frac{[det\mathbf{T}_i]^{(n+2)}}{(1+2n+1_{(n\geq 2)}\binom{n}{2})[det\mathbf{T}_i]^{(2)}}$$
$$- \left(1_{(n\geq 3)}\sum_{k=1}^{n-2}\binom{n}{k+2} + 1_{(n\geq 2)}\sum_{k=1}^{n-1}\left(\binom{n}{k+1}+\binom{n+1}{k+1}\right)\right)\mathbf{q}_i^{(n-k)}$$
$$\frac{[det\mathbf{T}_i]^{(k+2)}}{(1+2n+1_{(n\geq 2)}\binom{n}{2})[det\mathbf{T}_i]^{(2)}} \quad n\geq 1. \tag{4.22}$$

The statement of the theorem means that the factorial moments of the stationary number of i-customers ($\mathbf{q}_i(n)$) can be computed recursively by applying the formula (4.22) for $k=1,\ldots,n$. Furthermore $\mathbf{q}_i(n)$ depends on the factorial moments of the number of i-customers and the phase probability vectors of the i-BMAP at i-polling and i-departure epochs ($(\mathbf{f}_i^{(k)}-\mathbf{m}_i^{(k)})$, $0\leq k\leq n+1$), which are service discipline dependent.

Proof. Left multiplying $\mathbf{I}=\frac{\widehat{\mathbf{T}}_i(z)adj\widehat{\mathbf{T}}_i(z)}{det\widehat{\mathbf{T}}_i(z)}$ by $\frac{d^n\widehat{\mathbf{q}}_i(z^n)}{dz}$ and taking the limit for $z\to 1$ yields

$$\mathbf{q}_i^{(n)} = lim_{z\to 1}\frac{\frac{d^n\widehat{\mathbf{q}}_i(z)}{dz^n}\widehat{\mathbf{T}}_i\left(z\right)adj\widehat{\mathbf{T}}_i\left(z\right)}{det\widehat{\mathbf{T}}_i\left(z\right)}. \tag{4.23}$$

It is shown in proposition B.2 (see in Appendix B) that as $lim_{z\to 1}$ both the nominator and the denominator of the r.h.s. of (4.23) as well as also the first derivatives of the nominator and the denominator of the r.h.s. of (4.23) are 0. Furthermore it is also shown there that as $lim_{z\to 1}$ the second derivative of the denominator of the r.h.s. of (4.23) is nonzero.

Therefore we apply L'Hospital rule two times on (4.23). Using also that the terms in (4.23) are continuously differentiable, we get

$$\mathbf{q}_i^{(n)} = \frac{\left.\frac{d^2\left(\frac{d^n\widehat{\mathbf{q}}_i(z)}{dz^n}\widehat{\mathbf{T}}_i(z)adj\widehat{\mathbf{T}}_i(z)\right)}{dz^2}\right|_{z=1}}{[det\mathbf{T}_i]^{(2)}}. \tag{4.24}$$

We define vector $\widehat{\mathbf{y}}_i\left(z\right)$ as

$$\widehat{\mathbf{y}}_i\left(z\right) = \widehat{\mathbf{q}}_i\left(z\right)\widehat{\mathbf{T}}_i\left(z\right). \tag{4.25}$$

Using $\widehat{\mathbf{T}}_i\left(z\right)adj\widehat{\mathbf{T}}_i\left(z\right) = det\widehat{\mathbf{T}}_i\left(z\right)$ and (4.25) we rearrange the nominator of (4.24) as

$$\frac{d^n\widehat{\mathbf{q}}_i(z)}{dz^n}\widehat{\mathbf{T}}_i(z)\,adj\widehat{\mathbf{T}}_i(z) = \frac{d^n\widehat{\mathbf{q}}_i(z)}{dz^n}det\widehat{\mathbf{T}}_i(z) \tag{4.26}$$
$$= \frac{d^n\left(\widehat{\mathbf{q}}_i(z)\,det\widehat{\mathbf{T}}_i(z)\right)}{dz^n} - \sum_{k=1}^{n}\binom{n}{k}\frac{d^{n-k}\widehat{\mathbf{q}}_i(z)}{dz^{n-k}}\frac{d^k det\widehat{\mathbf{T}}_i(z)}{dz^k}$$
$$= \frac{d^n\left(\widehat{\mathbf{y}}_i(z)\,adj\widehat{\mathbf{T}}_i(z)\right)}{dz^n} - \sum_{k=1}^{n}\binom{n}{k}\frac{d^{n-k}\widehat{\mathbf{q}}_i(z)}{dz^{n-k}}\frac{d^k det\widehat{\mathbf{T}}_i(z)}{dz^k}.$$

Substituting (4.26) into (4.24) and rearranging yields

$$\mathbf{q}_i^{(n)} = \frac{\left.\frac{d^{n+2}\left(\widehat{\mathbf{y}}_i(z)adj\widehat{\mathbf{T}}_i(z)\right)}{dz^{n+2}}\right|_{z=1}}{[det\mathbf{T}_i]^{(2)}} - \frac{\left.\frac{d^2\left(\sum_{k=1}^{n}\binom{n}{k}\frac{d^{n-k}\widehat{\mathbf{q}}_i(z)}{dz^{n-k}}\frac{d^k det\widehat{\mathbf{T}}_i(z)}{dz^k}\right)}{dz^2}\right|_{z=1}}{[det\mathbf{T}_i]^{(2)}}$$
$$= \frac{\sum_{k=0}^{n+2}\binom{n+2}{k}\mathbf{y}_i^{(n+2-k)}[adj\mathbf{T}_i]^{(k)}}{[det\mathbf{T}_i]^{(2)}} - \frac{\sum_{k=1}^{n}\binom{n}{k}\mathbf{q}_i^{(n+2-k)}[det\mathbf{T}_i]^{(k)}}{[det\mathbf{T}_i]^{(2)}}$$
$$-\frac{2\sum_{k=1}^{n}\binom{n}{k}\mathbf{q}_i^{(n+1-k)}[det\mathbf{T}_i]^{(k+1)}}{[det\mathbf{T}_i]^{(2)}} - \frac{\sum_{k=1}^{n}\binom{n}{k}\mathbf{q}_i^{(n-k)}[det\mathbf{T}_i]^{(k+2)}}{[det\mathbf{T}_i]^{(2)}}.$$

Further manipulation and using $\mathbf{y}_i = \mathbf{q}_i\mathbf{T}_i = \boldsymbol{\pi}_i\mathbf{D}_i(\mathbf{I} - \mathbf{A}_i) = 0$ and $[det\mathbf{T}_i]^{(1)} = 0$ (see in the proof of proposition B.2 in Appendix B) leads to

$$\mathbf{q}_i^{(n)}[det\mathbf{T}_i]^{(2)} = \sum_{k=0}^{n+1}\binom{n+2}{k}\mathbf{y}_i^{(n+2-k)}[adj\mathbf{T}_i]^{(k)} - 1_{(n\geq 3)}\sum_{k=3}^{n}\binom{n}{k}\mathbf{q}_i^{(n+2-k)}[det\mathbf{T}_i]^{(k)}$$
$$-1_{(n\geq 2)}2\sum_{k=2}^{n}\binom{n}{k}\mathbf{q}_i^{(n+1-k)}[det\mathbf{T}_i]^{(k+1)} - \sum_{k=1}^{n}\binom{n}{k}\mathbf{q}_i^{(n-k)}[det\mathbf{T}_i]^{(k+2)}$$
$$-1_{(n\geq 2)}\binom{n}{2}\mathbf{q}_i^{(n)}[det\mathbf{T}_i]^{(2)} - 2\binom{n}{1}\mathbf{q}_i^{(n)}[det\mathbf{T}_i]^{(2)}. \tag{4.27}$$

Now we express the derivatives of $\widehat{\mathbf{y}}_i(z)$. The left-hand side (l.h.s.) of (4.19) is exactly $\widehat{\mathbf{y}}_i(z)$. Observe that for $n \geq 0$ the n-th derivative of the term $(z-1)$ vanishes at $z = 1$ except for $n = 1$. It follows that the n-th derivative of $\widehat{\mathbf{y}}_i(z)$ at $z = 1$ can be expressed as

$$\mathbf{y}_i^{(n)} = \frac{\lambda_i(1-\rho_i^S)}{f_i^{(1)} - m_i^{(1)}}\sum_{l=0}^{n-1}\binom{n}{1, l, n-1-l}\left(\mathbf{f}_i^{(l)} - \mathbf{m}_i^{(l)}\right)\mathbf{A}_i^{(n-1-l)}. \tag{4.28}$$

Applying (4.28) in (4.27) and rearranging gives

$$
\begin{aligned}
&(1+2n+1_{(n\geq 2)}\binom{n}{2})\mathbf{q}_i^{(n)}[det\mathbf{T}_i]^{(2)} \\
&= \frac{\lambda_i(1-\rho_i^S)}{f_i^{(1)}-m_i^{(1)}}\sum_{k=0}^{n+1}\binom{n+2}{k}\sum_{l=0}^{n+1-k}\binom{n+2-k}{1,l,n+1-k-l}\left(\mathbf{f}_i^{(l)}-\mathbf{m}_i^{(l)}\right)\mathbf{A}_i^{(n+1-k-l)}[adj\mathbf{T}_i]^{(k)} \\
&-\left(1_{(n\geq 3)}\sum_{k=1}^{n-2}\binom{n}{k+2}+1_{(n\geq 2)}2\sum_{k=1}^{n-1}\binom{n}{k+1}+\sum_{k=1}^{n}\binom{n}{k}\right)\mathbf{q}_i^{(n-k)}[det\mathbf{T}_i]^{(k+2)}.
\end{aligned}
$$

Separating the terms with $\mathbf{q}_i = \boldsymbol{\pi}_i$ and further rearranging leads to

$$
\begin{aligned}
&(1+2n+1_{(n\geq 2)}\binom{n}{2})\mathbf{q}_i^{(n)}[det\mathbf{T}_i]^{(2)} \\
&= \frac{\lambda_i(1-\rho_i^S)}{f_i^{(1)}-m_i^{(1)}}\sum_{l=0}^{n+1}\sum_{k=0}^{n+1-l}\binom{n+2}{1,l,n+1-k-l,k}\left(\mathbf{f}_i^{(l)}-\mathbf{m}_i^{(l)}\right)\mathbf{A}_i^{(n+1-k-l)}[adj\mathbf{T}_i]^{(k)} \\
&-\left(1_{(n\geq 3)}\sum_{k=1}^{n-2}\binom{n}{k+2}+1_{(n\geq 2)}2\sum_{k=1}^{n-1}\binom{n}{k+1}+1_{(n\geq 2)}\sum_{k=1}^{n-1}\binom{n}{k}\right)\mathbf{q}_i^{(n-k)}[det\mathbf{T}_i]^{(k+2)} \\
&-\boldsymbol{\pi}_i[det\mathbf{T}_i]^{(n+2)}.
\end{aligned} \tag{4.29}
$$

Rearranging (4.29) results in the statement. □

Remark 4.2 *(The vector mean formula.) The vector mean of the stationary number of i-customers at an arbitrary instant is a special case of 4.22 for $n = 1$. It can be rearranged to the following form:*

$$
\begin{aligned}
\mathbf{q}_i^{(1)} &= \frac{\mathbf{f}_i^{(2)}-\mathbf{m}_i^{(2)}}{\left(f_i^{(1)}-m_i^{(1)}\right)}\lambda_i\left(1-\rho_i^S\right)\mathbf{A}_i adj\mathbf{T}_i\frac{1}{[det\mathbf{T}_i]^{(2)}} \\
&+ 2\frac{\mathbf{f}_i^{(1)}-\mathbf{m}_i^{(1)}}{\left(f_i^{(1)}-m_i^{(1)}\right)}\lambda_i\left(1-\rho_i^S\right)\mathbf{A}_i^{(1)} adj\mathbf{T}_i\frac{1}{[det\mathbf{T}_i]^{(2)}} \\
&+ 2\frac{\mathbf{f}_i^{(1)}-\mathbf{m}_i^{(1)}}{\left(f_i^{(1)}-m_i^{(1)}\right)}\lambda_i\left(1-\rho_i^S\right)\mathbf{A}_i[adj\mathbf{T}_i]^{(1)}\frac{1}{[det\mathbf{T}_i]^{(2)}} \\
&+ \frac{\mathbf{f}_i-\mathbf{m}_i}{\left(f_i^{(1)}-m_i^{(1)}\right)}\lambda_i\left(1-\rho_i^S\right)\mathbf{A}_i^{(2)} adj\mathbf{T}_i\frac{1}{[det\mathbf{T}_i]^{(2)}} \\
&+ 2\frac{\mathbf{f}_i-\mathbf{m}_i}{\left(f_i^{(1)}-m_i^{(1)}\right)}\lambda_i\left(1-\rho_i^S\right)\mathbf{A}_i^{(1)}[adj\mathbf{T}_i]^{(1)}\frac{1}{[det\mathbf{T}_i]^{(2)}} \\
&+ \frac{\mathbf{f}_i-\mathbf{m}_i}{\left(f_i^{(1)}-m_i^{(1)}\right)}\lambda_i\left(1-\rho_i^S\right)\mathbf{A}_i[adj\mathbf{T}_i]^{(2)}\frac{1}{[det\mathbf{T}_i]^{(2)}} \\
&- \frac{1}{3}\pi_i[det\mathbf{T}_i]^{(3)}\frac{1}{[det\mathbf{T}_i]^{(2)}}.
\end{aligned} \tag{4.30}
$$

An alternative derivation of this form of $\mathbf{q}_i^{(1)}$ *can be found in [199].*

4.3.2 Factorial moments at i-customer departure epochs

Let $\mathbf{q}_i^{d(n)}$ denote the n-th ($n \geq 1$) factorial moment of $\widehat{\mathbf{q}}_i^d(z)$, i.e., $\mathbf{q}_i^{d(n)} = \frac{d^n}{dz^n}\widehat{\mathbf{q}}_i^d(z)|_{z=1}$ and $\mathbf{q}_i^d$ and $\mathbf{q}_i^{d(0)}$ denote its value at $z = 1$, i.e., $\mathbf{q}_i^d = \mathbf{q}_i^{d(0)} = \widehat{\mathbf{q}}_i^d(1)$.

Deriving the vector factorial moments of the stationary number of i-customers at i-customer departure epochs ($\mathbf{q}_i^{d(n)}$, $n \geq 1$) from the expression of $\widehat{\mathbf{q}}_i^d(z)$ (4.15) would lead to an expression depending on $\mathbf{q}_i^d$, the stationary i-BMAP phase probability vector at i-customer departure epochs, which is unknown. Therefore instead of it we utilize the relation (4.21) between $\widehat{\mathbf{q}}_i^d(z)$ and $\widehat{\mathbf{q}}_i(z)$ and derive a formula for relating the vector factorial moments ($\mathbf{q}_i^{d(n)}$, $n \geq 1$) to ($\mathbf{q}_i^{(n)}$, $n \geq 1$). Thus the vector factorial moments $\mathbf{q}_i^{d(n)}$, for $n \geq 1$ can be calculated from an appropriate set of $\mathbf{q}_i^{(n)}$-s.

Corollary 4.2 *(Relation between* $\mathbf{q}_i^{d(n)}$*,* $n \geq 1$ *and* $\mathbf{q}_i^{(n)}$*,* $n \geq 1$*.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds between the vector factorial moments of the stationary number of i-customers at i-customer departure epochs and at an arbitrary instant:*

$$\mathbf{q}_i^{d(n)} = \frac{1}{\lambda_i}\sum_{k=0}^{n+1}\binom{n+1}{k}\mathbf{q}_i^{(n+1-k)}\mathbf{D}_i^k, \quad n \geq 0. \tag{4.31}$$

Proof. Taking the n-th derivative of the stationary relationship (4.21), for $n \geq 1$, yields

$$\lambda_i\mathbf{q}_i^{d(n-1)} = \sum_{k=0}^{n}\binom{n}{k}\mathbf{q}_i^{(n-k)}\mathbf{D}_i^k, \quad n \geq 1. \tag{4.32}$$

Rearranging (4.32) gives the statement. □

4.4 Relations at i-BMAP state change epochs

From now on the superscript * in the notation of quantities denotes the BMAP state change epochs. Hence a quantity having superscript * in its notation stands either for a quantity defined just before the BMAP state changes (like e.g. $\widehat{\mathbf{q}}_i^*(z)$, which is the vector GF of the stationary number of i-customers just before i-BMAP state change epochs) or for a quantity defined as the number of BMAP state change epochs during the specified interval (like a_i^*, which is the mean stationary number of i-BMAP state changes during a polling cycle).

Using the identity

$$\sum_{n=0}^{\infty} E\left[1_{\left(N_i(t_i^f(\ell))=n\right)} 1_{\left(J_i(t_i^f(\ell))=j\right)}\right] z^n = E\left[z^{N_i\left(t_i^f(\ell)\right)} 1_{\left(J_i(t_i^f(\ell))=j\right)}\right]$$

for $j \in \{1, \ldots, L\}$ and $i = 1, \ldots, N$, the (4.1) definition of the $1 \times L$ vector $\widehat{\mathbf{f}}_i(z)$ can be rewritten as

$$\widehat{\mathbf{f}}_i(z) = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i\left(t_i^f(\ell)\right)} \mathbf{1}_{\left(J_i\left(t_i^f(\ell)\right)\right)}\right]}{m}, \quad |z| \le 1. \tag{4.33}$$

Similarly the definition of the $1 \times L$ vector $\widehat{\mathbf{m}}_i(z)$, the vector GF of the stationary number of i-customers at i-departure epochs, can be rewritten as

$$\widehat{\mathbf{m}}_i(z) = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i\left(t_i^m(\ell)\right)} \mathbf{1}_{\left(J_i\left(t_i^m(\ell)\right)\right)}\right]}{m}, \quad |z| \le 1. \tag{4.34}$$

Furthermore let $A_i^*(\ell, n)$ be the number of i-BMAP state changes, i.e. either batch arrival or phase change, during the service time of the n-th i-customer in the ℓ-th polling cycle, for $\ell \ge 1$ and $n = 1, \ldots, G_i(\ell) + 1$, where $A_i^*(\ell, G_i(\ell) + 1)$ is the number of i-BMAP state changes during the i-intervisit time in the ℓ-th polling cycle. We define the mean stationary number of i-BMAP state changes during a polling cycle as

$$a_i^* = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} \sum_{n=1}^{G_i(\ell)+1} A_i^*(\ell, n)\right]}{m}. \tag{4.35}$$

Let $t_i(\ell, n, k)$ be the time of the k-th i-BMAP state changes during the service time of the n-th i-customer in the ℓ-th polling cycle, for $\ell \ge 1$, $n = 1, \ldots, G_i(\ell) + 1$ and $k = 0, \ldots, A_i^*(\ell, n)$. By definition $t_i(\ell, n, 0) = t_i^d(\ell, n-1)$ for $\ell \ge 1$, $n = 1, \ldots, G_i(\ell)+1$. Here $t_i(\ell, G_i(\ell) + 1, k)$, for $k = 1, \ldots, A_i^*(\ell, n)$, is the time of the k-th i-BMAP state changes during the i-intervisit time in the ℓ-th polling cycle.

Some of these notations are illustrated on Fig. 4.2.

The set of i-BMAP state change epochs are small extension to the batch arriving epochs. Therefore we define the number of i-customers just before these epochs analogously to the case of "seen by batch arrivals". However the phase of the i-BMAP just before the actual i-BMAP state change equals the phase at the previous i-BMAP state

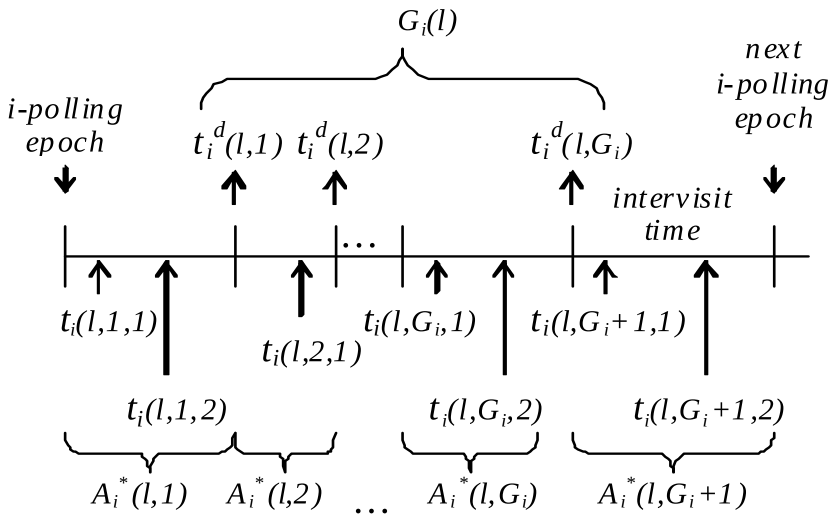

Figure 4.2: Notations

change. Similarly, assuming that there is no departure between the actual and previous i-BMAP state changes, the number of i-customers just before the actual i-BMAP state change equals the number of i-customers at the previous state change. Note that this condition holds at $t_i(\ell,n,k)$ epochs for fix ℓ and n. Using these arguments we define the $1 \times L$ vector $\widehat{\mathbf{q}}_i^*(z)$, the vector GF of the stationary number of i-customers just before i-BMAP state change epochs as

$$\widehat{\mathbf{q}}_i^*(z) = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1} A_i^*(\ell,n)\right]}, \qquad |z| \le 1. \tag{4.36}$$

We define the mean stationary i-BMAP phase probability vector just before i-BMAP state change epochs as

$$\boldsymbol{\pi}_i^* = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} \mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1} A_i^*(\ell,n)\right]}. \tag{4.37}$$

The defining expression (4.37) implies that $\boldsymbol{\pi}_i^* = \widehat{\mathbf{q}}_i^*(1)$. $\boldsymbol{\pi}_i^*$ can be uniquely determined from $\boldsymbol{\pi}_i^* \boldsymbol{\Psi}_i = \boldsymbol{\pi}_i^*$ and $\boldsymbol{\pi}_i^* \mathbf{e} = 1$. Applying the expression of $\boldsymbol{\Psi}_i$ from (2.6) in $\boldsymbol{\pi}_i^* \boldsymbol{\Psi}_i = \boldsymbol{\pi}_i^*$ results in

$$\boldsymbol{\pi}_i^* \left(-\mathbf{D}_{i,0}^d\right)^{-1} \mathbf{D}_i = 0. \tag{4.38}$$

Let $A_i(\ell, n, k)$ be the number of i-customers arrived at the k-th i-BMAP state change during the service time of the n-th i-customer in the ℓ-th polling cycle, for $\ell \geq 1$, $n = 1, \ldots, G_i(\ell) + 1$ and $k = 1, \ldots, A_i^*(\ell, n)$. Here $A_i(\ell, G_i(\ell) + 1, k)$ is the number of i-customers arrived in the batch at the k-th i-BMAP state changes during the i-intervisit time in the ℓ-th polling cycle. Now we define also the mean stationary number of i-customers arrived during a polling cycle as

$$a_i = \lim_{m \to \infty} \frac{E\left[\sum_{\ell=1}^{m} \sum_{n=1}^{G_i(\ell)+1} \sum_{k=1}^{A_i^*(\ell,n)} A_i(\ell, n, k)\right]}{m}. \tag{4.39}$$

Proposition 4.1 *(Expression of a_i^*.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the mean stationary number of i-customer arrivals during a polling cycle can be expressed as*

$$a_i^* = \frac{g_i}{\lambda_i \left(\boldsymbol{\pi}_i^* (-\mathbf{D}_{i,0}^d)^{-1} \mathbf{e}\right)}. \tag{4.40}$$

Proof. According to (2.1) the rank of $\mathbf{D}_i$ is $L - 1$ and hence, due to (4.38), $\boldsymbol{\pi}_i^* \left(-\mathbf{D}_{i,0}^d\right)^{-1}$ must be a constant multiple of $\boldsymbol{\pi}_i$. It follows, that

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_i^* \frac{(-\mathbf{D}_{i,0}^d)^{-1}}{\boldsymbol{\pi}_i^* (-\mathbf{D}_{i,0}^d)^{-1} \mathbf{e}}. \tag{4.41}$$

We remark here that $\boldsymbol{\pi}_i^* (-\mathbf{D}_{i,0}^d)^{-1} \mathbf{e}$ can be interpreted as the mean stationary length of an arbitrary i-BMAP state.

Now we continue with the defining expression of a_i. Using $\sum_{n=1}^{G_i(\ell)+1} = 1_{(n \leq G_i(\ell)+1)} \sum_{n=1}^{\infty}$ and $\sum_{k=1}^{A_i^*(\ell,n)} = 1_{\left(k \leq A_i^*(\ell,n)\right)} \sum_{k=1}^{\infty}$ yields

$$\begin{aligned} a_i = & \lim_{m \to \infty} \frac{1}{m} E\left[\sum_{\ell=1}^{m} \sum_{n=1}^{G_i(\ell)+1} \sum_{m=1}^{A_i^*(\ell,n)} A_i(\ell, n, k)\right] \\ = & \lim_{m \to \infty} \frac{1}{m} \sum_{\ell=1}^{m} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} E\left[1_{(n \leq G_i(\ell)+1)} 1_{\left(k \leq A_i^*(\ell,n)\right)} A_i(\ell, n, k)\right]. \end{aligned} \tag{4.42}$$

If the initial and final phases of the i-BMAP at an i-BMAP state are known then the number of i-customers arrived at that i-BMAP state change can be described by matrix $\widehat{\mathbf{\Psi}}(z)$. In order to utilize it we multiply the internal terms of the expected value on the r.h.s of (4.42) by $1 = \sum_{j^-=1}^{L} 1_{(J_i(t_i(\ell,n,k-1))=j^-)}$ and $1 = \sum_{j^+=1}^{L} 1_{(J_i(t_i(\ell,n,k))=j^+)}$ as well as insert a conditional expectation by applying the relation $E[H_1, H_2] = E[H_1, E[H_2|H_1]]$, which holds for generic random variables H_1 and H_2. This leads to

$$\begin{aligned} a_i = \quad & \lim_{m\to\infty} \frac{1}{m} \sum_{\ell=1}^{m} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} E\big[1_{(n\le G_i(\ell)+1)} 1_{\left(k\le A_i^*(\ell,n)\right)} \sum_{j^-=1}^{L} 1_{(J_i(t_i(\ell,n,k-1))=j^-)} A_i(\ell,n,k) \\ & \sum_{j^+=1}^{L} 1_{(J_i(t_i(\ell,n,k))=j^+)}\big] \\ & = \lim_{m\to\infty} \frac{1}{m} \sum_{\ell=1}^{m} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} E\Big[1_{(n\le G_i(\ell)+1)} 1_{\left(k\le A_i^*(\ell,n)\right)} \sum_{j^-=1}^{L} 1_{(J_i(t_i(\ell,n,k-1))=j^-)} \sum_{j^+=1}^{L} \\ & E\Big[A_i(\ell,n,k) 1_{(J_i(t_i(\ell,n,k))=j^+)} | 1_{(J_i(t_i(\ell,n,k-1))=j^-)}, 1_{\left(k\le A_i^*(\ell,n)\right)}, 1_{(n\le G_i(\ell)+1)}\Big]\Big]. \end{aligned}$$

If $k \le A_i^*(\ell,n)$ then the number of i-customers arrived at the k-th i-BMAP state change during the service time of the n-th i-customer in the ℓ-th polling cycle depends only on the phase of the i-BMAP just before and at the current state change. Using it yields

$$\begin{aligned} a_i = \quad & \lim_{m\to\infty} \frac{1}{m} \sum_{\ell=1}^{m} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} E\Big[1_{(n\le G_i(\ell)+1)} 1_{\left(k\le A_i^*(\ell,n)\right)} \sum_{j^-=1}^{L} 1_{(J_i(t_i(\ell,n,k-1))=j^-)} \sum_{j^+=1}^{L} \\ & E\left[A_i(\ell,n,k) 1_{(J_i(t_i(\ell,n,k))=j^+)} | 1_{(J_i(t_i(\ell,n,k-1))=j^-)}, (k\le A_i^*(\ell,n))\right]\Big] \\ = \quad & \lim_{m\to\infty} \frac{1}{m} E\Big[\sum_{\ell=1}^{m} \sum_{n=1}^{G_i(\ell)+1} \sum_{k=1}^{A_i^*(\ell,n)} \sum_{j^-=1}^{L} 1_{(J_i(t_i(\ell,n,k-1))=j^-)} \sum_{j^+=1}^{L} \\ & E\left[A_i(\ell,n,k) 1_{(J_i(t_i(\ell,n,k))=j^+)} | 1_{(J_i(t_i(\ell,n,k-1))=j^-)}, (k\le A_i^*(\ell,n))\right]\Big]. \end{aligned} \quad (4.43)$$

The conditional expectation term expresses the mean number of i-customers arrived at i-BMAP state change at $t_i(\ell,n,k)$ epoch for the given initial and end phases, which is given by $\mathbf{e}_{j^-} \mathbf{\Psi}_i^{(1)} \mathbf{e}_{j^+}^T$. Substituting it into (4.43) and rearranging yields

$$\begin{aligned} a_i = & \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)}\sum_{j^-=1}^{L} 1_{(J_i(t_i(\ell,n,k-1)=j^-))}\sum_{j^+=1}^{L}\mathbf{e}_{j^-}\mathbf{\Psi}_i^{(1)}\mathbf{e}_{j^+}^T\right]}{m} \\ = & \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} \mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{m}\mathbf{\Psi}_i^{(1)}\mathbf{e} \\ = & \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} \mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{E\left[\sum_{\ell=1}^{k}\sum_{n=1}^{G_i(\ell)+1} A_i^*(\ell,n)\right]} \\ & \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1} A_i^*(\ell,n)\right]}{m}\mathbf{\Psi}_i^{(1)}\mathbf{e}. \end{aligned} \tag{4.44}$$

Applying (4.37) and (4.35) in (4.44), expressing $\mathbf{\Psi}_i^{(1)}$ from (2.5) and using (4.41) and (2.3) leads to

$$\begin{aligned} a_i = & \quad a_i^*\boldsymbol{\pi}_i^*\mathbf{\Psi}_i^{(1)}\mathbf{e} = a_i^*\boldsymbol{\pi}_i^*(-\mathbf{D}_{i,0}^d)^{-1}\mathbf{D}_i^{(1)}\mathbf{e} \\ = & \quad a_i^*\boldsymbol{\pi}_i\mathbf{D}_i^{(1)}\mathbf{e}\left(\boldsymbol{\pi}_i^*(-\mathbf{D}_{i,0}^d)^{-1}\mathbf{e}\right) = a_i^*\lambda_i\left(\boldsymbol{\pi}_i^*(-\mathbf{D}_{i,0}^d)^{-1}\mathbf{e}\right). \end{aligned}$$

Using that in the stable polling model $g_i = a_i$ holds results in the statement of the proposition.

□

We define the mean stationary number of i-BMAP state changes during an i-station time as

$$a_i^{s*} = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)} A_i^*(\ell,n)\right]}{m}.$$

Similarly we also define the mean stationary number of i-BMAP state changes during an i-intervisit time as

$$a_i^{i*} = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} A_i^*(\ell,G_i(\ell)+1)\right]}{m}.$$

We define the $1 \times L$ vector $\widehat{\mathbf{q}}_i^{s*}(z)$, the vector GFs of the stationary number of i-customers just before i-BMAP state change epochs during an i-station time as

$$\widehat{\mathbf{q}}_i^{s*}(z) \quad = \quad \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)} A_i^*(\ell,n)\right]}, \quad |z|\leq 1.$$

Similarly we also define the $1 \times L$ vector $\widehat{\mathbf{q}}_i^{i*}(z)$, the vector GFs of the stationary number of i-customers just before i-BMAP state change epochs during an i-intervisit time as

$$\widehat{\mathbf{q}}_i^{i*}(z) \quad = \quad \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)} z^{N_i(t_i(\ell,G_i(\ell)+1,k-1))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell)+1,k-1)))}\right]}{E\left[\sum_{\ell=1}^{m} A_i^*(\ell,G_i(\ell)+1)\right]}, \quad |z|\leq 1.$$

Lemma 4.1 *(Expressions of $\widehat{\mathbf{q}}_i^{i*}(z)\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right)$ and $\widehat{\mathbf{q}}_i^{s*}(z)\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right)$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the vectors $\widehat{\mathbf{q}}_i^{i*}(z)$ and $\widehat{\mathbf{q}}_i^{s*}(z)$ can be expressed by means of the vector GFs of the stationary number of i-customers at different instants as*

$$a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z)\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right) = \widehat{\mathbf{m}}_i(z)-\widehat{\mathbf{f}}_i(z), \tag{4.45}$$

$$a_i^{s*}\widehat{\mathbf{q}}_i^{s*}(z)\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right) = (1-z)g_i\widehat{\mathbf{q}}_i^{d}(z)+\widehat{\mathbf{f}}_i(z)-\widehat{\mathbf{m}}_i(z). \tag{4.46}$$

Proof. The proof of the lemma can be found in the Appendix D.

