Universiteit Antwerpen

Faculteit Wetenschappen

Asymptotisch Gedrag van Wachtrijen

—

Asymptotic Behaviour of Queueing Systems

Proefschrift voorgelegd tot het behalen van de graad van doctor in de Wetenschappen aan de Universitaire Instelling Antwerpen te verdedigen door Tim Daniëls, aspirant FWO-Vlaanderen

Promotor: Prof. Dr. C. Blondia
Copromotor: Prof. Dr. R. Lowen

Antwerpen, 1999

Dankwoord

Langs deze weg wens ik iedereen te bedanken die heeft bijgedragen tot de realisatie van dit proefschrift. In de eerste plaats zijn dat het FWO-Vlaanderen en mijn promotoren Prof. Chris Blondia en Prof. Bob Lowen. Verder natuurlijk ook Prof. Jan Van Casteren en Floris Geerts voor de talrijke en nuttige discussies. Buitenshuis zijn vooral de contacten met de SMACS onderzoeksgroep onder leiding van Prof. Herwig Bruneel erg produktief gebleken. In het bijzonder denk ik hierbij aan Koen Laevens die mij op het spoor zette van de *tail transitions*.

Dank eveneens aan de talrijke collega's voor de toffe en motiverende werksfeer. Daarbij denk ik vooral aan mijn bureaugenoten Kathleen, Sandra en Benny, en natuurlijk ook aan Stefan Becuwe voor de talrijke LaTeX tips.

Tot slot dank ik Karin, ouders en familie, voor de blijvende aanmoediging.

Overview

Although the study of buffer asymptotics is of a mathematical nature, this domain is still closely related to issues arising in the telecom practice. Among other things, this is illustrated by the fact that the mathematical model used throughout this thesis is derived from a real-world network device, namely the ATM statistical multiplexer. Together with a formal description of the multiplexer model, a short and informal overview of ATM and Quality of Service is given in Chapter 1. As this thesis focuses on the influence of the arrival process, Chapter 1 closes with an overview of the characteristics of teletraffic.

All studied arrival processes are modelled as a Discrete-time Batch Markovian Arrival Process or DBMAP, most queueing systems are represented by DMAP-D-1 queues. Therefore both the DBMAP and the DMAP-D-1 queue are presented in detail in Chapter 2. In Section 2.7.1 the asymptotic behaviour for a large class of DMAP-D-1 queues is examined by applying the classical dominant pole approximation. How buffer asymptotics can be obtained for these DMAP-D-1 queues by large deviation techniques is demonstrated in Section 2.7.2.

Chapter 3 does not deal explicitly with buffer asymptotics. It is shown there that the Matrix-Analytical Approach (MAA) and the Functional Equation Approach (FEA) to solve discrete-time queueing system are essentially equivalent when applied to the DMAP-D-1 queue. It is furthermore indicated that this equivalence also holds for the DBMAP-G-1 queue and the DBMAP-D-c multi-server queue. This chapter is concluded with two case studies. The first one is merely an illustration of the general theory, the second one deals with an approximation for the occurring boundary probabilities. The accuracy of this approximation is assessed by numerical examples, where the emphasis is on its application to buffer asymptotics. This chapter is an extension of [18].

Tail transitions are the subject of study in Chapter 4. A tail transition is an abrupt change of the asymptotic behaviour which occurs when the parameters determining the system are only slightly varied. Two radically different queueing systems in which such transitions take place are presented. The first system describes two queues in tandem. The techniques used here to obtain the asymptotic behaviour can be seen as a generalisation of the dominant pole approximation. The second system is a multi-server queue having as input a traffic mix which consists of a Long Range Dependent (LRD) and a Short Range

Dependent (SRD) component. It turns out that two types of asymptotic behaviour can be observed for this queue, depending on the number of servers and on the sizes of the components of the total arrival process.

Chapter 5 aims at applying the philosophy behind pseudo self-similar processes to obtain buffer asymptotics for a queue with LRD input. Therefore an LRD arrival process is defined as the limit of a sequence of Markovian arrival processes. The correlation structure of this traffic model is determined in detail. Buffer asymptotics are obtained by using large deviation techniques. Most material presented here can also be found in [19].

In Chapter 6 a multiplexer queue with discrete-time LRD $M/G/\infty$ input is considered. An analytically tractable expression for the generating function associated with the stationary buffer distribution is derived. The exact buffer asymptotics are obtained by applying a Tauberian theorem to this generating function. The same approach is successfully applied to a queue having as input the superposition of a Markovian arrival process and an LRD $M/G/\infty$ process. The obtained results are also considered from a practical point of view. This chapter is based on both [17] and [20].

Contents

Chapter 1

Introduction

The main part of this thesis deals with buffer asymptotics, or more precisely stated, with the asymptotic behaviour of the stationary distribution associated with the occupancy of buffers in queueing systems. Although this study is of a mathematical nature, it has its motivation in the domain of today's telecommunication systems. This is illustrated by the fact that the mathematical model used throughout this text is derived from a rather generic network device, namely the ATM statistical multiplexer. Preceding the formal definition of the so-called multiplexer model given in Section 1.2, its importance is indicated in Section 1.1. Because the arrival process turns out to be the most important component of this model, Section 1.3 is devoted to the characteristics and the modelling of network traffic.

1.1 ATM and Quality of Service

For years telephone systems were the only telecommunication systems of great economical importance. This clearly changed during the last 20 years. Data networks emerged and the amount of traffic they are carrying is rapidly increasing, consider e.g. the Internet. The telecommunication infrastructure being deployed now is intended to transport different types of traffic: voice, data, video, ... Otherwise stated, the idea behind the Integrated Services Digital Network or ISDN is slowly becoming a reality. Consequently the Asynchronous Transfer Mode (ATM), the transport mechanism defined at the end of the eighties by the telecom community to deal with the traffic generated by integrated services, is gaining ground. Of course the involved technologies are evolving rapidly and some drawbacks of ATM are becoming clear. Nevertheless the most distinguished property of ATM, its notion of Quality of Service (QoS), is being adopted by other important transport mechanisms, e.g. the protocols running the Internet. As the research performed in this thesis is closely related with a specific type of QoS, we start with a short and informal overview of ATM and QoS (for a detailed exposition see e.g. [27]), [21], [30] and [38].

ATM is a connection-oriented and packet-switched transmission procedure using fixed-size data packets, called cells. An efficient use of network resources is made possible by statistical multiplexing. Roughly speaking this technique allows to put multiple traffic streams together on a single line, as long as on the average the total arrival rate does not exceed the capacity of the channel. When there arrive temporarily more cells than can be transmitted, those extra cells are buffered. The typical ATM device which enables statistical multiplexing is the ATM statistical multiplexer, studied in detail in Section 1.2.

Since ATM is connection-oriented, users have to set up a connection before they are allowed to transmit data. The ATM network can decide whether or not to accept the new connection through a procedure called Connection Admission Control (CAC). A connection can be refused, e.g. if the QoS requested by the new user cannot be guaranteed. Let us therefore present QoS in more detail.

QoS can be easily understood when considering video traffic. This type of data traffic has stringent real-time constraints: cells are not allowed to arrive with too big delays, which would otherwise result in an unacceptable degradation of the image quality. So limiting the delays for some types of traffic is a first kind of QoS a network should support. When transporting files, delay is only of limited importance. Here clearly the integrity of the data has to be conserved. Consequently the cell loss probability (CLP), which is the probability that cells get lost by e.g. buffer overflows, should be kept as small as possible.

Buffer asymptotics are related to this second type of QoS presented above. When considering large buffers the CLP can be approximated by using buffer asymptotics obtained for the appropriate queueing systems. Hence the study of buffer asymptotics is useful for both the design of buffers in network elements (e.g. switches, multiplexers, ...) and for the development of CAC algorithms.

1.2 The Multiplexer Model and Buffer Asymptotics

The ATM network device which — by its nature — gives rise to cell loss is the statistical multiplexer. For this device a mathematical model will be derived, the so-called multiplexer queue. It is a rather generic model, which can also be identified as a component of more complicated devices, such as switches.

The statistical multiplexer is the network element which multiplexes several incoming links on one outgoing link, as pictured in Figure 1.2. Since cells can arrive simultaneously at the multiplexer, and since only one cell can be transmitted to the outgoing link at a time, there is a buffer where cells can temporarily wait. If the buffer is full, newly arriving cells will be lost. Hence the CLP can be derived from the probability of buffer overflows.

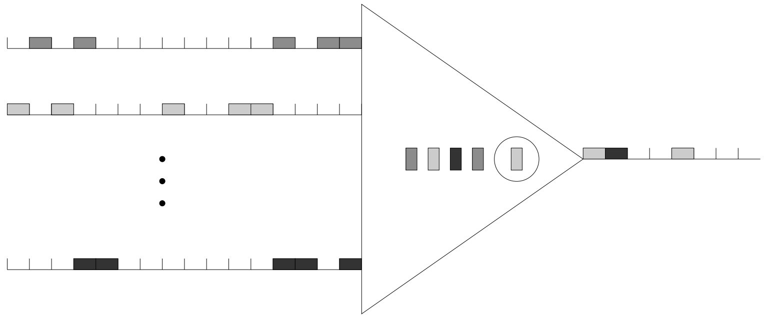

Figure 1.1: Statistical multiplexer

1.2.1 The multiplexer queue

The multiplexer model is defined by mapping the physical description given above onto a queueing system. Therefore we recall the three main components of a queueing system:

- the input or arrival process,
- the service mechanism,
- the queueing discipline.

The input process describes stochastically how the requests for service or customers arrive in time. Details on the characteristics of these processes can be found in 1.3. The service mechanism typically is a description of how long it takes to serve a customer. The queueing discipline is the order in which the customers are served: this can be according to a FIFO or LIFO discipline, some customers may have priority over others, ...

Notice that for the ATM multiplexer one essentially deals with a discrete-time model: the unit of time, also called slot, is defined as the time needed to transmit 1 cell to the outgoing link. Since it is assumed that the cells are served in FIFO order, we are left with identifying the arrival process and the service mechanism.

The arrival process is described by a discrete-time stochastic process X_k, with X_k representing the number of cells arriving at slot k. All arrival processes considered in this thesis will be modelled as DBMAPs or Discrete-time Batch Markovian Arrival Processes. This versatile class of processes is introduced in detail in Chapter 2. For now it suffices to remark that the process X_k can be very general. For example it does not necessarily represent a superposition of traffic streams, as is the case with a real multiplexer. This generality contributes to the generic character of the multiplexer model.

The service mechanism is straightforward: there is only one server, the service time is deterministic and by definition equal to 1 time slot. Although physically incorrect we assume that the buffer is infinitely long. An estimate for the CLP associated with a finite buffer is given in Section 1.2.2.

Whenever the arrival process of the multiplexer queue is modelled as a DBMAP, the queue itself corresponds to a DBMAP-D-1 queue. This class of queueing systems is presented in depth in Chapter 2.

1.2.2 Cell loss and buffer asymptotics

Let q_k denote the number of cells in the system, including the one in service, at time k. The stochastic evolution of the multiplexer queue is given by Lindley's equation

$$q_{k+1} = (q_k - 1)^+ + X_k,$$

with $a^+ = \max(a, 0)$. Hence if $\mathbf{E}[X_k] < 1$, there exists a stationary version q of the q_k. The random variable q will be said to correspond to the stationary queue length distribution, or to be the stationary buffer distribution. Note that the cell in service is also taken into account by the definition of q.

A straightforward and conservative approximation for the CLP of a buffer with length K is given by $\mathbf{P}\{q > K\}$. When long buffers are concerned, it is of interest to know the behaviour of $\mathbf{P}\{q > K\}$ for $K \to \infty$. Mathematical results about this asymptotic behaviour make it possible to derive approximations for the CLP, which are easy to compute compared with the determination of the whole distribution of q.

1.3 Traffic Characteristics and Modelling

The most important characteristic of an arrival process X_k, besides its mean arrival rate, is its autocorrelation, given by the autocorrelation function

$$r(k) = \frac{\mathrm{Cov}\big(X_1, X_{1+k}\big)}{\mathrm{Var}\big(X_1\big)}. \tag{1.1}$$

According to this second-order characteristic arrival processes can be divided in two classes: the Short Range Dependent (SRD) and the Long Range Dependent (LRD) arrival processes. We recall the formal definition of these two types of processes from [52].

Definition 1.3.1. A process X_k is called LRD if

$$\sum_{k=1}^{\infty} \mathrm{Var}\big(X_1, X_k\big) = \infty,$$

otherwise X_k is said to be SRD.

1.3.1 SRD arrival processes

Two main types of arrival processes can be identified within the class of the SRD processes. First of all one has processes without autocorrelation, here X_k is a sequence of i.i.d. random variables. The multiplexer queue with such input is actually a GI-G-1 queue. Buffer asymptotics for this type of queues can be studied by using the theory developed in e.g. [13].

The second important subclass contains the so-called Markovian arrival processes, having an exponentially decaying autocorrelation. For the continuous-time case these processes are typically modelled as a finite state Batch Markovian Arrival process or BMAP. When working in discrete-time, one can use the Discrete-time BMAP or DBMAP. The asymptotic behaviour of queues with such input can be obtained by using the dominant pole approximation, presented in Theorem 2.7.2.

1.3.2 LRD arrival processes and self-similarity

The careful statistical analysis of huge amounts of high quality traffic measurements, which started in the mid 90's (see e.g. [39] and [50]), revealed the prevalence of self-similar traffic patters in today's high-speed networks. The notion of self-similar traffic, which will be formalised later on, can intuitively be understood as traffic which is bursty over a wide range of time scales. Otherwise stated, the number of data packets passing through the network each 10s shows the same variability as the number of packets measured over 1s intervals, and so on. Hence the aggregation in time of such a traffic stream does not make it smoother. Among other things, self-similarity implies a high degree of autocorrelation, in sharp contrast with the commonly made modelling assumptions at that time. Most traffic models were essentially Markovian or SRD, having at worst an exponentially decaying autocorrelation. The autocorrelation of self-similar traffic is decaying much slower, i.e. according to a power-law, resulting in LRD.

In the literature the terms LRD and self-similarity are more or less used as synonyms. It will take us however some definitions and properties to state the precise relationship between these two notions. We start with defining self-similarity for continuous-time processes.

Definition 1.3.2. The real-valued process $\{X(t), t \geqslant 0\}$ is self-similar with Hurst parameter $0 < H < 1$ if for all $a > 0$, the finite-dimensional distributions of $\{X(at), t \geqslant 0\}$ are identical to the finite-dimensional distributions of the process $\{a^H X(t), t \geqslant 0\}$.

Fractional Brownian motion, constructed in e.g. [44], is probably the best known example of a self-similar process. Ordinary Brownian motion is a self-similar with $H = 1/2$. For discrete-time processes the definition of self-similarity is based on the notion of aggregated process.

Definition 1.3.3. With a discrete-time stochastic process $X = (X_1, X_2, \dots)$, one can associate a sequence of aggregated processes $X^{(m)} = (X_1^{(m)}, X_2^{(m)}, \dots)$, $m \geqslant 1$, defined by

$$X_k^{(m)} = \frac{1}{m}\big(X_{(k-1)m} + \ldots + X_{km}\big).$$

For discrete-time processes several different definitions for self-similarity can be found in the literature. The one most resembling Definition 1.3.2, presented in e.g. [39], is given below.

Definition 1.3.4. A process $X = (X_1, X_2, \dots)$ is called self-similar with Hurst parameter H, $0 < H < 1$, if for all $m \geqslant 1$,

$$X = m^{1-H} X^{(m)},$$

where the equality is understood in the sense of equality of the corresponding finite-dimensional distributions.

A less stringent notion of self-similarity considers only the second order characteristics of the process, see again [39].

Definition 1.3.5. A covariance-stationary process $X = (X_1, X_2, \dots)$, with variance σ and autocorrelation function $r(k)$, is called exactly second-order self-similar, with Hurst parameter $0 < H < 1$, if for all $m \geqslant 1$, X and $m^{1-H}X^{(m)}$ have identical second-order statistics; i.e. for all $m \geqslant 1$,

$$\mathrm{Var}(m^{1-H} X^{(m)}) = \sigma^2 \tag{1.2}$$

and

$$r^{(m)}(k) = r(k) \text{ for } k \geqslant 1,$$

with $r^{(m)}(k)$ the autocorrelation function of the aggregated process $X_k^{(m)}$.

In [59] it is shown that there is some redundancy in Definition 1.3.5, as can be concluded from the following proposition.

Proposition 1.3.6. *If the process X satisfies condition (1.2) then*

$$r(k) = g(k) \overset{def}{=} \frac{1}{2}\big((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\big).$$

In practice one will seldomly encounter a process having exactly $g(k)$ as autocorrelation function. Therefore the notion of asymptotically second-order self-similarity is introduced.

Definition 1.3.7. A covariance-stationary process X is called asymptotically second-order self-similar with Hurst parameter $0 < H < 1$ if for $k \geqslant 1$,

$$\lim_{m\to\infty} r^{(m)}(k) = g(k).$$

Most of the time it is shown that a process is asymptotically second-order self-similar by invoking the following proposition, proved in [59]. The notation $f(k) \sim g(k)$, with f, g real-valued functions, stands for $\lim_{k\to\infty} f(k)/g(k) = 1$.

Proposition 1.3.8. *If the autocorrelation function $r(k)$ of the covariance-stationary process X satisfies*

$$r(k) \sim ck^{-\beta}, \tag{1.3}$$

with $0 < \beta < 1$ and $c > 0$, then X_k is asymptotically second-order self-similar with Hurst parameter $H = 1 - \beta/2$.

From Proposition 1.3.6 it can be seen that an asymptotically second-order self-similar process is LRD if the Hurst parameter $H \in (1/2, 1]$. Based on this observation the following definition is most frequently used in the literature.

Definition 1.3.9. A covariance-stationary process X is LRD with Hurst parameter $H \in (1/2, 1]$, if for some $c > 0$,

$$\mathrm{Cov}(X_1, X_n) \sim cn^{2H-2}. \tag{1.4}$$

An equivalent way to express the property (1.4) is presented in [52].

Proposition 1.3.10. *The covariance-stationary process X is LRD with Hurst parameter $H \in (1/2, 1]$, if for some $c > 0$,*

$$\mathrm{Var}\big(X_1 + \ldots + X_n\big) \sim cn^{2H}.$$

Chapter 2

The DBMAP-D-1 Queue

Both the Discrete-time Batch Markovian Arrival Process (DBMAP) and the DBMAP-D-1 queue belong to the domain of matrix-analytical queueing analysis. The matrix-analytical approach to queueing systems, presented in detail in [49], is merely the generalisation of the way the classical M/G/1-queue is solved. This solution of the M/G/1-queue relies on the technique of the embedded Markov chain, which is a technique widely used in queueing theory. Here we explain the main idea by looking at a general queueing system, introduced in Section 1.2. Using the information contained in the arrival process, the service mechanism and the service discipline, one aims at solving the queueing system: the calculation of the steady-state distribution — or related characteristics — of the number of customers in the system. The most straightforward technique — of course only applicable to the most simple models — is to formulate, in an analytically tractable way, the stochastic process describing the system contents. Complicated queueing systems need to be tackled by more sophisticated techniques, because the process describing the number of customers in the system at arbitrary time instants is often very complex. The key idea to simplify things is not to keep track of the system whole the time, but to consider it only at certain well-defined time instants, called epochs. An example are the so-called departure epochs. Here one only observes the system immediately after the departure of a customer. The M/G/1-queue is the best known example of a queueing system that can be solved this way. For this queue the number of customers in the system at departure epochs constitutes a Markov chain, the so-called embedded Markov chain.

As demonstrated in [49] it is possible to obtain an embedded Markov chain for a wide range of arrival processes and service mechanisms. The general applicability of this technique comes however at the price of introducing auxiliary variables or states, resulting in embedded Markov chains of the so-called M/G/1-type. In [49] a rigid framework is developed to solve a wide variety of these M/G/1-type Markov chains associated with queueing systems. The arrival processes which lead in a natural way — both conceptually and notationally — to M/G/1-type queueing systems, are the Batch Markovian Arrival Processes or BMAPs. This type of arrival process, presented in [42], is essentially a more transparent reformulation of the versatile Markovian point process introduced in [48].

The DBMAP now is the discrete-time counterpart of the BMAP and was introduced in [9]. We recall four properties which make DBMAPs an attractive model for discrete-time arrival processes:

- the notation is simple and transparent,
- the superposition of DBMAPs is again a DBMAP,
- the output process of a queue with DBMAP input can also be described as a DBMAP,
- the related queueing systems can be solved by using efficient computational methods.

2.1 The DBMAP: Definition and Properties

Formally, a DBMAP is defined by an infinite set of positive $m \times m$ matrices $\big(\mathbf{D}_l\big)_{0 \leqslant l < \infty}$, with the property that

$$\mathbf{D} = \sum_{l=0}^{\infty} \mathbf{D}_l$$

is a transition matrix. A DBMAP will be denoted by the tuple $(\mathbf{D}, \mathbf{D}_l)$, which completely determines it. By definition the Markov chain J_n, associated with $\mathbf{D}$ and having $\{i; 1 \leqslant i \leqslant m\}$ as state space, is controlling the actual arriving process as follows. Suppose J is in state i at time n. By going to the next time instant $n+1$ there occurs a transition to another or possibly the same state, and a batch arrival may or may not occur. The entries $(\mathbf{D}_l)_{i,j}$ represent the probabilities of having a transition from state i to j and a batch arrival of size l. So a transition from state i to j without an arrival will occur with probability $(\mathbf{D}_0)_{i,j}$. Denote by X_n the number of arrivals generated at slot n. Define the process N_n by $N_n = \sum_{m=1}^{n} X_m$. The construction described above implies that (N_n, J_n) is a two-dimensional Markov chain with state space $\{(k,i); 0 \leqslant k < \infty, 1 \leqslant i \leqslant m\}$ and transition matrix

$$\mathbf{T} = \begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \dots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Although DBMAPs are always introduced with m finite, this is not necessary. Most calculations can also be performed if $m = \infty$. We will point out where an infinite state space really makes a difference.

Let $\boldsymbol{\pi}$ be the stationary probability vector of the Markov chain J_n, i.e.

$$\boldsymbol{\pi}\mathbf{D} = \boldsymbol{\pi},$$

and $\boldsymbol{\pi}\mathbf{e} = 1$ with $\mathbf{e}$ a column vector of 1's. The mean arrival rate $\rho = \mathbf{E}[X_n]$ of the DBMAP $(\mathbf{D}, \mathbf{D}_l)$ is given by

$$\rho = \boldsymbol{\pi}\left(\sum_{l=1}^{\infty} l\mathbf{D}_l\right)\mathbf{e}.$$

Many calculations within the field of discrete probability involve generating functions, the theory of DBMAPs being no exception. The generating function $\mathbf{D}(z)$ associated with a DBMAP is defined by

$$\mathbf{D}(z) = \sum_{l=0}^{\infty} \mathbf{D}_l z^l.$$

Notice that

$$\rho = \boldsymbol{\pi}\left(\frac{d}{dz}\mathbf{D}(z)\bigg|_{z=1}\right)\mathbf{e}.$$

Example 2.2 (A Markovian on-off source). An on-off source is said to be Markovian if the length of the off- and on-periods is geometrically distributed. The parameters are respectively denoted by α and β. When the source is in the off-state no arrivals are generated, when the source is in the on-state it generates arrivals according to a Bernouilli distribution with parameter p. Modelling such a source as a DBMAP requires only two states: one on-state and one off-state. The probability that a silent, resp. active, source becomes active, resp. silent, in the next time slot is $1 - \alpha$, resp. $1 - \beta$. By taking the first state to be the off-state, the DBMAP is completely determined by

$$\mathbf{D}(z) = \begin{pmatrix} \alpha & 1-\alpha \\ (1-\beta)(1-p+pz) & \beta(1-p+pz) \end{pmatrix}.$$

One easily derives that

$$\boldsymbol{\pi} = \begin{pmatrix} \dfrac{1-\beta}{2-\alpha-\beta} & \dfrac{1-\alpha}{2-\alpha-\beta} \end{pmatrix}.$$

The mean arrival rate is given by

$$\rho = p\frac{1-\alpha}{2-\alpha-\beta}.$$

2.2.1 Superposition of DBMAPs

One of the advantages of the DBMAP description of traffic streams is that a superposition of DBMAPs is again a DBMAP. The construction of this superposition involves the Kronecker product $\otimes$. Consider for its definition a matrix $\mathbf{A} = [a_{ij}]$ of order $m \times n$ and a matrix $\mathbf{B}$ of order $r \times s$. The Kronecker product of the two matrices, denoted by $\mathbf{A} \otimes \mathbf{B}$, is defined as the partitioned matrix,

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix}.$$

For the numerous properties of this product the reader is referred to [31].

Suppose two DBMAPs $(\mathbf{D}^1, \mathbf{D}_l^1)$ and $(\mathbf{D}^2, \mathbf{D}_l^2)$ are given. The superposition of these processes, generating $X_n^1 + X_n^2$ arrivals at slot n, is described by the DBMAP $(\mathbf{D}, \mathbf{D}_l)$ defined by

$$\mathbf{D}_l = \sum_{k+m=l} \mathbf{D}_k^1 \otimes \mathbf{D}_m^2.$$

Clearly $\mathbf{D} = \mathbf{D}^1 \otimes \mathbf{D}^2$, $\mathbf{D}(z) = \mathbf{D}^1(z) \otimes \mathbf{D}^2(z)$ and $\boldsymbol{\pi} = \boldsymbol{\pi}^1 \otimes \boldsymbol{\pi}^2$.