Proposition 4.2 *(Relation between $\widehat{\mathbf{q}}_i^{*}(z)$ and $\widehat{\mathbf{q}}_i^{d}(z)$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds between the vector GF of the stationary number of i-customers at i-BMAP state change and i-customer departure epochs:*

$$\widehat{\mathbf{q}}_i^{*}(z)\frac{\left(-\mathbf{D}_{i,0}^{d}\right)^{-1}}{\left(\boldsymbol{\pi}_i^{*}(-\mathbf{D}_{i,0}^{d})^{-1}\mathbf{e}\right)}\widehat{\mathbf{D}}_i(z) \quad = \quad \lambda_i(z-1)\widehat{\mathbf{q}}_i^{d}(z). \tag{4.47}$$

.

Proof. By the help of (4.35) and the definitions of $\widehat{\mathbf{q}}_i^{s*}(z)$, a_i^{s*}, $\widehat{\mathbf{q}}_i^{i*}(z)$ and a_i^{i*} we rearrange the defining expression of $\widehat{\mathbf{q}}_i^{*}(z)$ as

$$\begin{aligned}
\widehat{\mathbf{q}}_i^{*}(z) = & \frac{\lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{m}}{\lim_{m\to\infty}\frac{m}{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1} A_i^*(\ell,n)\right]}} \\
= & \frac{1}{a_i^*}\lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{m} \\
& +\frac{1}{a_i^*}\lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)} z^{N_i(t_i(\ell,G_i(\ell)+1,k-1))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell)+1,k-1)))}\right]}{m} \\
= & \frac{1}{a_i^*}\left(a_i^{s*}\widehat{\mathbf{q}}_i^{s*}(z) + a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z)\right).
\end{aligned} \tag{4.48}$$

Multiplying (4.48) by $\left(\mathbf{I} - \widehat{\mathbf{\Psi}}_i(z)\right)$ from right and applying the statements of lemma 4.1 leads to

$$\begin{aligned}
\widehat{\mathbf{q}}_i^{*}(z)\left(\mathbf{I} - \widehat{\mathbf{\Psi}}_i(z)\right) &= \frac{1}{a_i^*}\left((1-z)g_i\widehat{\mathbf{q}}_i^{d}(z) + \widehat{\mathbf{f}}_i(z) - \widehat{\mathbf{m}}_i(z) + \widehat{\mathbf{m}}_i(z) - \widehat{\mathbf{f}}_i(z)\right) \\
&= \frac{1}{a_i^*}\left((1-z)g_i\widehat{\mathbf{q}}_i^{d}(z)\right).
\end{aligned} \tag{4.49}$$

Substituting the expressions of $\widehat{\mathbf{\Psi}}_i(z)$ and a_i^* ((2.5) and (4.40)) into (4.49) gives the proposition. □

4.5 Relations at i-customer arrival epochs

4.5.1 Vector GF of the stationary number of i-customers

In the following the individual arrival times of the i-customers in the same batch are handled separately. In order to allow this treatment we virtually separate the individual arriving times of the i-customers in the same batch by applying the Orderliness Convention:

Definition 4.1 Orderliness Convention *(Wolff [190] p.388.): If customers arrive in a batch, we suppose that they make a line to enter the station instantaneously one after another.*

Now the individual arriving times can be described as virtually different times. Let $t_i(\ell, n, k, o)$ the arrival time of the v-th i-customer arrived at the k-th i-BMAP state change during the service time of the n-th i-customer in the ℓ-th polling cycle, for $\ell \geq 1$, $n = 1, \ldots, G_i(\ell) + 1$, $k = 1, \ldots, A_i(\ell, n)$ and $o = 1, \ldots, A_i(\ell, n, k)$. By definition $t_i(\ell, n, k, 0) = N_i\left(t_i(\ell, n, k-1)\right)$. $n = G_i(\ell) + 1$ means again the i-intervisit time in the ℓ-th polling cycle.

Hence the number of i-customers seen by the o-th i-customer arrived in the same batch is the number of i-customers just before its arrival time is $N_i\left(t_i(\ell, n, k, o-1)\right)$ for $o = 1, \ldots, A_i(\ell, n, k)$. The phases of the i-BMAP seen by the o-th i-customer arrived in the same batch are the same and it equals the phase at the actual state change. In other words $J_i\left(t_i(\ell, n, k, o-1)\right) = J_i\left(t_i(\ell, n, k)\right)$ for $o = 1, , \ldots, A_i(\ell, n, k)$.

We define the $1 \times L$ vector $\widehat{\mathbf{q}}_i^a(z)$, the vector GF of the stationary number of i-customers seen by the arriving i-customers as

$$\widehat{\mathbf{q}}_i^a(z) = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} \sum_{n=1}^{G_i(\ell)+1} \sum_{k=1}^{A_i^*(\ell,n)} \sum_{o=1}^{A_i(\ell,n,k)} z^{N_i(t_i(\ell,n,k,o-1))} \mathbf{1}_{(J_i(t_i(\ell,n,k,o-1)))}\right]}{E\left[\sum_{\ell=1}^{m} \sum_{n=1}^{G_i(\ell)+1} \sum_{k=1}^{A_i^*(\ell,n)} A_i(\ell,n,k)\right]}, \qquad |z| \leq 1.$$

Proposition 4.3 *(Relation between $\widehat{\mathbf{q}}_i^a(z)$ and $\widehat{\mathbf{q}}_i^*(z)$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds between the vector GF of the stationary number of i-customers at i-customer arrival and i-BMAP state change epochs:*

$$\lambda_i(z-1)\widehat{\mathbf{q}}_i^a(z) = \widehat{\mathbf{q}}_i^*(z) \frac{\left(-\mathbf{D}_{i,0}^d\right)^{-1}}{\left(\boldsymbol{\pi}_i^*(-\mathbf{D}_{i,0}^d)^{-1}\mathbf{e}\right)} \left(\widehat{\mathbf{D}}_i(z) - \mathbf{D}_i\right). \tag{4.50}$$

Proof. We consider the defining expression of vector $\widehat{\mathbf{q}}_i^a(z)$. Applying $J_i\left(t_i(\ell, n, k, v-1)\right) = J_i\left(t_i(\ell, n, k)\right)$ for $v = 1, , \ldots, A_i(\ell, n, k)$ and using also (4.39) $\widehat{\mathbf{q}}_i^a(z)$ can be rewritten as

$$\widehat{\mathbf{q}}_i^a(z) = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)}\sum_{o=1}^{A_i(\ell,n,k)} z^{N_i(t_i(\ell,n,k,o-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}\right]}{m} \times \lim_{m\to\infty} \frac{m}{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} A_i(\ell,n,k)\right]} \tag{4.51}$$

$$= \frac{1}{a_i}\lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)}\sum_{o=1}^{A_i(\ell,n,k)} z^{N_i(t_i(\ell,n,k,o-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}\right]}{m}$$

The random variable $N_i\left(t_i(\ell,n,k,o-1)\right)$ represents the number of i-customers seen by the o-th i-customer arrived at k-th state change of the i-BMAP during the the service time of the n-th i-customer in the ℓ-th polling cycle. This consists of, on one hand, the number of the i-customers seen just before the actual i-BMAP state change and, on the other hand, the $(o-1)$ previously arrived i-customers in the same batch. In other words $N_i\left(t_i(\ell,n,k,o-1)\right) = N_i\left(t_i(\ell,n,k-1)\right) + o - 1$. Applying it in (4.51) and rearranging yields

$$\widehat{\mathbf{q}}_i^a(z) = \frac{1}{a_i}\lim_{m\to\infty}\frac{1}{m}E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}\sum_{o=1}^{A_i(\ell,n,k)} z^{o-1}\right]$$

$$= \frac{1}{a_i}\lim_{m\to\infty}\frac{1}{m}E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}\frac{1-z^{A_i(\ell,n,k)}}{1-z}\right].$$

Rearranging yields

$$\widehat{\mathbf{q}}_i^a(z)a_i(1-z) = \lim_{m\to\infty}\frac{1}{m}E\Big[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\sum_{j=1}^{L} \tag{4.52}$$

$$1_{(J_i(t_i(\ell,n,k-1)=j)}\left(1-z^{A_i(\ell,n,k)}\right)\mathbf{1}_{(J_i(t_i(\ell,n,k)))}\Big]$$

$$= \lim_{m\to\infty}\frac{1}{m}E\Big[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\sum_{j=1}^{L}1_{(J_i(t_i(\ell,n,k-1)=j)}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}\Big]$$

$$-\lim_{m\to\infty}\frac{1}{m}E\Big[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\sum_{j=1}^{L}1_{(J_i(t_i(\ell,n,k-1)=j)}$$

$$z^{A_i(\ell,n,k)}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}\Big].$$

Now we insert conditional expectations inside of the expectations in both terms on the r.h.s. of (4.52) by applying the same technique as in the proof of proposition 4.1 (from (4.42) to (4.43)). This leads to

$$\widehat{\mathbf{q}}_i^a(z)a_i(1-z) = \lim_{m\to\infty}\frac{1}{m}E\Bigg[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\sum_{j=1}^{L}1_{(J_i(t_i(\ell,n,k-1))=j)} \tag{4.53}$$
$$E\left[1^{A_i(\ell,n,k)}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}|J_i(t_i(\ell,n,k-1))=j, k\le A_i^*(\ell,n)\right]\Bigg]$$
$$-\lim_{m\to\infty}\frac{1}{m}E\Bigg[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\sum_{j=1}^{L}1_{(J_i(t_i(\ell,n,k-1))=j)}$$
$$E\left[z^{A_i(\ell,n,k)}\mathbf{1}_{(J_i(t_i(\ell,n,k)))}|J_i(t_i(\ell,n,k-1))=j, k\le A_i^*(\ell,n)\right]\Bigg]$$
$$=\lim_{m\to\infty}\frac{1}{m}E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\sum_{j=1}^{L}1_{(J_i(t_i(\ell,n,k-1))=j)}\mathbf{e}_j\left(\mathbf{\Psi}_i-\widehat{\mathbf{\Psi}}_i(z)\right)\right]$$
$$=\lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)+1}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{m}\left(\mathbf{\Psi}_i-\widehat{\mathbf{\Psi}}_i(z)\right).$$

Applying the defining expressions (4.36) and (4.35) in (4.53) gives

$$\widehat{\mathbf{q}}_i^a(z)a_i(1-z) = \widehat{\mathbf{q}}_i^*(z)\left(\mathbf{\Psi}_i-\widehat{\mathbf{\Psi}}_i(z)\right)a_i^*. \tag{4.54}$$

Substituting the expressions of $\widehat{\mathbf{\Psi}}_i(z)$ and a_i^* ((2.5) and (4.40)) into (4.54) results in

$$\widehat{\mathbf{q}}_i^a(z)\frac{a_i}{g_i}\lambda_i(1-z) = \widehat{\mathbf{q}}_i^*(z)\frac{\left(-\mathbf{D}_{i,0}^d\right)^{-1}}{\left(\boldsymbol{\pi}_i^*(-\mathbf{D}_{i,0}^d)^{-1}\mathbf{e}\right)}\left(\mathbf{D}_i-\widehat{\mathbf{D}}_i(z)\right). \tag{4.55}$$

The proposition can be obtained by applying $g_i = a_i$ in (4.55). □

Proposition 4.4 *(Relation between $\widehat{\mathbf{q}}_i(z)$ and $\widehat{\mathbf{q}}_i^*(z)$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds between the vector GF of the stationary number of i-customers at an arbitrary instant and just before the i-BMAP state change epochs:*

$$\widehat{\mathbf{q}}_i(z) = \widehat{\mathbf{q}}_i^*(z)\frac{\left(-\mathbf{D}_{i,0}^d\right)^{-1}}{\left(\boldsymbol{\pi}_i^*(-\mathbf{D}_{i,0}^d)^{-1}\mathbf{e}\right)}. \tag{4.56}$$

.

Proof. Applying proposition 4.2 in the stationary relationship (4.21) leads to

$$\widehat{\mathbf{q}}_i(z)\widehat{\mathbf{D}}_i(z) = \widehat{\mathbf{q}}_i^*(z)\frac{\left(-\mathbf{D}_{i,0}^d\right)^{-1}}{\left(\boldsymbol{\pi}_i^*(-\mathbf{D}_{i,0}^d)^{-1}\mathbf{e}\right)}\widehat{\mathbf{D}}_i(z). \tag{4.57}$$

.

For values of z, for which $\widehat{\mathbf{D}}_i(z)$ is invertible (4.57) implies the statement. $\widehat{\mathbf{q}}_i(z)$, $\widehat{\mathbf{q}}_i^*(z)$ and $\widehat{\mathbf{D}}_i(z)$ are probability-generating functions in term of z, therefore they are continuously differentiable for $|z| \leq 1$. It follows that the statement also holds for the values of z, where matrix $\widehat{\mathbf{D}}_i(z)$ is singular, like e.g. in case of $z = 1$, which completes the proof. □

Theorem 4.5 *(Relation between $\widehat{\mathbf{q}}_i^a(z)$ and $\widehat{\mathbf{q}}_i(z)$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds between the vector GF of the stationary number of i-customers at i-customer arrival epochs and at an arbitrary instant:*

$$\lambda_i(z-1)\widehat{\mathbf{q}}_i^a(z) = \widehat{\mathbf{q}}_i(z)\left(\widehat{\mathbf{D}}_i(z) - \mathbf{D}_i\right). \tag{4.58}$$

Proof. The corollary can be obtained by applying proposition 4.4 in proposition 4.3. □

Remark 4.3 *For an alternative derivation of (4.58) we refer to Kim Chae and Lee [78]. They used "rate in = rate out" arguments to obtain this relation.*

4.5.2 The factorial moments of the stationary number of i-customers

Let $\mathbf{q}_i^{a(n)}$ denote the n-th ($n \geq 1$) factorial moment of $\widehat{\mathbf{q}}_i^a(z)$, i.e., $\mathbf{q}_i^{a(n)} = \frac{d^n}{dz^n}\widehat{\mathbf{q}}_i^a(z)|_{z=1}$ and $\mathbf{q}_i^a$ and $\mathbf{q}_i^{a(0)}$ denote its value at $z = 1$, i.e., $\mathbf{q}_i^a = \mathbf{q}_i^{a(0)} = \widehat{\mathbf{q}}_i^a(1)$.

We establish a formula for the vector factorial moments of the stationary number of i-customers at i-customer arrival epochs ($\mathbf{q}_i^{a(n)}$, $n \geq 1$) by utilizing the relation between $\widehat{\mathbf{q}}_i^a(z)$ and $\widehat{\mathbf{q}}_i(z)$. Thus the vector factorial moments $\mathbf{q}_i^{a(n)}$, for $n \geq 1$ can be calculated from an appropriate set of $\mathbf{q}_i^{(n)}$-s.

Corollary 4.3 *(Relation between $\mathbf{q}_i^{a(n)}$, $n \geq 1$ and $\mathbf{q}_i^{(n)}$, $n \geq 1$.) In the stable BMAP/G/1 cyclic polling model satisfying assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4** *the following relation holds between the vector factorial moments of the stationary number of i-customers at i-customer arrival epochs and at an arbitrary instant:*

$$\mathbf{q}_i^{a(n)} = \frac{1}{\lambda_i}\sum_{k=1}^{n+1}\binom{n+1}{k}\mathbf{q}_i^{(n+1-k)}\mathbf{D}_i^k, \quad n \geq 0. \tag{4.59}$$

Proof. Taking the n-th derivative of the stationary relationship (4.58), for $n \geq 1$, yields

$$\lambda_i \mathbf{q}_i^{a(n-1)} = \sum_{k=0}^{n} \binom{n}{k} \mathbf{q}_i^{(n-k)} \mathbf{D}_i^k - \mathbf{q}_i^{(n)} \mathbf{D}_i, \quad n \geq 1. \tag{4.60}$$

Rearranging (4.60) gives the statement. □

4.6 Validity scope of results

The proofs of the service discipline independent statements (all lemmas, propositions, theorems and corollaries in this chapter) did not utilize the independency of the switchover times according to assumption **A.3** and properties **P.3** and **P.4** of the model. Furthermore among the service discipline independent statements, the mutual independency of the arrival processes and the service times (part of assumption **A.3**) is used only for the proof of theorem 4.2 and as a consequence for the proof of the statements using this theorem, i.e. theorems 4.3 and 4.4.

Moreover all the service discipline independent results presented in this chapter are also valid for the zero-switchover-times counterpart of the considered model. In this model the server off periods, while the system is empty, can be taken into account e.g. by applying the argument presented in [202].

Hence the service discipline independent statements hold also under more general settings.

Remark 4.4 *(Extension to more general polling systems.) By the proper redefinition of the cycle time and the other relevant quantities the argument of the fundamental relation (theorem 4.1) can be applied also to more general polling systems, e.g. to periodic polling model or to polling model with Markovian server routing. Hence the stationary relationships based on the fundamental relation (corollary 4.1 and theorems 4.2, 4.3 and 4.4) can be extended also to these more general polling systems.*

Remark 4.5 *(Extension to handle set-up time or repair time.) The presented model can be extended to handle also other quantities like e.g., set-up time or repair time by relaxing the service discipline properties* **P.1**, **P.2** *and* **P.3**.

Chapter 5

Gated and exhaustive service system

In this chapter the determination of the vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs are discussed in the gated and exhaustive service systems. These are the quantities, which are needed for the application of the discipline independent results to these models.

By using Kronecker product notation we generalize the buffer occupancy method for the cyclic polling model with BMAP. This leads to the governing equations of the system in terms of joint PGFs of the stationary number of customers and the phases of the BMAPs at i-polling and i-departure epochs.

Different to the classical polling model with disciplines satisfying the branching property, like e.g., the exhaustive and gated ones no closed-form system of linear equations can be derived for the factorial moments of the stationary number of i-customers at i-polling and i-departure epochs for this model. This is due to the complexity introduced by BMAP. Instead a system of linear equations are derived for the joint probabilities of the stationary number of customers and the phases of the BMAPs at i-polling and i-departure epochs by taking the appropriate derivatives of the governing equations of the system. Apart from the higher complexity of the resulted system of linear equations this solution can be applied also to disciplines not satisfying the branching property, like e.g. the G-limited policy. After solving numerically the governing equations, the required vector factorial moments at i-polling and i-departure epochs are computed from the joint probabilities.

The last part of the chapter deals with the detailed discussion of the steps of the numerical solution. Besides of common steps for the cases of both disciplines generally the solution of exhaustive discipline is numerically more complex and it requires more computational steps. One of them is the computation of the LST of the time dependent first passage matrix at station i. This is performed by means of a newly introduced algorithm, which is based on the concept of uniformization. The convergence of the algorithm is also shown.

We remark here that the analysis and results presented in this chapter as well as the major part of the discussion on the numerical solution (in Section 5.5) have been published in the journal paper [199].

The joint probabilities of the stationary number of customers and the phases of the BMAPs at i-polling and i-departure epochs are described as hypervectors. Notation $\otimes$ stands the Kronecker product. The $1 \times L^N$ stationary probability hypervector $\mathbf{p}_i^f(n_1, \ldots, n_N)$ is defined as

$$\mathbf{p}_i^f(n_1, \ldots, n_N) = \lim_{m\to\infty} \sum_{j_1=1}^{L} \ldots \sum_{j_N=1}^{L} \mathbf{e}_{j_1} \otimes \ldots \otimes \mathbf{e}_{j_N}$$
$$Pr\{N_1(t_i^f(m)) = n_1, \ldots, N_N(t_i^f(m)) = n_N, J_1(t_i^f(m)) = j_1, \ldots, J_N(t_i^f(m)) = j_N\},$$
$$n_1, \ldots, n_N \in \{0, 1, \ldots\}; \quad i = 1, \ldots, N.$$

Similarly the $1 \times L^N$ stationary probability hypervector $\mathbf{p}_i^m(n_1, \ldots, n_N)$ is defined as:

$$\mathbf{p}_i^m(n_1, \ldots, n_N) = \lim_{m\to\infty} \sum_{j_1=1}^{L} \ldots \sum_{j_N=1}^{L} \mathbf{e}_{j_1} \otimes \ldots \otimes \mathbf{e}_{j_N}$$
$$Pr\{N_1(t_i^m(m)) = n_1, \ldots, N_N(t_i^m(m)) = n_N, J_1(t_i^m(m)) = j_1, \ldots, J_N(t_i^m(m)) = j_N\},$$
$$n_1, \ldots, n_N \in \{0, 1, \ldots\}; \quad i = 1, \ldots, N.$$

Based on these quantities we define the hypervector GFs of the stationary number of customers at i-polling and i-departure epochs as

$$\widehat{\mathbf{f}}_i(z_1, \ldots, z_N) = \sum_{n_1=0}^{\infty} \ldots \sum_{n_N=0}^{\infty} \mathbf{p}_i^f(n_1, \ldots, n_N)\, z_1^{n_1} \ldots z_N^{n_N}, \tag{5.1}$$
$$\widehat{\mathbf{m}}_i(z_1, \ldots, z_N) = \sum_{n_1=0}^{\infty} \ldots \sum_{n_N=0}^{\infty} \mathbf{p}_i^m(n_1, \ldots, n_N)\, z_1^{n_1} \ldots z_N^{n_N}, \tag{5.2}$$
$$i = 1, \ldots, N; \quad |z_1| \le 1, \ldots, |z_N| \le 1.$$

We remark here that the former definitions of the quantities $\widehat{\mathbf{f}}_i(z_i)$ and $\widehat{\mathbf{m}}_i(z_i)$ in Section 4.1 are slightly different from the marginal vector GFs $\widehat{\mathbf{f}}_i(1, \ldots, z_i, \ldots, 1)\, \mathbf{e} \otimes \ldots \otimes \mathbf{I} \otimes \ldots \otimes \mathbf{e}$ and $\widehat{\mathbf{m}}_i(1, \ldots, z_i, \ldots, 1)\, \mathbf{e} \otimes \ldots \otimes \mathbf{I} \otimes \ldots \otimes \mathbf{e}$, in which $\mathbf{I}$ is at the i-th position as they follow from the above definitions of the hypervector GFs. In spite of it $\widehat{\mathbf{f}}_i(z_i) = \widehat{\mathbf{f}}_i(1, \ldots, z_i, \ldots, 1)\, \mathbf{e} \otimes \ldots \otimes \mathbf{I} \otimes \ldots \otimes \mathbf{e}$ and $\widehat{\mathbf{m}}_i(z_i) = \widehat{\mathbf{m}}_i(1, \ldots, z_i, \ldots, 1)\, \mathbf{e} \otimes \ldots \otimes \mathbf{I} \otimes \ldots \otimes \mathbf{e}$ due to the finite memory properties of the Markov chains describing the number of customers and the phases of the BMAPs at the embedded i-polling and i-departure epochs, respectively (see in Appendix C).

We define a notation also for substituting an $L^N \times L^N$ hypermatrix A into the defining series of the $1 \times L^N$ hypervector GF $\widehat{\mathbf{f}}_i(z_1, \ldots, z_N)$:

$$\begin{aligned}\widehat{\mathbf{f}}_i(z_1, \ldots, z_{i-1}, A, z_{i+1}, \ldots, z_N) \quad = \quad & \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^f\,(n_1, \ldots, n_N) \\ & z_1^{n_1} \ldots z_{i-1}^{n_{i-1}} A^{n_i} z_{i+1}^{n_{i+1}} \ldots z_N^{n_N}, \;\; i = 1, \ldots, N,\end{aligned}$$

This results in also an $1 \times L^N$ hypervector.

We use notation $\oplus$ for the Kronecker sum and $\oplus_{k=1}^N \widehat{\mathbf{D}}_k(z_k)$ stands for $\widehat{\mathbf{D}}_1(z_1) \oplus \ldots \oplus \widehat{\mathbf{D}}_N(z_N)$. Additionally we introduce one further notation as follows:

$$\widehat{\mathbf{A}}_i(z_1, \ldots, z_N) = \int_0^\infty e^{t \oplus_{k=1}^N \widehat{\mathbf{D}}_k(z_k)} dB_i(t),$$

$$\widehat{\mathbf{U}}_i(z_1, \ldots, z_N) = \int_0^\infty e^{t \oplus_{k=1}^N \widehat{\mathbf{D}}_k(z_k)} dR_i(t). \tag{5.3}$$

Note that both $\widehat{\mathbf{A}}_i(z_1, \ldots, z_N)$ and $\widehat{\mathbf{U}}_i(z_1, \ldots, z_N)$ are $L^N \times L^N$ hypermatrices.

From now on the station index i is to be understood as modulo N, i.e. for $i = N$ station $i + 1$ means station 1.

5.1 Gated service system

5.1.1 Governing equations of the system

Theorem 5.1 *(Governing equations of the system.) The governing equations of the stable BMAP/G/1 cyclic nonzero-switchover-times polling model with gated service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4**, *are given in terms of the hypervector GFs* $\widehat{\mathbf{f}}_i(z_1, \ldots, z_N)$ *and* $\widehat{\mathbf{m}}_i(z_1, \ldots, z_N)$ *for* $i = 1, \ldots, N$ *as*

$$\begin{aligned}\widehat{\mathbf{m}}_i\,(z_1, \ldots, z_N) &= \widehat{\mathbf{f}}_i \left(z_1, \ldots, z_{i-1}, \widehat{\mathbf{A}}_i(z_1, \ldots, z_N), z_{i+1}, \ldots, z_N\right), \\ \widehat{\mathbf{f}}_{i+1}\,(z_1, \ldots, z_N) &= \widehat{\mathbf{m}}_i\,(z_1, \ldots, z_N)\,\widehat{\mathbf{U}}_i(z_1, \ldots, z_N).\end{aligned} \tag{5.4}$$

Proof. We generalize the buffer occupancy method (see e.g. in Cooper and Murray [39], Cooper [38]) for our model with BMAPs. Under gated discipline only those i-customers are served during the service of station i, which already present at i-polling epoch. Hence those i-customers present at i-departure epoch who arrived during the service of station i. The matrix GF of the number of i-customers arriving during a service of one i-customer is given as $\widehat{\mathbf{A}}_i(z_i)$. Assuming, that the number of i-customers present at i-polling epoch is n_i, the matrix GF of the number of i-customers present at i-departure epoch can be expressed by:

$$\left(\widehat{\mathbf{A}}_i(z_i)\right)^{n_i} = \left(\int_0^\infty e^{\widehat{\mathbf{D}}_i(z_i)t} dB_i(t)\right)^{n_i}.$$

To describe the evolution of the k-customers ($k \neq i$), we have to take into account also the number of k-customers present at the i-polling epoch. The number of k-customers ($k \neq i$) at i-departure epoch equals the number of k-customers present at the i-polling epoch plus the number of k-customers arriving during the service of station i, which are independent once the number of i-customers and k-customers present at the i-polling epoch are given. Hence assuming, that the number of k-customers present at i-polling epoch is n_k, and the number of i-customers present at i-polling epoch is n_i, the matrix GF of the number of k-customers present at i-departure epoch can be expressed by:

$$z_k^{n_k} \left(\int_0^\infty e^{\widehat{\mathbf{D}}_k(z_k)t} dB_i(t)\right)^{n_i}.$$

To describe the evolution of both the i-customers and the k-customers ($k \neq i$) at the same time, we need the hypermatrix GF of the number of i-customers and k-customers, which arrive during the service of one i-customer.

To this end we need the following property for the Kronecker product of exponential functions of matrices:

$$e^{\widehat{\mathbf{D}}_i(z_i)} \otimes e^{\widehat{\mathbf{D}}_k(z_k)} = e^{\widehat{\mathbf{D}}_i(z_i)\otimes\mathbf{I}+\mathbf{I}\otimes\widehat{\mathbf{D}}_k(z_k)} = e^{\widehat{\mathbf{D}}_i(z_i)\oplus\widehat{\mathbf{D}}_k(z_k)}, \tag{5.5}$$

Using (5.5) we can express the hypermatrix GF of the number of i-customers and k-customers, which arrive during the service of one i-customer by

$$\begin{aligned} \int_0^\infty \left(e^{\widehat{\mathbf{D}}_i(z)t} \otimes e^{\widehat{\mathbf{D}}_k(z)t}\right) dB_i(t) &= \int_0^\infty e^{\left(\widehat{\mathbf{D}}_i(z_i)\otimes\mathbf{I}+\mathbf{I}\otimes\widehat{\mathbf{D}}_k(z_k)\right)t} dB_i(t) \\ = \int_0^\infty e^{\left(\widehat{\mathbf{D}}_i(z_i)\oplus\widehat{\mathbf{D}}_k(z_k)\right)t} dB_i(t) &= \widehat{\mathbf{A}}_i(z_i, z_k). \end{aligned} \tag{5.6}$$

Assuming that the number of k-customers present at i-polling epoch is n_k, and the number of i-customers present at i-polling epoch is n_i, using again the previous arguments of the buffer occupancy method for the evolution of both the i-customers and k-customers at the same time and applying (5.6) results in the GF of the number of i-customers and k-customers present at i-departure epoch as

$$z_k^{n_k}\left(\widehat{\mathbf{A}}_i(z_i,z_k)\right)^{n_i}.$$

Repeating the same arguments now for the i-customers and for every k-customers ($k \neq i$) at the same time, after unconditioning we get the relation for the transition $f_i \to m_i$ of the gated polling model with N stations :

$$\begin{aligned}&\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}\mathbf{p}_i^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_{i-1}^{n_{i-1}}\left(\widehat{\mathbf{A}}_i(z_1,\dots,z_N)\right)^{n_i}z_{i+1}^{n_{i+1}}\dots z_N^{n_N}\\ &=\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}\mathbf{p}_i^m(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N},\quad i=1,\dots,N.\end{aligned} \tag{5.7}$$

Using similar arguments as before, for transition $m_i \to f_{i+1}$ of the nonzero-switchover-times gated polling model with N stations we get

$$\begin{aligned}&\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}\mathbf{p}_i^m(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N}\widehat{\mathbf{U}}_i(z_1,\dots,z_N)\\ &=\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}\mathbf{p}_{i+1}^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N},\quad i=1,\dots,N.\end{aligned} \tag{5.8}$$

Applying notation $\widehat{\mathbf{f}}_i(z_1,\dots,z_i-1,A,z_{i+1},\dots,z_N)$ and the defining equations of the hypervector GFs $\widehat{\mathbf{f}}_i(z_1,\dots,z_N)$ and $\widehat{\mathbf{m}}_i(z_1,\dots,z_N)$ in (5.7) and (5.8), after rearranging we get the stated equations as compact forms of the governing relations of the system.
□

5.1.2 Vector moments of the stationary number of i-customers at i-polling and i-departure epochs

To compute the vector moments of the stationary number of i-customers at i-polling and i-departure epochs, for $i = 1,\dots,N$, we need the $1 \times L$ marginal stationary probability vectors

$$\begin{aligned}\mathbf{p}_i^f(n_i)&=\sum_{n_1=0}^{\infty}\dots\sum_{n_{i-1}=0}^{\infty}\sum_{n_{i+1}=0}^{\infty}\dots\sum_{n_N=0}^{\infty}\mathbf{p}_i^f(n_1,\dots,n_N)\,\mathbf{e}\otimes\dots\otimes\mathbf{I}\otimes\dots\otimes\mathbf{e},\\ \mathbf{p}_i^m(n_i)&=\sum_{n_1=0}^{\infty}\dots\sum_{n_{i-1}=0}^{\infty}\sum_{n_{i+1}=0}^{\infty}\dots\sum_{n_N=0}^{\infty}\mathbf{p}_i^m(n_1,\dots,n_N)\,\mathbf{e}\otimes\dots\otimes\mathbf{I}\otimes\dots\otimes\mathbf{e},\end{aligned}$$

where **I** is at the i-th position. However the structure of (5.7) and (5.8) does not allow the direct determination of these quantities. Instead we can determine the stationary probability hypervectors $\mathbf{p}_i^f(n_1,\ldots,n_N)$ and $\mathbf{p}_i^m(n_1,\ldots,n_N)$ for $i=1,\ldots,N$.

We set an upper limit X for $n_1,\ldots,n_N$ in (5.7) and (5.8). Then we take their x_1-th, …, x_N-th derivatives ($x_1,\ldots,x_N \in \{0,\ldots,X\}$) at $z_1=\ldots=z_N=1$, respectively. This results in the following system of linear equations for $i=1,\ldots,N$ and $x_1,\ldots,x_N \in \{0,\ldots,X\}$:

$$\begin{aligned}&\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^f(n_1,\ldots,n_N)\\&\times\left.\frac{d^{x_1}\ldots d^{x_N}\left(z_1^{n_1}\ldots z_{i-1}^{n_{i-1}}\left(\widehat{\mathbf{A}}_i(z_1,\ldots,z_N)\right)^{n_i}z_{i+1}^{n_{i+1}}\ldots z_N^{n_N}\right)}{d{z_1}^{x_1}\ldots d{z_N}^{x_N}}\right|_{\mathbf{z}=1}\\&=\sum_{n_1=x_1}^{X}\cdots\sum_{n_N=x_N}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\frac{n_1!}{(n_1-x_1)!}\cdots\frac{n_N!}{(n_N-x_N)!},\end{aligned} \tag{5.9}$$

$$\begin{aligned}&\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\left.\frac{d^{x_1}\ldots d^{x_N}\left(z_1^{n_1}\ldots z_N^{n_N}\widehat{\mathbf{U}}_i(z_1,\ldots,z_N)\right)}{d{z_1}^{x_1}\ldots d{z_N}^{x_N}}\right|_{\mathbf{z}=1}\\&=\sum_{n_1=x_1}^{X}\cdots\sum_{n_N=x_N}^{X}\mathbf{p}_{i+1}^f(n_1,\ldots,n_N)\frac{n_1!}{(n_1-x_1)!}\cdots\frac{n_N!}{(n_N-x_N)!},\end{aligned} \tag{5.10}$$

where $\mathbf{z}=1$ stands for $z_1=\ldots=z_N=1$.

These linear equations relate quantities, which are essentially close to the factorial moments of the stationary number of customers at polling and departure epochs. Based on this system of linear equations a numerical method can be developed for computing the stationary probability hypervectors $\mathbf{p}_i^f(n_1,\ldots,n_N)$ and $\mathbf{p}_i^m(n_1,\ldots,n_N)$ for $i=1,\ldots,N$. In a basic realization X is increased until $\left(1-\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^f(n_1,\ldots,n_N)\right)$ or/and $\left(1-\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\right)$ becomes less than the allowed error according to the required precision.

The number of equations and the number of unknowns in the system of linear equations ((5.9) and (5.10)) is $2NL^N(X+1)^N$.

5.2 Exhaustive service system

Before the analysis we define matrix $\mathbf{G}_i(t)$, since we use it for the description of the polling model with disciplines having exhaustive characteristic.

We define the homogenous bivariate Markov chain $\left\{(N_i(t_i^d(n)),J_i(t_i^d(n)));n\in\{1,\ldots\}\right\}$ on the state space $\{0,1,\ldots\}\times\{1,2,\ldots,L\}$, where

$t_i^d(n)$ denotes the n-th i-customer departure epoch during the same server visit at station i for $n \geq 1$. We define matrix $\mathbf{G}_i(t)$, $t \geq 0$, whose (j,ℓ)-th element is given as the probability that the first passage starting from state $(n+1, j)$ in the Markov chain to the level n, $n \in 0, 1, 2, \ldots$, $1 \leq j, \ell \leq L$, occurs no later than time t, and the first state visited in level n is (n, ℓ).

Additionally we introduce the notation

$$\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N) = \int_0^\infty e^{t \oplus_{k=1}^{i-1} \widehat{\mathbf{D}}_k(z_k)} \otimes d\mathbf{G}_i(t) \otimes e^{t \oplus_{k=i+1}^{N} \widehat{\mathbf{D}}_k(z_k)}.$$

Note that $\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)$ is an $L^N \times L^N$ hypermatrix.

5.2.1 Governing equations of the system

Theorem 5.2 *(Governing equations of the system.) The governing equations of the stable BMAP/G/1 cyclic nonzero-switchover-times polling model with exhaustive service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4***, are given in terms of the hypervector GFs* $\widehat{\mathbf{f}}_i(z_1, \ldots, z_N)$ *and* $\widehat{\mathbf{m}}_i(z_1, \ldots, z_N)$ *for* $i = 1, \ldots, N$ *as*

$$\begin{aligned} &\widehat{\mathbf{m}}_i\left(z_1, \ldots, z_{i-1}, 1, z_{i+1}, \ldots, z_N\right) \\ &= \widehat{\mathbf{f}}_i\left(z_1, \ldots, z_{i-1}, \widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N), z_{i+1}, \ldots, z_N\right), \\ &\widehat{\mathbf{f}}_{i+1}\left(z_1, \ldots, z_N\right) = \widehat{\mathbf{m}}_i\left(z_1, \ldots, z_{i-1}, 1, z_{i+1}, \ldots, z_N\right) \widehat{\mathbf{U}}_i(z_1, \ldots, z_N). \end{aligned} \quad (5.11)$$

Proof. The transition $m_i \to f_{i+1}$ is discipline independent. However in case of exhaustive discipline $n_i = 0$ at i-departure epochs. Applying it to the second part of (5.4) results in the second relation of (5.11).

To describe the transition $f_i \to m_i$ we apply a similar line of arguments as we used for the model with gated discipline. Concerning the evolution of i-customers only the evolution of the phase of i-th BMAP arises, as due to the exhaustive discipline the number of i-customers is 0 at i-departure epoch.

To describe the evolution of both the phase of i-th BMAP and the k-customers $(k \neq i)$ at the same time, we need the hypermatrix transform describing the phase of i-th BMAP and the number of k-customers, which arrive during the first passage time at station i. For the sake of notation simplicity we assume here that $k < i$. This transform is given by

$$\int_0^\infty e^{\widehat{\mathbf{D}}_k(z_k)t} \otimes d\mathbf{G}_i(t).$$

Assuming that the number of k-customers present at i-polling epoch is n_k, and the number of i-customers present at i-polling epoch is n_i, using again the arguments of the buffer occupancy method for the evolution of both the phase of i-th BMAP and k-customers at the same time results in the phase of i-th BMAP and k-customers present at i-departure epoch as

$$z_k^{n_k}\left(\widehat{\mathbf{H}}_i(z_k)\right)^{n_i}.$$

Repeating the same arguments now for the phase of i-th BMAP and for every k-customers ($k \neq i$) at the same time, after unconditioning we get the relation for the transition $f_i \to m_i$ of the exhaustive polling model with N stations:

$$\sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}\mathbf{p}_i^f(n_1,\ldots,n_N)\, z_1^{n_1}\ldots z_{i-1}^{n_{i-1}}\left(\widehat{\mathbf{H}}_i(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_N)\right)^{n_i} z_{i+1}^{n_{i+1}}\ldots z_N^{n_N}$$
$$=\sum_{n_1=0}^{\infty}\cdots\sum_{n_{i-1}=0}^{\infty}\sum_{n_{i+1}=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}\mathbf{p}_i^m(n_1,\ldots,n_{i-1},0,n_{i+1},\ldots,n_N)$$
$$z_1^{n_1}\ldots z_{i-1}^{n_{i-1}} z_{i+1}^{n_{i+1}}\ldots z_N^{n_N},\quad i=1,\ldots,N. \tag{5.12}$$

Applying notation $\widehat{\mathbf{f}}_i(z_1,\ldots,z_i-1,A,z_{i+1},\ldots,z_N)$ and the defining equations of the hypervector GFs $\widehat{\mathbf{f}}_i(z_1,\ldots,z_N)$ and $\widehat{\mathbf{m}}_i(z_1,\ldots,z_N)$ in (5.12), after rearranging we get the first governing relation of the system. □

5.2.2 Vector moments of the stationary number of i-customers at i-polling and i-departure epochs

To compute the vector moments of the stationary number of i-customers at i-polling and i-departure epochs we follow the same way as before for the gated discipline. We set an upper limit X for $n_1,\ldots,n_N$ in (5.12) and (5.8). Then we take their x_1-th, …, x_N-th derivatives ($x_1,\ldots,x_N \in \{0,\ldots,X\}$) at $z_1=\ldots=z_N=1$, respectively. This results in the following system of linear equations for $i=1,\ldots,N$ and $x_1,\ldots,x_N\in\{0,\ldots,X\}$:

$$\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^f(n_1,\ldots,n_N)\frac{d^{x_1}\ldots d^{x_{i-1}}d^{x_{i+1}}\ldots d^{x_N}}{dz_1{}^{x_1}\ldots dz_{i-1}{}^{x_{i-1}}\; dz_{i+1}{}^{x_{i+1}}\ldots dz_N{}^{x_N}}$$
$$\times\left(z_1^{n_1}\ldots z_{i-1}^{n_{i-1}}\left(\widehat{\mathbf{H}}_i(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_N)\right)^{n_i} z_{i+1}^{n_{i+1}}\ldots z_N^{n_N}\right)\Big|_{\mathbf{z}=1} \tag{5.13}$$
$$=\sum_{n_1=x_1}^{X}\cdots\sum_{n_{i-1}=x_{i-1}}^{X}\sum_{n_{i+1}=x_{i-1}}^{X}\cdots\sum_{n_N=x_N}^{X}\mathbf{p}_i^m(n_1,\ldots,n_{i-1},0,n_{i+1},\ldots,n_N)$$
$$\times\frac{n_1!}{(n_1-x_1)!}\cdots\frac{n_{i-1}!}{(n_{i-1}-x_{i-1})!}\frac{n_{i+1}!}{(n_{i+1}-x_{i+1})!}\cdots\frac{n_N!}{(n_N-x_N)!},$$

$$\sum_{n_1=0}^{X} \cdots \sum_{n_{i-1}=0}^{X} \sum_{n_{i+1}=0}^{X} \cdots \sum_{n_N=0}^{X} \mathbf{p}_i^m (n_1, \ldots, n_{i-1}, 0, n_{i+1}, \ldots, n_N)$$
$$\times \left. \frac{d^{x_1} \ldots d^{x_N} \left(z_1^{n_1} \ldots z_{i-1}^{n_{i-1}} z_{i+1}^{n_{i+1}} \ldots z_N^{n_N} \widehat{\mathbf{U}}_i(z_1, \ldots, z_N) \right)}{dz_1{}^{x_1} \ldots dz_N{}^{x_N}} \right|_{\mathbf{z}=1} \tag{5.14}$$
$$= \sum_{n_1=x_1}^{X} \cdots \sum_{n_N=x_N}^{X} \mathbf{p}_{i+1}^f (n_1, \ldots, n_N) \frac{n_1!}{(n_1 - x_1)!} \cdots \frac{n_N!}{(n_N - x_N)!}.$$

The number of equations and the number of unknowns in the system of linear equations ((5.13) and (5.14)) is $NL^N \left((X+1)^N + (X+1)^{N-1} \right)$.

5.3 Symmetric system

In the symmetric system $\widehat{\mathbf{D}}_i(z)$, B_i and R_i are the same for $i = 1, \ldots, N$ and therefore the behavior of each station is the same. Hence the set of unknowns $\mathbf{p}_i^f (n_1, \ldots, n_N)$ and $\mathbf{p}_i^m (n_1, \ldots, n_N)$ are the same for each station i. Similarly, the set of equations (5.9) and (5.10) as well as (5.13) and (5.14) for $x_1, \ldots, x_N = 0, \ldots, X$ are also the same for each station i, respectively. It follows, that the sizes of the linear systems of equations reduces to $2L^N (X+1)^N$ for the system with gated discipline and to $L^N \left((X+1)^N + (X+1)^{N-1} \right)$ for the system with exhaustive discipline.

5.4 Numerical considerations

5.4.1 Rearrangement of the equations

In order to further simplify the equations (5.9) and (5.10) (as well as (5.13) and (5.14) for the system with exhaustive discipline) the number of equations and the number of unknowns are reduced by eliminating the quantities $\mathbf{p}_i^f (n_1, \ldots, n_N)$ for $i = 1, \ldots, N$ from them. This results in a system of linear equations for stationary probability hypervectors $\mathbf{p}_i^m (n_1, \ldots, n_N)$, which has a form $\mathbf{p}_i^m (n_1, \ldots, n_N) \to \mathbf{p}_{i+1}^m (n_1, \ldots, n_N)$ for $i = 1, \ldots, N$ and $x_1, \ldots, x_N \in \{0, \ldots, X\}$. To make this system of linear equations tractable it is arranged into a hypermatrix form. We show this rearrangement for the system with gated discipline. Starting from equations (5.13) and (5.14) the same rearrangements can be also done for the system with exhaustive discipline.

We define the hypermatrix $\mathbf{R}_i(l_1, \ldots, l_N)$ for $l_1, \ldots, l_N \in \{0, 1, \ldots\}$ as

$$\mathbf{R}_i(l_1, \ldots, l_N) = \int_0^\infty \mathbf{P}_1(l_1, t) \otimes \ldots \otimes \mathbf{P}_N(l_N, t) \; dR_i(t).$$

$\mathbf{R}_i(l_1, \ldots, l_N)$ is interpreted as the transition probability hypermatrix of the number of simultaneously arriving k-customers for every $k = 1, \ldots, N$ during the switchover time R_i when the number of simultaneously arriving $k_1, \ldots, k_N$-customers are exactly $l_1, \ldots, l_N$, respectively.

Utilizing the convolution form of the service discipline independent relation for transition $m_i \to f_{i+1}$ (5.8), it can be rearranged as

$$\mathbf{p}_{i+1}^{f}(n_1, \ldots, n_N) = \sum_{k_1=0}^{n_1} \ldots \sum_{k_N=0}^{n_N} \mathbf{p}_i^m(k_1, \ldots, k_N)\, \mathbf{R}_i(n_1 - k_1, \ldots, n_N - k_N). \quad (5.15)$$

Applying (5.15) in equation (5.9) with station index $i+1$ leads to

$$\sum_{n_1=0}^{X} \ldots \sum_{n_N=0}^{X} \sum_{k_1=0}^{n_1} \ldots \sum_{k_N=0}^{n_N} \mathbf{p}_i^m(k_1, \ldots, k_N)\, \mathbf{R}_i(n_1 - k_1, \ldots, n_N - k_N)$$
$$\times \left. \frac{d^{x_1} \ldots d^{x_N} \left(z_1^{n_1} \ldots z_i^{n_i} \left(\widehat{\mathbf{A}}_{i+1}(z_1, \ldots, z_N) \right)^{n_{i+1}} z_{i+2}^{n_{i+2}} \ldots z_N^{n_N} \right)}{dz_1{}^{x_1} \ldots dz_N{}^{x_N}} \right|_{\mathbf{z}=1}$$
$$= \sum_{n_1=x_1}^{X} \ldots \sum_{n_N=x_N}^{X} \mathbf{p}_{i+1}^m(n_1, \ldots, n_N) \frac{n_1!}{(n_1 - x_1)!} \cdots \frac{n_N!}{(n_N - x_N)!}. \quad (5.16)$$

We remark here that an appropriate value of X depends on the required precision level, at which the probabilities $\mathbf{p}_i^m(k_1, \ldots, k_N)$ can be neglected for $k_i > X$ at least for one $i = 1, \ldots, N$. In the following these probabilities are set 0.

Rearranging (5.16) yields

$$\sum_{k_1=0}^{X} \ldots \sum_{k_N=0}^{X} \sum_{n_1-k_1=0}^{X-k_1} \ldots \sum_{n_N-k_N=0}^{X-k_N} \mathbf{p}_i^m(k_1, \ldots, k_N)\, \mathbf{R}_i(n_1 - k_1, \ldots, n_N - k_N)$$
$$\times \left. \frac{d^{x_1} \ldots d^{x_N} \left(z_1^{n_1} \ldots z_i^{n_i} \left(\widehat{\mathbf{A}}_{i+1}(z_1, \ldots, z_N) \right)^{n_{i+1}} z_{i+2}^{n_{i+2}} \ldots z_N^{n_N} \right)}{dz_1{}^{x_1} \ldots dz_N{}^{x_N}} \right|_{\mathbf{z}=1}$$
$$= \sum_{n_1=x_1}^{X} \ldots \sum_{n_N=x_N}^{X} \mathbf{p}_{i+1}^m(n_1, \ldots, n_N) \frac{n_1!}{(n_1 - x_1)!} \cdots \frac{n_N!}{(n_N - x_N)!}. \quad (5.17)$$

Replacing the indices k_i by n_i and $n_i - k_i$ by k_i for $i = 1, \ldots, N$ in the l.h.s. of (5.17) results in the $\mathbf{p}_i^m(n_1, \ldots, n_N) \to \mathbf{p}_{i+1}^m(n_1, \ldots, n_N)$ relation for the system with gated discipline for $i = 1, \ldots, N$ and $x_1, \ldots, x_N \in \{0, \ldots, X\}$ as

$$\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\sum_{k_1=0}^{X-n_1}\cdots\sum_{k_N=0}^{X-n_N}\mathbf{R}_i(k_1,\ldots,k_N)$$
$$\times\left.\frac{d^{x_1}\ldots d^{x_N}\left(z_1^{n_1+k_1}\ldots z_i^{n_i+k_i}\left(\widehat{\mathbf{A}}_{i+1}(z_1,\ldots,z_N)\right)^{n_{i+1}+k_{i+1}}z_{i+2}^{n_{i+2}+k_{i+2}}\ldots z_N^{n_N+k_N}\right)}{d{z_1}^{x_1}\ldots d{z_N}^{x_N}}\right|_{\mathbf{z}=1}$$
$$=\sum_{n_1=x_1}^{X}\cdots\sum_{n_N=x_N}^{X}\mathbf{p}_{i+1}^m(n_1,\ldots,n_N)\frac{n_1!}{(n_1-x_1)!}\cdots\frac{n_N!}{(n_N-x_N)!}. \quad (5.18)$$

In the next the system of linear equations (5.18) is rearranged into a hypermatrix form. Let $\mathbf{e}_\ell^{X+1}=(0,\ldots,0,1,0,\ldots,0)$ denote the $1\times(X+1)$ vector with 1 at the ℓ-th position. We define the $1\times L^N(X+1)^N$ hypervector $\boldsymbol{\theta}_i$, representing the unknowns of the system of linear equations (5.18), as

$$\boldsymbol{\theta}_i=\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\otimes\mathbf{e}_{n_1+1}^{X+1}\otimes\ldots\otimes\mathbf{e}_{n_N+1}^{X+1}\quad i=1,\ldots,N.$$

Note that each element of $\boldsymbol{\theta}_i$ is a probability. We also introduce the $L^N(X+1)^N\times L^N(X+1)^N$ hypermatrices $\boldsymbol{\Phi}_{i+1}$ and $\boldsymbol{\Upsilon}_{i+1}$ representing the coefficients on the l.h.s. and on the right-hand side (r.h.s.) of the above system of linear equations, respectively. Let $p(n_1,\ldots,n_N)$ denote the position of the 1 in the unit hypervector $\mathbf{e}_{n_1+1}^{X+1}\otimes\ldots\otimes\mathbf{e}_{n_N+1}^{X+1}$ for $n_1,\ldots,n_N\in\{0,\ldots,X\}$. Let $L^N\times L^N$ matrices $\boldsymbol{\Delta}_{i+1}^l(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N))$ and $\boldsymbol{\Delta}_{i+1}^r(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N))$ stand for the coefficients of $\mathbf{p}_i^m(n_1,\ldots,n_N)$ on the l.h.s. and $\mathbf{p}_{i+1}^m(n_1,\ldots,n_N)$ on the r.h.s. of the system of linear equations (5.18) for $x_1,\ldots,x_N$, for $n_1,\ldots,n_N\in\{0,\ldots,X\}$ and $x_1,\ldots,x_N\in\{0,\ldots,X\}$, respectively. Applying these notations in (5.18) yields the compact form of the above system of linear equations for $i=1,\ldots,N$ and $x_1,\ldots,x_N\in\{0,\ldots,X\}$ as

$$\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\,\boldsymbol{\Delta}_{i+1}^l(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N))$$
$$=\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_{i+1}^m(n_1,\ldots,n_N)\,\boldsymbol{\Delta}_{i+1}^r(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N)). \quad (5.19)$$

Matrices $\boldsymbol{\Phi}_{i+1}$ and $\boldsymbol{\Upsilon}_{i+1}$ are defined as

$$\boldsymbol{\Phi}_{i+1}=\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\sum_{x_1=0}^{X}\cdots\sum_{x_N=0}^{X}\boldsymbol{\Delta}_{i+1}^l(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N))\otimes$$
$$\left(\mathbf{e}_{n_1+1}^{X+1}\otimes\ldots\otimes\mathbf{e}_{n_N+1}^{X+1}\right)^T\left(\mathbf{e}_{x_1+1}^{X+1}\otimes\ldots\otimes\mathbf{e}_{x_N+1}^{X+1}\right)\quad i=1,\ldots,N, \quad (5.20)$$

$$\mathbf{\Upsilon}_{i+1} = \sum_{n_1=0}^{X} \ldots \sum_{n_N=0}^{X} \sum_{x_1=0}^{X} \ldots \sum_{x_N=0}^{X} \mathbf{\Delta}^r_{i+1}(p(n_1,\ldots,n_N),p(x_1,\ldots,x_N)) \otimes \left(\mathbf{e}^{X+1}_{n_1+1} \otimes \ldots \otimes \mathbf{e}^{X+1}_{n_N+1}\right)^T \left(\mathbf{e}^{X+1}_{x_1+1} \otimes \ldots \otimes \mathbf{e}^{X+1}_{x_N+1}\right) \quad i=1,\ldots,N. \tag{5.21}$$

In both hypermatrices the values of $p(n_1,\ldots,n_N)$ and $p(x_1,\ldots,x_N)$ specify the row and the column indices of the corresponding $L^N \times L^N$ block matrix. Hence the block structure of hypermatrices $\mathbf{\Phi}_{i+1}$ and $\mathbf{\Upsilon}_{i+1}$, for $i=1,\ldots,N$, can be given as

$$\mathbf{\Phi}_{i+1} = \begin{pmatrix} \mathbf{\Delta}^l_{i+1}(1,1) & \mathbf{\Delta}^l_{i+1}(1,2) & \cdots & \mathbf{\Delta}^l_{i+1}(1,X+1^N) \\ \mathbf{\Delta}^l_{i+1}(2,1) & \mathbf{\Delta}^l_{i+1}(2,2) & \cdots & \mathbf{\Delta}^l_{i+1}(2,X+1^N) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Delta}^l_{i+1}(X+1^N,1) & \mathbf{\Delta}^l_{i+1}(X+1^N,2) & \cdots & \mathbf{\Delta}^l_{i+1}(X+1^N,X+1^N) \end{pmatrix},$$

$$\mathbf{\Upsilon}_{i+1} = \begin{pmatrix} \mathbf{\Delta}^r_{i+1}(1,1) & \mathbf{\Delta}^r_{i+1}(1,2) & \cdots & \mathbf{\Delta}^r_{i+1}(1,X+1^N) \\ \mathbf{\Delta}^r_{i+1}(2,1) & \mathbf{\Delta}^r_{i+1}(2,2) & \cdots & \mathbf{\Delta}^r_{i+1}(2,X+1^N) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Delta}^r_{i+1}(X+1^N,1) & \mathbf{\Delta}^r_{i+1}(X+1^N,2) & \cdots & \mathbf{\Delta}^r_{i+1}(X+1^N,X+1^N) \end{pmatrix}.$$

Using the definitions (5.20) and (5.21) the hypermatrix form of the system of linear equation (5.19) can be given as

$$\boldsymbol{\theta}_i \mathbf{\Phi}_{i+1} = \boldsymbol{\theta}_{i+1} \mathbf{\Upsilon}_{i+1}, \quad i=1,\ldots,N.$$

The sums on the l.h.s. and on the r.h.s. of (5.19) realize the product of hypervectors $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{i+1}$ by hypermatrix $\mathbf{\Phi}_{i+1}$ and $\mathbf{\Upsilon}_{i+1}$, respectively.

5.4.2 Exhaustive discipline specific handling of the hypermatrix form of the system of linear equation

Due to zero elements in the hypervectors and hypermatrices, the hypermatrix form of the system of linear equations requires a specific handling for the exhaustive service system.

Starting from equations (5.13) and (5.14) and repeating the rearrangements described in the previous Subsection leads to the hypermatrix form of the system of linear equations for the exhaustive service system. We apply an e in superscript to differentiate the hypervectors and hypermatrices for the the exhaustive service system from those of the gated service system. Using this notation the hypermatrix form of the system of linear equations for the exhaustive service system is given as

$$\boldsymbol{\theta}_i^e \boldsymbol{\Phi}_{i+1}^e = \boldsymbol{\theta}_{i+1}^e \boldsymbol{\Upsilon}_{i+1}^e, \quad i = 1, \ldots, N, \tag{5.22}$$

In the following we give the solution of (5.22) for $\boldsymbol{\theta}_{i+1}^e$. The exhaustive discipline of station $i+1$ implies that the elements of $\boldsymbol{\theta}_{i+1}^e$ belonging to $n_{i+1} > 0$ are 0. Furthermore the governing equation of station $i+1$ implies that the columns of $\boldsymbol{\Phi}_{i+1}^e$ and $\boldsymbol{\Upsilon}_{i+1}^e$ representing the derivatives $x_{i+1} > 0$ are also 0. Additionally the values in those rows of $\boldsymbol{\Upsilon}_{i+1}^e$, which have the same indices as the zero columns of $\boldsymbol{\Upsilon}_{i+1}^e$, have no any role, since in the term $\boldsymbol{\theta}_{i+1}^e \boldsymbol{\Upsilon}_{i+1}^e$ they are multiplied with the zero elements of $\boldsymbol{\theta}_{i+1}^e$. In order to take all these zero elements into account in the solution of (5.22) for $\boldsymbol{\theta}_{i+1}^e$ we define several operators.

We define the *column reduction operator* $\mathcal{R}(\mathbf{V})$ as an operator, which deletes the zero columns of the argument matrix $\mathbf{V}$. Similarly we define the *matrix reduction operator* $\mathcal{M}(\mathbf{Y})$ as an operator, which deletes the zero columns of the quadratic matrix $\mathbf{Y}$ and those rows, which have the same indices as the zero columns. Thus the resulted matrix remains quadratic. We also use the *inverse column reduction operator* $\mathcal{R}_{\mathbf{V}}^{-1}(\mathbf{W})$, which inserts zero columns into matrix $\mathbf{W}$ at those positions, at which the *column reduction operator* $\mathcal{R}(\mathbf{V})$ would delete zero columns from matrix $\mathbf{V}$. Thus

$$\mathcal{R}_{\mathbf{V}}^{-1}(\mathcal{R}(\mathbf{V})) = \mathbf{V}. \tag{5.23}$$

Applying the column reduction operator on the l.h.s. of (5.22) gives

$$\mathcal{R}(\boldsymbol{\theta}_i^e \boldsymbol{\Phi}_{i+1}^e) = \boldsymbol{\theta}_i^e \mathcal{R}(\boldsymbol{\Phi}_{i+1}^e). \tag{5.24}$$

Similarly applying $\mathcal{R}()$ on the r.h.s. of (5.22) and utilizing also the zero elements of $\boldsymbol{\theta}_{i+1}^e$ leads to

$$\mathcal{R}(\boldsymbol{\theta}_{i+1}^e \boldsymbol{\Upsilon}_{i+1}^e) = \boldsymbol{\theta}_{i+1}^e \mathcal{R}(\boldsymbol{\Upsilon}_{i+1}^e) = \mathcal{R}(\boldsymbol{\theta}_{i+1}^e) \mathcal{M}(\boldsymbol{\Upsilon}_{i+1}^e). \tag{5.25}$$

Thus applying $\mathcal{R}()$ on (5.22) and using (5.24) and (5.25) yields

$$\boldsymbol{\theta}_i^e \mathcal{R}(\boldsymbol{\Phi}_{i+1}^e) = \mathcal{R}(\boldsymbol{\theta}_{i+1}^e) \mathcal{M}(\boldsymbol{\Upsilon}_{i+1}^e). \tag{5.26}$$

The exhaustive discipline specific handling realized by (5.26) can be illustrated in sketch form as

$$\begin{pmatrix} * & * & \boldsymbol{\theta}_i^e & * & * \end{pmatrix} \begin{pmatrix} * & * & * & 0 & 0 \\ * & * & * & 0 & 0 \\ * & \mathcal{R}(\boldsymbol{\Phi}_{i+1}^e) & * & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & 0 & 0 \end{pmatrix} =$$

$$\begin{pmatrix} * & \mathcal{R}(\boldsymbol{\theta}_{i+1}^e) & * & 0 & 0 \end{pmatrix} \begin{pmatrix} * & * & * & 0 & 0 \\ * & \mathcal{M}(\boldsymbol{\Upsilon}_{i+1}^e) & * & 0 & 0 \\ * & * & * & 0 & 0 \\ + & + & + & 0 & 0 \\ + & + & + & 0 & 0 \end{pmatrix}.$$

$\mathcal{R}(\boldsymbol{\theta}_{i+1}^e)$ can be expressed from (5.26) as

$$\boldsymbol{\theta}_i^e \mathcal{R}(\boldsymbol{\Phi}_{i+1}^e) \left(\mathcal{M}(\boldsymbol{\Upsilon}_{i+1}^e)\right)^{-1} = \mathcal{R}(\boldsymbol{\theta}_{i+1}^e). \tag{5.27}$$

Applying the inverse column reduction operator on both sides of (5.27) and using (5.23) results in the solution of (5.22) for $\boldsymbol{\theta}_{i+1}^e$ as

$$\boldsymbol{\theta}_i^e \mathcal{R}^{-1}_{\boldsymbol{\theta}_{i+1}^e}\left(\mathcal{R}(\boldsymbol{\Phi}_{i+1}^e) \left(\mathcal{M}(\boldsymbol{\Upsilon}_{i+1}^e)\right)^{-1}\right) = \mathcal{R}^{-1}_{\boldsymbol{\theta}_{i+1}^e}(\mathcal{R}(\boldsymbol{\theta}_{i+1}^e)) = \boldsymbol{\theta}_{i+1}^e. \tag{5.28}$$

5.4.3 Computation of matrix $\widetilde{\mathbf{G}}_i(s)$

The recursive algorithm

The LSTs of the matrices $\mathbf{G}_i(t)$, for $i = 1, \ldots, N$, are used in the numerical solution for the exhaustive discipline. They are defined as

$$\widetilde{\mathbf{G}}_i(s) = \int_0^\infty e^{-st} d\mathbf{G}_i(t), \quad Re(s) \geq 0, \quad i = 1, \ldots, N.$$

It is shown in Lucantoni [114] that $\widetilde{\mathbf{G}}_i(s)$ satisfies the equation

$$\widetilde{\mathbf{G}}_i(s) = \int_0^\infty e^{-st} e^{\widehat{\mathbf{D}}_i(\widetilde{\mathbf{G}}_i(s))t} dB_i(t). \tag{5.29}$$

We provide a recursive algorithm for computing the matrices $\widetilde{\mathbf{G}}_i(s)$, for every $i = 1, \ldots, N$. This is a generalization of the algorithm provided for computing matrices

$\widetilde{\mathbf{G}}_i(0)$, which is based on the application of the concept of uniformization, see in Lucantoni [114], pp. 27-28. This concept leads to the following relation for $\widetilde{\mathbf{G}}_i(s)$,

$$\widetilde{\mathbf{G}}_i(s) = \sum_{n=0}^{\infty} \gamma_{i,n}(s)(\mathbf{I} + \vartheta_i^{-1}\widehat{\mathbf{D}}_i(\widetilde{\mathbf{G}}_i(s)))^n, \tag{5.30}$$

where $\vartheta_i = max_j(-[\mathbf{D}_{i,0}]_{jj})$ and

$$\gamma_{i,n}(s) = \int_0^{\infty} e^{-(s+\vartheta_i)t} \frac{(\vartheta_i t)^n}{n!} dB_i(t), \quad n \geq 0. \tag{5.31}$$

The advantage of (5.30) is that it involves only powers of quadratic matrices, which is simpler to compute than calculating the matrix exponential $e^{\widehat{\mathbf{D}}_i(\widetilde{\mathbf{G}}_i(s))t}$ in (5.29).