Generally speaking a disadvantage of this construction is that the superposition of a large number of sources has to be described by an enormous number of states. This phenomenon is called state space explosion. In e.g. [40] and [54] strategies to cope with this problem are developed.

2.3 The Correlation Structure of a DBMAP

Although DBMAPs are able to cover a very wide range of correlation structures, the calculations are rather straightforward as shown in [9]. In the next section the most important results are recalled.

2.3.1 Calculating the correlation structure

Define the matrix $\mathbf{f}(n_1, \dots, n_k)$ as the joint distribution matrix of $(X_1, \dots, X_k)$. Formally:

$$\mathbf{f}_{i,j}(n_1, \dots, n_k) = \mathbf{P}\{X_1 = n_1, \dots, X_k = n_k, J_k = j | J_0 = i\}, \tag{2.1}$$

with $1 \leqslant i, j \leqslant m$ and $0 \leqslant n_l < \infty$. Denote the corresponding z-transform by $\hat{\mathbf{f}}(z_1, \dots, z_k)$. Clearly

$$\hat{\mathbf{f}}(z_1, \dots, z_k) = \prod_{j=1}^{k} \mathbf{D}(z_k).$$

To be able to calculate the scalar covariance $\mathrm{Cov}(X_1, X_k)$ we need to introduce the so-called covariance matrix $\mathbf{COV}(X_1, X_k)$, the entries of which are defined by

$$\mathbf{COV}(X_1, X_k)_{i,j} = \mathbf{E}\big[(X_1 - \lambda)(X_k - \lambda)1_{\{J_k=j\}}\big|J_0 = i\big].$$

Since

$$\mathbf{E}\big[X_1 1_{\{J_k=j\}}\big|J_0 = i\big] = \frac{\partial}{\partial z_1}\hat{\mathbf{f}}(z_1, \ldots, z_k)\Big|_{z_i=1} = \left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{D}^{k-1},$$

$$\mathbf{E}\big[X_k 1_{\{J_k=j\}}\big|J_0 = i\big] = \frac{\partial}{\partial z_k}\hat{\mathbf{f}}(z_1, \ldots, z_k)\Big|_{z_i=1} = \mathbf{D}^{k-1}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right],$$

and

$$\begin{aligned}\mathbf{E}\big[X_1 X_k 1_{\{J_k=j\}}\big|J_0 = i\big] &= \frac{\partial^2}{\partial z_1 \partial z_k}\hat{\mathbf{f}}(z_1, \ldots, z_k)\Big|_{z_i=1} \\ &= \left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{D}^{k-2}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right],\end{aligned}$$

one obtains

$$\mathbf{COV}(X_1, X_k) = \left[\sum_{l=1}^{\infty} l\mathbf{D}_l - \lambda\mathbf{D}\right]\mathbf{D}^{k-2}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l - \lambda\mathbf{D}\right].$$

Consequently the scalar covariance $\mathrm{Cov}(X_1, X_k)$ is given by

$$\begin{aligned}\mathrm{Cov}(X_1, X_k) &= \boldsymbol{\pi}\mathbf{COV}(X_1, X_k)\mathbf{e} \\ &= \boldsymbol{\pi}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{D}^{k-2}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{e} - \lambda^2.\end{aligned} \tag{2.2}$$

From this we derive the autocorrelation function $r(k) = \mathrm{Cov}(X_1, X_k)/\,\mathrm{Var}(X_1)$:

$$r(k) = \frac{\boldsymbol{\pi}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{D}^{k-2}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{e} - \lambda^2}{\boldsymbol{\pi}\left[\sum_{l=1}^{\infty} l^2\mathbf{D}_l\right]\mathbf{e} - \lambda^2}.$$

If $\mathbf{D}$ is finite-dimensional and diagonalisable, (2.2) implies that the autocorrelation of the DBMAP is exponentially decreasing. Indeed, using the spectral decomposition of $\mathbf{D}$, see e.g. [14],

$$\mathbf{D} = \lambda_1\mathbf{B}_1 + \lambda_2\mathbf{B}_2 + \ldots \lambda_n\mathbf{B}_n,$$

with the λ_i the eigenvalues of $\mathbf{D}$ and the matrices $\mathbf{B}_i$ such that $\mathbf{B}_i\mathbf{B}_j = \delta_j^i \mathbf{B}_i$, with δ_j^i the Kronecker δ. Since $\mathbf{D}$ is a stochastic matrix, $\lambda_1 = 1$ and $\mathbf{B}_1 = \mathbf{e}\boldsymbol{\pi}$. Hence

$$\begin{aligned}\mathrm{Cov}(X_1, X_k) &= \boldsymbol{\pi}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\left(\mathbf{D}^{k-2} - \mathbf{e}\boldsymbol{\pi}\right)\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{e} \\ &= \boldsymbol{\pi}\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\left(\lambda_2^{k-2}\mathbf{B}_2 + \ldots \lambda_n^{k-2}\mathbf{B}_n\right)\left[\sum_{l=1}^{\infty} l\mathbf{D}_l\right]\mathbf{e}.\end{aligned}$$

Since the $|\lambda_j| < 1$, a consequence of the Perron-Frobenius theorem (see [53]), the autocorrelation decays exponentially (a similar reasoning can be found in [10]). If the state space is infinite this is not necessarily true. In general

$$\sum_{i=0}^{\infty} \alpha_i \gamma_i^k,$$

with the α_i, γ_i real, can expose any behaviour. By choosing a right set of values for γ_i and α_i, the series can decay according to a power-law. This observation initiated the research on the so-called pseudo self-similar models. More information on this subject can be found in Chapter 5.

Example 2.4. We reuse the Markovian on-off source introduced in Example 2.2. Making use of the spectral decomposition we immediately have:

$$(\mathbf{D} - \mathbf{e}\boldsymbol{\pi})^{k-1} = (\alpha + \beta - 1)^{k-1} \begin{pmatrix} \dfrac{1-\alpha}{2-\alpha-\beta} & \dfrac{\alpha-1}{2-\alpha-\beta} \\ \dfrac{\beta-1}{2-\alpha-\beta} & \dfrac{1-\beta}{2-\alpha-\beta} \end{pmatrix}.$$

This results in

$$\mathrm{Cov}(X_1, X_{k+1}) = (\alpha + \beta - 1)^k \frac{(1-\alpha)(1-\beta)}{(2-\alpha-\beta)^2},$$

and

$$r(k+1) = (\alpha + \beta - 1)^k.$$

2.4.1 DBMAPs and the IDC

The burstiness of a traffic stream can be characterised in several different ways, one of them is the Index of Dispersion for Counts (IDC). Properties and applications of the IDC can be found in e.g. [32] and [55]. The IDC at time k, denoted by $I(k)$, is defined as the variance of the number of arrivals in the interval $(1, k]$, divided by the mean number of arrivals in this interval. Formally, with X_j denoting the number of arrivals at slot j,

$$I(k) = \frac{\mathrm{Var}\big(\sum_{j=1}^{k} X_j\big)}{\mathbf{E}\big[\sum_{j=1}^{k} X_j\big]} = \frac{k\,\mathrm{Cov}(X_1, X_k) + 2\sum_{j=1}^{k-1}(k-j)\,\mathrm{Cov}(X_1, X_{j+1})}{k\lambda}.$$

It is well known, see e.g. [55], that for a renewal process $I(k) = c_1^2$, where c_1^2 is the squared coefficient of variation (variance divided by the square of the mean) of the number of arrivals in a slot. In particular for a Poisson process, $I(k) = 1$. For DBMAPs it is possible to derive a closed formula for the limit of the IDC as shown in [10].

Using the formulas derived in the previous section and recalling [36, pg. 101 corollary 5.1.5] it follows that

$$\lim_{k\to\infty} \sum_{j=1}^{k-1} \frac{k-j-1}{k-1}\big(\mathbf{D}^j - \boldsymbol{\pi}\mathbf{e}\big) = \mathbf{Z} - \mathbf{I},$$

with

$$\mathbf{Z} = \big(\mathbf{I} - (\mathbf{D} - \mathbf{e}\boldsymbol{\pi})\big)^{-1},$$

the fundamental matrix of $\mathbf{D}$. One finds

$$\lim_{k\to\infty} I(k) = \frac{\big(\boldsymbol{\pi}\sum_{l=1}^{\infty} l^2\mathbf{D}_l\big)\mathbf{e} - 3\lambda^3 + 2\boldsymbol{\pi}\big(\sum_{l=1}^{\infty} l\mathbf{D}_l\big)\mathbf{Z}\big(\sum_{l=1}^{\infty} l\mathbf{D}_l\big)\mathbf{e}}{\lambda}. \tag{2.3}$$

Example 2.5. For the Markovian on-off source of Example 2.2 one obtains, taking $p = 1$,

$$\lim_{k\to\infty} I(k) = \frac{(1-\beta)(\alpha+\beta)}{(2-\alpha-\beta)^2}.$$

2.6 The DBMAP-D-1 Queue

The DBMAP-D-1 queue is a single-server queue with an infinite waiting room and a deterministic service time equal to 1 time slot. Its input process is a DBMAP. Hence it is an instance of the multiplexer model introduced in Section 1.2. Denote the number of customers in the system at time k, including the one in service, by q_k. The evolution of this queueing system is expressed by

$$q_{k+1} = [q_k - 1]^+ + X_k,$$

with X_k as before representing the number of arrivals generated by the DBMAP at slot k. This relationship alone is insufficient to obtain the steady-state version of the q_k, one should also incorporate the evolution of the DBMAP. Indeed, the number of customers in the system, together with the state of the DBMAP, is a two-dimensional Markov chain with state space $\{(k,i) \mid 0 \leqslant k < \infty; 0 \leqslant i \leqslant m\}$ and transition matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \dots \\ \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \dots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Because of its special structure the matrix $\mathbf{Q}$ belongs to the class of the M/G/1-type transition matrices, see e.g. [49]. The invariant probability vector of $\mathbf{Q}$ is denoted by $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots)$. It satisfies $\mathbf{xQ} = \mathbf{x}$ and $\mathbf{xe} = 1$, with $(\mathbf{x}_k)_i$ the steady state probability of having k customers in the system and the DBMAP being in state i. Theoretically speaking the problem of determining the stationary buffer distribution is solved: all that is needed is the invariant probability vector of some Markov chain. In practice, the calculation of all these probabilities turns out to be rather difficult. There exists however a huge amount of literature concerning the numerical solution of M/G/1-type Markov chains. We will touch upon the main results after recalling the Pollachek-Kinchin equation, introduced in e.g. [42] and [49].

The generating function associated with the the invariant probability vector $\mathbf{x}$ is defined by $\mathbf{X}(z) = \sum_{k=0}^{\infty} \mathbf{x}_k z^k$. From the relationship

$$\mathbf{x}_i = \mathbf{x}_0 \mathbf{D}_i + \sum_{\nu=1}^{i+1} x_\nu \mathbf{D}_{i+\nu-1} \text{ for } i \geqslant 1,$$

one easily obtains the so-called Pollachek-Kinchin equation

$$\mathbf{X}(z) = (z-1)\mathbf{x}_0 \mathbf{D}(z)\big(z\mathbf{I} - \mathbf{D}(z)\big)^{-1}. \tag{2.4}$$

One immediately observes that $\mathbf{X}(z)$ is completely determined by $\mathbf{x}_0$ and $\mathbf{D}(z)$. The same is true with regard to the actual calculation of the $\mathbf{x}_i$, which are obtained by means of an iterative procedure needing $\mathbf{x}_0$ as starting point. The details of this algorithm can be found in e.g. [49] or [51]. Consequently, determining $\mathbf{x}_0$ is the task one has to perform before being able to calculate the total vector $\mathbf{x}$. Sometimes $\mathbf{x}_0$ can be derived by probabilistic reasoning. If this is not possible, one has to calculate the stochastic matrix $\mathbf{G}$ describing the first passage times from level i to level $i-1$. The entries of $\mathbf{G}$ are probabilistically defined as follows. The entry (i,j) of $\mathbf{G}$ is the probability that starting from state (k,i), with $k \geqslant 1$, the Markov chain defined by $\mathbf{Q}$ will first appear at level $k-1$ in state $(k-1,j)$. Note that k can be chosen arbitrarily by the homogeneity of the matrix $\mathbf{Q}$. The matrix $\mathbf{G}$ satisfies the key relationship

$$\mathbf{G} = \sum_{k=0}^{\infty} \mathbf{D}_k \mathbf{G}^k, \tag{2.5}$$

from which it can be calculated iteratively as described in [49] and [51]. As shown in [49] and [42], $\mathbf{x}_0 = (1-\rho)\mathbf{g}$, with $\mathbf{g}$ the invariant probability vector of $\mathbf{G}$. Note that the complexity of the method depends heavily on the dimension of $\mathbf{G}$. If too many states are involved in the DBMAP description of the arrival process, the computation may become intractable.