Based on (5.30) $\widetilde{\mathbf{G}}_i(s)$ can be computed by successively iterating for $k \geq 0$ in the following recursion,

$$\begin{aligned} &\mathbf{H}_{i,n+1,k}(s) = (\mathbf{I} + \vartheta_i^{-1}\widehat{\mathbf{D}}_i(\widetilde{\mathbf{G}}_{i,k}(s)))\mathbf{H}_{i,n,k}(s) \quad n \geq 0, \\ &\widetilde{\mathbf{G}}_{i,k+1}(s) = \sum_{n=0}^{\infty} \gamma_{i,n}(s)\mathbf{H}_{i,n,k}(s), \text{ where} \\ &\mathbf{H}_{i,0,k}(s) = \mathbf{I} \text{ for } k \geq 0, \text{ and } \widetilde{\mathbf{G}}_{i,0}(s) = \mathbf{0}. \end{aligned} \tag{5.32}$$

We have found that starting with $\widetilde{\mathbf{G}}_{i,0}(s) = \mathbf{e}\boldsymbol{\pi}$ leads to a faster convergence. This is in accordance with the proposal in Lucantoni [114] for computing $\widetilde{\mathbf{G}}_i(0)$, since $\mathbf{e}\boldsymbol{\pi}$ is close to $\widetilde{\mathbf{G}}_i(0)$. Therefore the setting $\widetilde{\mathbf{G}}_{i,0}(s) = \mathbf{e}\boldsymbol{\pi}$ is proposed.

The matrix $\widehat{\mathbf{D}}_i(\widetilde{\mathbf{G}}_{i,k}(s))$ is defined as

$$\widehat{\mathbf{D}}_i(\widetilde{\mathbf{G}}_{i,k}(s)) = \sum_{\ell=0}^{\infty} \mathbf{D}_{i,\ell}\widetilde{\mathbf{G}}_{i,k}^{\ell}(s), \tag{5.33}$$

and can be computed for M as truncation index on the matrices $\mathbf{D}_{i,\ell}$, by using Horner's method as

$$\mathbf{O}_{i,0}(s) = \mathbf{D}_{i,M}, \quad \mathbf{O}_{i,\ell}(s) = \mathbf{D}_{i,M-\ell} + \mathbf{O}_{i,\ell-1}(s)\widetilde{\mathbf{G}}_{i,k}(s), \text{ for } 1 \leq \ell \leq M.$$

This gives $\mathbf{O}_{i,M}(s) = \sum_{\ell=0}^{M} \mathbf{D}_{i,\ell}\widetilde{\mathbf{G}}_{i,k}^{\ell}(s)$ as the approximation of $\widehat{\mathbf{D}}_i(\widetilde{\mathbf{G}}_{i,k}(s))$.

The infinite sum in (5.32) is approximated by finite number of elements. The number of elements is set dynamically by checking the achieved accuracy iteratively.

The convergence of the algorithm

It was proven in Lucantoni [113] that starting with $\widetilde{\mathbf{G}}_{i,0}(s) = \mathbf{0}$ and successively iterating in (5.29) it goes to $\widetilde{\mathbf{G}}_i(s)$. Matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$ given by (5.32) implements the computation based on (5.30), which can be interpreted as the numerical evaluation of (5.29). Therefore it follows that the above mentioned proof in [113] implies that also the matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$ converges to $\widetilde{\mathbf{G}}_i(s)$.

Here we provide an alternative proof of the convergence of matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$ for $Re(s) \geq 0$ as $k \to \infty$.

Lemma 5.1 *(Properties of matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$ for real $s \geq 0$.) For real $s \geq 0$ the matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$, given by (5.32), is monotone increasing and it is convergent as $k \to \infty$.*

Proof. (5.32) can be rearranged as

$$\widetilde{\mathbf{G}}_{i,k+1}(s) = \sum_{n=0}^{\infty} \gamma_{i,n}(s)(\mathbf{I} + \vartheta_i^{-1}\mathbf{D}_{i,0} + \vartheta_i^{-1}\sum_{\ell=1}^{\infty}\mathbf{D}_{i,l}\widetilde{\mathbf{G}}_{i,k}^{l}(s))^n. \tag{5.34}$$

According to (5.34) the first two members of $\widetilde{\mathbf{G}}_{i,k}(s)$ are given as

$$\widetilde{\mathbf{G}}_{i,0}(s) = \mathbf{0}, \tag{5.35}$$
$$\widetilde{\mathbf{G}}_{i,1}(s) = \sum_{n=0}^{\infty} \gamma_{i,n}(s)(\mathbf{I} + \vartheta_i^{-1}\mathbf{D}_{i,0})^n.$$

It can be seen from (5.31) that $\gamma_{i,n}(s) > 0$ for real $s \geq 0$. Additionally matrices $(\mathbf{I} + \vartheta_i^{-1}\mathbf{D}_{i,0})$ and $\mathbf{D}_{i,l}$ for $l \geq 1$ are nonnegative. Thus for real $s \geq 0$ matrix $\widetilde{\mathbf{G}}_{i,1}(s)$ is nonnegative and

$$\widetilde{\mathbf{G}}_{i,1}(s) \geq \widetilde{\mathbf{G}}_{i,0}(s). \tag{5.36}$$

Furthermore replacing $\widetilde{\mathbf{G}}_{i,k}(s)$ in (5.34) by $\widetilde{\mathbf{G}}_{i,k+1}(s) \geq \widetilde{\mathbf{G}}_{i,k}(s)$ results in $\widetilde{\mathbf{G}}_{i,k+2}(s) \geq \widetilde{\mathbf{G}}_{i,k+1}(s)$ for $k \geq 0$. Applying it recursively by starting with relation (5.36) yields

$$\widetilde{\mathbf{G}}_{i,k+1}(s) \geq \widetilde{\mathbf{G}}_{i,k}(s), \quad k \geq 0.$$

Hence matrices $\widetilde{\mathbf{G}}_{i,k}(s)$ are nonnegative and monotone increasing for $k \geq 0$.

It follows from expression (5.31) that $\gamma_{i,n}(s)$ is monotone decreasing function of s for real $s \geq 0$. Applying it recursively in (5.34) shows that also matrices $\widetilde{\mathbf{G}}_{i,k}(s)$, for $k \geq 0$, are monotone decreasing function of s for real $s \geq 0$. Hence for real $s \geq 0$

$$\widetilde{\mathbf{G}}_{i,k}(s) \leq \widetilde{\mathbf{G}}_{i,k}(0), \quad k \geq 0. \tag{5.37}$$

It is shown (see Lucantoni [114]) that matrix series $\widetilde{\mathbf{G}}_{i,k}(0)$ is convergent as $k \to \infty$. Using it (5.37) implies that matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$, $k \geq 0$ is upper limited. Thus matrix series $\widetilde{\mathbf{G}}_{i,k}(0)$, for $k \geq 0$, is monotone increasing and upper limited, which ensures the convergence. □

Theorem 5.3 *(Convergence of matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$ for $Re(s) \geq 0$.) If $B_i(t)$ is continuously differentiable on $t \geq 0$ then for $Re(s) \geq 0$ the matrix series $\widetilde{\mathbf{G}}_{i,k}(s)$, given by (5.32), is convergent as $k \to \infty$.*

Proof. Let $b_i(t)$ denote the first derivative of $B_i(t)$, i.e. $b_i(t) = \frac{dB_i(t)}{dt}$. Using it the expression (5.31) can be rewritten as

$$\gamma_{i,n}(s) = \int_0^\infty e^{-(s+\vartheta_i)t} \frac{(\vartheta_i t)^n}{n!} b_i(t)dt, \quad n \geq 0, \tag{5.38}$$

The form of expression (5.38) shows that $\gamma_{i,n}(s)$ is a Laplace transform. In the following we utilize that any linear combination and any product of Laplace transforms results in also a Laplace transform, which follows from the elementary properties of the Laplace transform. Therefore starting with $\gamma_{i,n}(s)$ as a Laplace transform each step of the successive iteration (5.32) leads also to a Laplace transform. Thus matrices $\widetilde{\mathbf{G}}_{i,k}(s)$ are also a Laplace transforms for $k \geq 1$. Let matrices $\boldsymbol{G}_{i,k}(t)$ be the inverse Laplace transform of $\widetilde{\mathbf{G}}_{i,k}(s)$ for $k \geq 1$, i.e.

$$\widetilde{\mathbf{G}}_{i,k}(s) = \int_0^\infty e^{-st} \boldsymbol{G}_{i,k}(t)dt, \quad Re(s) \geq 0. \tag{5.39}$$

We define $\Delta_k(s)$ for $k \geq 1$ as

$$\Delta_k(s) = \widetilde{\mathbf{G}}_{i,k+1}(s) - \widetilde{\mathbf{G}}_{i,k}(s), \quad Re(s) \geq 0. \tag{5.40}$$

Using (5.39) we rewrite (5.40) for $k \geq 1$ as

$$\begin{aligned} \Delta_k(s) &= \int_0^\infty e^{-st}(\boldsymbol{G}_{i,k}(t) - \boldsymbol{G}_{i,k}(t))dt \qquad (5.41) \\ &= \int_0^\infty e^{-Re(s)t}(cos(-Im(s)t) + i\ sin(-Im(s)t))\ (\boldsymbol{G}_{i,k}(t) - \boldsymbol{G}_{i,k}(t))dt \quad Re(s) \geq 0, \end{aligned}$$

where i stands for the imaginary unit. Using $-1 \leq cos(-Im(s)t) \leq 1$, $-1 \leq sin(-Im(s)t) \leq 1$ and (5.41) we have

$$\begin{aligned} -\Delta_k(Re(s)) &\leq Re(\Delta_k(s)) \leq \Delta_k(Re(s)), \\ -\Delta_k(Re(s)) &\leq Im(\Delta_k(s)) \leq \Delta_k(Re(s)). \end{aligned} \tag{5.42}$$

It follows from lemma 5.1 that matrices $\Delta_k(Re(s))$, for $k \geq 1$, are nonnegative and they converge to 0 as $k \to \infty$. Using them the theorem comes from (5.42). $\square$

5.4.4 Computation of matrix $\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)$

The matrices $\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)$, for $i = 1, \ldots, N$, are used in the numerical solution for the exhaustive discipline. They can be determined directly from matrices $\widetilde{\mathbf{G}}_i(s)$, for $i = 1, \ldots, N$, for every $i = 1, \ldots, N$, without taking their inverse transform.

We define an i-dependent position rearrangement operator, which rearranges the positions of the elements of an $L^N \times L^N$ hypermatrix $e^{t \oplus_{k=1}^{i-1} \widehat{\mathbf{D}}_k(z_k)} \otimes e^{t \oplus_{k=i+1}^{N} \widehat{\mathbf{D}}_k(z_k)} \otimes \mathbf{G}_i(t)$ as

$$\mathcal{P}_i\left(e^{t \oplus_{k=1}^{i-1} \widehat{\mathbf{D}}_k(z_k)} \otimes e^{t \oplus_{k=i+1}^{N} \widehat{\mathbf{D}}_k(z_k)} \otimes \mathbf{G}_i(t)\right) = e^{t \oplus_{k=1}^{i-1} \widehat{\mathbf{D}}_k(z_k)} \otimes \mathbf{G}_i(t) \otimes e^{t \oplus_{k=i+1}^{N} \widehat{\mathbf{D}}_k(z_k)}.$$

Using this position rearrangement operator and taking into account $e^{t \oplus_{k=1}^{i-1} \widehat{\mathbf{D}}_k(z_k)} \otimes e^{t \oplus_{k=i+1}^{N} \widehat{\mathbf{D}}_k(z_k)} \otimes \mathbf{G}_i(t) = e^{t \oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k)} \otimes \mathbf{G}_i(t)$ the hypermatrix $\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)$ can be expressed as

$$\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N) = \mathcal{P}_i\left(\int_0^\infty e^{t \oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k)} \otimes d\mathbf{G}_i(t)\right). \tag{5.43}$$

The part of $\int_0^\infty e^{t \oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k)} \otimes d\mathbf{G}_i(t)$, which is determined by $[\mathbf{G}_i(t)]_{j,l}$, for $j, l = 1, \ldots, L$, can be rearranged as

$$\begin{aligned} \int_0^\infty e^{t \oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k)} d[\mathbf{G}_i(t)]_{j,l} &= \int_0^\infty \sum_{n=0}^\infty \frac{t^n}{n!} (\oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k))^n d[\mathbf{G}_i(t)]_{j,l} \\ &= \mathbf{I} + \sum_{n=1}^\infty \frac{1}{n!} \int_0^\infty (-1)^n t^n d[\mathbf{G}_i(t)]_{j,l} \; ((-1) \oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k))^n \\ &= \mathbf{I} + \sum_{n=1}^\infty \frac{1}{n!} \int_0^\infty \left. \frac{d^n(e^{-st})}{ds^n} \right|_{s=0} d[\mathbf{G}_i(t)]_{j,l} \; ((-1) \oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k))^n \\ &= \mathbf{I} + \sum_{n=1}^\infty \frac{1}{n!} \left. \frac{d^n [\widetilde{\mathbf{G}}_i(s)]_{j,l}}{ds^n} \right|_{s=0} ((-1) \oplus_{k \neq i} \widehat{\mathbf{D}}_k(z_k))^n, \end{aligned} \tag{5.44}$$

where the interchange of the integral and sum is justified by the absolute convergence of $\int_0^\infty \sum_{n=0}^\infty \frac{t^n}{n!}(\oplus_{k\neq i}\widehat{\mathbf{D}}_k(z_k))^n d[\mathbf{G}_i(t)]_{j,l}$. This follows from the existence of an upper limit for $\int_0^\infty \sum_{n=0}^\infty |\frac{t^n}{n!}(\oplus_{k\neq i}\widehat{\mathbf{D}}_k(z_k))^n| d[\mathbf{G}_i(t)]_{j,l}$, which can be shown as

$$\int_0^\infty \sum_{n=0}^\infty |\frac{t^n}{n!}(\oplus_{k\neq i}\widehat{\mathbf{D}}_k(z_k))^n| d[\mathbf{G}_i(t)]_{j,l} \leq \int_0^\infty \sum_{n=0}^\infty \frac{t^n}{n!}| \oplus_{k\neq i} \widehat{\mathbf{D}}_k(z_k)|^n d[\mathbf{G}_i(t)]_{j,l}$$
$$= \int_0^\infty e^{t|\oplus_{k\neq i}\widehat{\mathbf{D}}_k(z_k)|} d[\mathbf{G}_i(t)]_{j,l}.$$

Applying (5.44) for each $j, l = 1, \ldots, L$ and using (5.43) the hypermatrix $\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)$ can be computed from $\widetilde{\mathbf{G}}_i(s)$. The number of moments $\left.\frac{d^n[\widetilde{\mathbf{G}}_i(s)]_{j,l}}{ds^n}\right|_{s=0}$ in (5.44) depends on the given parameter settings and the required precision. Hence it is set dynamically by checking the achieved precision level iteratively.

5.5 Numerical solution

In this Section we present the steps of the computational procedure for the vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs ($\mathbf{f}_i^{(n)}$ and $\mathbf{m}_i^{(n)}$ for $n \geq 1$), for which the numerical considerations discussed in the previous Section are also used. The numerical solution for the systems with gated and exhaustive disciplines is based on the the equations (5.9) and (5.10) as well as (5.13) and (5.14), respectively.

5.5.1 Steps of the numerical solution for the gated service system

The computation of the vector moments of the stationary number of i-customers at i-polling and i-departure epochs consists of the following steps:

1. Computation of matrices $\widehat{\mathbf{A}}_i(z_1, \ldots, z_N)$ for every $i = 1, \ldots, N$. They can be computed from their definitions by applying numerical integration technique.
2. Building up the large hypermatrices of the system. The system of linear equations (5.9) and (5.10), for each $i = 1, \ldots, N$, is rearranged into a hypermatrix form as described in Subsection 5.4.1. This form consists of two large $1 \times L^N(X+1)^N$ hypervectors $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{i+1}$ as well as two large $L^N(X+1)^N \times L^N(X+1)^N$ hypermatrices $\boldsymbol{\Phi}_{i+1}$ and $\boldsymbol{\Upsilon}_{i+1}$ representing the unknowns and the coefficients on the l.h.s and on the r.h.s of the $\mathbf{p}_i^m(n_1, \ldots, n_N) \rightarrow \mathbf{p}_{i+1}^m(n_1, \ldots, n_N)$ form of system of linear equations, respectively. The hypermatrix form is given as

$$\boldsymbol{\theta}_i\boldsymbol{\Phi}_{i+1} = \boldsymbol{\theta}_{i+1}\boldsymbol{\Upsilon}_{i+1}, \quad i = 1, \ldots, N. \tag{5.45}$$

3. Composing relations $\boldsymbol{\theta}_i \to \boldsymbol{\theta}_i$ for every $i = 1, \ldots, N$. The solution of (5.45) for $\boldsymbol{\theta}_{i+1}$ is given as

$$\boldsymbol{\theta}_i \boldsymbol{\Phi}_{i+1} \boldsymbol{\Upsilon}_{i+1}^{-1} = \boldsymbol{\theta}_{i+1}, \quad i = 1, \ldots, N.$$

Applying it recursively for $i, i+1, \ldots, N, 1, \ldots, i-1$ leads to the formula

$$\boldsymbol{\theta}_i \boldsymbol{\Phi}_{i+1} \boldsymbol{\Upsilon}_{i+1}^{-1} \boldsymbol{\Phi}_{i+2} \boldsymbol{\Upsilon}_{i+2}^{-1} \ldots \boldsymbol{\Phi}_i \boldsymbol{\Upsilon}_i^{-1} = \boldsymbol{\theta}_i, \quad i = 1, \ldots, N. \tag{5.46}$$

We define $\boldsymbol{\Xi}_i$ as the hypermatrix on the l.h.s. of (5.46), i.e.

$$\boldsymbol{\Xi}_i = \boldsymbol{\Phi}_{i+1} \boldsymbol{\Upsilon}_{i+1}^{-1} \boldsymbol{\Phi}_{i+2} \boldsymbol{\Upsilon}_{i+2}^{-1} \ldots \boldsymbol{\Phi}_i \boldsymbol{\Upsilon}_i^{-1}. \tag{5.47}$$

Applying (5.47) in (5.46) gives a system of linear equations for $\boldsymbol{\theta}_i$ as

$$\boldsymbol{\theta}_i \boldsymbol{\Xi}_i = \boldsymbol{\theta}_i, \quad i = 1, \ldots, N. \tag{5.48}$$

Hypermatrix $\boldsymbol{\Xi}_i$ relates the probabilities of hypervector $\boldsymbol{\theta}_i$ as it represents the governing equations of the system. Thus hypermatrix $\boldsymbol{\Xi}_i$ can be interpreted as transition probability hypermatrix. It follows that $\boldsymbol{\Xi}_i$ is stochastic and similar to (2.10) $rank\,(\mathbf{I} - \boldsymbol{\Xi}_i)$ is one less than the dimension of $\boldsymbol{\Xi}_i$. Therefore (5.48) does not determine $\boldsymbol{\theta}_i$ uniquely. To make the system of linear equations complete we add the normalization condition for $i = 1, \ldots, N$ as

$$\boldsymbol{\theta}_i \mathbf{e}^{L^N (X+1)^N} = \sum_{n_1=0}^{X} \ldots \sum_{n_N=0}^{X} \mathbf{p}_i^m (n_1, \ldots, n_N) \ \mathbf{e} \otimes \ldots \otimes \mathbf{e} = 1. \tag{5.49}$$

Here $\mathbf{e}^{L^N (X+1)^N}$ denotes the $1 \times (L^N (X+1)^N)$ column vector having all elements equal to one and the number of $L \times 1$ vectors $\mathbf{e}$ in the Kronecker product is N.

This step involves inverse computation of N large hypermatrices
($N \left(L^{3N} (X+1)^{3N} \right)$ elementary computational steps) and for all stations a total of $N^2 + N$ times multiplication of large hypermatrices ($(N^2 + N) \left(L^{3N} (X+1)^{3N} \right)$ elementary computational steps).

4. Solving the linear system of equations for relation $\boldsymbol{\theta}_i \rightarrow \boldsymbol{\theta}_i$ ((5.48) and (5.49)) for $i = 1, \ldots, N$. This takes $L^{3N}(X+1)^{3N}$ elementary computational steps for station i. It gives the stationary probability distribution represented by hypervectors $\mathbf{p}_i^m(n_1, \ldots, n_N)$.

5. Computation of stationary probability distribution $\mathbf{p}_i^f(n_1, \ldots, n_N)$. The hypervectors $\mathbf{p}_i^f(n_1, \ldots, n_N)$ are computed from $\mathbf{p}_{i-1}^m(n_1, \ldots, n_N)$ by using the service discipline independent relation for transition $m_i \rightarrow f_{i+1}$ (5.8) for every $i = 1, \ldots, N$.

6. Computation of vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs. The vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs ($\mathbf{f}_i^{(n)}$ and $\mathbf{m}_i^{(n)}$ for $n \geq 1$) are calculated from $\mathbf{p}_i^f(n_1, \ldots, n_N)$ and $\mathbf{p}_i^m(n_1, \ldots, n_N)$ on elementary way for every $i = 1, \ldots, N$.

5.5.2 Steps of the numerical solution for the exhaustive service system

Similarly to the case with gated discipline the computation of the vector moments of the stationary number of i-customers at i-polling and i-departure epochs consists of several steps:

1. Computation of matrices $\widetilde{\mathbf{G}}_i(s)$ for every $i = 1, \ldots, N$. They are calculated by means of the recursive algorithm described in Subsection 5.4.3.

2. Computation of matrices $\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)$ for every $i = 1, \ldots, N$. They can be computed directly from matrices $\widetilde{\mathbf{G}}_i(s)$ as described in Subsection 5.4.4.

3. Building up the large hypermatrices of the system. The system of linear equations (5.13) and (5.14), for each $i = 1, \ldots, N$, can be rearranged into the hypermatrix form, as described in Subsection 5.4.2, leading to

$$\boldsymbol{\theta}_i^e \boldsymbol{\Phi}_{i+1}^e = \boldsymbol{\theta}_{i+1}^e \boldsymbol{\Upsilon}_{i+1}^e, \quad i = 1, \ldots, N,$$

   where $\boldsymbol{\theta}_i^e$ and $\boldsymbol{\theta}_{i+1}^e$ are $1 \times L^N(X+1)^N$ hypervectors and $\boldsymbol{\Phi}_{i+1}^e$ and $\boldsymbol{\Upsilon}_{i+1}^e$ are $L^N(X+1)^N \times L^N(X+1)^N$ hypermatrices.

4. Composing relations $\mathcal{R}(\boldsymbol{\theta}_i^e) \rightarrow \mathcal{R}(\boldsymbol{\theta}_i^e)$ for every $i = 1, \ldots, N$. According to the zero element specific handling described in Subsection 5.4.2 the expression of $\boldsymbol{\theta}_{i+1}^e$ in terms of the hypermatrices of the above hypermatrix form system of linear

equations is given by (5.28). We define $\mathbf{\Theta}^e_{i+1}$ as the hypermatrix on the l.h.s. of (5.28), i.e.

$$\mathbf{\Theta}^e_{i+1} = \mathcal{R}^{-1}_{\boldsymbol{\theta}^e_{i+1}}(\mathcal{R}(\mathbf{\Phi}^e_{i+1})\left(\mathcal{M}(\mathbf{\Upsilon}^e_{i+1})\right)^{-1}). \tag{5.50}$$

Applying (5.50) in (5.28) gives the solution of the hypermatrix form system of linear equations (5.22) for $\boldsymbol{\theta}^e_{i+1}$ as

$$\boldsymbol{\theta}^e_i \mathbf{\Theta}^e_{i+1} = \boldsymbol{\theta}^e_{i+1}, \quad i = 1, \ldots, N. \tag{5.51}$$

Hence the computation of $L^N(X+1)^N \times L^N(X+1)^N$ hypermatrix $\mathbf{\Theta}^e_{i+1}$ involves an inverse computation of an $L^N(X+1)^{N-1} \times L^N(X+1)^{N-1}$ matrix $\mathcal{M}(\mathbf{\Upsilon}^e_{i+1})$. Applying (5.51) recursively for $i, i+1, \ldots, N, 1, \ldots, i-1$ leads to the formula

$$\boldsymbol{\theta}^e_i \mathbf{\Theta}^e_{i+1} \mathbf{\Theta}^e_{i+2} \ldots \mathbf{\Theta}^e_i = \boldsymbol{\theta}^e_i, \quad i = 1, \ldots, N. \tag{5.52}$$

We define $\mathbf{\Xi}^e_i$ as the matrix on the l.h.s. of (5.52), i.e.

$$\mathbf{\Xi}^e_i = \mathbf{\Theta}^e_{i+1} \mathbf{\Theta}^e_{i+2} \ldots \mathbf{\Theta}^e_i, \quad i = 1, \ldots, N. \tag{5.53}$$

Applying (5.53) in (5.52) gives a system of linear equations for $\boldsymbol{\theta}^e_i$ as

$$\boldsymbol{\theta}^e_i \mathbf{\Xi}^e_i = \boldsymbol{\theta}^e_i, \quad i = 1, \ldots, N. \tag{5.54}$$

The columns of matrix $\mathbf{\Xi}^e_i$ representing the derivatives $x_i > 0$ are 0. This is due to the right most position of hypermatrix $\mathbf{\Theta}^e_i$ in (5.53), whose definition implies zero columns at the same positions. Furthermore the elements of $\boldsymbol{\theta}^e_i$ belonging to $n_i > 0$ are 0. Taking them into account the system of linear equations (5.54) can be reduced as

$$\mathcal{R}(\boldsymbol{\theta}^e_i)\mathcal{M}(\mathbf{\Xi}^e_i) = \mathcal{R}(\boldsymbol{\theta}^e_i), \quad i = 1, \ldots, N. \tag{5.55}$$

As expected the dimension of $\mathcal{M}(\mathbf{\Xi}^e_i)$ is $L^N(X+1)^{N-1}$. Similar to the case of gated discipline the hypermatrix $\mathcal{M}(\mathbf{\Xi}^e_i)$ can be interpreted as transition probability hypermatrix. It follows that $\mathcal{M}(\boldsymbol{\Xi}_i^e)$ is stochastic and $rank\,(\mathbf{I} - \mathcal{M}(\boldsymbol{\Xi}_i^e))$ is one less than the dimension of $\mathcal{M}(\boldsymbol{\Xi}_i^e)$. Thus (5.55) does not determine $\mathcal{R}(\boldsymbol{\theta}_i^e)$ uniquely. Therefore we add the normalization condition for $i = 1, \ldots, N$

$$\mathcal{R}(\boldsymbol{\theta}_i^e)\mathbf{e}^{L^N(X+1)^{N-1}} = \sum_{n_1=0}^{X} \cdots \sum_{n_{i-1}=0}^{X} \sum_{n_{i+1}=0}^{X} \cdots \sum_{n_N=0}^{X} \mathbf{p}_i^m\,(n_1, \ldots, n_{i-1}, 0, n_{i+1}, \ldots n_N) \times \mathbf{e} \otimes \ldots \otimes \mathbf{e} = 1, \tag{5.56}$$

where the number of $L \times 1$ vectors $\mathbf{e}$ in the Kronecker product is N. This makes the reduced system of linear equations complete.

This step involves inverse computation of N large hypermatrices ($N\left(L^{3N}(X+1)^{3N-3}\right)$ elementary computational steps) and for all stations a total of $N^2 + N$ times multiplication of large hypermatrices (taking into account the non-used rows of $\boldsymbol{\Xi}_i^e$ due to the matrix reduction operator this requires $(N^2 + N)\left(L^{3N}(X+1)^{3N-3}\right)$ elementary computational steps).

5. Solving the linear system of equations for relation $\mathcal{R}(\boldsymbol{\theta}_i^e) \to \mathcal{R}(\boldsymbol{\theta}_i^e)$ ((5.55) and (5.56)) for $i = 1, \ldots, N$. This takes $L^{3N}(X+1)^{3N-3}$ elementary computational steps for station i. After getting the solution for $\mathcal{R}(\boldsymbol{\theta}_i^e)$ the hypervector $\boldsymbol{\theta}_i^e$ can be retained by using the inverse column reduction operator as

$$\boldsymbol{\theta}_i^e = \mathcal{R}^{-1}_{\boldsymbol{\theta}_i^e}(\mathcal{R}(\boldsymbol{\theta}_i^e))$$

$\boldsymbol{\theta}_i^e$ gives the stationary probability distribution represented by hypervectors $\mathbf{p}_i^m\,(n_1, \ldots, n_{i-1}, 0, n_{i+1}, \ldots n_N)$.

6. Computation of stationary probability distribution $\mathbf{p}_i^f\,(n_1, \ldots, n_N)$. The hypervectors $\mathbf{p}_i^f\,(n_1, \ldots, n_N)$ are computed from $\mathbf{p}_{i-1}^m\,(n_1, \ldots, n_{i-2}, 0, n_i, \ldots n_N)$ by using the service discipline independent relation for transition $m_i \to f_{i+1}$ (5.8) for every $i = 1, \ldots, N$.

7. Computation of vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs. The vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs ($\mathbf{f}_i^{(n)}$ and $\mathbf{m}_i^{(n)}$ for $n \geq 1$) are calculated from $\mathbf{p}_i^f\,(n_1, \ldots, n_N)$ and $\mathbf{p}_i^m\,(n_1, \ldots, n_N)$ on elementary way for every $i = 1, \ldots, N$.

5.5.3 Numerical characteristics

The overall numerical complexity of the above computation procedure depends on the way of realizing the steps. The most computational intensive parts of the procedure are the computation of matrices $\widehat{\mathbf{A}}_i(z_1,\ldots,z_N)$ (or matrices $\widehat{\mathbf{H}}_i(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_N)$ for the system with exhaustive discipline), building up the large hypermatrices of the system and composing and solving the relations $\boldsymbol{\theta}_i \to \boldsymbol{\theta}_i$ (or $\mathcal{R}(\boldsymbol{\theta}_i^e) \to \mathcal{R}(\boldsymbol{\theta}_i^e)$ for the system with exhaustive discipline). The total number of operations required by the whole numerical procedure is in the magnitude of $N^2\ L^{3N}(X+1)^{3N}$, while it is $N^2\ L^{3N}(X+1)^{3N-3}$ for the system with exhaustive discipline.

The total number of required elementary computational steps increases with X, L and with N. Hence the solution of (5.9) and (5.10) ((5.13) and (5.14) for the system with exhaustive discipline) becomes difficult when the server utilization is high, the number of BMAP phases is high or the system is large.

To reduce the numerical complexity of the solution for high server utilization we selected $\mathbf{p}_i^m(n_1,\ldots,n_N)$-s as unknowns instead of $\mathbf{p}_i^f(n_1,\ldots,n_N)$-s. This enables a lower value of X, since for several disciplines including also the gated and the exhaustive one the stationary number of i-customers at i-departure epochs is less and hence much closer to 0 than at i-polling epochs.

The memory need and run time of the above computation procedure totally depend on the way of implementation, on the used program language and on the hardware characteristics of the used computer. The numerical procedure is also suitable for parallel computing, since it consists of matrix operations. The implementation dependent investigation of these numerical characteristics is left for future task.

5.6 Numerical examples

We provide two simple numerical examples to illustrate the numerical solution of the $BMAP/G/1$ nonzero-switchover-times polling model with gated and exhaustive disciplines. We investigate a polling model with two stations, i.e. $N = 2$.

The form of matrix GFs of the arrival processes at both stations ($\widehat{\mathbf{D}}_1(z)$ and $\widehat{\mathbf{D}}_2(z)$) are given by

$$\begin{pmatrix} -\alpha_1-\beta_1 & \alpha_1 \\ 0 & -\alpha_2-\beta_2 \end{pmatrix} + z \begin{pmatrix} 0 & \beta_1 \\ \alpha_2 & \beta_2 \end{pmatrix}.$$

Hence the number of phases of both $BMAP$s is $L = 2$.

For both examples the customer service times B_1, B_2 are constant and exponential with parameters τ_1 and γ_2, respectively. The switchover times R_1, R_2 are both exponential with parameters δ_1 and δ_2, respectively. It follows

$$\begin{aligned}\widehat{\mathbf{A}}_1(z_1,z_2) &= e^{\left(\widehat{\mathbf{D}}_1(z_1)\oplus\widehat{\mathbf{D}}_2(z_2)\right)\tau_1},\\ \widehat{\mathbf{A}}_2(z_1,z_2) &= \int_0^\infty e^{\left(\widehat{\mathbf{D}}_1(z_1)\oplus\widehat{\mathbf{D}}_2(z_2)\right)t}\gamma_2 e^{-\gamma_2 t}dt.\end{aligned}$$

$$\begin{aligned}\widehat{\mathbf{U}}_1(z_1,z_2) &= \int_0^\infty e^{\left(\widehat{\mathbf{D}}_1(z_1)\oplus\widehat{\mathbf{D}}_2(z_2)\right)t}\delta_1 e^{-\delta_1 t}dt,\\ \widehat{\mathbf{U}}_2(z_1,z_2) &= \int_0^\infty e^{\left(\widehat{\mathbf{D}}_1(z_1)\oplus\widehat{\mathbf{D}}_2(z_2)\right)t}\delta_2 e^{-\delta_2 t}dt.\end{aligned}$$

Example 5.1 *Polling model with gated discipline - asymmetric load*

We set the following parameter values:

Station 1: $\alpha_1 = 1, \quad \alpha_2 = 3, \quad \beta_1 = 2, \quad \beta_2 = 5, \quad \tau_1 = 1/50, \quad \delta_1 = 10,$

Station 2: $\alpha_1 = 2, \quad \alpha_2 = 5, \quad \beta_1 = 3, \quad \beta_2 = 0, \quad \gamma_2 = 12, \quad \delta_2 = 20.$

The stationary arrival rates are normalized to have $\lambda_1 = 3$, $\lambda_2 = 2$ *and hence* $\rho_1 = 0.06$ *and* $\rho_2 = 0.166$.

For this example $X = 3$ *results in a sufficient precision.*

Based on the steps of the numerical solutions described in Subsection 5.5.1 we get $(L(X+1))^N = 64$ *equations for each station. From their solutions we get the following vector factorial moments:*

$$\begin{aligned}\mathbf{f}_1 &= (0.50794, \quad 0.49206), \qquad \mathbf{m}_1 = (0.512944, \quad 0.487056),\\ \mathbf{f}_1^{(1)} &= (0.216573, \quad 0.364495), \qquad \mathbf{m}_1^{(1)} = (0.0134251, \quad 0.0263906),\\ \mathbf{f}_1^{(2)} &= (0.240737, \quad 0.381594), \qquad \mathbf{m}_1^{(2)} = (0.00202023, \quad 0.00385997),\end{aligned}$$

$$\begin{aligned}\mathbf{f}_2 &= (0.49141, \quad 0.50859), \qquad \mathbf{m}_2 = (0.483488, \quad 0.516512),\\ \mathbf{f}_2^{(1)} &= (0.221901, \quad 0.165864), \quad \mathbf{m}_2^{(1)} = (0.034574, \quad 0.0284277),\\ \mathbf{f}_2^{(2)} &= (0.117678, \quad 0.103332), \quad \mathbf{m}_2^{(2)} = (0.0144472, \quad 0.0118866).\end{aligned}$$

Applying them in vector mean formula (4.30) gives the means of the stationary number of customers as ,

$$\mathbf{q}_1^{(1)} = (0.196923, \quad 0.33781), \qquad \mathbf{q}_2^{(1)} = (0.284536, \quad 0.223775).$$

Example 5.2 *Polling model with exhaustive discipline - moderate asymmetric load*
We set the following parameter values:

$$\text{Station 1: } \alpha_1 = 1, \quad \alpha_2 = 3, \quad \beta_1 = 2, \quad \beta_2 = 5, \quad \tau_1 = 0.15, \quad \delta_1 = 30,$$
$$\text{Station 2: } \alpha_1 = 2, \quad \alpha_2 = 5, \quad \beta_1 = 3, \quad \beta_2 = 0, \quad \gamma_2 = 10, \quad \delta_2 = 40.$$

The stationary arrival rates are normalized to have $\lambda_1 = 2$, $\lambda_2 = 1$ *and hence* $\rho_1 = 0.3$ *and* $\rho_2 = 0.1$. *The iterative computation returns matrices* $\widetilde{\mathbf{G}}_1(s)$ *and* $\widetilde{\mathbf{G}}_2(s)$ *at* $s = 0$ *as*

$$\widetilde{\mathbf{G}}_1(s)\Big|_{s=0} = \begin{pmatrix} 0.866976 & 0.133024 \\ 0.174611 & 0.825389 \end{pmatrix},$$

$$\widetilde{\mathbf{G}}_2(s)\Big|_{s=0} = \begin{pmatrix} 0.90405 & 0.0959496 \\ 0.0917977 & 0.908202 \end{pmatrix}.$$

For this example $X = 2$ *results in a sufficient precision.*

Based on the steps of the numerical solutions described in Subsection 5.5.2 we get $L^N(X+1)^{N-1} = 12$ *equations for each station. From their solutions we get the following vector factorial moments:*

$$\begin{aligned}
\mathbf{f}_1 &= (0.562999, \quad 0.437001), & \mathbf{f}_2 &= (0.488475, \quad 0.511525), \\
\mathbf{f}_1^{(1)} &= (0.0397073, \quad 0.086862), & \mathbf{f}_2^{(1)} &= (0.053116, \quad 0.0345903), \\
\mathbf{f}_1^{(2)} &= (0.0152189, \quad 0.0266408), & \mathbf{f}_2^{(2)} &= (0.0104823, \quad 0.00958865), \\
\mathbf{m}_1 &= (0.572489, \quad 0.427511), & \mathbf{m}_2 &= (0.486567, \quad 0.513433).
\end{aligned}$$

Applying them in vector mean formula (4.30) gives the means of the stationary number of customers as

$$\mathbf{q}_1^{(1)} = (0.186013, \quad 0.331504), \qquad \mathbf{q}_2^{(1)} = (0.127842, \quad 0.0908439).$$

5.7 Validity scope of the analysis

The proofs of the governing equations (theorems 5.1 and 5.2) utilized both the independency of the switchover times according to assumption **A.3** and properties **P.3** and **P.4** of the model. Therefore, taking into account also the validity scope of the discipline independent results (see Section 4.6), the application of the unified analysis with the two-step methodology requires all the model assumptions (**A.1 - A.3**) and discipline properties (**P.1 - P.4**).

Remark 5.1 *The unified analysis with the two-step methodology can be also applied to the binomial-gated, the binomial-exhaustive and the Bernoulli-gated disciplines, as they satisfy all the required model assumptions (***A.1** - **A.3***) and discipline properties (***P.1** - **P.4***), but it is out of scope of this thesis.*

Chapter 6

System with general disciplines

This chapter deals with the computation of the quantities required for the application of the discipline independent results to polling model with the G-limited and decrementing-K disciplines. These quantities are the vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs, which can be determined on discipline specific way.

In the discipline specific determination of the above- mentioned quantities we apply the method described in chapter 5 for the case with gated discipline. According to this the governing equations of the system are set up in terms of joint PGFs of the stationary number of customers and the phases of the BMAPs at i-polling and i-departure epochs. They are numerically solved by means of the corresponding system of linear equations and afterwards the required quantities at i-polling and i-departure epochs are computed.

6.1 Polling model with G-limited discipline

Theorem 6.1 *(Governing equations of the system.)The governing equations of the stable BMAP/G/1 cyclic nonzero-switchover-times polling model with G-limited service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4**, *are given, for* $i = 1, \dots, N$,

for transition $f_i \to m_i$ *as*

$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^m (n_1, \dots, n_N) z_1^{n_1} \dots z_N^{n_N} \tag{6.1}$$
$$= \sum_{n_1=0}^{\infty} \cdots \sum_{n_i=0}^{K-1} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^f (n_1, \dots, n_N) z_1^{n_1} \dots z_{i-1}^{n_{i-1}} \left(\widehat{\mathbf{A}}_i(z_1, \dots, z_N)\right)^{n_i} z_{i+1}^{n_{i+1}} \dots z_N^{n_N}$$
$$+ \sum_{n_1=0}^{\infty} \cdots \sum_{n_i=K}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^f (n_1, \dots, n_N) z_1^{n_1} \dots z_{i-1}^{n_{i-1}} \left(\widehat{\mathbf{A}}_i(z_1, \dots, z_N)\right)^{K} z_i^{n_i - K} z_{i+1}^{n_{i+1}} \dots z_N^{n_N},$$

for transition $m_i \to f_{i+1}$ *as*

$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_{i+1}^f (n_1, \dots, n_N) z_1^{n_1} \dots z_N^{n_N} \tag{6.2}$$
$$= \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^m (n_1, \dots, n_N) z_1^{n_1} \dots z_N^{n_N} \widehat{\mathbf{U}}_i(z_1, \dots, z_N).$$

Proof. Under G-limited discipline the number of i-customers can be served during the service of station i among those, who already present at i-polling epoch, is limited by K. Hence the i-customers present at i-departure epoch are those who arrived during the service of station i and those who were already present at the i-polling epoch but are not served during the actual service of station i.

Once the number of k-customers for $k = 1, \ldots, N$ and the phases of every BMAP-s at i-polling epoch are given, the simultaneous evolution of k-customers for $k = 1, \ldots, N$ can be described (property **P.3**). The hypermatrix GF of the number of simultaneously arriving k-customers for $k = 1, \ldots, N$ during the interval $(0, t)$, where t is independent of the arrival processes, is given as

$$e^{\widehat{\mathbf{D}}_1(z_1)t} \otimes \ldots \otimes e^{\widehat{\mathbf{D}}_N(z_N)t} = e^{\left(\widehat{\mathbf{D}}_1(z_1) \oplus \ldots \oplus \widehat{\mathbf{D}}_N(z_N)\right)t}.$$

It follows that the hypermatrix GF of the number of simultaneously arriving k-customers for $k = 1, \ldots, N$ during the service of one i-customer can be expressed as

$$\int_0^\infty \left(e^{t \oplus_{k=1}^N \widehat{\mathbf{D}}_k(z_k)}\right) dB_i(t) = \widehat{\mathbf{A}}_i(z_1, \ldots, z_N).$$

Each i-customer, among those present at the i-polling epoch, who is served during the current service of station i generates a random population of k-customers, for $k = 1, \ldots, N$, which arrive during its service time and depend on the phases of the BMAP-s. It follows that if the number of i-customers present at i-polling epoch is given by n_i then the number of customers served during the current service of station i is $min(n_i, K)$, where $min(n_i, K)$ stands for the smallest value of a set $\{n_i, K\}$. Hence the hypermatrix GF of the number of simultaneously arriving k-customers for $k = 1, \ldots, N$ during the service of station i is given as

$$\left(\widehat{\mathbf{A}}_i(z_1, \ldots, z_N)\right)^{min(n_i, K)}.$$

The number of i-customers, among those present at i-polling epoch, but not served during the current service of station i is $max(0, n_i - K)$, where $max(0, n_i - K)$ stands for the largest value of a set $\{0, n_i - K\}$. Now assuming that the number of $1, \ldots, N$-customers present at i-polling epoch is $n_1, \ldots, n_N$, respectively, the hypermatrix GF of the number of $1, \ldots, N$-customers present at i-departure epoch can be expressed as

$$z_1^{n_1} \ldots z_{i-1}^{n_{i-1}} \left(\widehat{\mathbf{A}}_i(z_1, \ldots, z_N)\right)^{min(n_i, K)} z_i^{max(0, n_i - K)} z_{i+1}^{n_{i+1}} \ldots z_N^{n_N}. \tag{6.3}$$

Unconditioning (6.3) leads to the first statement, which describes the transition $f_i \to m_i$ of the G-limited polling model. Applying similar arguments as before yields the second statement, the relation for the transition $m_i \to f_{i+1}$ of the G-limited polling model. □

The stationary probability hypervectors $\mathbf{p}_i^f(n_1,\ldots,n_N)$ and $\mathbf{p}_i^m(n_1,\ldots,n_N)$, for $i=1,\ldots,N$, can be determined from the equations, which can be obtained by setting an upper limit X for $n_1,\ldots,n_N$ in (6.1) and (6.2) and taking their x_1-th, …, x_N-th derivatives ($x_1,\ldots,x_N \in \{0,\ldots,X\}$) at $z_1=\ldots=z_N=1$, respectively. This results in the following system of linear equations for $i=1,\ldots,N$ and $x_1,\ldots,x_N \in \{0,\ldots,X\}$:

$$\sum_{n_1=x_1}^{X}\cdots\sum_{n_N=x_N}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\frac{n_1!}{(n_1-x_1)!}\cdots\frac{n_N!}{(n_N-x_N)!} \tag{6.4}$$
$$=\sum_{n_1=0}^{X}\cdots\sum_{n_i=0}^{min(X,K-1)}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^f(n_1,\ldots,n_N)$$
$$\left.\frac{d^{x_1}\ldots d^{x_N}\left(z_1^{n_1}\ldots z_{i-1}^{n_{i-1}}\left(\widehat{\mathbf{A}}_i(z_1,\ldots,z_N)\right)^{n_i}z_{i+1}^{n_{i+1}}\ldots z_N^{n_N}\right)}{d{z_1}^{x_1}\ldots d{z_N}^{x_N}}\right|_{\mathbf{z}=1}$$
$$+1_{(X\geq K)}\left(\sum_{n_1=0}^{X}\cdots\sum_{n_i=K}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^f(n_1,\ldots,n_N)\right.$$
$$\left.\left.\frac{d^{x_1}\ldots d^{x_N}\left(z_1^{n_1}\ldots z_{i-1}^{n_{i-1}}\left(\widehat{\mathbf{A}}_i(z_1,\ldots,z_N)\right)^{K}z_i^{n_i-K}z_{i+1}^{n_{i+1}}\ldots z_N^{n_N}\right)}{d{z_1}^{x_1}\ldots d{z_N}^{x_N}}\right|_{\mathbf{z}=1}\right),$$

$$\sum_{n_1=x_1}^{X}\cdots\sum_{n_N=x_N}^{X}\mathbf{p}_{i+1}^f(n_1,\ldots,n_N)\frac{n_1!}{(n_1-x_1)!}\cdots\frac{n_N!}{(n_N-x_N)!} \tag{6.5}$$
$$=\sum_{n_1=0}^{X}\cdots\sum_{n_N=0}^{X}\mathbf{p}_i^m(n_1,\ldots,n_N)\left.\frac{d^{x_1}\ldots d^{x_N}\left(z_1^{n_1}\ldots z_N^{n_N}\widehat{\mathbf{U}}_i(z_1,\ldots,z_N)\right)}{d{z_1}^{x_1}\ldots d{z_N}^{x_N}}\right|_{\mathbf{z}=1}.$$

Based on this system of linear equations a numerical method can be developed for computing the stationary probability hypervectors $\mathbf{p}_i^f(n_1,\ldots,n_N)$ and $\mathbf{p}_i^m(n_1,\ldots,n_N)$ for $i=1,\ldots,N$.

6.2 Polling model with decrementing-K discipline

Let $\mathbf{G}_i$ denote the value of $\widetilde{\mathbf{G}}_i(s)$ at $s=0$, i.e. $\mathbf{G}_i=\widetilde{\mathbf{G}}_i(0)$, see the definition of $\widetilde{\mathbf{G}}_i(s)$ at the beginning of Section 5.2.

Theorem 6.2 *(Governing equations of the system.) The governing equations of the stable BMAP/G/1 cyclic nonzero-switchover-times polling model with decrementing-K* *service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4**, *are given, for* $i = 1, \ldots, N$,

for transition $f_i \to m_i$ *as*

$$\begin{aligned}
&\sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^m (n_1, \ldots, n_N)\, z_1^{n_1} \ldots z_N^{n_N} \\
&= \sum_{n_1=0}^{\infty} \cdots \sum_{n_i=0}^{K-1} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^f (n_1, \ldots, n_N)\, z_1^{n_1} \ldots z_{i-1}^{n_{i-1}} \\
&\Big(\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)\Big)^{n_i} z_{i+1}^{n_{i+1}} \ldots z_N^{n_N} \\
&+ \sum_{n_1=0}^{\infty} \cdots \sum_{n_i=K}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^f (n_1, \ldots, n_N)\, z_1^{n_1} \ldots z_{i-1}^{n_{i-1}} \\
&\Big(\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)\Big)^{K} z_i^{n_i - K} z_{i+1}^{n_{i+1}} \ldots z_N^{n_N},
\end{aligned} \tag{6.6}$$

for transition $m_i \to f_{i+1}$ *as*

$$\begin{aligned}
&\sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_{i+1}^f (n_1, \ldots, n_N)\, z_1^{n_1} \ldots z_N^{n_N} \\
&= \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \mathbf{p}_i^m (n_1, \ldots, n_N)\, z_1^{n_1} \ldots z_N^{n_N} \widehat{\mathbf{U}}_i(z_1, \ldots, z_N).
\end{aligned} \tag{6.7}$$

Proof. Under decrementing-K discipline either the server continues the service of station i until the number of i-customers is decremented by K compared to those present at i-polling epoch or the station i becomes empty before.

The i-customers arriving during the service of station i are considered in the evolution of the system, through matrix $\mathbf{G}_i$, which describes the evolution of the system during the decrement of the number of i-customers by one. Such decrement corresponds to the busy period of an i-customer originally present at the i-polling epoch, because it is a level independent quantity.

However, besides those i-customers who were already present at the i-polling epoch but are not served during the actual service of station i, the number of k-customers for every $k \neq i$ and the phases of every BMAP-s takes part in evolution of the system. Therefore we describe the hypermatrix GF of the number of simultaneously arriving k-customers for every $k = 1, \ldots, i-1, i+1, \ldots N$ and the phase changes of every BMAP-s during the decrement of the number of i-customers by one, which can be expressed as

$$\int_0^{\infty} e^{t \oplus_{k=1}^{i-1} \widehat{\mathbf{D}}_k(z_k)} \otimes d\mathbf{G}_i(t) \otimes e^{t \oplus_{k=i+1}^{N} \widehat{\mathbf{D}}_k(z_k)} = \widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N)$$

The decrement of each i-customer, among those present at the i-polling epoch, who is served during the current service of station i generates a random population of k-

customers, for every $k = 1, \dots, i-1, i+1, \dots N$, which arrive during the current decrement and depend on the phases of every BMAP-s. It follows that if the number of i-customers present at i-polling epoch is given by n_i then the number of i-customer decrements during the current service of station i is $min(n_i, K)$. Hence the hypermatrix GF of the number of simultaneously arriving k-customers for every $k = 1, \dots, i-1, i+1, \dots N$ and the phase changes of every BMAP-s during the service of station i is given as

$$\left(\widehat{\mathbf{H}}_i(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_N)\right)^{min(n_i,K)}.$$

The number of i-customers, among those present at i-polling epoch, but not served during the current service of station i is $max(0, n_i - K)$. Now assuming that the number of $1, \dots, N$-customers present at i-polling epoch is $n_1, \dots, n_N$, respectively, the hypermatrix GF of the number of $1, \dots, N$-customers present at i-departure epoch can be expressed as

$$z_1^{n_1} \dots z_{i-1}^{n_{i-1}} \left(\widehat{\mathbf{H}}_i(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_N)\right)^{min(n_i,K)} z^{max(0,n_i-K)} z_{i+1}^{n_{i+1}} \dots z_N^{n_N}. \tag{6.8}$$

Unconditioning (6.8) leads to the first statement, which describes the transition $f_i \to m_i$ of the decrementing-K polling model. The transition $m_i \to f_{i+1}$ is discipline independent. Hence the second statement, the relation for the transition $m_i \to f_{i+1}$ of the decrementing-K polling model, equals the relation (6.2). □

The stationary probability hypervectors $\mathbf{p}_i^f(n_1, \dots, n_N)$ and $\mathbf{p}_i^m(n_1, \dots, n_N)$, for $i = 1, \dots, N$, can be determined from the following system of linear equations for $i = 1, \dots, N$ and $x_1, \dots, x_N \in \{0, \dots, X\}$:

$$\begin{aligned}
&\sum_{n_1=x_1}^{X} \dots \sum_{n_N=x_N}^{X} \mathbf{p}_i^m(n_1, \dots, n_N) \frac{n_1!}{(n_1-x_1)!} \dots \frac{n_N!}{(n_N-x_N)!} \qquad (6.9)\\
&= \sum_{n_1=0}^{X} \dots \sum_{n_i=0}^{min(X,K-1)} \dots \sum_{n_N=0}^{X} \mathbf{p}_i^f(n_1, \dots, n_N)\\
&\left.\frac{d^{x_1} \dots d^{x_N}\left(z_1^{n_1} \dots z_{i-1}^{n_{i-1}} \left(\widehat{\mathbf{H}}_i(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_N)\right)^{n_i} z_{i+1}^{n_{i+1}} \dots z_N^{n_N}\right)}{dz_1{}^{x_1} \dots dz_N{}^{x_N}}\right|_{\mathbf{z}=1}\\
&+1_{(X \geq K)}\Bigg(\sum_{n_1=0}^{X} \dots \sum_{n_i=K}^{X} \dots \sum_{n_N=0}^{X} \mathbf{p}_i^f(n_1, \dots, n_N)\\
&\left.\frac{d^{x_1} \dots d^{x_N}\left(z_1^{n_1} \dots z_{i-1}^{n_{i-1}} \left(\widehat{\mathbf{H}}_i(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_N)\right)^{K} z_i^{n_i-K} z_{i+1}^{n_{i+1}} \dots z_N^{n_N}\right)}{dz_1{}^{x_1} \dots dz_N{}^{x_N}}\right|_{\mathbf{z}=1}\Bigg),
\end{aligned}$$

$$\sum_{n_1=x_1}^{X} \cdots \sum_{n_N=x_N}^{X} \mathbf{p}_{i+1}^{f}(n_1,\ldots,n_N) \frac{n_1!}{(n_1-x_1)!} \cdots \frac{n_N!}{(n_N-x_N)!} \tag{6.10}$$
$$= \sum_{n_1=0}^{X} \cdots \sum_{n_N=0}^{X} \mathbf{p}_{i}^{m}(n_1,\ldots,n_N) \left. \frac{d^{x_1}\ldots d^{x_N}\left(z_1^{n_1}\ldots z_N^{n_N}\widehat{\mathbf{U}}_i(z_1,\ldots,z_N)\right)}{dz_1{}^{x_1}\ldots dz_N{}^{x_N}} \right|_{\mathbf{z}=1} .$$

Again these equations can be obtained by setting an upper limit X for $n_1,\ldots,n_N$ in (6.6) and (6.7) and taking their x_1-th, …, x_N-th derivatives ($x_1,\ldots,x_N \in \{0,\ldots,X\}$) at $z_1 = \ldots = z_N = 1$, respectively.

6.3 Numerical solution

In this Section we summarize the computational procedure for the vector factorial moments of the stationary number of i-customers at i-polling and i-departure epochs ($\mathbf{f}_i^{(n)}$ and $\mathbf{m}_i^{(n)}$ for $n \geq 1$) and characterize its overall numerical complexity. The numerical solution for the systems with the G-limited and with the decrementing-K disciplines are based on the the equations (6.4) and (6.5) as well as (6.9) and (6.10), respectively.

For the system with G-limited discipline the steps of the computational procedure are the same as the ones described for the gated service system in Subsection 5.5.1.

For the system with decrementing-K discipline, for finite K, the first two steps of the computational procedure are the same as the ones for the exhaustive service systems. These are the computation of matrices $\widetilde{\mathbf{G}}_i(s)$ and $\widehat{\mathbf{H}}_i(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_N)$ for every $i = 1,\ldots,N$ as described in Subsection 5.5.2. The remaining steps are taken from the gated service system starting with the step of building up the large hypermatrices of the system. The description of these five steps can be found in Subsection 5.5.1.

According to this the total number of operations required by the whole numerical procedure is in the magnitude of $N^2\ L^{3N}(X+1)^{3N}$ for the system with both the G-limited and the decrementing-K disciplines.

Setting $K = \infty$ in decrementing-K discipline results in the special case of exhaustive discipline. For the polling model with this discipline the computational procedure and the overall numerical complexity are different as discussed in the Subsections 5.5.2 and 5.5.3.

Chapter 7

Special cases

In this chapter BMAP vacation models and classical cyclic polling models are considered as the most important special cases of the BMAP cyclic polling models. Specialized results are provided for all the previously analyzed disciplines, namely for the gated, for the exhaustive, for the G-limited and for the decrementing-K disciplines.

7.1 Vacation models

Setting $N = 1$ in the polling model leads to a vacation model. To reflect the difference in the model throughout this Section the index i is suppressed in the notation. In the vacation model the only switchover time (r) becomes the vacation time, which is denoted by v throughout this Section. Note that according to the model assumption **A.3**, in this vacation model the vacation time is independent of the arrival process and the service time.

7.1.1 Stationary number of customers

In the vacation model the discipline independent relation for the transition $m \to f$ (second part of (5.4)) reduces to

$$\widehat{\mathbf{f}}(z) = \widehat{\mathbf{m}}(z)\widehat{\mathbf{U}}(z). \tag{7.1}$$

Applying (7.1) in the expression of the vector GF of the stationary number of customers at an arbitrary instant (4.19) yields

$$\widehat{\mathbf{q}}(z)\widehat{\mathbf{D}}(z)\left(z\mathbf{I} - \widehat{\mathbf{A}}(z)\right) = \lambda(1-\rho^S)(z-1)\frac{\widehat{\mathbf{m}}(z)(\widehat{\mathbf{U}}(z)-\mathbf{I})}{f^{(1)} - m^{(1)}}\widehat{\mathbf{A}}(z). \tag{7.2}$$

Using (4.18) and the notation v for the vacation time in (7.2) results in the vacation model specific simplified form of the vector GF of the stationary number of customers at an arbitrary instant as

$$\widehat{\mathbf{q}}(z)\widehat{\mathbf{D}}(z)\left(z\mathbf{I} - \widehat{\mathbf{A}}(z)\right) = \frac{\widehat{\mathbf{m}}(z)(\widehat{\mathbf{U}}(z)-\mathbf{I})}{v}(1-\rho)(z-1)\widehat{\mathbf{A}}(z). \tag{7.3}$$

Remark 7.1 *This result was also published in [203].*

Similarly applying the n-th derivatives of (7.1) at $z = 1$ for $n \geq 1$ in (4.22) and using (4.18) leads to the the vacation model specific form of the recursive formula for computing the factorial moments of the stationary number of customers at an arbitrary instant as

$$\begin{aligned}\mathbf{q}^{(n)} = & \frac{(1-\rho)}{v}\sum_{l=0}^{n+1}\sum_{k=0}^{n+1-l}\binom{n+2}{1,l,n+1-k-l,k}\left(\sum_{o=0}^{l}\binom{l}{o}\mathbf{m}^{(l-o)}\mathbf{U}^{o}-\mathbf{m}^{(l)}\right)\mathbf{A}^{(n+1-k-l)} \\ & \frac{[adj\mathbf{T}]^{(k)}}{(1+2n+1_{(n\geq 2)}\binom{n}{2})[det\mathbf{T}]^{(2)}} - \boldsymbol{\pi}\frac{[det\mathbf{T}]^{(n+2)}}{(1+2n+1_{(n\geq 2)}\binom{n}{2})[det\mathbf{T}]^{(2)}} \\ - & \left(1_{(n\geq 3)}\sum_{k=1}^{n-2}\binom{n}{k+2}+1_{(n\geq 2)}\sum_{k=1}^{n-1}\left(\binom{n}{k+1}+\binom{n+1}{k+1}\right)\right)\mathbf{q}^{(n-k)} \\ & \frac{[det\mathbf{T}]^{(k+2)}}{(1+2n+1_{(n\geq 2)}\binom{n}{2})[det\mathbf{T}]^{(2)}} \quad n \geq 1. \end{aligned} \tag{7.4}$$

7.1.2 Discipline specific relations

In this Subsection the governing equations are provided for the vacation models with the gated, the exhaustive, the G-limited and the decrementing-K disciplines. The special case of the BMAP/G/1 queue is also discussed.

Model with gated discipline

The quantities $\mathbf{m}_n$ for $n \geq 0$ are defined by means of the relation $\widehat{\mathbf{m}}(z) = \sum_{n=0}^{\infty}\mathbf{m}_n z^n$ for $|z| \leq 1$.

Corollary 7.1 *(Governing equations of the vacation model with gated discipline.) The governing equation of the stable BMAP/G/1 vacation model with gated service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4***, is given as*

$$\widehat{\mathbf{m}}(z) = \widehat{\mathbf{m}}(\widehat{\mathbf{A}}(z))\widehat{\mathbf{U}}(\widehat{\mathbf{A}}(z)), \tag{7.5}$$

where

$$\widehat{\mathbf{m}}(\widehat{\mathbf{A}}(z)) = \sum_{n=0}^{\infty}\mathbf{m}_n\widehat{\mathbf{A}}(z)^n \quad \textit{and} \quad \widehat{\mathbf{U}}(\widehat{\mathbf{A}}(z)) = \sum_{n=0}^{\infty}\mathbf{U}_n\widehat{\mathbf{A}}(z)^n.$$

Proof. Setting the number of stations one in the first part of (5.4) and using (7.1) gives the statement. □

Remark 7.2 *This result was also published in [203]. Furthermore as the gated discipline is the special case of the binomial-gated one, an alternative solution of this model can be found in [201], in which closed-form result is derived for the BMAP/G/1 vacation model with binomial-gated discipline.*

Model with exhaustive discipline

We introduce the notation

$$\mathbf{G} = \int_0^\infty d\mathbf{G}(t). \tag{7.6}$$

Corollary 7.2 *(Governing equations of the vacation model with exhaustive discipline.) The governing equation of the stable BMAP/G/1 vacation model with exhaustive service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4**, *is given as*

$$\mathbf{m} = \mathbf{m}\widehat{\mathbf{U}}(\mathbf{G}), \tag{7.7}$$

where

$$\widehat{\mathbf{U}}(\mathbf{G}) = \sum_{n=0}^{\infty} \mathbf{U}_n \mathbf{G}^n.$$

Proof. In case of this discipline, no customer present at start of vacation, i.e. $\widehat{\mathbf{m}}(z) = \mathbf{m}$. Taking it into account after setting the number of stations one in the first part of (5.11), using (7.1) and (7.6) gives the statement. □

Corollary 7.3 *For the stable BMAP/G/1 queue, the solution of system of linear equations (7.7) and* $\mathbf{me} = 1$ *for* $\mathbf{m}$ *is given explicitly as*

$$\mathbf{m} = \mathbf{e_L}\Big(\left(\mathbf{I} - \widehat{\mathbf{U}}\left(\mathbf{G}\right)\right) \mid\mid \mathbf{e}\Big)^{-1}. \tag{7.8}$$

Proof. Due to stability of the model, (7.7) and $\mathbf{me} = 1$ uniquely determine $\mathbf{m}$, which implies that matrix $\left(\mathbf{I} - \widehat{\mathbf{U}}\left(\mathbf{G}\right)\right)$ has rank $L - 1$. Let $\mathbf{Y} \mid\mid \mathbf{x}$ denote the matrix $\mathbf{Y}$ with the last column replaced by the column vector $\mathbf{x}$. With this notation the solution of the system of linear equations (7.7) and $\mathbf{me} = 1$ for $\mathbf{m}$ is given by (7.8). □

Remark 7.3 *Another numerical solution is provided for this model by Chang and Takine in [33]. An alternative derivation of the results (7.7) and (7.8) can be found in [201], in which this model with the exhaustive discipline is treated as a special case of the BMAP/G/1 vacation model with binomial-exhaustive discipline.*

Model with G-limited discipline

The quantities $\mathbf{f}_n$ for $n \geq 0$ are defined by means of the relation $\widehat{\mathbf{f}}(z) = \sum_{n=0}^{\infty} \mathbf{f}_n z^n$ for $|z| \leq 1$. Similarly the definitions of $\mathbf{U}_n$ for $n \geq 0$ are given by the relation $\widehat{\mathbf{U}}(z) = \sum_{n=0}^{\infty} \mathbf{U}_n z^n$ for $|z| \leq 1$.