2.7 The DBMAP-D-1 Queue and the Dominant Pole Approximation

Before LRD showed up in the domain of queueing analysis, most theorems concerning asymptotic behaviour were proved by some version of the dominant pole approximation. Here we present an application of this classical technique to the case of the DBMAP-D-1 queue. The reason for doing so is twofold. In Chapter 3 the obtained result will be used for its intended purposes, and secondly, in Section 4.2 a generalisation of the dominant pole approximation is presented for a class of tandem queues.

Today buffer asymptotics are frequently derived by using large deviation techniques. When applied to the DBMAP-D-1 queue, many interesting connections with the dominant pole approximation appear, as can be observed in the outline presented in Section 2.7.2.

2.7.1 The dominant pole approximation

The dominant pole approximation is a widely used technique to obtain buffer asymptotics. It can be applied to a large number of queueing systems, as long as the generating function corresponding to the buffer distribution is available. Several different proofs and heuristics, all formulated in their own context, can be found in the literature. The papers [1] and [25] focus on the asymptotics of stationary distributions of M/G/1-type Markov chains, whereas [6] and [7] specialise on the MAP/G/1/K-queue. For the dominant pole approximation in the field of fluid flow modelling the reader is referred to [3]. In [60] the dominant pole approximation is put into a wider perspective. Although the theory developed in [11] does not involve the notion of dominant pole, its application sheds an interesting light on the dominant pole approximation in case of heavy traffic.

A proof for the dominant pole approximation applied to the DBMAP-D-1 queue will be given below. It follows more or less the outline of the proof given in [25]. But whereas in [25] the author relies on the notion of asymptotically geometric sequences, we invoke Darboux's theorem because this theorem yields also potential applications outside the domain of the dominant pole approximation, as will be shown in Section 4.2. Furthermore it allows for a more direct and transparent formulation of the proof. The version of Darboux's theorem formulated here deals with algebraic singularities and can be found in [8]. A singularity α of a complex function f is called algebraic if $f(z)$ can be written as a function which is analytic in a neighbourhood of α, plus a finite sum of terms of the form

$$(1 - z/\alpha)^{-\omega} g(z), \tag{2.6}$$

where g is a function which is analytic and non-zero near α and ω is a complex number not equal to 0, -1, -2, ... The weight of the singularity in (2.6) is the real part of ω.

Theorem 2.7.1 (Darboux). *Suppose $A(z) = \sum_{n \geqslant 0} a_n z^n$ is analytic near 0 and has only algebraic singularities on its circle of convergence. Let w be the maximum of the weights at these singularities. Denote by α_k, ω_k and g_k the values of α, ω and g for those terms of the form (2.6) of weight w. Then*

$$a_n - \frac{1}{n} \sum_k \frac{g_k(\alpha_k) n^{\omega_k}}{\Gamma(\omega_k) \alpha_k^n} = o(r^{-n} n^{w-1}), \tag{2.7}$$

where $r = |\alpha_k|$ is the radius of convergence of $A(z)$, and $\Gamma(s)$ denotes the Gamma function. The notation $f(n) = o\big(g(n)\big)$ stands for $\lim_{n \to \infty} f(n)/g(n) = 0$.

We consider the DBMAP-D-1 queue with as input a DBMAP $(\mathbf{D}, \mathbf{D}_l)$. The matrix $\mathbf{D}$ is supposed to be irreducible, which makes $\mathbf{D}(z)$ irreducible for real and strictly positive z. The radius of convergence of $\mathbf{D}(z)$, which is the minimum of the radii of convergence of its entries, will be denoted by $R_{\mathbf{D}}$. It is supposed that $R_{\mathbf{D}} > 1$, which is a necessary condition. The eigenvalues of $\mathbf{D}(z)$ will be denoted by $\lambda_i(z)$, and $\lambda_1(z)$ is chosen such that it represents the Perron-Frobenius eigenvalue of $\mathbf{D}(z)$ for z real and positive. The left and right Perron-Frobenius eigenvectors are denoted by $\mathbf{u}(z)$ and $\mathbf{v}(z)$. These vectors are normalised, i.e. $\mathbf{u}(z)\mathbf{e} = 1$ and $\mathbf{u}(z)\mathbf{v}(z) = 1$. Hence $\lambda_1(1) = 1$, $\mathbf{u}(1) = \boldsymbol{\pi}$ and $\mathbf{v}(1) = \mathbf{e}$. Starting from $\lambda_1(z)\mathbf{v}(z) = \mathbf{D}(z)\mathbf{v}(z)$ one derives that

$$\frac{d}{dz} \lambda_1(z) \Big|_{z=1} = \mathbf{u}(z) \mathbf{D}'(z) \mathbf{v} \Big|_{z=1} = \boldsymbol{\pi} \mathbf{D}'(1) \mathbf{e} = \rho.$$

For details see [25]. The dominant pole approximation can now be stated and proved for the DBMAP-D-1 queue.

Theorem 2.7.2. *If*

1. *there exists some $\tau \in (1, R_{\mathbf{D}})$ such that $\lambda_1(\tau) = \tau$ and*
2. *$|\mathbf{D}(z)| < \mathbf{D}(\tau)$ for all z with $|z| = \tau$ and $z \neq \tau$,*

then

$$\mathbf{x}_n = \frac{(\tau - 1)}{\lambda_1'(\tau) - 1} \mathbf{x}_0 \mathbf{v}(\tau) \frac{1}{\tau^n} \mathbf{u}(\tau) + o\big(\tau^{-n} \mathbf{u}(\tau)\big). \tag{2.8}$$

Proof. First note that the statement (2.8) is well-defined since the solution τ is unique in $(1, R_{\mathbf{D}})$. This uniqueness can be derived from the convexity of the function $s \mapsto \log \lambda_1(e^s)$, defined on the domain $(-\infty, \log R_{\mathbf{D}})$. The convexity of this function is proved in Section 2.7.2, which deals with the application of large deviation techniques to the DBMAP-D-1 queue.

The proof of (2.8) is based on the Pollachek-Kinchin equation (2.4) which can be rewritten as

$$\mathbf{X}(z) = (z-1)\mathbf{x}_0\mathbf{D}(z)\frac{1}{\Phi(z)}\operatorname{Adj}\bigl(z\mathbf{I}-\mathbf{D}(z)\bigr), \tag{2.9}$$

with $\Phi(z) = \operatorname{Det}\bigl(z\mathbf{I}-\mathbf{D}(z)\bigr)$. Clearly $\Phi(z) = \prod_{i=1}^{m}\bigl(z-\lambda_i(z)\bigr)$. Since we want to use Darboux's theorem we are looking for singularities of $\mathbf{X}(z)$. Clearly only the factor $1/\Phi(z)$ can introduce singularities in the r.h.s. of (2.9). Since $\mathbf{X}(z)$ is a generating function the singularities introduced within the unit disk are removable. The so-called dual subinvariance theorem [53, ex. 1.16 pg. 29] implies $|\lambda_j(z)| \leqslant \lambda_1(|z|)$, with equality only if $|\mathbf{D}(z)| = \mathbf{D}(|z|)$. Hence $|z| = \tau$ and $z \neq \tau$ imply $\lambda_i(z) \neq z$ by condition (2). Together with the uniqueness of τ this property implies that $\lambda_i(z) \neq z$ for $1 < |z| \leqslant \tau$ and $z \neq \tau$. As a consequence $\Phi(z) = 0$ for $1 < |z| \leqslant \tau$ only if $z = \tau$. It remains to show that τ really is a singularity of $\mathbf{X}(z)$. By [53, corollary 2 pg. 8],

$$\operatorname{Adj}\bigl(\tau\mathbf{I}-\mathbf{D}(\tau)\bigr) = K\mathbf{v}(\tau)\mathbf{u}(\tau),$$

with $K \neq 0$. Hence τ is a pole of the r.h.s. of (2.9), since the numerator evaluated in τ equals

$$(\tau-1)\mathbf{x}_0\tau\mathbf{v}(\tau)\mathbf{u}(\tau),$$

which is strictly positive. Consequently τ is the radius of convergence of $\mathbf{X}(z)$. Furthermore τ is the only singularity on this radius of convergence.

We now rewrite $1/\Phi(z)$ to be able to apply Darboux's theorem as formulated above:

$$\frac{1}{\Phi(z)} = (1-z/\tau)^{-1}\frac{1}{\tau}\frac{\tau-z}{z-\lambda_1(z)}\frac{z-\lambda_1(z)}{\Phi(z)}.$$

First we determine $\lim_{z\to\tau}(z-\lambda_1(z))/\Phi(z)$. For $1 < |z| < \tau$,

$$\begin{aligned}
\operatorname{Adj}\bigl(z\mathbf{I}-\mathbf{D}(z)\bigr)\mathbf{v}(z) &= \Phi(z)\bigl(z-\mathbf{D}(z)\bigr)^{-1}\mathbf{v}(z) \\
&= \Phi(z)\frac{1}{z}\sum_{k=0}^{\infty}\frac{\mathbf{D}(z)^k}{z^k}\mathbf{v}(z) \\
&= \Phi(z)\frac{1}{z}\sum_{k=0}^{\infty}\frac{\lambda_1(z)^k}{z^k}\mathbf{v}(z) \\
&= \Phi(z)\frac{1}{z-\lambda_1(z)}\mathbf{v}(z).
\end{aligned}$$

Hence

$$\lim_{z\to\tau} \frac{z-\lambda_1(z)}{\Phi(z)} = \frac{1}{K},$$

because all the entries of $\mathbf{v}(\tau)$ are strictly positive. The limit $\lim_{z\to\tau}(\tau-z)\big(z-\lambda_1(z)\big)$ is evaluated by l'Hôpital's rule:

$$\lim_{z\to\tau} \frac{\tau-z}{z-\lambda_1(z)} = \frac{-1}{\mathbf{u}(\tau)\mathbf{D}'(\tau)\mathbf{v}(\tau)-1}.$$

One is now able to conclude that

$$\begin{aligned}
&\lim_{z\to\tau}\left[(z-1)\mathbf{x}_0\mathbf{D}(z)\frac{1}{\tau}\frac{\tau-z}{z-\lambda_1(z)}\frac{z-\lambda_1(z)}{\Phi(z)}\operatorname{Adj}\big(z\mathbf{I}-\mathbf{D}(z)\big)\right]\\
&\quad= \frac{\tau-1}{1-\mathbf{u}(\tau)\mathbf{D}'(\tau)\mathbf{v}(\tau)}\frac{1}{\tau K}\mathbf{x}_0\mathbf{D}(\tau)K\mathbf{v}(\tau)\mathbf{u}(\tau)\\
&\quad= \frac{\tau-1}{1-\mathbf{u}(\tau)\mathbf{D}'(\tau)\mathbf{v}(\tau)}\big(\mathbf{x}_0\mathbf{v}(\tau)\big)\mathbf{u}(\tau).
\end{aligned}$$

The statement (2.8) now follows directly from Theorem 2.7.1 since there is only one singularity with $\omega = 1$, $\alpha = \tau$ and $g(\alpha)$ as calculated above. □

Remark For the dominant pole approximation one still needs to know the vector $\mathbf{x}_0$. Considering the singularities of the r.h.s. of (2.9) reveals another way to calculate $\mathbf{x}_0$. Since the singularities z_i inside the unit disk are removable, one has for each z_i,

$$\mathbf{x}_0\mathbf{D}(z_i)\operatorname{Adj}\big(z_i\mathbf{I}-\mathbf{D}(z_i)\big) = \mathbf{0},$$

resulting in a system of equations determining $\mathbf{x}_0$. For more details see e.g. [56]. An equivalent formulation of the same principle goes as follows: let γ and $\mathbf{v}$ be an eigenvalue and the corresponding left-eigenvector of $\mathbf{G}$. Since $\mathbf{G}$ is a stochastic matrix we have $|\gamma| \leqslant 1$. Recalling the implicit definition (2.5) for $\mathbf{G}$ one has

$$\begin{aligned}
\gamma\mathbf{v} = \mathbf{G}\mathbf{v} &= \sum_k \mathbf{D}_k\mathbf{G}^k\mathbf{v}\\
&= \Big(\sum_k \mathbf{D}_k\gamma^k\Big)\mathbf{v}\\
&= \mathbf{D}(\gamma)\mathbf{v}.
\end{aligned}$$

Hence γ is an eigenvalue of $\mathbf{D}(\gamma)$ with corresponding eigenvector $\mathbf{v}$. By determining all those z_i with $|z_i| \leqslant 1$, such that z_i is an eigenvalue of $\mathbf{D}(z_i)$, together with the corresponding left eigenvectors, one can reconstruct $\mathbf{G}$ and hence compute $\mathbf{x}_0$.