Using the convolution form of (7.1) it can be rearranged as

$$\mathbf{f}_n = \sum_{k=0}^{n} \mathbf{m}_k \mathbf{U}_{n-k}, \quad n \geq 0. \tag{7.9}$$

Corollary 7.4 *(Governing equations of the vacation model with G-limited discipline.) The governing equation of the stable BMAP/G/1 vacation model with G-limited service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4**, *is given as*

$$\sum_{n=0}^{\infty} \mathbf{p}^m(n) z^n = \sum_{n=0}^{K-1} \sum_{k=0}^{n} \mathbf{m}_k \mathbf{U}_{n-k} \left(\widehat{\mathbf{A}}(z)\right)^n + \sum_{n=K}^{\infty} \sum_{k=0}^{n} \mathbf{m}_k \mathbf{U}_{n-k} \left(\widehat{\mathbf{A}}(z)\right)^K z^{n-K}. \tag{7.10}$$

Proof. Setting the number of stations one in (6.1) and using (7.9) gives the statement.
□

Remark 7.4 *This result can be also found in [203].*

Model with decrementing-K discipline

Corollary 7.5 *(Governing equations of the vacation model with decrementing-K discipline.) The governing equation of the stable BMAP/G/1 vacation model with decrementing-K service discipline, supposing that the model satisfies assumptions* **A.1** - **A.3** *and properties* **P.1** - **P.4**, *is given as*

$$\sum_{n=0}^{\infty} \mathbf{p}^m(n) z^n = \sum_{n=0}^{K-1} \sum_{k=0}^{n} \mathbf{m}_k \mathbf{U}_{n-k} (\mathbf{G})^n + \sum_{n=K}^{\infty} \sum_{k=0}^{n} \mathbf{m}_k \mathbf{U}_{n-k} (\mathbf{G})^K z^{n-K}. \tag{7.11}$$

Proof. Setting the number of stations one in (6.6) and using (7.9) gives the statement.
□

Remark 7.5 *This result was presented in [204].*

Special case of BMAP/G/1 queue

There is a huge literature on the solution of the BMAP/G/1 queue, see e.g. the tutorial of Lucantoni [115]. In the next we provide an alternative solution based on the unified analysis framework of the thesis and utilizing that the BMAP/G/1 queue is a special case of the BMAP/G/1 polling model.

The BMAP/G/1 queue can be considered as the special case of the BMAP/G/1 vacation model with exhaustive service, in which the vacation time is the idle period. In this case the independency of the vacation time according to assumption **A.3** does not hold, since the idle period depends on the arrival process. However, as it is mentioned in Section 4.6, this independency was not used in proofs of the service discipline independent statements. Therefore all the discipline independent results and the result (7.7), except the definition of $\widehat{\mathbf{U}}(z)$ (according to (5.3), hold also for the BMAP/G/1 queue. Thus the solution of the system of linear equations (7.7) and $\mathbf{me} = 1$ for $\mathbf{m}$ for the BMAP/G/1 queue requires the determination of $\widehat{\mathbf{U}}(z)$ for the idle period.

Corollary 7.6 *For the stable BMAP/G/1 queue, the solution of the system of linear equations (7.7) and* $\mathbf{me} = 1$ *for* $\mathbf{m}$ *is given as*

$$\mathbf{m} = \mathbf{e_L}\left(\left(\mathbf{I} - \sum_{n=1}^{\infty} (-\mathbf{D}_0)^{-1} \mathbf{D}_n \mathbf{G}^n \right) || \, \mathbf{e} \right)^{-1}. \tag{7.12}$$

Proof. It can be shown (e.g. by applying (16) of Lucantoni [114]) that for the idle period $\widehat{\mathbf{U}}(z)$ is given explicitly as

$$\widehat{\mathbf{U}}(z) = \mathbf{I} - (\mathbf{D}_0)^{-1} \widehat{\mathbf{D}}(z). \tag{7.13}$$

$\mathbf{U}_n$ for $n \geq 1$ follows directly from (7.13) as

$$\mathbf{U}_n = (-\mathbf{D}_0)^{-1} \mathbf{D}_n, \quad n \geq 1. \tag{7.14}$$

Applying (7.14) in (7.8) gives the corollary.

Corollary 7.7 *For the special case of the stable BMAP/G/1 queue, for which* $\mathbf{A}_0$ *is a product of two vectors, i.e., each row of* $\mathbf{A}_0$ *is a constant multiplication of* $\boldsymbol{\phi}$ *and* $\boldsymbol{\phi}\mathbf{e} = 1$*, the solution of system of linear equations (7.7) and* $\mathbf{me} = 1$ *for* $\mathbf{m}$ *is given explicitly as*

$$\mathbf{m} = \boldsymbol{\phi}. \tag{7.15}$$

Proof. In this special case matrix G is given explicitly as $\mathbf{G} = \mathbf{e}\boldsymbol{\phi}$, and $\mathbf{G}^n = \mathbf{G}$ (see Liu and Zhao [110] and Riska and Smirni [144]). Applying them in (7.7) and using that $\mathbf{U}$ is stochastic gives the corollary.

7.2 Polling models with Poisson arrival process

Applying Poisson arrival processes as special BMAPs in the polling model leads to the classical cyclic M/G/1 polling model as special case of the BMAP/G/1 polling model.

7.2.1 Stationary number of i-customers

For this model with Poisson arrival process $\widehat{\mathbf{q}}_i(z)$, $\widehat{\mathbf{f}}_i(z)$ and $\widehat{\mathbf{m}}_i(z)$ become $\widehat{q}_i(z)$, $\widehat{f}_i(z)$ and $\widehat{m}_i(z)$, respectively. Furthermore this specialization implies the settings $\widehat{\mathbf{D}}_i(z) = -(\lambda_i - \lambda_i z)$, $\widehat{\mathbf{A}}_i(z) = \widetilde{B}_i(\lambda_i - \lambda_i z)$, $\widehat{\mathbf{T}}_i(z) = \lambda_i(z-1)\left(z - \widetilde{B}_i(\lambda_i - \lambda_i z)\right)$, $\rho_i^S = \rho_i$, $\mathbf{A}_i = 1$, $\mathbf{A}_i^{(k)} = \lambda_i^k b_i^{(k)}$ for $k \geq 1$ and $adj\widehat{\mathbf{T}}_i(z) = 1$. In addition $\frac{1}{[det\mathbf{T}_i]^{(2)}} = \frac{1}{2\lambda_i(1-\rho_i)}$ and $[det\mathbf{T}_i]^{(k)} = -k\lambda_i^k b_i^{(k-1)}$ for $k \geq 3$.

Substituting them into (4.19) and rearranging it leads to the expression of the PGF of the stationary number of i-customers in the classical cyclic $M/G/1$ polling model as

$$\widehat{q}_i(z) = \frac{(1-\rho_i)(1-z)\widetilde{B}_i(\lambda_i - \lambda_i z)}{\left(z - \widetilde{B}_i(\lambda_i - \lambda_i z)\right)} \frac{\widehat{f}_i(z) - \widehat{m}_i(z)}{(f_i^{(1)} - m_i^{(1)})(1-z)}. \tag{7.16}$$

Remark 7.6 *Alternative derivations of the expression (7.16) can be found e.g. in Borst and Boxma [19] or in [204].*

It is well known that in this classical queueing model $\widehat{q}_i^a(z) = \widehat{q}_i^d(z)$ (see e.g. chap. 5 in Kleinrock [80]) and due to PASTA (see Wolff [189]) $\widehat{q}_i^a(z) = \widehat{q}_i(z)$. Hence (7.16) gives also the expression of the PGF of the stationary number of i-customers at i-customer arrival epochs and at i-customer departure epochs in the classical cyclic $M/G/1$ polling model.

Similarly applying the settings for the special case of Poisson arrival process in (4.22) and rearranging it yields the expression of the factorial moments of the stationary number of i-customers in the classical cyclic $M/G/1$ polling model as

$$\begin{aligned} q_i^{(n)} = {} & \frac{(n+2)\lambda_i^{n+1} b_i^{(n+1)}}{2(1+2n+1_{(n\geq 2)}\binom{n}{2})(1-\rho_i)} \\ & + \sum_{l=1}^{n} \binom{n+2}{1, l, n+1-l} \frac{\lambda_i^{n+1-l} b_i^{(n+1-l)}}{(1+2n+1_{(n\geq 2)}\binom{n}{2})} \frac{\left(f_i^{(l)} - m_i^{(l)}\right)}{2(f_i^{(1)} - m_i^{(1)})} \\ & + \frac{(n+2)}{(1+2n+1_{(n\geq 2)}\binom{n}{2})} \frac{\left(f_i^{(n+1)} - m_i^{(n+1)}\right)}{2(f_i^{(1)} - m_i^{(1)})} \\ & + \left(1_{(n\geq 3)} \sum_{k=1}^{n-2} \binom{n}{k+2} + 1_{(n\geq 2)} \sum_{k=1}^{n-1} \left(\binom{n}{k+1} + \binom{n+1}{k+1}\right)\right) \\ & \frac{(k+2)\lambda_i^{k+1} b_i^{(k+1)}}{2(1+2n+1_{(n\geq 2)}\binom{n}{2})(1-\rho_i)} \, q_i^{(n-k)} \quad n \geq 1. \end{aligned} \tag{7.17}$$

Remark 7.7 *(Dependency of the factorial moments of the stationary number of i-customers on service discipline specific terms in polling model with Poisson arrivals and in polling model with BMAPs.) In the classical cyclic polling model with Poisson arrivals* $q_i^{(n)}$, *for* $n \geq 1$, *depends on the applied service discipline via* $f_i^{(k)} - m_i^{(k)}$ *for* $k = 1, \ldots, n+1$. *In the polling model with BMAPs in the corresponding relation (4.22)* $\mathbf{q}_i^{(n)}$, *for* $n \geq 1$, *depends on* $(\mathbf{f}_i^{(k)} - \mathbf{m}_i^{(k)})$ *for* $k = 1, \ldots, n+1$ *and* $(\mathbf{f}_i - \mathbf{m}_i)$, *where* $(\mathbf{f}_i - \mathbf{m}_i)\mathbf{e} = 0$.

Setting $n = 1$ in (7.17) leads to the expression of the mean stationary number of i-customers in the classical cyclic M/G/1 polling model as

$$q_i^{(1)} = \rho_i + \frac{\lambda_i^2 b_i^{(2)}}{2(1-\rho_i)} + \frac{f_i^{(2)} - m_i^{(2)}}{2\left(f_i^{(1)} - m_i^{(1)}\right)}. \tag{7.18}$$

7.2.2 Discipline specific relations

In the polling model with Poisson arrival processes $\mathbf{p}_i^f(n_1, \ldots, n_N)$, $\mathbf{p}_i^m(n_1, \ldots, n_N)$, $\widehat{\mathbf{f}}_i(z_1, \ldots, z_N)$ and $\widehat{\mathbf{m}}_i(z_1, \ldots, z_N)$ become $p_i^f(n_1, \ldots, n_N)$, $p_i^m(n_1, \ldots, n_N)$, $\widehat{f}_i(z_1, \ldots, z_N)$ and $\widehat{m}_i(z_1, \ldots, z_N)$, respectively. Furthermore this specialization implies the settings $\widehat{\mathbf{A}}_i(z_1, \ldots, z_N) = \widetilde{B}_i\left(\sum_{k=1}^N \lambda_k - \lambda_k z_k\right)$, $\widehat{\mathbf{U}}_i(z_1, \ldots, z_N) = \widetilde{R}_i\left(\sum_{k=1}^N \lambda_k - \lambda_k z_k\right)$ and $\widehat{\mathbf{H}}_i(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_N) = \widetilde{H}_i\left(\sum_{k=1,k\neq i}^N \lambda_k - \lambda_k z_k\right)$. Here $\widetilde{H}_i(s)$ is the LST of the busy period at station i of the M/G/1 polling model, for which
$\widetilde{H}_i(s) = \widetilde{B}_i\left(s + \lambda_i - \lambda_i \widetilde{H}_i(s)\right)$ (see in Kleinrock [80]).

The discipline specific governing equations for the classical cyclic polling model with Poisson arrival processes can be obtained by substituting the specialized settings into the governing equations for the BMAP/G/1 polling model with the same discipline. We provide them for the classical cyclic polling model with the gated, the exhaustive, the G-limited and the decrementing-K disciplines.

Gated service system

$$\widehat{m}_i(z_1, \ldots, z_N) = \widehat{f}_i\left(z_1, \ldots, z_{i-1}, \widetilde{B}_i\left(\sum_{k=1}^N \lambda_k - \lambda_k z_k\right), z_{i+1}, \ldots, z_N\right). \tag{7.19}$$

$$\widehat{f}_{i+1}(z_1, \ldots, z_N) = \widehat{m}_i(z_1, \ldots, z_N)\, \widetilde{R}_i\left(\sum_{k=1}^N \lambda_k - \lambda_k z_k\right). \tag{7.20}$$

Exhaustive service system

$$\widehat{m}_i(z_1,\dots,z_N)=\widehat{f}_i\left(z_1,\dots,z_{i-1},\widetilde{H}_i\left(\sum_{k=1,k\neq i}^{N}\lambda_k-\lambda_k z_k\right),z_{i+1},\dots,z_N\right). \tag{7.21}$$

$$\widehat{f}_{i+1}(z_1,\dots,z_N)=\widehat{m}_i(z_1,\dots,z_N)\,\widetilde{R}_i\left(\sum_{k=1}^{N}\lambda_k-\lambda_k z_k)\right). \tag{7.22}$$

System with G-limited discipline

Transition $f_i \to m_i$:

$$\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}p_i^m(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N} \tag{7.23}$$

$$=\sum_{n_1=0}^{\infty}\dots\sum_{n_i=0}^{K-1}\dots\sum_{n_N=0}^{\infty}p_i^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_{i-1}^{n_{i-1}}\widetilde{B}_i\left(\sum_{k=1}^{N}\lambda_k-\lambda_k z_k)\right)^{n_i}z_{i+1}^{n_{i+1}}\dots z_N^{n_N}$$

$$+\sum_{n_1=0}^{\infty}\dots\sum_{n_i=K}^{\infty}\dots\sum_{n_N=0}^{\infty}p_i^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_{i-1}^{n_{i-1}}\widetilde{B}_i\left(\sum_{k=1}^{N}\lambda_k-\lambda_k z_k)\right)^{K}z_i^{n_i-K}z_{i+1}^{n_{i+1}}\dots z_N^{n_N}.$$

Transition $m_i \to f_{i+1}$:

$$\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}p_{i+1}^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N} \tag{7.24}$$

$$=\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}p_i^m(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N}\widetilde{R}_i\left(\sum_{k=1}^{N}\lambda_k-\lambda_k z_k)\right).$$

System with decrementing-K discipline

Transition $f_i \to m_i$:

$$\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}p_i^m(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N} \tag{7.25}$$

$$=\sum_{n_1=0}^{\infty}\dots\sum_{n_i=0}^{K-1}\dots\sum_{n_N=0}^{\infty}p_i^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_{i-1}^{n_{i-1}}\widetilde{H}_i\left(\sum_{k=1,k\neq i}^{N}\lambda_k-\lambda_k z_k\right)^{n_i}z_{i+1}^{n_{i+1}}\dots z_N^{n_N}$$

$$+\sum_{n_1=0}^{\infty}\dots\sum_{n_i=K}^{\infty}\dots\sum_{n_N=0}^{\infty}p_i^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_{i-1}^{n_{i-1}}\widetilde{H}_i\left(\sum_{k=1,k\neq i}^{N}\lambda_k-\lambda_k z_k\right)^{K}z_i^{n_i-K}z_{i+1}^{n_{i+1}}\dots z_N^{n_N}.$$

Transition $m_i \to f_{i+1}$:

$$\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}p_{i+1}^f(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N} \tag{7.26}$$

$$=\sum_{n_1=0}^{\infty}\dots\sum_{n_N=0}^{\infty}p_i^m(n_1,\dots,n_N)\,z_1^{n_1}\dots z_N^{n_N}\widetilde{R}_i\left(\sum_{k=1}^{N}\lambda_k-\lambda_k z_k)\right).$$

Remark 7.8 *For an alternative derivations of the governing equations for the classical cyclic polling model with the gated and the exhaustive disciplines ((7.19), (7.20) and (7.21), (7.22)) the reader is referred to Takagi [167].*

Remark 7.9 *The classical cyclic polling models with the G-limited and the decrementing-K disciplines were previously solved by using the PSA method only for Markovian (e.g. Coxian distributed) customer service times and switchover times (see Blanc [13]). Hence the results (7.23), (7.24), (7.25) and (7.26) are new as they are valid for the model with general customer service times and switchover times. Due to the Poisson arrivals the only computational intensive part of the numerical procedure is the solution of the system of linear equations. Hence the total number of operations required by the numerical solution is in the magnitude of* $N\ (X+1)^{3N} + N^2$*. The comparison of the applicability of the current method and PSA for Markovian customer service times and switchover times is left for future task.*

Chapter 8

Conclusions

In this chapter we summarize the methodological contributions and the new results presented in the thesis. It is followed by a brief discussion of the application of the results. Finally we identify several open questions for future research.

8.1 Summary of contributions

In this thesis a methodology is proposed for the analysis of the cyclic polling model with BMAP. Using it general and discipline specific new results are obtained. The general results are valid for a broad class of service disciplines. Discipline specific results are presented for the gated, for the exhaustive, for the G-limited and for the decrementing-K disciplines. The model assumptions and discipline properties, which are necessary for the application of the methodology are also specified.

8.1.1 Methodological contributions

The first contribution in methodology is the applied stability methodology, which is based on identification of 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 like the work of Fricker and Jaïbi in [56]).

The second contribution is the proposed two-step methodology, in which the analysis is separated into two parts based on quantities at polling and departure epochs. Although this methodology incorporates several elements of the matrix analytic-methods, it can be seen rather as the generalization of the method used for the analysis of the classical cyclic polling model by Borst and Boxma in [19]. The application of the two-step methodology opens the way for analyzing polling models with BMAPs as such queueing models have not yet been previously analyzed. Treating the two major analytical difficulties of BMAP polling models (discussed in Section 1.5) separately results in simplification in the overall analysis. In the service discipline independent part the closed-form of the established factorization results makes the analysis considerably easier. In the service discipline dependent part the simplified description of the system dynamic due to its necessity only at polling and departure epochs results in a simpler mathematical structure compared to the possible descriptions at other system epochs (like e.g. at customer departure times) or to the application of other methods (like e.g. the supplementary variable technique).

In the course of deriving closed-form factorization results, the argument of Eisenberg [47] has been also generalized to the case of BMAP. The buffer occupancy method is also generalized to the model with BMAP. Besides of utilizing the Kronecker product notation the system of linear equations are derived for the joint probabilities of the stationary number of customers and the phases of the BMAPs at polling and departure epochs by taking the appropriate derivatives of the governing equations of the system. This element of the solution is new compared to the classical buffer occupancy method. It results in more comprehensive applicability of the method and hence it can be applied to broad class of service disciplines including e.g. the G-limited or the decrementing-K disciplines. However the numerical solution of the above-mentioned system of linear equations has higher computational need when the server utilization is high, the number of BMAP phases is high or the system is large, as it is discussed in Subsection 5.5.3. In some extent it can be considered as the "price" of the applicability of the method to broad class of service disciplines.

8.1.2 New results

New results are obtained in the stability analysis and in the stationary analysis of the number of customers in the system.

The new stability results include the characterization of global stability, overview of stability regions of a particular station, order of instability of stations, conditions for partial stability and the necessary and sufficient conditions for the stability states of the system. The discipline specific forms of the condition for the whole stability of the system are also given for the gated, for the exhaustive and for the G-limited disciplines.

In the discipline independent part of the stationary analysis, the main results are the new expression of the vector GF of the stationary number of customers at an arbitrary epoch. This is the most significant theoretical result of the thesis. Based on it the rather practical new formula is derived for the vector factorial moments of the stationary number of customers at an arbitrary epoch. Moreover further formulas determining the vector GF of the stationary number of customers at customer departure and arrival epochs are also presented. For the vector factorial moments of the stationary number of customers at customer departure and arrival epochs relations are established, which relate them to the proper set of the vector factorial moment of the stationary number of customers at an arbitrary instant.

In the discipline specific part of the stationary analysis, the vector factorial moments of the stationary number of customers at polling and departure epochs are determined in the polling model with gated, exhaustive, G-limited and decrementing-K disciplines. Numerical procedures are provided for determining these quantities, which are needed for the application of the discipline independent results to these models.

8.1.3 Model assumptions and discipline properties

The stability methodology can be applied under model assumptions **A.1 - A.3** and discipline properties **P.1 - P.7**. According to this, the general stability results are valid for disciplines, which are limited by discipline properties **P.1 - P.7**. This allows the unlimited and the gated-type limited disciplines. We remark here that the stability methodology can not be applied to the exhaustive-type limited disciplines (see in Subsectiom1.1.2) in its current form as they do not satisfy the discipline property **P.6**.

The two-step methodology of the stationary analysis is applied under model assumptions **A.1 - A.3** and discipline properties **P.1 - P.4**. In fact, as it is mentioned in Section 4.6, the discipline independent part of the stationary analysis is valid under even more general setting, which is limited only by discipline properties **P.1 - P.2**. Moreover this part of the analysis holds also for the zero-switchover-times polling models. The discipline specific part of the stationary analysis can be applied for disciplines, which are limited by discipline properties **P.1 - P.4**. This allows a broad class of service policies including all disciplines discussed in Subsection 1.1.2.

8.1.4 Grouping of the contributions

The contributions of this thesis reflecting to the thesis objectives (Section 1.4) are grouped in a related thesis summary as follows. The first group of theses deals with the stability results and it is based on chapter 3 ([198]). The second group of theses is about the service discipline independent analysis and it is based on the major part of chapter 4 ([199]). The third group of theses discusses the application of the service discipline independent results to the gated, the exhaustive and the G-limited disciplines. This is based on chapter 5 ([199]) and on the first part of chapter 6.

8.2 Application of results

The major application area of the presented results for cyclic polling models with BMAP is at the field of modern telecommunication networks, in which so far e.g. the classical polling models have been applied (see Subsection 1.2.3). This is motivated by achieving more precise queueing model and more accurate results in performance evaluation. Such potential application examples are performance modeling to IEEE 802.11 and IEEE 802.16 BWA systems or analysis of power saving mechanisms.

There are at least two clear ways of exploiting the advanced traffic modeling capabilities of BMAP in the practical applications of the results presented in this thesis. The first one is the fitting of BMAP to correlated traffic models in order to apply them in the performance analysis of the system under consideration. Such correlated traffic models have been elaborated for various data, voice and video traffic types e.g. for simulation based performance analysis of the IEEE 802.16 [159], [187]. In the second case the BMAP is constructed from given traffic parameters to allow traffic characteristics dependent performance evaluation of the studied system.

An important current limitation on the application of the results of the thesis is that both the above-mentioned fitting task and the problem of constructing an arrival process from given traffic characteristics are actually solved only for special classes of BMAPs, e.g. for two-phase MAPs (see Bodrog, Heindl, Horvath and Telek in [15]).

An example for the investigation of the effect of traffic parameters in the performance evaluation of the considered system is the application of BMAP vacation model to IEEE 802.16e sleep mode mechanism ([208], [200]). In the work [200] optimization examples are also presented for determining the optimal sleep mode parameters under different constraints. This includes enforcing an upper bound on mean delay and a cost model taking into account more general Quality of Service (QoS) requirements. These optimizations facilitate the tuning of the sleep mode parameters to the requirements of the actual application scenario and thus they can be applied in network control.

8.3 Future research directions

The methodology and the analysis presented in this thesis may be extended or generalized in several directions towards some potential research topics that are not considered in this thesis.

As already stated, the stability methodology can not be applied to the exhaustive-type limited disciplines including several policies, like e.g. the decrementing-K discipline. This is because in this disciplines the maximum of the mean number of customers that can be served at a station also depends on the phase of BMAP at that station at its polling epoch. It implies that the discipline property **P.6** is not guaranteed. Thus it remains a future research topic to investigate the possible generalizations of the presented stability methodology to the exhaustive-type limited disciplines.

Another challenge is to extend the presented analysis and results to more general polling systems like the periodic polling model or the polling model with Markovian server routing.

The polling model presented in this thesis can be extended to handle also other quantities, like e.g. set-up time or repair time, which requires relaxing several service discipline properties. This is left for future research.

The analysis of the waiting time is usually more complicated than that of the number of customers and it requires more model assumptions. The analysis of waiting time of polling models with BMAP is a challenging and still completely open problem for future research.

Appendix A

Special cases of BMAP

Several useful arrival processes are special cases of BMAP:

- The *Markovian arrival process (MAP)*, defined by D. M. Lucantoni, K. S. Meier-Hellstern and M.F. Neuts [116], is a special BMAP, in which the arrival batch size is 1. For MAP hold $\mathbf{D}_j = \mathbf{0}, j \geq 2$, i.e. it has the following infinitesimal generator:

$$\begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

- The *Markov-modulated Poisson process (MMPP)* (see e.g. in K.S. Meier-Hellstern and W. Fischer [130]) represented by infinitesimal generator matrix $\mathbf{\Sigma}$ of the modulating Markov chain and by the arrival rate matrix $\mathbf{\Lambda} = diag(\lambda_1, \dots, \lambda_m)$ is a special MAP, for which $\mathbf{D}_0 = \mathbf{\Sigma} - \mathbf{\Lambda}$ and $\mathbf{D}_1 = \mathbf{\Lambda}$. Thus the infinitesimal generator of the MMPP is given as

$$\begin{pmatrix} \mathbf{\Sigma} - \mathbf{\Lambda} & \mathbf{\Lambda} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{\Sigma} - \mathbf{\Lambda} & \mathbf{\Lambda} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma} - \mathbf{\Lambda} & \mathbf{\Lambda} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma} - \mathbf{\Lambda} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

  The modulation introduces correlation between the consecutive interarrival times and hence the MMPP is a useful non-renewal process.

- The *phase type renewal process (PH-renewal process)* (M. F. Neuts [136], [137]) represented by $(\boldsymbol{\alpha}, \mathbf{A})$ is a MAP, in which $\mathbf{D}_0 = \mathbf{A}$ and $\mathbf{D}_1 = -\mathbf{A}\mathbf{e}\boldsymbol{\alpha}$. Thus the infinitesimal generator of the PH-renewal process is given as

$$\begin{pmatrix} \mathbf{A} & -\mathbf{A}\mathbf{e}\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{A} & -\mathbf{A}\mathbf{e}\boldsymbol{\alpha} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{A} & -\mathbf{A}\mathbf{e}\boldsymbol{\alpha} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

- The arrival process having *Coxian distributed interarrival times* is a special PH-renewal process. The random variable with Coxian distribution of order k goes through up to at most k exponential phases. It starts at phase 1. After phase n, for $n = 1, \dots, k$ it ends with probability $1 - p_n$ otherwise it jumps into the next phase with probability p_n. For the last phase $p_k = 0$. The Coxian distribution is also called by several papers (e.g. in Bertsimas and Mourtzinou [9]) as mixed generalized Erlang (MGE) distribution.

  Hence $\boldsymbol{\alpha} = (1, 0, \dots, 0)$ and

$$\mathbf{A} = \begin{pmatrix} -\lambda_1 & \lambda_1 p_1 & 0 & \dots & 0 \\ 0 & -\lambda_2 & \lambda_2 p_2 \dots & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda_m \end{pmatrix}.$$

- The Erlang-k and the hyperexponential arrival processes are also special cases of the PH-renewal process. The interarrival time is hyperexponentially distributed if with probability p_i it is distributed exponentially with parameter λ_i for $i = 1, \dots, k$. The Erlang-k distribution is the sum of k independent exponential random variables with the same mean.

  The arrival process, whose interarrival time distribution is a finite mixture of the above ones, is also a special cases of the PH-renewal process. An example for it is e.g. the *mixed Erlang distribution of order k*, which is with probability p_n, for $n = 1, \dots, k$, the sum of n exponentials with the same mean.

  Hence $\boldsymbol{\alpha} = (p_1, \dots, p_k)$ and

$$\mathbf{A} = \begin{pmatrix} -\lambda & 0 & 0 & \dots & 0 \\ \lambda & -\lambda & 0 \dots & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda & -\lambda \end{pmatrix}.$$

- The Poisson process is a very special MAP with only 1 phase, for which $\mathbf{D}_0 = -\lambda$ and $\mathbf{D}_1 = \lambda$.

Appendix B

Properties of model specific key matrices

In the following we establish the properties of model specific key matrices $\widehat{\mathbf{D}}_i(z)$ and $\left(z\mathbf{I}-\widehat{\mathbf{A}}_i(z)\right)$ at $z=1$.

Lemma B.1 *(First derivative of determinant.) The following relation holds for the determinant of the $L \times L$ matrix $\mathbf{Y}(z)$, which is differentiable when $|z| \leq 1$:*

$$\frac{d\ det\mathbf{Y}(z)}{dz} = Tr(adj\mathbf{Y}(z)\frac{d\mathbf{Y}}{dz}), \quad |z| \leq 1. \tag{B.1}$$

Proof. Jacobi's formula expresses the differential of the determinant of matrix $\mathbf{Y}(z)$ as

$$d\ det\mathbf{Y}(z) = Tr(adj\mathbf{Y}(z)\,d\mathbf{Y}). \tag{B.2}$$

Dividing (B.2) by dz results in the lemma. □

Lemma B.2 *(Properties of $adj\mathbf{D}_i$ and $adj(\mathbf{I}-\mathbf{A}_i)$.) Each row of $adj\mathbf{D}_i$ is the same and it differs from $\boldsymbol{\pi}_i$ only in a multiplication constant. Similarly, each row of $adj(\mathbf{I}-\mathbf{A}_i)$ is the same and it differentiates from $\boldsymbol{\pi}_i$ only in a multiplication constant. In other words, there exist real values $\kappa_1 \neq 0$ and $\kappa_2 \neq 0$ so, that*

$$\begin{aligned} adj\mathbf{D}_i &= \kappa_1 \mathbf{e}\boldsymbol{\pi}_i, \\ adj(\mathbf{I}-\mathbf{A}_i) &= \kappa_2 \mathbf{e}\boldsymbol{\pi}_i. \end{aligned} \tag{B.3}$$

Proof. $rank(\mathbf{D}_i) = L-1$ and so $\mathbf{D}_i$ is singular. Hence $adj\mathbf{D}_i\ \mathbf{D}_i = det\mathbf{D}_i = 0$ and every rows of $adj\mathbf{D}_i$ can differ from $1 \times L$ size row vector $\mathbf{x_r}$ only in a multiplication constant, where $\mathbf{x_r}\mathbf{D}_i = 0$. Similarly $\mathbf{D}_i\ adj\mathbf{D}_i = det\mathbf{D}_i = 0$ and every columns of $adj\mathbf{D}_i$ can differ from $L \times 1$ size column vector $\mathbf{x_c}$ only in a multiplication constant, where $\mathbf{D}_i\mathbf{x_c} = 0$. Additionally $\boldsymbol{\pi}_i\mathbf{D}_i = 0$ and $\mathbf{D}_i\mathbf{e} = 0$. It follows that $adj\mathbf{D}_i = \kappa_1\mathbf{e}\boldsymbol{\pi}_i$.

Now we come to the second statement. According to (2.10) $rank(\mathbf{I}-\mathbf{A}_i) = L-1$ and thus $(\mathbf{I}-\mathbf{A}_i)$ is singular. It can be seen from the Taylor expansion of $\mathbf{A}_i$ that $\boldsymbol{\pi}_i\mathbf{A}_i = \boldsymbol{\pi}_i$

and hence $\boldsymbol{\pi}_i (\mathbf{I} - \mathbf{A}_i) = 0$. Furthermore $(\mathbf{I} - \mathbf{A}_i)\mathbf{e} = 0$, since $\mathbf{A}_i$ is stochastic. Using the same argument as before for matrix $(\mathbf{I} - \mathbf{A}_i)$ results in the second statement. □

Let $[det\mathbf{D}_i]^{(1)}$ denote the first derivative of $det\widehat{\mathbf{D}}_i(z)$ at $z = 1$, i.e., $[det\mathbf{D}_i]^{(1)} = \frac{d\left(det\widehat{\mathbf{D}}_i(z)\right)}{dz}\Big|_{z=1}$. Similarly $[det(\mathbf{I} - \mathbf{A}_i)]^{(1)}$ denotes the first derivative of $det\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right)$ at $z = 1$, i.e., $[det(\mathbf{I} - \mathbf{A}_i)]^{(1)} = \frac{d\left(det\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right)\right)}{dz}\Big|_{z=1}$.

Proposition B.1 *(First derivatives of $det\widehat{\mathbf{D}}_i(z)$ and $det\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right)$ at $z = 1$.) The following statements hold for matrices $\widehat{\mathbf{D}}_i(z)$ and $\left(z\mathbf{I} - \widehat{\mathbf{A}}_i(z)\right)$ at $z = 1$:*

$$\begin{aligned} &[det\mathbf{D}_i]^{(1)} \neq 0, \\ &[det(\mathbf{I} - \mathbf{A}_i)]^{(1)} \neq 0. \end{aligned} \tag{B.4}$$

Proof. Applying Lemma B.1 for determinant $det\mathbf{D}_i(z)$ at $z = 1$ leads to

$$[det\mathbf{D}_i]^{(1)} = Tr(adj\mathbf{D}_i\ \mathbf{D}_i^{(1)}). \tag{B.5}$$

Replacing $adj\mathbf{D}_i$ by $\kappa_1\mathbf{e}\boldsymbol{\pi}_i$ in (B.5), where $\kappa_1 \neq 0$ (Lemma B.2), and applying (2.3) yields

$$\begin{aligned} [det\mathbf{D}_i]^{(1)} &= Tr(\kappa_1\mathbf{e}\boldsymbol{\pi}_i\ \mathbf{D}_i^{(1)}) = \sum_{\ell=1}^{L}\sum_{j=1}^{L}\kappa_1[\boldsymbol{\pi}_i]_j\ [\mathbf{D}_i^{(1)}]_{j,\ell} \\ &= \sum_{j=1}^{L}\kappa_1[\boldsymbol{\pi}_i]_j\sum_{\ell=1}^{L}[\mathbf{D}_i^{(1)}]_{j,\ell} = \kappa_1\sum_{j=1}^{L}[\boldsymbol{\pi}_i]_j\mathbf{e}_j\mathbf{D}_i^{(1)}\mathbf{e} = \kappa_1\boldsymbol{\pi}_i\mathbf{D}_i^{(1)}\mathbf{e} = \kappa_1\lambda_i. \end{aligned} \tag{B.6}$$

Using $\kappa_1 \neq 0$ and $\lambda_i > 0$ the first statement comes from (B.6).

Starting with determinant $det(z\mathbf{I} - \widehat{\mathbf{A}}_i(z))$ at $z = 1$ and using the same argument as before, we get:

$$[det(\mathbf{I} - \mathbf{A}_i)]^{(1)} = \kappa_2\boldsymbol{\pi}_i\left(\mathbf{I} - \mathbf{A}_i^{(1)})\right)\mathbf{e}. \tag{B.7}$$

Using (2.9) and the Taylor expansion of $\boldsymbol{\pi}_i\frac{d\widehat{\mathbf{A}}_i(z)}{dz}\Big|_{z=1}\mathbf{e}$ as well as $\boldsymbol{\pi}_i\mathbf{D}_i = 0$, we get

$$\begin{aligned} &\boldsymbol{\pi}_i\frac{d\widehat{\mathbf{A}}_i(z)}{dz}\Big|_{z=1}\mathbf{e} = \boldsymbol{\pi}_i\frac{dE\left(e^{\widehat{\mathbf{D}}_i(z)B_i}\right)}{dz}\Big|_{z=1}\mathbf{e} = \boldsymbol{\pi}_i E\left(\sum_{k=0}^{\infty}\frac{d\left(\widehat{\mathbf{D}}_i(z)^k\right)}{dz}\Bigg|_{z=1}\mathbf{e}\ \frac{B_i^k}{k!}\right) = \\ &E\left(\sum_{k=1}^{\infty}\boldsymbol{\pi}_i\mathbf{D}_i^{k-1}\ \frac{d\widehat{\mathbf{D}}_i(z)}{dz}\Bigg|_{z=1}\mathbf{e}\ \frac{B_i^k}{k!}\right) = \boldsymbol{\pi}_i\mathbf{D}_i^{(1)}\mathbf{e}\ E(B_i) = \lambda_i b_i = \rho_i, \end{aligned}$$

from which

$$\boldsymbol{\pi}_i \left(\mathbf{I} - \mathbf{A}_i^{(1)})\right) \mathbf{e} = 1 - \rho_i. \tag{B.8}$$

Substituting (B.8) into (B.7) yields

$$\left[det\left(\mathbf{I} - \mathbf{A}_i\right)\right]^{(1)} = \kappa_2 \left(1 - \rho_i\right). \tag{B.9}$$

$\kappa_2 \neq 0$ and from stability $(1 - \rho) > 0$ and hence (B.9) gives the second statement. □

The next proposition deals with the properties of the r.h.s. of the expression (4.23):

$$\mathbf{q}_i^{(n)} = lim_{z\to 1} \frac{\frac{d^n \widehat{\mathbf{q}}_i(z)}{dz^n} \widehat{\mathbf{T}}_i\left(z\right) adj \widehat{\mathbf{T}}_i\left(z\right)}{det \widehat{\mathbf{T}}_i\left(z\right)}.$$

Proposition B.2 *(Properties of the r.h.s. of the expression (4.23).) As $lim_{z\to 1}$ the properties of the r.h.s. of (4.23) can be summarized as follows:*

- *the nominator and the denominator of the right hand side of (4.23) and their first derivatives are 0,*
- *the second derivative of the denominator of the right hand side of (4.23) differs from 0.*

Proof. Equations (2.2) and (2.8) imply that $\widehat{\mathbf{D}}_i(z)$ and $\widehat{\mathbf{A}}_i(z)$ are continuously differentiable when $|z| \leq 1$. Using the definition of $\widehat{\mathbf{T}}_i\left(z\right)$ it follows that $\widehat{\mathbf{T}}_i\left(z\right)$, $adj\widehat{\mathbf{T}}_i\left(z\right)$ and $det\widehat{\mathbf{T}}_i\left(z\right)$ are also continuously differentiable. Therefore instead of $lim_{z\to 1}$ we consider the corresponding values at $z = 1$.

In the next first we show that as $lim_{z\to 1}$ both the nominator and the denominator of the r.h.s. of (4.23) are 0. Afterwards we show that as $lim_{z\to 1}$ also the first derivatives of the nominator and the denominator of the r.h.s. of (4.23) are 0.

Due to (2.1) and (2.10) both $\mathbf{D}_i$ and $(\mathbf{I} - \mathbf{A}_i)$ are singular. Hence $det\mathbf{D}_i det(\mathbf{I}-\mathbf{A}_i) = det\mathbf{T}_i = \mathbf{T}_i adj\mathbf{T}_i = 0$, from which follows that as $lim_{z\to 1}$ both the nominator and denominator of the r.h.s. of (4.23) are 0.

Rearranging the first derivative of the denominator of the r.h.s. of (4.23) at $z = 1$ and using $det(\mathbf{I} - \mathbf{A}_i) = 0$ and $det\mathbf{D}_i = 0$ we get

$$[det\mathbf{T}_i]^{(1)} = [det\mathbf{D}_i]^{(1)} det(\mathbf{I} - \mathbf{A}_i) + det\mathbf{D}_i[det(\mathbf{I} - \mathbf{A}_i)]^{(1)} = 0,$$

i.e. as $lim_{z\to 1}$ the first derivative of the denominator of the r.h.s. of (4.23) is 0. Applying $\widehat{\mathbf{T}}_i\,(z)\,adj\widehat{\mathbf{T}}_i\,(z) = det\widehat{\mathbf{T}}_i\,(z)$ in the nominator of the r.h.s. of (4.23) we get $\frac{d^n\widehat{\mathbf{q}}_i(z)}{dz^n}det\widehat{\mathbf{T}}_i\,(z)$. Taking its first derivative as $z \to 1$ and using $[det\mathbf{T}_i]^{(1)} = det\mathbf{T}_i = 0$ leads to

$$\mathbf{q}_i^{(n)}[det\mathbf{T}_i]^{(1)} + \mathbf{q}_i^{(n+1)}det\mathbf{T}_i = 0,$$

i.e. as $lim_{z\to 1}$ the first derivative of the nominator of the r.h.s. of (4.23) is also 0.

Now we investigate the second derivative of the denominator of the r.h.s. of (4.23) as $lim_{z\to 1}$. Taking the second derivative of the denominator of the r.h.s. of (4.23) as $lim_{z\to 1}$, rearranging and applying $det(\mathbf{I} - \mathbf{A}_i) = 0$ and $det\mathbf{D}_i = 0$ we get

$$\begin{aligned} [det\mathbf{T}_i]^{(2)} &= \left.\frac{d^2\left(det\widehat{\mathbf{D}}_i\,(z)\,det(z\mathbf{I} - \widehat{\mathbf{A}}_i(z))\right)}{dz^2}\right|_{z=1} \\ &= 2[det\mathbf{D}_i]^{(1)}[det(\mathbf{I} - \mathbf{A}_i)]^{(1)}. \end{aligned} \tag{B.10}$$

Applying proposition B.1 in (B.10) shows that the second derivative of the denominator of the r.h.s. of (4.23) does not equal 0 as $lim_{z\to 1}$. □

Appendix C

Equivalence of the different definitions of $\widehat{\mathbf{f}}_i(z)$ and $\widehat{\mathbf{m}}_i(z)$

The different definitions of $\widehat{\mathbf{f}}_i(z)$ and $\widehat{\mathbf{m}}_i(z)$ represent pointwise limits and corresponding limiting averages in the Markov chains embedded at i-polling epochs (in the case of $\widehat{\mathbf{f}}_i(z)$) and at i-departure epochs (in the case of $\widehat{\mathbf{m}}_i(z)$) of the polling cycles. Their equivalence follows from the Markov property of these Markov chains.

Let $\{\mathbf{Z}(m); m \in \{1, \ldots\}\}$ a homogeneous discrete-time Markov chain (DTMC) on the state space Ω. We introduce the notations for the probabilities and for the transition probabilities of this Markov chain as $p_\omega(m) = Pr\{\mathbf{Z}(m) \in \omega\}$ for every $\omega \in \Omega$ and $p_{\omega_1,\omega_2} = Pr\{\mathbf{Z}(m+1) \in \omega_2 \mid \mathbf{Z}(m) \in \omega_1\}$ for every $\omega_1, \omega_2 \in \Omega$ and for $m \geq 1$, respectively.

Lemma C.1 *(Equivalence of pointwise limit and the corresponding limiting average in the DTMC.) If* $\lim_{m\to\infty} p_\omega(m)$ *exists for every* $\omega \in \Omega$ *then*

$$\lim_{m\to\infty} p_\omega(m) = \lim_{m\to\infty} \frac{\sum_{\ell=1}^{m} p_\omega(\ell)}{m}. \tag{C.1}$$

Proof. We introduce the following short notations for the pointwise limit and the corresponding limiting average:

$$p_\omega^p = \lim_{m\to\infty} p_\omega(m),$$
$$p_\omega^{la} = \lim_{m\to\infty} \frac{\sum_{\ell=1}^{m} p_\omega(\ell)}{m}.$$

Using the Markov property of the chain in the definition of the pointwise limit leads to

$$p_\omega^p = \lim_{m+1\to\infty} \sum_{\omega_1\in\Omega} p_{\omega_1}(m) p_{\omega_1,\omega} = \sum_{\omega_1\in\Omega} \lim_{m\to\infty} p_{\omega_1}(m) p_{\omega_1,\omega} = \sum_{\omega_1\in\Omega} p_{\omega_1}^p p_{\omega_1,\omega}. \tag{C.2}$$

Similarly for the corresponding limiting average we have

$$p_\omega^{la} = \lim_{m+1\to\infty} \frac{m}{m+1} \frac{p_\omega(1) + \sum_{\ell=2}^{m+1} p_\omega(\ell)}{m} = \lim_{m+1\to\infty} \frac{\sum_{\ell=2}^{m+1} \sum_{\omega_1\in\Omega} p_{\omega_1}(\ell-1) p_{\omega_1,\omega}}{m}$$
$$= \sum_{\omega_1\in\Omega} \lim_{m\to\infty} \frac{\sum_{\ell=1}^{m} p_{\omega_1}(\ell)}{m} p_{\omega_1,\omega} = \sum_{\omega_1\in\Omega} p_{\omega_1}^{la} p_{\omega_1,\omega}. \tag{C.3}$$

The existence of the pointwise limit p_ω^p for every $\omega \in \Omega$ means that the limiting distribution of the Markov chain exists. In this case it is the unique solution of the equilibrium equation of the Markov chain. Using it the lemma comes from (C.2) and (C.3) as they show that both p_ω^p and p_ω^{la} satisfies the equilibrium equation of the Markov chain. □

The equivalence of the different definitions of $\widehat{\mathbf{f}}_i(z)$ ((4.1) and (5.1)) comes by applying lemma C.1 to $\mathbf{Z}(m) = \big(N_1(t_i^f(m)), \ldots, N_N(t_i^f(m)), J_1(t_i^f(m)), \ldots, J_N(t_i^f(m))\big)$ and taking into account $Pr\{N_i(t_i^f(\ell)) = n, J_i(t_i^f(\ell)) = j\} = E\left[1_{\left(N_i(t_i^f(\ell))=n\right)} 1_{\left(J_i(t_i^f(\ell))=j\right)}\right]$ for $n \in \{0, 1, \ldots\}$, $j \in \{1, \ldots, L\}$ and $i = 1, \ldots, N$. Thus m in $\mathbf{Z}(m)$ here is the index of the polling cycles. Similarly the equivalence of the different definitions of $\widehat{\mathbf{m}}_i(z)$ ((4.2) and (5.2)) comes by applying lemma C.1 to $\mathbf{Z}(m) = \big(N_1(t_i^m(m)), \ldots, N_N(t_i^m(m)),$ $J_1(t_i^m(m)), \ldots, J_N(t_i^m(m))\big)$ and taking into account $Pr\{N_i(t_i^m(\ell)) = n, J_i(t_i^m(\ell)) = j\} = E\left[1_{\left(N_i(t_i^m(\ell))=n\right)} 1_{\left(J_i(t_i^m(\ell))=j\right)}\right]$ for $n \in \{0, 1, \ldots\}$, $j \in \{1, \ldots, L\}$ and $i = 1, \ldots, N$.

Note that here the existence of the pointwise limits ($\lim_{m\to\infty} Pr\big\{N_1(t_i^f(m)) = n_1, \ldots, N_N(t_i^f(m)) = n_N, J_1(t_i^f(m)) = j_1, \ldots, J_N(t_i^f(m)) = j_N\big\}$ and $\lim_{m\to\infty}$ $Pr\big\{N_1(t_i^m(m)) = n_1, \ldots, N_N(t_i^m(m)) = n_N, J_1(t_i^m(m)) = j_1, \ldots, J_N(t_i^m(m)) = j_N\big\}$ for $n_1, \ldots, n_N \in \{0, 1, \ldots\}$, $j_1, \ldots, j_N \in \{1, \ldots, L\}$ and $i = 1, \ldots, N$) follows from the stability of the polling system as described in Subsection 1.1.6.

Appendix D

Proof of lemma 4.1

Let $\Lambda_i(\ell,n,k) = \Lambda_i\left(t_i(\ell,n,k)\right) - \Lambda_i\left(t_i(\ell,n-1)\right)$ for $\ell \geq 1$, $n = 1,\ldots,G_i(\ell)+1$ and $k = 1,\ldots,A_i^*(\ell,n)$. Thus $\Lambda_i(\ell,n,k)$ is the number of i-BMAP arrivals from the service completion time of the $n-1$-th i-customer to the k-th i-BMAP state changes during the next (the n-th) service time of i-customer in the ℓ-th polling cycle. Furthermore $\Lambda_i\left(\ell, G_i(\ell)+1, k\right)$ is the number of i-BMAP arrivals from the i-departure epoch in the ℓ-th polling cycle to the k-th i-BMAP state changes during the next i-intervisit time. Note that $\Lambda_i(\ell,n,k)$ includes the number of i-customers arrived at the last specified (the k-th) i-BMAP state change in the given interval, since $\Lambda_i(t)$ is right continuous.

By the help of the definitions of $\widehat{\mathbf{q}}_i^{s*}(z)$, a_i^{s*}, $\widehat{\mathbf{q}}_i^{i*}(z)$ and a_i^{i*} the quantities $a_i^{s*}\widehat{\mathbf{q}}_i^{s*}(z)$ and $a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z)$ can be expressed as

$$
\begin{aligned}
a_i^{s*}\widehat{\mathbf{q}}_i^{s*}(z) &= \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)}\sum_{k=1}^{A_i^*(\ell,n)} z^{N_i(t_i(\ell,n,k-1))}\mathbf{1}_{(J_i(t_i(\ell,n,k-1)))}\right]}{m},\\
a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z) &= \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)} z^{N_i(t_i(\ell,G_i(\ell)+1,k-1))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell)+1,k-1)))}\right]}{m}.
\end{aligned}
$$

First we consider the vector $a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z)$. The number of i-customers just before the k-th state change of the i-BMAP during the i-intervisit time in the ℓ-th polling cycle consists of, on one hand, the number of the i-customers left at that i-departure epoch and, on the other hand, the number of the i-customers arrived during the first $k-1$ state transitions of the i-BMAP during that i-intervisit time. It yields

$$
\begin{aligned}
a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z) &= \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m}\sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)} z^{N_i(t_i(\ell,G_i(\ell)+1,k-1))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell)+1,k-1)))}\right]}{m} \qquad \text{(D.1)}\\
&= \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)} z^{\Lambda_i(\ell,G_i(\ell)+1,k-1)}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell)+1,k-1)))}\right]}{m}.
\end{aligned}
$$

Now we insert conditional expectation inside of the expectation on the r.h.s. of (D.1) by applying the same technique as in the proof of proposition 4.1 (from (4.42) to (4.43)). This leads to

$$a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z) = \lim_{m\to\infty}\frac{1}{m}E\Bigg[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\sum_{j=1}^{L} 1_{(J_i(t_i(\ell,G_i(\ell)))=j)} \sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)} \tag{D.2}$$
$$E\left[z^{\Lambda_i(\ell,G_i(\ell)+1,k-1)}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell)+1,k-1)))}|J_i(t_i\left(\ell,G_i(\ell)\right))=j,k\le A_i^*\left(\ell,G_i(\ell)+1\right)\right]\Bigg]$$
$$= \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\sum_{j=1}^{L} 1_{(J_i(t_i(\ell,G_i(\ell)))=j)}\sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)}\mathbf{e}_j\widehat{\mathbf{\Psi}}_i(z)^{k-1}\right]}{m}.$$

Multiplying (D.2) by $\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right)$ from right results in

$$a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z)\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right) \tag{D.3}$$
$$= \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell))))}\sum_{k=1}^{A_i^*(\ell,G_i(\ell)+1)}\widehat{\mathbf{\Psi}}_i(z)^{k-1}\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right)\right]}{m}$$
$$= \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell))))}\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)^{A_i^*(\ell,G_i(\ell)+1)}\right)\right]}{m}.$$

Using again the appropriate conditional expectation and simplifying by $\sum_{j=1}^{L}1_{(J_i(t_i(\ell,G_i(\ell)))=j)}=1$ we rearrange the second term on the r.h.s. of (D.3) as

$$\lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell))))}\widehat{\mathbf{\Psi}}_i(z)^{A_i^*(\ell,G_i(\ell)+1)}\right]}{m} \tag{D.4}$$
$$= \lim_{m\to\infty}\frac{1}{m}E\Bigg[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\sum_{j_1=1}^{L}1_{(J_i(t_i(\ell,G_i(\ell)))=j_1)}\mathbf{e}_{j_1}$$
$$\sum_{j_2=1}^{L}\mathbf{e}_{j_2}^T E\left[z^{\Lambda_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)}\mathbf{1}_{\left(J_i\left(t_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)\right)\right)}|J_i\left(t_i\left(\ell,G_i(\ell)\right)\right)=j_2\right]\Bigg]$$

$$= \lim_{m\to\infty} \frac{1}{m} E\Big[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))} \sum_{j=1}^{L} 1_{(J_i(t_i(\ell,G_i(\ell)))=j)}$$
$$E\left[z^{\Lambda_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)} \mathbf{1}_{\left(J_i\left(t_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)\right)\right)} | J_i\left(t_i\left(\ell,G_i(\ell)\right)\right) = j \right] \Big]$$
$$= \lim_{m\to\infty} \frac{1}{m} E\Big[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))} \sum_{j=1}^{L} 1_{(J_i(t_i(\ell,G_i(\ell)))=j)}$$
$$z^{\Lambda_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)} \mathbf{1}_{\left(J_i\left(t_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)\right)\right)} \Big]$$
$$= \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))+\Lambda_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)} \mathbf{1}_{\left(J_i\left(t_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)\right)\right)} \right]}{m}.$$

The sum in the nominator on the r.h.s. of (D.4) is the number of i-customers at the i-departure epoch in the ℓ-th polling cycle and those who arrive during the next i-intervisit time. This sum equals the number of i-customers at the last i-BMAP state change during that i-intervisit time. Applying it in (D.4) and using (D.3) gives:

$$a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z)\left(\mathbf{I} - \widehat{\mathbf{\Psi}}_i(z)\right) = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))} \mathbf{1}_{(J_i(t_i(\ell,G_i(\ell))))} \right]}{m} \tag{D.5}$$
$$- \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i\left(t_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)\right)} \mathbf{1}_{\left(J_i\left(t_i\left(\ell,G_i(\ell)+1,A_i^*(\ell,G_i(\ell)+1)\right)\right)\right)} \right]}{m}.$$

The number of i-customers and the phase of the i-BMAP at the last i-BMAP state change during the i-intervisit time in the ℓ-th polling cycle equals the number of i-customers and the phase of the i-BMAP at the next i-polling epoch, respectively. Applying it in (D.5) leads to

$$a_i^{i*}\widehat{\mathbf{q}}_i^{i*}(z)\left(\mathbf{I} - \widehat{\mathbf{\Psi}}_i(z)\right) = \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))} \mathbf{1}_{(J_i(t_i(\ell,G_i(\ell))))} \right]}{m} \tag{D.6}$$
$$- \lim_{m\to\infty} \frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell+1,0))} \mathbf{1}_{(J_i(t_i(\ell+1,0)))} \right]}{m}.$$

Taking into account the alternative notations $t_i(\ell, G_i(\ell)) = t_i^m(\ell)$ and $t_i(\ell+1, 0) = t_i^f(\ell+1)$ as well as applying the definitions (4.34) and (4.33) in (D.6) results in the first statement.

Starting from $a_i^{s*}\widehat{\mathbf{q}}_i^{s*}(z)$ and using the same line of arguments analogously until (D.5) we get the following relation:

$$a_i^{s*}\widehat{\mathbf{q}}_i^{s*}(z)\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right) = \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)} z^{N_i(t_i(\ell,n-1))}\mathbf{1}_{(J_i(t_i(\ell,n-1)))}\right]}{m} - \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)} z^{N_i\left(t_i\left(\ell,n,A_i^*(\ell,n)\right)\right)}\mathbf{1}_{\left(J_i\left(t_i\left(\ell,n,A_i^*(\ell,n)\right)\right)\right)}\right]}{m}. \tag{D.7}$$

The number of i-customers at the last i-BMAP state change epoch during the service time of the n-th i-customer in the ℓ-th polling cycle equals the number of i-customers just before the service completion time of the n-th i-customer, i.e it is one more than the number of i-customers seen by the n-th departing i-customer. Additionally the phase of the i-BMAP at the last i-BMAP state change epoch during the service time of the n-th i-customer in the ℓ-th polling cycle equals the phase of the i-BMAP at service completion time of the n-th i-customer. Applying them in (D.7) and rearranging leads to

$$\begin{aligned} a_i^{s*}\widehat{\mathbf{q}}_i^{s*}(z)\left(\mathbf{I}-\widehat{\mathbf{\Psi}}_i(z)\right) &= \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)} z^{N_i(t_i(\ell,n))}\mathbf{1}_{(J_i(t_i(\ell,n)))}\right]}{m} \\ &+ \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,0))}\mathbf{1}_{(J_i(t_i(\ell,0)))}\right]}{m} \\ &- \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m} z^{N_i(t_i(\ell,G_i(\ell)))}\mathbf{1}_{(J_i(t_i(\ell,G_i(\ell))))}\right]}{m} \\ &- \lim_{m\to\infty}\frac{E\left[\sum_{\ell=1}^{m}\sum_{n=1}^{G_i(\ell)} z^{N_i(t_i(\ell,n))+1}\mathbf{1}_{(J_i(t_i(\ell,n)))}\right]}{m}. \end{aligned} \tag{D.8}$$

Taking into account the alternative notations $t_i\left(\ell,G_i(\ell)\right)=t_i^m(\ell)$ and $t_i\left(\ell+1,0\right)=t_i^f(\ell+1)$ as well as applying the necessary definitions in (D.8) results in the second statement. □

Bibliography

References

[1] I. J. B. F. Adan, O. J. Boxma, and J. A. C. Resing. Queuing Models with Multiple Waiting Lines. *Queuing Syst.*, 37(1-3):65–98, 2001.

[2] S. Alouf, E. Altman, and A. Azad. Analysis of an M/G/1 Queue with Repeated Inhomogeneous Vacations with Application to IEEE 802.16e Power Saving Mechanism. In *Fifth International Conference on Quantitative Evaluation of Systems (QEST'08)*, pages 27–36, sep 2008.

[3] E. Altman. Analysing timed-token ring protocols using the power series algorithm. In *Proc. 14th Int. Teletraffic Cong.*, pages 212–226. North-Holland Elsevier Science B.V, jun 1994.

[4] E. Altman and U. Yechiali. Cyclic Bernoulli Polling. *Mathematical Methods of Operations Research*, 38(1):55–76, feb 1993.

[5] E. Altman and U. Yechiali. Polling in a Closed Network. *Prob. Engineering Inform. Sci.*, 8:327–343, 1994.

[6] E. Altman, P. Konstantopoulos, and Z. Liu. Stability, monotonicity and invariant Quantities in General Polling Systems. *Queueing Systems*, 11:35–57, 1992.

[7] J. E. Baker and I. Rubin. Polling with a general service-order table. *IEEE Trans. on Commun.*, COM-35:283–288, mar 1987.

[8] A. D. Banik, U. C. Gupta, and S. S. Pathak. BMAP/G/1/N queue with vacations and limited service discipline. *Applied Mathematics and Computation*, 180:707–721, 2006.

[9] D. Bertsimas and G. Mourtzinou. Decomposition results for general polling systems and their applications. *Queueing Systems*, 31:295–316, 1999.

[10] D. Bertsimas and D. Nakazato. The distributional Little's law and its application. *Operations Research*, 43:298–310, 1995.

[11] J. Blanc and L. Lenzini. Analysis of communication systems with timed token protocols using the power-series algorithm. *Performance Evaluation*, 27-28:391–409, oct 1996.

[12] J. P. C. Blanc. A numerical approach to cyclic-service queueing models. *Queueing Systems*, 6:173–188, 1990.

[13] J. P. C. Blanc. An algorithmic solution of polling models with limited service disciplines. *IEEE Trans. on Comm.*, 40(7):1152–1155, jul 1992.

[14] J. P. C. Blanc. Optimization of Periodic Polling Systems with Non-Preemptive, Time-Limited Service, CentER Discussion Paper, 1996-63. Technical report, 1996.

[15] L. Bodrog, A. Heindl, G. Horváth, and M. Telek. A Markovian Canonical Form of Second-Order Matrix-Exponential Processes. *European Journal of Operational Research*, 190(2):459–477, 2008.

[16] M. A. A. Boon and I. J. B. F. Adan. Mixed gated/exhaustive service in a polling model with priorities. *Queueing Systems*, 63(1-4):383–399, dec 2009.

[17] M. A. A. Boon, I. J. B. F. Adan, and O. J. Boxma. A polling model with multiple priority levels, EURANDOM report 2008-029. Technical report, EURANDOM, 2008.

[18] A. A. Borovkov and R. Schassberger. Ergodicity of polling network. *Stochastic Processes and their Applications*, 50:253–262, 1994.

[19] S. C. Borst and O. J. Boxma. Polling models with and without switchover times. *Operations Research*, 45:536–543, 1997.

[20] O. J. Boxma. Workloads and waiting times in Single-server systems with multiple customer classes. *Queueing Systems*, 5:185–214, 1989.

[21] O. J. Boxma. Analysis and optimization of polling systems. In *Queueing, performance and control in ATM, Proc. 13th Telegr. Congr., Copenhagen Denmark, North-Holland Stud. Telecomm.*, volume 15, 1991.

[22] O. J. Boxma and W. P. Groenendijk. Pseudo-conservation-laws in cyclic queues. *J. Appl. Prob.*, 24:949–964, 1987.

[23] O. J. Boxma and B. Meister. Waiting-time approximations for cyclic-service systems with switchover times. In *Performance '86 and ACM SIGMETRICS 1986, Perform. Eval. Rev., Also in Perform. Eval. 7,4 (Nov.),299-308, 1987.*, volume 14(1), pages 254–262, 1986.

[24] O. J. Boxma, W. P. Groenendijk, and J. A. Weststrate. A pseudoconservation law for service systems with a polling table. In *SIGMETRICS 1989*, page 223, 1989.

[25] O. J. Boxma, H. Levy, and J. A. Weststrate. Optimization of polling systems. Technical report, Cent. Math. and Comp. Sci., Amsterdam, also in Performance'90s, eds. P. J. B. King, I. Mitrani, and R. J. Pooley, Eds. Amsterdam, The Netherlands: North-Holland, 1990, pp.349-361., 1989.

[26] O. J. Boxma, H. Levy, and J. A. Weststrate. Efficient visit frequencies for polling tables: minimization of waiting cost, Report BS-R9111. Technical report, Report Centre for Mathematics and Computer Science, Amsterdam, 1991.

[27] O. J. Boxma, H. Levy, and U. Yechiali. Cyclic reservation schemes for efficient operation of multiple-queue single-server systems. *Annals of Operations Research*, 35:187–208, 1992.

[28] O. J. Boxma, S. Schlegel, and U. Yechiali. Two-Queue Polling Models with a Patient Server. *Annals of Operations Research*, 112(1-4):101–121, apr 2002.

[29] S. Browne and U. Yechiali. Dynamic priority rules for cyclic-type queues. *Adv. Appl. Prob.*, 21:432–450, 1989.

[30] H. Bruneel and B. G. Kim. *Discrete-time models for communication systems including ATM.* Boston: Kluwer, 1993.

[31] W. Bux and H. L. Truong. Mean-delay approximation for cyclic service queueing systems. *Performance Evaluation*, 3(3):187–196, aug 1983.