Numerical example

This numerical example demonstrates the dominant pole approximation for two DBMAP-D-1 queues with as input the superposition of 6 homogeneous Markovian on-off sources (see Example 2.2). For the first queue, shown in Figure 2.1, the parameters are $\alpha = 0.03$, $\beta = 0.04$ and $p = 0.25$. For the second queue, shown in Figure 2.2, the parameters are $\alpha = 0.0075$, $\beta = 0.01$ and $p = 0.25$. For both queueing systems the offered load equals 0.64, but the second DBMAP has a much stronger degree of autocorrelation than the first one, as can be seen from Example 2.4. The influence of the autocorrelation is clear: the tail probabilities of the second queue decay much slower, and it also takes longer to reach the asymptotic regime.

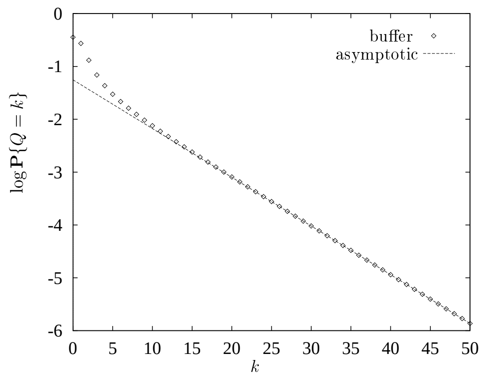

Figure 2.1: Queue 1

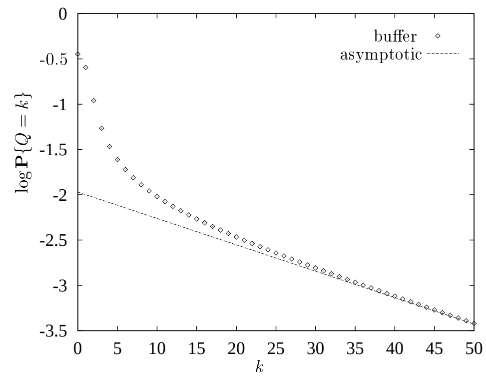

Figure 2.2: Queue 2

2.7.2 The application of large deviation techniques

Large deviation techniques are nowadays frequently used to obtain the asymptotic behaviour of queueing systems. The main advantage of these techniques, as illustrated in [22], is their general applicability, which comes however at the price of sometimes rather coarse bounds. A detailed study of the application of large deviation techniques to discrete-time queueing systems can be found in [29].

In this section we apply the main result of [29] to the case of the DBMAP-D-1 queue. For the sake of completeness, the theorem of [29] we will use is recalled first. In [29] a sequence $\{Y_n, n \geqslant 0\}$ of real-valued random variables is considered. With this sequence the waiting-time sequence $\{W_n, n \geqslant 0\}$ is associated as follows: $W_0 = 0$ and

$$W_{n+1} = [W_n + Y_n]^+, n \geqslant 0.$$

Here $Y_n = X_n - 1$, with X_n the number of arrivals generated by the DBMAP $(\mathbf{D}, \mathbf{D}_l)$ at slot n. Let $S_0 = 0$, $S_n = Y_0 + \ldots + Y_{n-1}$ and let $\Rightarrow$ denote convergence in distribution.

Theorem 2.7.3 (Glynn and Whitt). *Let $\{Y_n, n \geqslant 0\}$ be strictly stationary. If there exists a function φ and positive constants θ^* and ϵ^* such that*

1. $\frac{1}{n} \log \mathbf{E}[e^{\theta S_n}] \to \varphi(\theta)$ *as* $n \to \infty$ *for* $|\theta - \theta^*| < \epsilon^*$,
2. φ *is finite in a neighbourhood of* θ^* *and differentiable at* θ^* *with* $\varphi(\theta^*) = 0$ *and* $\varphi'(\theta^*) > 0$, *and*
3. $\mathbf{E}[e^{\theta^* S_n}] < \infty$ *for* $n \geqslant 1$,

then $W_n \Rightarrow W$ *and*

$$\frac{1}{k} \log \mathbf{P}\{W > k\} \to -\theta^* \text{ as } k \to \infty. \tag{2.10}$$

Suppose there exists some τ having the properties mentioned in Theorem 2.7.2. We demonstrate that Theorem 2.7.3 holds for the sequence Y_n associated with the DBMAP $(\mathbf{D}, \mathbf{D}_l)$.

First an expression for the function φ is derived. Denote the m eigenvalues of $\mathbf{D}(z)$ again by $\lambda_i(z)$, with $\lambda_1(z)$ as before corresponding to the Perron-Frobenius eigenvalue. The corresponding normalised left and right eigenvectors are denoted by $\mathbf{u}_i(z)$ and $\mathbf{v}_i(z)$. Define the scalar function $c_i(z) = \boldsymbol{\pi}\mathbf{v}_i(z)\mathbf{u}_i(z)\mathbf{e}$. By the spectral decomposition theorem one obtains

$$\begin{aligned}
\mathbf{E}[e^{\theta S_n}] &= \boldsymbol{\pi}\frac{\mathbf{D}(e^{n\theta})}{e^{\theta n}}\mathbf{e} \\
&= \boldsymbol{\pi}\frac{1}{e^{\theta n}}\left(\sum_{j=1}^{m} \lambda_j^n(e^\theta)\mathbf{v}_j(e^\theta)\mathbf{u}_j(e^\theta)\right)\mathbf{e} \\
&= \frac{1}{e^{\theta n}}\left(\sum_{j=1}^{m} \lambda_j^n(e^\theta)c_j(e^\theta)\right) \\
&= \frac{\lambda_1^n(e^\theta)}{e^{\theta n}}\left(c_1(e^\theta) + \sum_{j=2}^{m} \frac{\lambda_j^n(e^\theta)}{\lambda_1^n(e^\theta)}c_j(e^\theta)\right).
\end{aligned}$$

Hence the Perron-Frobenius theorem (see [53]) implies

$$\varphi(\theta) = \log \frac{\lambda_1(e^\theta)}{e^\theta}.$$

By considering the expression for $\mathbf{E}[e^{\theta S_n}]$ it is clear that condition (3) is satisfied. For condition (2) we rely on the convexity of $\varphi(\theta)$, which implies the convexity of $\log \lambda_1(e^\theta)$, needed in the proof of Theorem 2.7.2. By Hölder's inequality the function $\mathbf{E}[e^{\theta S_n}]$ is convex, hence φ is convex as a limit of convex functions. Let $\theta^* = \log \tau$, hence $\varphi(\theta^*) = 0$. Clearly there exists an ϵ^* satisfying condition (3). The function φ is differentiable since λ_1 is differentiable, which is a consequence of the implicit function theorem. Furthermore $\log \lambda_1(e^\theta)$ is a convex function with $\log \lambda_1(e^0) = 0$ and $\log \lambda_1(e^{\theta^*}) = \theta^*$. Since

$$\frac{d}{d\theta} \log \lambda_1(e^\theta)\Big|_{\theta=0} = \rho < 1,$$

with ρ the mean arrival rate of the DBMAP, one concludes by the mean value theorem that

$$\frac{d}{d\theta} \log \lambda_1(e^\theta)\Big|_{\theta=\theta^*} > 1,$$

hence $\varphi'(\theta^*) > 0$. Therefore (2.10) holds, which is also implied by Theorem 2.7.2. Observe that less information is obtained by the large deviations approach than by the dominant pole approximation.

Chapter 3

On the FEA-MAA Relationship

Both the Functional Equation Approach (FEA) and the Matrix-Analytical Approach (MAA) designate a methodology to solve queueing systems: the FEA applicable to discrete-time queueing systems, the MAA can be used for continuous-time and discrete-time systems. When both methods are applied to the same queueing system, the same analytical results are obtained — provided that the necessary calculations can be performed. Consequently there has to be some kind of equivalence between the two methods. The mathematical clarification of this relationship is the subject of this chapter.

Because the MAA was already introduced in Chapter 2, only the outline of the FEA presented in Section 3.1 is needed to settle the stage. To demonstrate the FEA-MAA relationship, the FEA is applied in Section 3.2.1 to a general DBMAP-D-1 queue. It turns out that the functional equation obtained by the FEA essentially follows from the Pollachek-Kinchin equation. The ways in which both methods extract information from this equation nevertheless differ to a large extent. It is however again possible to link the used techniques, as is shown in Section 3.2.2.

The chapter concludes with two case studies. The first one is mainly intended as an illustration of the general theory. The second case study deals with an approximation proposed by the FEA for the boundary probabilities, which correspond in the MAA language to the vector $\mathbf{x}_0$ (see Section 2.6). The accuracy of this approximation, with an emphasis on its application to buffer asymptotics, is examined in two numerical examples.

3.1 The Functional Equation Approach

3.1.1 Introduction

In contrast to the MAA, which uses a mixture of probabilistic reasoning, matrix algebra and z-transforms, the FEA explicitly focuses on manipulating generating functions. The main reference on the FEA is [13].

In Section 3.1.2 we sketch, as general as possible, the FEA applied to the multiplexer model. At a certain point we have to conclude that this generality prevents us from proceeding further. The typical mathematical remedy is the introduction of a reasonable assumption. As shown in Section 3.2.1, it turns out that it is natural to assume that the system under study is a DBMAP-D-1 queue.

3.1.2 The FEA applied to the multiplexer queue

Consider the multiplexer model of Section 1.2, i.e. a discrete-time single-server queue with an infinite waiting room and a deterministic service time equal to 1 time slot. For this class of systems Lindley's equation

$$s_{k+1} = (s_k - 1)^+ + e_k, \tag{3.1}$$

is valid with s_k representing the number of customers in the system at time k, and with e_k denoting the number of customers having arrived during slot k.

Applied to such a queueing system the FEA proceeds as follows. First the arrival process, which determines e_k, is modelled by a multi-dimensional stochastic process $\theta = \theta_k$. The components of θ_k are denoted by $\theta_k^{(i)}$, $i \in I$, with I a possibly infinite index set. The actual definition of θ depends on the characteristics of the system under study. For the further development it is necessary that the state space of each component $\theta^{(i)}$ is a subset of the positive integers. The relationship between e_k and θ will be written formally as $e_k = \Theta_k\big((\theta^{(i)})_{i\in I}\big)$. This notation should however not be read as indicating that e_k can only be a mere functional of θ. Relationships like

$$e_k = \sum_{i\in I} \sum_{j=0}^{\theta_{k-1}^{(i)}} X_i,$$

with $\{X_i, i \in I\}$ some collection of integer-valued random variables, are also possible.

The goal of the FEA is to obtain a functional equation involving the generating function associated with both the steady-state versions of s_k and θ_k. Since the steady-state behaviour of a stochastic process results from its evolution, the procedure starts with defining the sequence of the joint probability generating functions P_k, representing the system at time instant k:

$$P_k\big((x_i)_{i\in I}, z\big) = \mathbf{E}\left[\left(\prod x_i^{\theta_{k-1}^i}\right) z^{s_k}\right].$$

It is clear that $S_k(z) = P_k\big((1)_{i\in I}, z\big)$ represents the generating function of s_k.

Using (3.1) one can rewrite P_{k+1} as follows:

$$P_{k+1}\big((x_i)_{i\in I}, z\big) = \mathbf{E}\left[\prod x_i^{\theta_k^{(i)}} z^{(s_k-1)^+ + \Theta_k\left((\theta^{(i)})_{i\in I}\right)}\right]. \tag{3.2}$$

Unfortunately, the generality of the formulation used hitherto prevents us to proceed any further in a formal way. If more detailed information concerning the arrival process is available, i.e. the behaviour of θ and its relationship with e_k can be written down explicitly, one can try to manipulate the r.h.s. of (3.2) such that this equation transforms into a relationship between P_k and P_{k+1} of the form

$$P_{k+1}\big((x_i)_{i\in I}, z\big) = \Phi\Big(P_k\Big(\big[\phi_i\big((x_j)_{j\in I}, z\big)\big]_{i\in I}, z\Big), (x_i)_{i\in I}, z\Big), \tag{3.3}$$

with Φ, ϕ_i and ϕ functions depending on the description of the queueing system. By taking the limit $k \to \infty$ — actually this corresponds to the replacement of both P_{k+1} and P_k with the steady state version P — one obtains the so-called functional equation. Most of the time it is impossible to obtain an explicit expression for P from this implicit equation. However this does not prevent one from deriving several interesting quantities concerning $P(z)$ and $S(z) = P\big((1)_{i\in I}, z\big)$. In some particular cases, see e.g. Section 3.5, it is even possible to obtain an explicit expression for $S(z)$.

3.2 The FEA and the DBMAP-D-1 Queue

Many queueing models which are solved in the literature by the FEA have an arrival process which can be modelled as a DBMAP. By carefully studying these examples one can see that their solutions share a common pattern. The main property of this pattern is that the tuple (e_k, θ_k) is actually modelled as a DBMAP, with the driving process θ corresponding to the underlying Markov chain. This observation enables us to derive the functional equation for the DBMAP-D-1 queue. Halfway these calculations, at the end of Section 3.2.1, the Pollachek-Kinchin equation appears, clearly indicating that the MAA and the FEA are interchangeable.