[32] S. R. Chakravarthy and S. Thiagarajan. Two Parallel Finite Queues with Simultaneous Services and Markovian Arrivals. *J. Appl. Math. Stochastic Anal.*, 10(10): 383–405, 1997.

[33] S. H. Chang and T. Takine. Factorization and stochastic decomposition properties in bulk queues with generalized vacations. *Queueing Systems*, 50:165–183, 2005.

[34] S. H. Chang, T. Takine, K. C. Chae, and H. W. Lee. A unified queue length formula for BMAP/G/1 queue with generalized vacations. *Stochastic Models*, 18: 369–386, 2002.

[35] G. Choudhury and W. Whitt. Computing transient and steady state distributions in polling models by numerical transform inversion. *Perf. Eval.*, 25:267–292, 1996.

[36] E. G. Coffman, A. A. Puhalskii, and M. I. Reiman. Polling systems with zero switch-over times: a heavy-traffic principle. *Ann. Appl. Probab.*, 5:681–719, 1995.

[37] E. G. Coffman, A. A. Puhalskii, and M. I. Reiman. Polling systems in heavy-traffic: a Bessel process limit. *Math. Oper. Res.*, 23:257–304, 1998.

[38] R. B. Cooper. Queues served in cyclic order: waiting times. *The Bell System Technical Journal*, 49:399–413, 1970.

[39] R. B. Cooper and G. Murray. Queues served in cyclic order. *Bell Syst. Tech. J.*, 48(3):675–689, 1969.

[40] R. B. Cooper, S.-C. Niu, and M. M. Srinivasan. A Decomposition Theorem for Polling Models: The Schwitchover Times are Effectively Additive. *Operations Research*, 44:629–633, 1996.

[41] R. B. Cooper, S.-C. Niu, and M. M. Srinivasan. Setups in polling models: does it make sense to set up if no work is waiting? *J. Appl. Probab.*, 36(2):585–592, 1999.

[42] O. Czerniak and U. Yechiali. Fluid polling systems. *Queueing Systems*, 63(1-4): 401–435, dec 2009.

[43] B. T. Doshi. Queueing systems with vacations - a survey. *Queueing Systems*, 1: 29–66, 1986.

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

[45] H. Dror and U. Yechiali. Closed Polling Models with Failing Nodes. *Queuing Syst.*, 35(1-4):55–81, 2000.

[46] I. M. Dukhovny. An approximate model of motion of urban passenger transportation over annular rules. *Eng. Cybern.*, 1:223–231, 1979.

[47] M. Eisenberg. Queues with Periodic Service and Changeover Time. *Operations Research*, 20:440–451, 1972.

[48] D. Everitt. Simple approximations for token rings. *IEEE Transactions on Communications*, COM-34(7):719–721, 1986.

[49] A. Fallahi, E. Hossain, and A. Alfa. QoS and energy trade off in distributed energy-limited mesh/relay networks: a queuing analysis. *IEEE Transactions on Parallel and Distributed Systems*, 17(6), jun 2006.

[50] M. J. Ferguson and Y. J. Aminetzah. Exact results for nonsymmetric token ring systems. *IEEE Trans. Commun.*, COM-33(3):223–231, 1985.

[51] J. M. Ferrandiz. The BMAP/G/1 queue with server set-up times and server vacations. *Adv. Appl. Prob.*, 25:235–254, 1993.

[52] S. Foss and N. Chernova. Dominance theorems and ergodic properties of polling systems. *translated from Problemy Predachi Informatsii*, 32(4):46–71, oct/dec 1996.

[53] S. Foss and A. Kovalevskii. A stability criterion via fluid limits and its application to a polling system. *Queueing Systems*, 32:131–168, 1999.

[54] S. Foss and G. Last. Stability of Polling Systems with Exhaustive Service Policies and State-Dependent Routing. *The Annals of Applied Probability*, 6(1):116–137, 1996.

[55] 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*, pages 220–227, oct 1996.

[56] C. Fricker and M. R. Jaïbi. Monotonicity and stability of periodic polling models. *Queueing Systems*, 15:211–238, 1994.

[57] I. Frigui and A.-S. Alfa. Analysis of a time-limited polling system. *Computer Communications*, 21:558–571, 1998.

[58] S. W. Fuhrmann. Performance analysis of a class of cyclic schedules, Technical Memorandum 81-59531-1. Technical report, Bell Laboratories, 1981.

[59] S. W. Fuhrmann. A Decomposition Result for a Class of Polling Models. *Queueing Systems*, 11:109–120, 1992.

[60] S. W. Fuhrmann and R. B. Cooper. Stochastic Decompositions in the M/G/1 Queue with Generalized Vacations. *Operations Research*, 33:1117–1129, 1985.

[61] L. Georgiadis and W. Szpankowski. Stability of token passing rings. *Queueing Systems*, 11:7–34, 1992.

[62] R. Ghazizadeh and P. Fan. Queuing Analysis of HCCA for Multi-Rate Wireless LANs with Truncated ARQ Protocol. *Wireless Personal Communications, DOI: 10.1007/s11277-009-9822-0*, 2009.

[63] W. P. Groenendijk. Waiting-time approximations for cyclic service systems with mixed service strategies. In M. Bonatti, editor, *Teletraffic Science for New Cost-Effective Systems, Networks and Services, In Proc. ITC 12*, pages 1434–1441. North-Holland, Amsterdam, 1989.

[64] D. Gupta and M. M. Srinivasan. Polling systems with state-dependent setup times. *Queueing Systems*, 22(3-4):403–423, sep 1996.

[65] R. Haan, R. J. Boucherie, and J.-K. van Ommeren. A polling model with an autonomous server. *Queueing Systems*, 62(3):279–308, jul 2009.

[66] J. Hammond and P. O'Reilly. *Performance analysis of Local Computer Networks.* Addison-Wesley, Reading, Mass., 1986.

[67] O. Hashida. Gating multiqueues served in cyclic order. *Systems-Computers-Controls*, 1:1–8, 1970.

[68] D. P. Heymann and M. J. Sobel. *Stochastic models in operations research, Volume I: Stochastic processes and Operating Characteristics.* McGraw-Hill Book Company, 1982.

[69] T. Hirayama. Multiclass polling systems with Markovian feedback. *J Oper Res Soc Jpn*, 48(3):226–255, 2005.

[70] T. Hirayama, S. J. Hong, and M. Krunz. A new approach to analysis of polling systems. *Queueing systems*, 48(1-2):135–158, 2004.

[71] M. Hofri and K. W. Ross. On the optimal control of two queues with server set-up times and its analysis. *Siam J. Comp.*, 16(2):399–419, 1987.

[72] G. Hooghiemstra, M. Keane, and S. van de Ree. Power series for stationary distributions of coupled processor models. *SIAM Journal on Applied Mathematics*, 48:1159–1166, 1988.

[73] A. Horváth. *Approximating non-Markovian Behaviour by Markovian Models. Ph. D. dissertation.* PhD thesis, Department of Telecomunications, Budapest University of Technology and Economics, Hungary, 2003.

[74] W. Y. Jung and C. K. Un. Analysis of Finite-Buffer Polling System with Exhaustive Service Based on Virtual Buffering. *IEEE Transactions on Communications*, 42(12):3144–3149, 1994.

[75] S. Kasahara, T. Takine, Y. Takahashi, and T. Hasegawa. MAP/G/1 queues under N-policy with and without vacations. *Research Society of Japan*, 39(2):188–212, 1996.

[76] T. Katayama and K. Kobayashi. A Pseudoconservation Law for a Time-Limited Service Polling System with Structured Batch Poisson Arrivals. *Computers and Mathematics with Applications*, 51:171–178, 2006.

[77] J. Keilson and L. D. Servi. Oscillating random walk models for GI/G/l vacation systems with Bernoulii schedules. *J. Appl. Prob.*, 23:790–802, sep 1986.

[78] N. Kim, K. Chae, and H. Lee. On the relationships among queue lengths at a random instant,at an arrival,and at a departure in the finite-capacity queue with BMAP arrivals. Tech. Rep. 2000-08. Technical report, IE,KAIST, 2000.

[79] N. K. Kim, K. C. Chae, and M. L. Chaudhry. An Invariance Relation and a Unified Method to Derive Stationary Queue-Length Distributions. *Operations Research*, 52:756–764, 2004.

[80] L. Kleinrock. *Queuing Systems. Vol I: Theory.* John Wiley, 1975.

[81] L. Kleinrock. *Queueing Systems, Vol. 2, Computer applications.* Wiley, New York, 1976.

[82] L. Kleinrock and H. Levy. The analysis of random polling systems. *Operations Research*, 36(5):716–732, sep/oct 1988.

[83] A. Klemm, C. Lindemann, and M. Lohmann. Traffic Modeling of IP Networks Using the Batch Markovian Arrival Process. In *Computer Performance Evaluation: Modelling Techniques and Tools. LNCS*, volume 2324, pages 129–182, 2002.

[84] G. P. Klimov. Time-sharing service systems I. *Theory Prob. Appl.*, 19:532–551, 1974.

[85] G. P. Klimov. Time-sharing service systems II. *Theory Prob. Appl.*, 23:314–321, 1978.

[86] D. Kofman. Blocking probability, throughput and waiting time in finite capacity polling systems. *Queueing Systems*, 14(3-4):385–411, 1993.

[87] A. G. Konheim, H. Levy, and M. M. Srinivasan. Descendant set: an efficient approach for the analysis of polling systems. *IEEE Transactions on Communications*, 42(2,3,4):1245–1253, 1994.

[88] P. J. Kuehn. Multiqueue Systems with Nonexhaustive Cyclic Service. *The Bell System Technical Journal*, 58:671–698, 1979.

[89] V. G. Kulkarni. *Modeling and Analysis of Stochastic Systems.* Chapman & Hall, 1995.

[90] R. O. LaMaire. An M/G/l vacation model of an FDDI station, RC 15747. Technical report, IBM.' Res. Rep., revised Jan. 1991., 1991.

[91] R. Landry and I. Stavrakakis. Queueing study of a 3-priority policy with distinct service strategies. *IEEE/ACM Transactions on Networking*, 1(5):576–589, oct 1993.

[92] C. Langaris. A Polling Model with Retrial Customers. *J. Oper. Res. Soc. Japan*, 40(4):489–508, 1997.

[93] C. Langaris. Gated Polling Models with Customers in Orbit. *Math. Comput. Modeling*, 30(3-4):171–187, 1999.

[94] T. Y. S. Lee. Models for Design and Control of Single Server Polling Computer and Communication Systems. *Operations Research*, 46(4):515–531, jul-aug 1998.

[95] T. Y. S. Lee and J. Sunjaya. Exact Analysis of Asymmetric Random Polling Systems with Single Buffers and Correlated Input Process. *Queueing Syst.*, 23: 131–156, 1996.

[96] K. K. Leung. Cyclic-Service Systems with Probabilistically-Limited Service. *IEEE Journal on Selected Areas in Communications*, 9(2):185–193, feb 1991.

[97] H. Levy. *Non-uniform structures and synchronization patterns in shared channel communication networks. Ph. D. dissertation, UCLA-CSD-840049.* PhD thesis, Computer Science Dept., School of Engineering and Applied Science, Univ. of California, Los Angeles, Calif., 1984.

[98] H. Levy. Optimization of polling systems: The fractional exhaustive service method, technical report. Technical report, Department of Computer Science, Tel Aviv University, 1988.

[99] H. Levy. Analysis of cyclic polling systems with binomial-gated service. In T. Hasegawa, H. Takagi, and Y. Takahashi, editors, *Performance of Distributed and Parallel Systems*, pages 127–139. Elsevier Science Publishers, North-Holland, 1989.

[100] H. Levy. Delay computation and dynamic behavior of non-symmetric polling systems. *Performance Evaluation*, 10(1):35–51, 1989.

[101] H. Levy. Binomial-gated service: a method for effective operation and optimization of polling systems. *IEEE Transactions on Communications*, 39(9):1341 – 1350, sep 1991.

[102] H. Levy and M. Sidi. Polling systems: applications, modeling, and optimization. *IEEE Transactions on Communications*, 38(10):1750 – 1760, oct 1990.

[103] H. Levy and M. Sidi. Polling systems with simultaneous arrivals. *IEEE Transactions on communications*, 39(6):823–827, jun 1991.

[104] H. Levy, M. Sidi, and O. J. Boxma. Dominance relations in polling systems. *Queueing Systems*, 6:155–172, 1990.

[105] H. Levy, M. Sidi, and O. J. Boxma. Dominance relations in polling systems. *Queueing systems*, 6:155–172, 1990.

[106] F. Li and O. Yang. Queue length performance of some non-exhaustive polling models with Bernoulli feedback. In *Military Communications Conference, MILCOM 98. Proceedings., IEEE, vol.3*, pages 888 – 892, 1998.

[107] F. Li, R. Sun, and Z. Liu. Measurement estimation approach to efficiency evaluation of bandwidth allocation scheme in ATM networks. In *Proceeding of ICC'97, Montreal, Canada*, pages 1138–1142, jun 1997.

[108] 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 (1):144–162, 2005.

[109] J. D. C. Little. A proof of the queueing formula L = λ W. *Operations Research*, 9:383–387, 1961.

[110] D. Liu and Y. Q. Zhao. Determination of explicit solution for a genaral class of markov processes Matrix-Analytic Methods in Stochastic Models. In S. Chakravarthy and A. S. Alfa, editors, *EPEW 2008. LNCS*, pages 343–357. Marcel Dekker, 1996.

[111] L.Massoulie. Stability of non-Markovian polling systems. *Queueing Systems*, 21: 67–96, 1995.

[112] B. L. Long, E. Hossain, and A. Alfa. Queueing analysis and admission control for multi-rate wireless networks with opportunistic scheduling and ARQ-based error control. In *IEEE International Conference on Communications (ICC 2005)*, volume 5, pages 3329–3333, may 2005.

[113] D. L. Lucantoni. *An algorithmic analysis of a communication model with retransmission of flawed messages.* London:Pitman, 1983.

[114] D. L. Lucantoni. New results on the single server queue with a batch markovian arrival process. *Stochastic Models*, 7:1–46, 1991.

[115] D. L. Lucantoni. The BMAP/G/1 queue: A tutorial. In L. Donatiello and R. Nelson, editors, *Models and Techniques for Performance Evaluation of Computer and Communications Systems*, pages 330–358. Springer Verlag, 1993.

[116] D. M. Lucantoni, K. S. Meier-Hellstern, and M. Neuts. A single server queue with server vacations and a class of non-renewal arrival arrival processes. *Adv. Appl. Prob.*, sep 1989.

[117] A. D. M. Karakul. An approximation method to analyse polling models of pull-type production systems. *European J. of Industrial Engineering*, 1(2):200 – 222, 2007.

[118] C. Mack. The efficiency of N machines uni-directionally patrolled by one operative when walking time is constant and repair times are variable. *J. Roy. Stat. Soc. Series B*, 19(1):173–178, 1957.

[119] I. MacPhee, M. Menshikov, D. Petritis, and S. Popov. A Markov chain model of a polling system with parameter regeneration. *The Annals of Applied Probability*, 17(5/6):1447–1473, 2007.

[120] M. A. Marsan, L. F. de Moraes, S. Donatelli, and F. Neri. Analysis of symmetric nonexhaustive polling with multiple servers. In *INFOCOM*, pages 284–295, 1990.

[121] R. D. V. Mei. Polling systems in heavy traffic: higher moments of the delay. *Queueing Systems*, 31:265–294, 1999.

[122] R. D. V. Mei. Distributions of the delay in polling systems in heavy traffic. *Perf. Eval.*, 38:133–148, 1999.

[123] R. D. V. Mei. Polling systems with switch-over times under heavy load: moments of the delay. *Queueing Systems*, 36:381–404, 2000.

[124] R. D. V. Mei. Waiting-time distributions in polling systems with simultaneous batch arrivals. *Ann. of Oper. Res.*, 113:157–173, 2002.

[125] R. D. V. Mei. Towards a unifying theory on branching-type polling models in heavy traffic. *Queueing Systems*, 57:29–46, 2007.

[126] R. D. V. Mei and S. C. Borst. Analysis of multiple-server polling systems by means of the power-series algorithm. *Stochastic Models*, 13(2):339 – 369, 1997.

[127] R. D. V. Mei and H. Levy. Pollingsystems in heavy traffic: exhaustiveness of service policies. *QueueingSystems*, 27:227–250, 1997.

[128] R. D. V. Mei and H. Levy. Expected delay in polling systems in heavy traffic. *Adv. Appl. Prob.*, 30:586–602, 1998.

[129] R. D. V. Mei and E. M. M. Winands. Polling models with renewal arrivals: A new method to derive heavy-traffic asymptotics. *Performance Evaluation*, 64(9-12): 1029–1040, oct 2007.

[130] K. Meier-Hellstern and W. Fischer. The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation*, 18:149–171, 1992.

[131] J. Misic and V. B. Misic. Modeling Bluetooth Piconet Performance. *IEEE Comm. Letters*, 7:18–20, jan 2003.

[132] L. F. M. Moraes and I. Rubin. Analysis and comparison of message queueing delays in token-rings and token-buses local area networks. In *IEEE International Conference on Communications (ICC'84), Amsterdam*, pages 130–135, may 1984.

[133] M. Murata, K. Shiomoto, and H. Miyahara. Performance Analysis of Token Ring Networks with a Reservation Priority Discipline. *IEEE Trans. Commun.*, 38(10): 1844–1853, 1990.

[134] S. Nahmias and M. H. Rothkopf. Stochastic models of internal mail delivery systems. *Management Science*, 30(9):1113–1120, sep 1984.

[135] O. Nakdimon and U. Yechiali. Polling systems with breakdowns and repairs. *European Journal of Operational Research*, 149(3):588–613, 2001.

[136] M. F. Neuts. Probability distributions of phase-type. In *Liber Amicorum Prof. Emeritus H. Florin, Department of Mathematics. Belgium: University of Louvain*, pages 173–206, 1975.

[137] M. F. Neuts. Renewal processes of phase type. *Nav. Res. Logist. Quart.*, 25: 445–54, 1978.

[138] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach.* The John Hopkins University Press, Baltimore, 1981.

[139] M. F. Neuts. *Structured stochastic matrices of M/G/1 type and their applications.* Marcel Dekker, New York, 1989.

[140] S. Nishimura. A spectral method for a nonpreemtive priority BMAP/G/1/N queue. *Stochastic Models*, 21:579–597, 2005.

[141] T. L. Olsen and R. D. van der Mei. Polling systems with periodic server routingin heavy traffic: renewal arrivals. *Operations Research Letters*, 33:17–25, 2005.

[142] M. I. Reiman and L. M. Wein. Heavy traffic analysis of polling systems in tandem. *Operations Research*, 47(4):524–534, jul/aug 1999.

[143] J. A. C. Resing. Polling systems and multitype branching processes. *Queueing Systems*, 13:409–426, 1993.

[144] A. Riska and E. Smirni. M/G/1-type Markov processes: A tutorial. In *Tutorials of the IFIPWWG7.3. Int. Symposium on Computer Performance Modeling, Measurements and Evaluation. LNCS*, volume 2459, pages 36–63, 2002.

[145] J. S. Stidham. Optimal control of a signalized intersection. Part I: Introduction: Structure of intersection models; Part II: Determining the optimal switching policies; Part III: Descriptive stochastic models. Tech. Reps. 94, 95, and 96. Technical report, Dept. of Operations Research, Cornell Univ., Ithaca, New York, 1969.

[146] D. Sarkar and W. I. Zangwill. Expected waiting time for nonsymmetric cyclic queueing systems - exact results and applications. *Management Science*, 35:1463–1474, 1989.

[147] M. Schwartz. *Telecommunication Networks: Protocols, Modeling and Analysis.* Addison-Wesley Publishing Company, Reading, Massachusetts, 1987.

[148] L. D. Servi. Average delay approximation of M/G/l cyclic service queues with Bernoulli schedules. *IEEE J. Select. Areas Commun.*, SAC-4:813–822, sep 1986.

[149] Y. W. Shin and C. E. M. Pearce. The BMAP/G/1 vacation queue with queue-length dependent vacation schedule. *J. Austral. Math. Soc.*, Ser. B 40:207–221, 1998.

[150] M. Sidi, H. Levy, and S. W. Fuhrmann. A Queuing Network with a Single Cyclically Roving Server. *Queuing Syst.*, 11(1-2):121–144, 1992.

[151] M. L. Simoes, P. M. Oliveira, and A. P. da Costa. Queues with server vacations in urban traffic control. In J. Janssen and P. Lenca, editors, *11th ASMDA2005. Proceedings ASMDA 2005. Proceedings of the 11th Symposium on ASMDA, Brest, France*, may 2005.

[152] Specification Bluetooth SIG. Specification of the Bluetooth System - Version 1.1., 2001.

[153] M. M. Srinivasan and H.-S. Lee. An analysis of a prioritized polling mechanism for cyclic server systems. Tech. Rep. 87-6. Technical report, Dept. of Industrial and Operations Engineering, Univ. of Michigan, Ann Arbor, Mich, 1987.

[154] M. M. Srinivasan, S.-C. Niu, and R. B. Cooper. Relating Polling Models with Zero and Nonzero Schwitchover Times. *Queueing Systems*, 19:149–168, 1995.

[155] M. M. Srinivasan, H. Levy, and A. G. Konheim. The individual station technique for the analysis of polling systems. *Naval Research Logistics*, 43(1):79–101, 1996.

[156] R. Srinivasan and J. S. Baras. Analyzing the Performance of Data Users in Packet Switched Wireless Systems with Prioritized Voice Traffic. In *Wired/Wireless Internet Communications. LNCS*, volume 2957, 2004.

[157] Standard IEEE 802.11. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, jan 1997.

[158] Standard IEEE 802.16-2009. Part 16: Air Interface for Broadband Wireless Access Systems, Standard for Local and Metropolitan Area Networks, may 2009.

[159] Standard IEEE 802.16.3c-01/30r1. Traffic Model for 802.16 TG3 MAC/PHY Simulations, mar 2001.

[160] S. Stidham. Regenerative processes in the theory of queues, with applications to the alternating-priority queue. *Adv. Appl. Prob.*, 4:542–577, 1972.

[161] S. Suzuki and H. Yamashita. Mean Waiting Times of the Alternating Traffic Starting Delays. *J. Oper. Res. Soc. Japan.*, 41(3):442–454, 1998.

[162] J. Sykes. Analysis of the communications aspects of an inquiry-response system. In *AFIPS Conference Proceedings, 1969 Fall Joint Computer Conference, Las Vegas, Nevada*, volume 35, pages 655–667, nov 1969.

[163] P. T.-Gia and T. Raith. Performance analysis of finite capacity polling systems with nonexhaustive service. *Performance Evaluation*, 9:1–16, 1988.

[164] R. D. v. d. M. T. L. Olsen. Polling systems with periodic server routing in heavy traffic: distribution of the delay. *J. Appl. Probab.*, 40:305–326, 2002.

[165] L. Takács. A single-server queue with feedback. *Bell System Techn. Journ.*, 42: 505–519, 1963.

[166] L. Takács. Two queues attended by a single server. *Operations Research*, 16(3): 639–650, 1968.

[167] H. Takagi. *Analysis of Polling Systems.* MIT Press, 1986.

[168] H. Takagi. Queueing analysis of polling systems with zero switchover times. TRL Res. Rep. TR87-0034. Technical report, IBM Japan, Tokyo, 1987.

[169] H. Takagi. Exact analysis of round-robin scheduling of services. *IBM J. Res. Dev.*, 31(4):484–488, jul 1987.

[170] H. Takagi. Queueing analysis of polling models. *ACM Comput. Surveys*, 20:5–28, 1988.

[171] H. Takagi. Analysis of Finite Capacity Polling Systems. *Advanced Applied Probability*, 23:373–387, 1991.

[172] H. Takagi. *Queueing Analysis - A Foundation of Performance Evaluation, Vacation and Prority Systems, vol.1.* North-Holland, New York, 1991.

[173] H. Takagi. Analysis and Application of Polling Models. In *Lecture Notes In Computer Science: Performance Evaluation: Origins and Directions*, volume 1769, pages 423–442, 2000.

[174] T. Takine. The nonpreemtive priority MAP/G/1 queue. *Operations Research*, 47: 917–927, 1999.

[175] T. Takine and Y. Takahashi. On the relationship between queue lengths at a random instant and at a departure in the stationary queue with BMAP arrivals. *Stochastic Models*, 14:601–610, 1998.

[176] T. Takine, Y. Takahashi, and T. Hasegawa. Average message delay of an asymmetric single buffer polling system with round-robin scheduling of services. Tech. Rep. 87020. Technical report, Dept. of Applied Mathematics and Physics, Kyoto Univ.,Japan., 1987.

[177] T. Takine, Y. Takahashi, and T. Hasegawa. Analysis of an asymmetric polling system with single buffers. In P.-J. Courtois and G. Latouche, editors, *Performance '87*, pages 241–251. Elsevier North-Holland, Amsterdam, 1988.

[178] T. Takine, H. Takagi, and T. Hasegawa. Sojourn Times in Vacation and Polling Systems with Bernoulli Feedback. *Journal of Applied Probability*, 28(2):422–432, jun 1991.

[179] T. E. Tedijanto. Exact Results for the Cyclic-Service Queue with a Bernoulli Schedule. *Perform. Eval.*, 11(2):107–115, 1990.

[180] V. M. Vishnevskii and O. V. Semenova. Mathematical methods to study the polling systems. *Automation and Remote Control*, 67:173–220, 2006.

[181] V. Vishnevsky and O. Semenova. Adaptive Dynamical Polling in Wireless Networks. *Cybernetics and Information Technologies*, 8(1), 2008.

[182] V. Vishnevsky and O. Semenova. The power-series algorithm for two-queue polling system with impatient customers. In *15th International Conference on Telecommunications (ICT 2008 and MACOM 2008 Workshop Proceedings), Saint-Petersburg, Russia*, jun 2008.

[183] V. Vishnevsky, O. Semenova, A. Dudin, and V. Klimenok. Queueing model with gated service and adaptive vacations. In *IEEE International Conference on Communications (IEEE ICC-2009 and MACOM-2009 Workshop Proceedings), Dresden, Germany*, jun 2009.

[184] V. M. Vishnevsky, A. N. Dudin, V. I. Klimenok, O. V. Semenova, and S. A. Shpilev. Approximate Analysis for M/G/1-Polling System with Adaptive Polling Mechanism. In *IEEE International Conference on Communications (IEEE ICC-2009 and MACOM-2009 Workshop Proceedings), Dresden, Germany*, jun 2009.

[185] Y. T. Wang and R. J. T. Morris. Load sharing in distributed systems. *IEEE Transactions on Computers*, C-34(3):204–217, mar 1985.

[186] A. Wierman, E. M. M. Winands, and O. J. Boxma. Scheduling in polling systems. *Performance Evaluation*, 64:1009–1028, 2007.

[187] WiMAX Forum. WiMAX System Evaluation Methodology V2.0, dec 2007.

[188] E. M. M. Winands, I. J. B. F. Adan, and G.-J. van Houtum. Mean value analysis for polling systems. *Queueing Systems*, 54:35–44, 2006.

[189] R. W. Wolff. Poisson Arrivals See Times Averages. *Operations Research*, 30: 223–231, 1982.

[190] R. W. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[191] Y. Wu, T. Shu, Z. Niu, and J. Zheng. A vacation model with setup and close-down times for transmitter buffer of arq schemes. In *PIMRC 2003. 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, volume 3, pages 2054–2058, sept 2003.

[192] H. Yang and B. Sikdar. Performance Analysis of Polling based TDMA MAC Protocols with Sleep and Wakeup Cycles. In *IEEE International Conference on Communications (ICC'07)*, pages 241–246, jun 2007.

[193] S.-R. Yang. Dynamic power saving mechanism for 3G UMTS system. *Mobile Networks and Applications*, 12(1):5–14, 2007.

[194] U. Yechiali. Optimal dynamic control of polling systems. In *Queueing, performance and control in ATM, Proc. 13th Telegr. Congr., Copenhagen Denmark, North-Holland Stud. Telecomm.*, volume 15, pages 205–217, 1991.

[195] U. Yechiali. Optimal server scheduling and dynamic control in polling systems. In *5th Australian Teletraffic Research Seminar, Melbourne*, dec 2004.

[196] Y. Yi, J. Zhangy, and M. Chiang. Delay and Effective Throughput of Wireless Scheduling in Heavy Traffic Regimes: Vacation Model for Complexity. In *MobiHoc'09, New Orleans, Louisiana, USA*, may 2009.

[197] G. Zussman, U. Yechiali, and A. Segall. Exact Probabilistic Analysis of the Limited Scheduling Algorithm for Symmetrical Bluetooth Piconets. In *PersonalWireless Communications. LNCS*, volume 2775, pages 276–290, 2003.

Journal papers

[198] Zs. Saffer and M. Telek. Stability of periodic polling system with BMAP arrivals. *European Journal of Operational Research*, 197(1):188–195, 2009.

[199] Zs. Saffer and M. Telek. Unified analysis of BMAP/G/1 cyclic polling models. *Queueing Systems*, 64(1):69–102, 2010.

[200] Zs. Saffer and M. Telek. Analysis of BMAP vacation queue and its application to IEEE 802.16e sleep mode. *Journal of Industrial and Management Optimization (JIMO)*, 6(3):661–690, aug 2010.

[201] Zs. Saffer and M. Telek. Closed form results for BMAP/G/1 vacation model with binomial type disciplines. *Publ. Math. Debrecen*, 76(3):359–378, 2010.

Conference and workshop papers

[202] Zs. Saffer. An introduction to classical cyclic polling model. In *Proc. of the 14th Int. Conf. on Analytical and Stochastic Modelling Techniques and Applications (ASMTA'07)*, pages 59–64, 2007.

[203] Zs. Saffer and M. Telek. Analysis of BMAP/G/1 vacation model of non-M/G/1-type. In N. Thomas and C. Juiz, editors, *EPEW 2008. LNCS*, volume 5261, pages 212–226. Springer Verlag, Sept. 2008.

[204] Zs. Saffer and M. Telek. Analysis of BMAP/GI/1 phase dependent vacation model with decrementing-K discipline. In *Queues: flows, systems, networks: Proceedings of the International Conference "Mathematical Methods for Analysis and Optimization of Information Telecommunication Networks" Minsk*, pages 193–198, jan 2009.

[205] Zs. Saffer and S. Andreev. Delay analysis of IEEE 802.16 wireless metropolitan area network. In *15th International Conference on Telecommunications (ICT 2008 and MACOM 2008 Workshop Proceedings), Saint-Petersburg, Russia*, jun 2008.

[206] S. Andreev, Zs. Saffer, and A. Anisimov. Overall Delay Analysis of IEEE 802.16 Network. In *IEEE International Conference on Communications (IEEE ICC-2009 and MACOM-2009 Workshop Proceedings, Dresden, Germany)*, June 2009.

[207] S. Andreev, Zs. Saffer, A. Turlikov, and A. Vinel. Overall Delay in IEEE 802.16 with Contention-Based Random Access. In K. Al-Begain, D. Fiems, and G. Horváth, editors, *16th International Conference, ASMTA 2009. LNCS*, volume 5513, pages 89–102. Springer Verlag, June 2009.

[208] Zs. Saffer and M. Telek. Waiting time analysis of BMAP vacation queue and its application to IEEE 802.16e sleep mode. In *4th International Conference on Queueing Theory and Network Applications (QTNA 2009). Singapore*, jul 2009.

[209] S. Andreev, Zs. Saffer, A. Turlikov, and A. Vinel. Upper Bound on Overall Delay in Wireless Broadband Networks with Non Real-Time Traffic. In K. Al-Begain, D. Fiems, and W. Knottenbelt, editors, *to appear in 17th International Conference, ASMTA 2010. LNCS.* Springer Verlag, June 2010.

[210] Zs. Saffer, S. Andreev, and Y. Koucheryavy. Modeling the influence of the Real-Time Traffic on the Delay of the Non Real-Time Traffic in IEEE 802.16 Network. *to appear in MACOM-2010 Workshop Proceedings, Barcelona, Spain.*, Sept. 2010.