In Section 3.2.2 it is demonstrated how the functional equation is derived from the Pollachek-Kinchin equation. Furthermore it is shown how this functional equation is used to obtain the performance measures of interest. Also the MAA counterparts of the used techniques are presented.

3.2.1 The Pollachek-Kinchin equation revisited

From now on the tuple (e_k, θ_k) is assumed to correspond to the DBMAP $(\mathbf{D}, \mathbf{D}_l)$. More specifically, θ is supposed to be the underlying Markov chain, having as state space $\prod_{i\in I} J_i$, with J_i denoting the state space of $\theta^{(i)}$. The entries of the matrices $\mathbf{D}_l$ are denoted by $(\mathbf{D}_l)_{(j_i)_{i\in I},(j'_i)_{i\in I}}$, with $(j_i)_{i\in I}, (j'_i)_{i\in I} \in \prod_{i\in I} J_i$. They are given by

$$(\mathbf{D}_l)_{(j_i)_{i\in I},(j'_i)_{i\in I}} = \mathbf{P}\big\{e_k = l, \theta_k^{(i)} = j'_i; i \in I | \theta_{k-1}^{(i)} = j_i; i \in I\big\}. \tag{3.4}$$

Use will be made of conditioning on the events, more formally the σ-algebra, generated by $\big((\theta_{k-1}^{(i)})_{i\in I}, s_k\big)$.
As outlined in Section 3.1.2, the starting point is the joint probability generating function P_{k+1}:

$$\begin{aligned}
&P_{k+1}\big((x_i)_{i\in I}, z\big)\\
&\quad= \mathbf{E}\Bigg[\bigg(\prod_{i\in I} x_i^{\theta_k^i}\bigg) z^{s_{k+1}}\Bigg]\\
&\quad= \mathbf{E}\Bigg[\bigg(\prod_{i\in I} x_i^{\theta_k^i}\bigg) z^{(s_k-1)^+ + e_k}\Bigg]\\
&\quad= \mathbf{E}\Bigg[\mathbf{E}\Bigg[\bigg(\prod_{i\in I} x_i^{\theta_k^i}\bigg) z^{(s_k-1)+e_k} \;\Big|\; \sigma\big((\theta_{k-1}^{(i)})_{i\in I}, s_k\big)\Bigg]\Bigg]\\
&\quad= \mathbf{E}\Bigg[\mathbf{E}\Bigg[\frac{1}{z}\bigg(\prod_{i\in I} x_i^{\theta_k^i}\bigg) z^{s_k+e_k} 1_{\{s_k>0\}} \;\Big|\; \sigma\big((\theta_{k-1}^{(i)})_{i\in I}, s_k\big)\Bigg]\Bigg]\\
&\qquad+ \mathbf{E}\Bigg[\mathbf{E}\Bigg[\bigg(\prod_{i\in I} x_i^{\theta_k^i}\bigg) z^{e_k} 1_{\{s_k=0\}} \;\Big|\; \sigma\big((\theta_{k-1}^{(i)})_{i\in I}, s_k\big)\Bigg]\Bigg]\\
&\quad= \sum_{\substack{\mathbf{j},\mathbf{j}'\in\mathbf{J}\\ l\geqslant 0, m>0}} \bigg(\prod_{i\in I} x_i^{j_i'}\bigg) z^{m+l-1}\mathbf{P}\Big\{e_k = l, s_k = m, \theta_{k-1}^{(i)} = j_i, \theta_k^{(i)} = j_i'; i\in I\Big\}\\
&\qquad+ \sum_{\substack{\mathbf{j},\mathbf{j}'\in\mathbf{J}\\ l\geqslant 0}} \bigg(\prod_{i\in I} x_i^{j_i'}\bigg) z^{l}\mathbf{P}\Big\{e_k = l, s_k = 0, \theta_{k-1}^{(i)} = j_i, \theta_k^{(i)} = j_i'; i\in I\Big\},
\end{aligned} \tag{3.5}$$

with $\mathbf{j} = (j_i)_{i\in I}$, $\mathbf{j}' = (j_i')_i$ and $\mathbf{J} = \prod_{i\in I} J_i$. At this point we invoke (3.4):

$$\mathbf{P}\Big\{s_k = m, e_k = l, \theta_{k-1}^{(i)} = j_i, \theta_k^{(i)} = j_i'; i\in I\Big\} = \\ \mathbf{P}\Big\{s_k = m, \theta_{k-1}^{(i)} = j_i; i\in I\Big\}(\mathbf{D}_l)_{(j_i)_{i\in I},(j_i')_{i\in I}},$$

this equality holds because $\big(e_k, \theta_k^{(i)}\big)$ and s_k are conditionally independent given $(\theta_{k-1}^{(i)})_{i\in I}$ and $(\theta_k^{(i)})_{i\in I}$. For each k we introduce the sequence of row vectors $\mathbf{x}_m^{(k)}$, $m \geqslant 0$, each one of them indexed by $\prod_{i\in I} J_i$, and with entries

$$\big(\mathbf{x}_m^{(k)}\big)_{(j_i)_{i\in I}} = \mathbf{P}\Big\{s_k = m, \theta_{k-1}^{(i)} = j_i; i\in I\Big\}.$$

Furthermore we need the column vectors $\mathbf{w}\big((x_i)_{i\in I}\big)$, also indexed by $\prod_{i\in I} J_i$, and with entries

$$\mathbf{w}\big((x_i)_{i\in I}\big)_{(j_i)_{i\in I}} = \prod_{i\in I} x_i^{j_i}.$$

The vectors $\mathbf{x}_m^{(k)}$ and $\mathbf{w}\big((x_i)_{i\in I}\big)$ enable us to rewrite (3.5) as

$$\begin{aligned}
&P_{k+1}\big((x_i)_{i\in I}, z\big) \\
&\quad = \frac{1}{z} \sum_{\substack{\mathbf{j},\mathbf{j}'\in\mathbf{J} \\ l\geqslant 0, m>0}} \big(\mathbf{x}_m^{(k)}\big)_{(j_i)_{i\in I}} z^m (\mathbf{D}_l)_{\mathbf{j},\mathbf{j}'} z^l \mathbf{w}\big((x_i)_{i\in I}\big)_{\mathbf{j}'} \\
&\qquad + \sum_{\substack{\mathbf{j},\mathbf{j}'\in\mathbf{J} \\ l\geqslant 0}} \big(\mathbf{x}_0^{(k)}\big)_{\mathbf{j}} (\mathbf{D}_l)_{\mathbf{j},\mathbf{j}'} z^l \mathbf{w}\big((x_i)_{i\in I}\big)_{\mathbf{j}'}.
\end{aligned}$$

Define for each k the generating function $\mathbf{X}^{(k)}(z) = \sum_{m=0}^{\infty} \mathbf{x}_m^{(k)} z^m$. This definition implies

$$P_k\big((x_i)_{i\in I}, z\big) = \mathbf{X}^{(k)}(z)\mathbf{w}\big((x_i)_{i\in I}\big).$$

Hence we can resume:

$$\begin{aligned}
&\mathbf{X}^{(k+1)}(z)\mathbf{w}\big((x_i)_{i\in I})\big) = \\
&\qquad\qquad \frac{1}{z}\big(\mathbf{X}^{(k)}(z) - \mathbf{x}_0^{(k)}\big)\mathbf{D}(z)\mathbf{w}\big((x_i)_{i\in I}\big) + \mathbf{x}_0^{(k)}\mathbf{D}(z)\mathbf{w}\big((x_i)_{i\in I}\big),
\end{aligned}$$

or

$$\begin{aligned}
&z\mathbf{X}^{(k+1)}(z)\mathbf{w}\big((x_i)_{i\in I})\big) = \\
&\qquad\qquad \mathbf{X}^{(k)}(z)\mathbf{D}(z)\mathbf{w}\big((x_i)_{i\in I}\big) + (z-1)\mathbf{x}_0^{(k)}\mathbf{D}(z)\mathbf{w}\big((x_i)_{i\in I}\big). \qquad (3.6)
\end{aligned}$$

Denote the steady-state version of the $\mathbf{x}_m^{(k)}$ by $\mathbf{x}_m$ and let $\mathbf{X}(z) = \sum_{m=0}^{\infty} \mathbf{x}_m z^m$. Replace in (3.6) both $\mathbf{X}^{(k+1)}(z)$ and $\mathbf{X}^{(k)}(z)$ by $\mathbf{X}(z)$. Hence the Pollachek-Kinchin equation

$$\mathbf{X}(z) = \mathbf{x}_0(z-1)\mathbf{D}(z)\big(z\mathbf{I} - \mathbf{D}(z)\big)^{-1},$$

introduced in Section 2.6, is obtained.

3.2.2 Derivation and use of the functional equation

To obtain the functional equation, the FEA proceeds with the following steady-state version of (3.6):

$$\mathbf{X}(z)\mathbf{w}\big((x_i)_{i\in I})\big) = \mathbf{X}(z)\mathbf{D}(z)\mathbf{w}\big((x_i)_{i\in I}\big) + (z-1)\mathbf{x}_0\mathbf{D}(z)\mathbf{w}\big((x_i)_{i\in I}\big). \qquad (3.7)$$

To continue we need to make an assumption which will be called the FEA condition. This is not a condition on the arrival process, but on its DBMAP representation. As will be shown in Section 3.2.3, each DBMAP can be replaced by an equivalent DBMAP, describing the same arrival process, which satisfies the FEA condition.

Definition 3.2.1. A DBMAP is said to satisfy the FEA condition if there exist functions ϕ_i and Λ such that

$$\mathbf{D}(z)\mathbf{w}\big((x_i)_{i\in I}\big) = \Lambda\big((x_i)_{i\in I}, z\big)\mathbf{w}\Big(\big(\phi_i\big((x_j)_{j\in I}, z\big)\big)_{i\in I}\Big).$$

From now on we assume that the DBMAP defined by (3.4) satisfies the FEA condition. As mentioned above this can be done without loss of generality. Using Definition 3.2.1 we rewrite (3.7):

$$\begin{aligned} zP\big((x_i)_{i\in I}, z\big) &= \Lambda\big((x_i)_{i\in I}, z\big)P\Big(\phi_i\big((x_j)_{j\in I}, z\big), z\Big) \\ &\quad + (z-1)\Lambda\big((x_i)_{i\in I}, z\big) \sum_{(j_i)_{i\in I}\in\prod_{i\in I} J_i} (\mathbf{x}_0)_{(j_i)_{i\in I}} \prod_{i\in I} \phi_i(x_j, z)^{j_i}. \end{aligned} \tag{3.8}$$

This equation is essentially in the form of the functional equation given in (3.3). Of course the functional equation is only a means to obtain useful information about the queueing system under study. Extracting this information directly from (3.8) is rather difficult because of its complexity. A first step in reducing this complexity is eliminating the variables x_i. A standard way is solving the system of equations

$$x_i = \phi_i((x_j)_{j\in I}, z); \quad i \in I,$$

for x_i. Clearly the solutions for the x_i will depend on z. Therefore they are denoted by $x_i(z)$. In the matrix-analytical setting this system of equations corresponds to an eigenvalue problem. Indeed $\mathbf{w}\big((x_i(z))_{i\in I}\big)$ is an eigenvector of $\mathbf{D}(z)$ with eigenvalue $\Lambda\big((x_i(z))_{i\in I}, z\big)$. Different solutions for the $x_i(z)$ are possible as can be concluded from Section 3.2.3. It is appropriate, as will be shown later on, to choose these solutions $x_i(z)$ such that $\Lambda\big((x_i(z))_{i\in I}, z\big)$ is the Perron-Frobenius eigenvalue of $\mathbf{D}(z)$ for z real and positive. This corresponds to the solutions $x_i(z)$ which satisfy $x_i(1) = 1$, simply because $\mathbf{w}\big((1)_{i\in I}\big)$ is the right Perron-Frobenius eigenvector of the stochastic matrix $\mathbf{D}$. To shorten the notation we will denote $\Lambda\big((x_i(z))_{i\in I}, z\big)$ by $\lambda(z)$ for this choice of the $x_i(z)$. Hence

$$P\big((x_i(z))_{i\in I}, z\big) = \frac{(z-1)\lambda(z)\sum_{(j_i)_{i\in I}\in\prod_{i\in I} J_i}(\mathbf{x}_0)_{(j_i)_{i\in I}}\prod_{i\in I}\phi_i(x_j, z)^{j_i}}{z-\lambda(z)}. \tag{3.9}$$

Because $\lambda(z)$ is the Perron-Frobenius eigenvalue one can calculate, starting from (3.9), the moments of the stationary buffer distribution, which has $S(z) = P((1)_{i\in I}, z)$ as generating function. This is demonstrated in e.g. [64]. Let us recall the general principle. Differentiating the r.h.s. of (3.9) to z results in

$$\frac{\partial}{\partial z}P\big((x_i(z))_{i\in I}, z\big)\Big|_{z=1} + \sum_{i\in I}\frac{\partial}{\partial x_i}P\big((x_i(z))_{i\in I}, z\big)\Big|_{z=1}. \tag{3.10}$$

Furthermore

$$\frac{d}{dz}S(z)\big|_{z=1} = \frac{\partial}{\partial z}P\big((x_i(z))_{i\in I}, z\big)\Big|_{z=1},$$

because $x_i(1) = 1$ for all i. Since it is also possible to differentiate the l.h.s. of (3.9) to z, one can calculate $S'(1)$. This procedure can be repeated for the higher moments. Although the method is straightforward, the calculations quickly become very cumbersome, which is not surprising in view of the recursive moment formulas derived for M/G/1-type queues in [49, pg. 143–148].

The choice of the Perron-Frobenius makes it also possible to obtain the dominant pole approximation. It is argued in e.g. [65] that the dominant pole of $S(z)$ equals the dominant pole of $P\big(x_i(z), z\big)$. Hence this pole can be derived from (3.9). Obtaining the residue for $S(z)$ requires some heuristic reasoning, described in detail in e.g. [63]. However, by invoking the relationship between the FEA and the MAA as described above, these heuristics can be avoided by applying Theorem 2.7.2.

3.2.3 On the FEA condition

As mentioned in Section 3.2.2, we will show that one can associate with every DBMAP an equivalent DBMAP, which satisfies the FEA condition given in Definition 3.2.1. By equivalent is meant that both DBMAPs model the same arrival process. We construct the associated DBMAP for the homogeneous superposition of N DBMAPs. The case of a single DBMAP follows directly by taking $N = 1$.

Consider an arrival process which is modelled by the DBMAP $(\mathbf{D}, \mathbf{D}_l)$. Suppose this DBMAP has K states, which are denoted by $1, \dots, K$. In the matrix-analytical context the superposition of N such DBMAPs would correspond, as pointed out in Section 2.2.1, to the DBMAP $(\mathbf{D}^N, \mathbf{D}_l^N)$, with $\mathbf{D}_l^N$ defined by

$$\mathbf{D}_l^N = \sum_{l_1+\dots l_N = l} \mathbf{D}_{l_1} \otimes \dots \otimes \mathbf{D}_{l_N}.$$

The FEA prefers another way to describe this superposition. Construct the driving process θ as follows: take its components $\theta_k^{(i)}$, for $i = 1, \dots, K$, to be the number of sources in state i at time k. Starting from θ we now construct the DBMAP $(\mathbf{D}^\star, \mathbf{D}_l^\star)$. Its state space consists of the tuples $(p_1, \dots, p_K) \in \prod_{i=1}^K \{0, 1, \dots, N\}$ with $\sum_i p_i = N$. The entries of the matrices $(\mathbf{D}^\star)_l$ are defined by

$$(\mathbf{D}_l^\star)_{(p_1,\dots,p_K),(q_1,\dots,q_K)} = \sum_{r\in \mathrm{CI}(p_j), s\in \mathrm{CI}(q_j)} (\mathbf{D}_l^N)_{r,s}.$$

The expression $\mathrm{CI}(p_j)$ denotes the set of the tuples $(r_1, ..., r_n) \in \prod_{i=1}^N \{1, \dots, K\}$ which have the property that the number of occurrences of state m in $(r_1, ..., r_n)$

equals p_m. (The abbreviation CI stands for corresponding indices.) It is clear by its construction that $(\mathbf{D}^\star, \mathbf{D}_l^\star)$ is a DBMAP, and that it describes the same arrival process as $(\mathbf{D}^N, \mathbf{D}_l^N)$. So we only need to show that $(\mathbf{D}^\star, \mathbf{D}_l^\star)$ satisfies the FEA condition. Let

$$\mathbf{w}^\star(1, x_2, \ldots, x_K)_{p_1,\ldots,p_K} = \prod_{i=1}^{K} x_i^{p_i}.$$

The replacement of x_1 by 1 is possible because for each tuple $(p_1, \ldots, p_K)$ of the state space one has $p_1 + \ldots p_K = N$, so there are only $K - 1$ degrees of freedom. Of course we could replace any of the x_i by 1, but x_1 is chosen for convenience. Define also

$$\mathbf{w}(x_1, \ldots, x_K) = \begin{pmatrix} 1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}.$$

Because the x_i have exponent 1, it follows that

$$\mathbf{D}(z)\mathbf{w}\big((x_i)_i\big) = \begin{pmatrix} \varphi_1(x_2, \ldots, x_K, z) \\ \varphi_2(x_2, \ldots, x_K, z) \\ \vdots \\ \varphi_K(x_2, \ldots, x_K, z) \end{pmatrix},$$

hence

$$\mathbf{D}(z)\mathbf{w}\big((x_i)_i\big) = \varphi_1(x_2, \ldots, x_K, z)\mathbf{w}\left(1, \frac{\varphi_2(x_2, \ldots, x_K, z)}{\varphi_1(x_2, \ldots, x_K, z)}, \ldots, \frac{\varphi_K(x_1, \ldots, x_K, z)}{\varphi_1(x_2, \ldots, x_K, z)}\right).$$

Furthermore, since $\mathbf{D}^\star(z)\mathbf{w}^\star\big((x_i)_i\big) = \mathbf{D}^N(z) \bigotimes_{i=1}^N \mathbf{w}\big((x_i)_i\big)$, one has

$$\mathbf{D}^\star(z)\mathbf{w}^\star(x_i) = \varphi_1^N(x_2, \ldots, x_K, z)\mathbf{w}^\star\left(\left[\frac{\varphi_i(x_1, \ldots, x_K, z)}{\varphi_1(x_1, \ldots, x_K, z)}\right]_i\right).$$

Hence the functions Λ and ϕ_i, introduced in Definition 3.2.1, are given by $\Lambda = \varphi_1^N$ and $\phi_i = \varphi_i/\varphi_1$.

3.3 Generalisations

The FEA-MAA relationship, presented above for the DBMAP-D-1 queue, is valid for a much wider range of queueing systems. First we take a look at the DBMAP-G-1 queue, where G refers to a generally distributed service time. Secondly we consider the DBMAP-D-c multi-server queue, with $c \geqslant 1$ denoting the number of parallel servers. Summarising we can say that the link between the FEA and the MAA is again essentially the Pollachek-Kinchin equation, which exists for both types of queueing systems.

3.3.1 Generally distributed service times

A DBMAP-G-1 queue has a generally distributed service time, with the corresponding z-transform denoted by $G(z) = \sum_{k=1}^{\infty} g_k z^k$. Although we believe that it is possible to apply the FEA to the DBMAP-G-1 queue in a way similar to the FEA solution of the GI-G-1 queue (see [13, Section 1.2]), we did not perform the actual calculations because of the expected complexity. So we contend ourselves to recall from [9] the derivation of the Pollachek-Kinchin equation for the DBMAP-G-1 queue. Next we use this equation to derive the DBMAP-G-1 counterpart of (3.9). Note that we limit the outline to the analysis of the DBMAP-G-1 queue observed at departure epochs.

Define $\left[\mathbf{A}_n^{(k)}\right]_{i,j}$ as the conditional probability that there are n arrivals during a period k slots long, and that at the end the DBMAP is in state j, given that it started in state i. The matrices $\mathbf{A}_n^{(k)}$ can be calculated by noting that

$$\mathbf{A}^{(k)}(z) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \mathbf{A}_n^{(k)} z^n = \left(\mathbf{A}^{(1)}(z)\right)^k = \left(\mathbf{D}(z)\right)^k. \tag{3.11}$$

Let $[\mathbf{A}_n]_{i,j}$ be the probability that during a service there are n arrivals and that at the end of the service the DBMAP is in state j, given that it was in state i at the start of the service, hence

$$\mathbf{A}_n = \sum_{k=1}^{\infty} g_k \mathbf{A}_n^{(k)}.$$

For the z-transform $\mathbf{A}(z) = \sum_{n=0}^{\infty} \mathbf{A}_n(z)$ it is clear that

$$\mathbf{A}(z) = \sum_{k=1}^{\infty} g_k \mathbf{D}(z)^k. \tag{3.12}$$

Let furthermore $[\mathbf{B}_n]_{i,j}$ be the probability that, given a departure which leaves the system empty and the arrival process in state i, at the next departure the DBMAP is in state j and there have been $n+1$ arrivals meanwhile. Straightforward probabilistic reasoning shows that

$$\mathbf{B}_n = \left(\mathbf{I} - \mathbf{D}_0\right)^{-1} \sum_{j=0}^{n} \mathbf{D}_{j+1} \mathbf{A}_{n-j}.$$

For the generating function $\mathbf{B}(z) = \sum_{n=0}^{\infty} \mathbf{B}_n z^n$ one has

$$\mathbf{B}(z) = \frac{1}{z}\left(\mathbf{I} - \mathbf{D}_0\right)^{-1}\left(\mathbf{D}(z) - \mathbf{D}_0\right)\mathbf{A}(z).$$

From the definitions of $\mathbf{A}_n$ and $\mathbf{B}_n$ it can be understood that at departure epochs the invariant probability vector $\mathbf{x} = \begin{pmatrix}\mathbf{x}_0 & \mathbf{x}_1 & \ldots\end{pmatrix}$ is determined by

$$\mathbf{x} = \mathbf{x}\begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \ldots \\ \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \ldots \\ \mathbf{0} & \mathbf{A}_0 & \mathbf{A}_1 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

and $\mathbf{xe} = 1$. One can derive the Pollachek-Kinchin equation

$$\mathbf{X}(z)\big(z\mathbf{I} - \mathbf{A}(z)\big) = \mathbf{x}_0\big(z\mathbf{B}(z) - \mathbf{A}(z)\big).$$

Starting from this Pollachek-Kinchin equation we will derive the DBMAP-G-1 counterpart of (3.9). The notation introduced in Section 3.2 will be reused, furthermore we assume the DBMAP satisfies the FEA condition. Hence

$$P\big((x_i)_{i\in I}, z\big) = \mathbf{X}(z)\mathbf{w}\big((x_i)_{i\in I}\big),$$

and $\mathbf{w}\big((x_i(z))_{i\in I}\big)$ represents the Perron-Frobenius left eigenvector of $\mathbf{D}(z)$. Equation (3.12) implies

$$\mathbf{A}(z)\mathbf{w}\big(x_i(z)\big) = \sum_{k=1}^{\infty} g_k \lambda(z)^k \mathbf{w}(x_i(z)),$$

hence one obtains

$$\mathbf{P}\big(x_i(z), z\big) = \frac{\mathbf{x}_0[z\mathbf{B}(z) - \mathbf{A}(z)]}{z - G(\lambda(z))}\mathbf{w}(x_i(z)). \tag{3.13}$$

Example 3.4 (The GI-G-1 queue). We apply (3.13) to the GI-G-1 queue, extensively studied in [13]. Note that the matrices denote in fact scalars. One obtains

$$P(1, z) = \mathbf{x}_0 \frac{\mathbf{D}(z) - 1}{1 - \mathbf{D}_0} \frac{\mathbf{A}(z)}{z - G(\lambda(z))},$$

the value for $\mathbf{x}_0$ can be derived according to the outline presented in [42]:

$$\mathbf{x}_0 = (1 - \mathbf{D}_0)\big(\mathbf{D}'(1)\big)^{-1}\big(1 - \mathbf{D}'(1)\mathbf{G}'(1)\big).$$

With this value for $\mathbf{x}_0$ the same result is obtained as in [13, section 1.2.3.2].

3.4.1 The DBMAP-D-c queue

The queueing system considered here has an infinite waiting room and c parallel servers, each having a service time equal to 1 slot. Its input is a DBMAP $(\mathbf{D}, \mathbf{D}_l)$. For the FEA solution of this type of queueing system we refer the reader to [56]. Just as with the DBMAP-G-1 queue we will not go into detail. We limit ourselves to presenting the Pollachek-Kinchin equation for the DBMAP-D-c queue, which can be found in [49, pg. 310–329]. The interested reader can compare the Pollachek-Kinchin equation with its FEA counterpart (Equation (20)) in [56], and conclude that they are essentially the same. A second reason for presenting the Pollachek-Kinchin equation for the DBMAP-D-c queue is that we will deal with such a system in Section 4.3.3.

The DBMAP-D-c queue is determined by the Markov chain with transition matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \dots \\ \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \dots \\ \mathbf{0} & \mathbf{D}_0 & \mathbf{D}_1 & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Here the first $c+1$ rows are equal to $\begin{pmatrix}\mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \dots\end{pmatrix}$, which is a consequence of the fact that up to c cells can be served in 1 slot. For the invariant probability vector $\mathbf{x} = \begin{pmatrix}\mathbf{x}_0 & \mathbf{x}_1 & \dots\end{pmatrix}$ of $\mathbf{Q}$ this results in

$$\mathbf{x}_i = \mathbf{x}_0\mathbf{D}_i + \mathbf{x}_1\mathbf{D}_i + \dots \mathbf{x}_c\mathbf{D}_i + \sum_{\nu=c+1}^{i+c+1} \mathbf{x}_\nu \mathbf{D}_{i+c+1-\nu}.$$

Hence the Pollachek-Kinchin equation for the DBMAP-D-c queue is given by

$$\mathbf{X}(z) = \sum_{j=0}^{c-1} \mathbf{x}_j \mathbf{D}(z)(z^c - z^j)\big(z^c\mathbf{I} - \mathbf{D}(z)\big)^{-1}.$$

For details about the numerical calculation of the $\mathbf{x}_i$ the reader is referred to [49, pg. 310–329].

3.5 Case study 1: Geometric Message Lengths

This first case study deals with a rather simple arrival process, which is presented in [61]. The simplicity of the model allows us to focus on clarifying the general theory of Section 3.2. Furthermore, it is possible to obtain from the functional equation an explicit expression for $S(z)$, the generating function of the stationary buffer distribution. The determination of $S(z)$ is however not trivial and will be looked at from both the FEA and the MAA point of view. To conclude we apply Theorem 2.7.2 to obtain the asymptotic behaviour of the tail probabilities.

3.5.1 System description

Consider a discrete-time queueing system with infinite storage capacity, to which variable-length messages consisting of multiple fixed-length packets arrive at the rate of one packet per slot. The message lengths (in terms of packets) are assumed to be independent, geometrically distributed, random variables with generating function

$$A(z) = \frac{(1-\alpha)z}{1-\alpha z}.$$

The numbers of new messages generated during a slot are assumed to be i.i.d. random variables with generating function

$$M(z) = \sum_{k=1}^{\infty} m_k z^k.$$

3.5.2 The FEA derivation of the functional equation

Denote by b_k the number of new messages generated during slot k. The total number of packets arriving at slot k, denoted by e_k, is given by

$$e_k = b_k + \sum_{i=1}^{e_{k-1}} c_i, \tag{3.14}$$

with the c_i a collection of i.i.d. random variables with generating function $C(z) = 1 - \alpha + \alpha z$. Consequently the driving process θ consists of only one component: $\theta_k = e_k$. Specialising Definition 3.1.2, one obtains the collection of joint probability generating functions

$$P_k(x,z) = \mathbf{E}\big[x^{e_{k-1}} z^{s_k}\big]. \tag{3.15}$$

From (3.1) and (3.14) it follows that

$$\begin{aligned} P_{k+1}(x,z) &= \mathbf{E}\Big[\mathbf{E}\Big[(xz)^{m_k + \sum_{i=1}^{e_{k-1}} c_i} z^{(s_k-1)^+} \;\Big|\; \sigma\big(e_{k-1}, m_k\big)\Big]\Big] \\ &= M(xz)\mathbf{E}\big[C(xz)^{e_{k-1}} z^{(s_k-1)^+}\big] \\ &= \frac{M(xz)}{z}\big(P_k(1-\alpha+\alpha xz, z) + (z-1)\mathbf{P}[s_k = 0]\big). \end{aligned} \tag{3.16}$$

When $k \to \infty$, both P_k and P_{k+1} converge to the steady state version P. Using (3.16) one obtains the functional equation

$$P(x,z) = \frac{M(xz)}{z}\big(P(1-\alpha+\alpha xz, z) + (z-1)p_0\big), \tag{3.17}$$

with p_0 the steady state probability of having an empty system.

3.5.3 The MAA derivation of the functional equation

According to the outline of Section 3.2 we construct the DBMAP corresponding to the arrival process described above. The state space of this DBMAP, being the state space of θ, consists of the non-negative integers. State k corresponds to the generation of k packets, or otherwise stated, to k active messages. The transition matrix $\mathbf{D}$, with row i containing the probabilities determining the transitions starting from state $i-1$, is given by

$$\mathbf{D} = \begin{pmatrix} m_0 & m_1 & \dots \\ m_0(1-\alpha) & m_0\alpha + m_1(1-\alpha) & \dots \\ m_0(1-\alpha) & m_1(1-\alpha)^2 + m_0 2\alpha(1-\alpha) & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{3.18}$$

It follows from (3.14) that row i of $\mathbf{D}$ corresponds to the probability density function of the convolution $\mathbf{P}_M \otimes \bigotimes_{j=1}^{i-1} \mathbf{P}_C$, with $\mathbf{P}_M$ and $\mathbf{P}_C$ the distributions corresponding to the generating functions $M(z)$ and $C(z)$. Applying Definition 3.4 we have

$$(\mathbf{D}_l)_{i,j} = \mathbf{P}\{e_k = l, \theta_k = i | \theta_{k-1} = j\} = \delta_l^j \mathbf{D}_{i,j},$$

with δ_l^j the Kronecker δ. Hence we have to take each column of $\mathbf{D}_i$ equal to $\mathbf{0}$, except for column $i-1$, which equals the $i-1$-th column of $\mathbf{D}$. Doing so the following expression for the z-transform $\mathbf{D}(z)$ is obtained:

$$\mathbf{D}(z) = \mathbf{D} \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & z & 0 & \dots \\ 0 & 0 & z^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{3.19}$$

One easily derives from $\mathbf{D}(z)$ the mean arrival rate $\rho = M'(1)/(1-\alpha)$.

Having defined the DBMAP $(\mathbf{D}, \mathbf{D}_l)$ we now proceed with noticing that in this setting

$$\mathbf{w}(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \end{pmatrix}. \tag{3.20}$$

Straightforward algebra shows that

$$\mathbf{D}(z)\mathbf{w}(x) = M(xz)\mathbf{w}(C(xz)),$$

hence the FEA-condition is satisfied. We end up with the functional equation

$$zP(x,z) = M(xz)P(C(xz),z) + (z-1)M(xz)\sum_{k=0}^{\infty}(\mathbf{x}_0)_k C(xz)^k.$$

Fortunately the vector $\mathbf{x}_0$ can be derived by probabilistic reasoning. Clearly the queueing system can only be empty at slot k if no arrivals occurred during the previous slot. So it is necessary that $e_{k-1} = \theta_{k-1} = 0$, hence one can conclude that $\mathbf{x}_0 = \begin{pmatrix} 1-\rho & 0 & 0 & \dots \end{pmatrix}$. Using this additional information one obtains again (3.17).

3.5.4 Using the functional equation

To use the functional equation one needs to replace x by a function $x(z)$ as explained in Section 3.2.2. The system of equations introduced there reduces to the single equation

$$x(z) = C(zx(z)),$$

with solution

$$x(z) = \frac{1-\alpha}{1-\alpha z}. \tag{3.21}$$

It is clear that $M(zx(z))$ and $\mathbf{w}(x(z))$ represent the Perron-Frobenius eigenvalue and the corresponding right eigenvector of $\mathbf{D}(z)$. Substituting $x(z)$ for x reduces the functional equation to

$$P\Big(\frac{1-\alpha}{1-\alpha z}, z\Big) = \frac{(z-1)p_0 M\left[\frac{1-\alpha}{1-\alpha z}\right]}{z - M\left[\frac{1-\alpha}{1-\alpha z}\right]}. \tag{3.22}$$

By l'Hôpital's rule one can calculate $p_0 = 1-\rho$ from this equation. Furthermore, as shown in [61] or [62], the moments of buffer distribution can be calculated from (3.22).

3.5.5 The generating function of the buffer distribution

For the queueing system under study one can derive an explicit expression for $S(z)$, starting from the functional equation (3.17), or by matrix-analytical arguments. We start with recalling from [61] the procedure according to the FEA. Here the determination of $S(z) = \sum_{k=0}^{\infty} s_k z^k$ requires an iterative procedure. Define $R_i(z) = C\big(zR_{i-1}(z)\big)$ and take $R_0(z) = 1$. Repeated use of (3.17) results in

$$\begin{aligned}
&S(z)\\
&\quad = P(1,z)\\
&\quad = \frac{M(z)}{z}\big(P(R_1(z),z) + (z-1)(1-\rho)\big)\\
&\quad = \frac{M(z)}{z}\Big[\frac{M(R_1(z)z)}{z}P(R_2(z),z) + (z-1)(1-\rho)\Big(1 + \frac{M(R_1(z)z)}{z}\Big)\Big]\\
&\quad = \prod_{i=0}^{\infty}\frac{M(R_i(z)z)}{z}P\Big(\frac{1-\alpha}{1-\alpha z}, z\Big) + (z-1)(1-\rho)\sum_{n=0}^{\infty}\prod_{i=0}^{n}\frac{M(R_i(z)z)}{z}.
\end{aligned} \tag{3.23}$$

Invoking (3.22) gives

$$\begin{aligned} S(z) &= (z-1)(1-\rho)\sum_{n=0}^{\infty}\prod_{i=0}^{n}\frac{M(R_i(z)z)}{z} \\ &+ (z-1)(1-\rho)\frac{M\left[\frac{(1-\alpha)z}{1-\alpha z}\right]\prod_{i=0}^{\infty}\frac{M(R_i(z)z)}{z}}{z-M\left[\frac{(1-\alpha)z}{1-\alpha z}\right]}. \end{aligned} \tag{3.24}$$

The matrix-analytical counterpart of this reasoning goes as follows. Denote the Perron-Frobenius eigenvalue $M(zx(z))$ of $\mathbf{D}(z)$ by $\phi(z)$. Clearly there exists a $z_0 \in (1,\infty)$ such that $\phi(z_0) = z_0$ and $\phi(z) < z$ on $(1, z_0)$. Hence, if $1 < z < z_0$,

$$\big(z\mathbf{I} - \mathbf{D}(z)\big)^{-1} = \frac{1}{z}\sum_{k=0}^{\infty}\frac{\mathbf{D}(z)^k}{z^k}. \tag{3.25}$$

Using $S(z) = \mathbf{X}(z)\mathbf{e}$ we obtain

$$S(z) = (z-1)(1-\rho)\sum_{n=0}^{\infty}\prod_{i=0}^{n}\frac{M(R_i(z)z)}{z}. \tag{3.26}$$

Since for $z \in (1, z_0)$,

$$\prod_{i=0}^{\infty}\frac{M(R_i(z)z)}{z} = 0,$$

both (3.24) and (3.26) yield the same result on the considered interval.

3.5.6 Asymptotic behaviour

The asymptotic behaviour of the stationary distribution of the buffer occupation can be derived from (3.24), as shown in [61], or in [64] for a more general model. The extra term in (3.24), when compared to (3.26), is explicitly used by this approach. This is however not surprising because

$$\prod_{i=0}^{\infty}\frac{M(R_i(z_0)z_0)}{z_0} \neq 0,$$

since $M(R_i(z_0)z_0)/z_0$ converges fast enough to 1 for $i \to \infty$.
The conditions of Theorem 2.7.2 are satisfied, hence

$$\mathbf{x}_n \sim \frac{(z_0-1)}{\phi'(z_0)-1}\mathbf{x}_0\mathbf{v}(z_0)z_0^{-n}\mathbf{u}(z_0). \tag{3.27}$$

with z_0 and ϕ as defined in Section 3.5.5. The vectors $\mathbf{v}(z)$ and $\mathbf{u}(z)$ are respectively the right and left Perron-Frobenius eigenvector of $\mathbf{D}(z)$ such that

$\mathbf{u}(z)\mathbf{e} = 1$ and $\mathbf{v}(z)\mathbf{u}(z) = 1$. The Perron-Frobenius theorem [53, Theorem 1.1] implies

$$\lim_{k\to\infty} \frac{\mathbf{D}(z)^k}{\phi(z)^k}\mathbf{e} = \mathbf{v}(z). \tag{3.28}$$

Using $\lim_{k\to\infty} R_k(z) = x(z)$ we obtain

$$\prod_{i=0}^{\infty} \frac{M(R_i(z)z)}{M(x(z)z)}\mathbf{w}(x(z)) = \mathbf{v}(z). \tag{3.29}$$

Finally we are able to conclude

$$s_n = \mathbf{x}_n\mathbf{e} \sim \frac{z_0 - 1}{\phi'(z_0) - 1} \prod_{i=0}^{\infty} \frac{M(R_i(z_0)z_0)}{z_0}, \tag{3.30}$$

which is the same result as in [61, pg. 30–31].

3.6 Case Study 2: General On-Off Sources

This second case study deals with the multiplexer queue having as input an homogeneous superposition of on-off sources with generally distributed on- and off-periods. The main reason for studying this example is that it allows for the introduction and the examination of an approximation for the vector $\mathbf{x}_0$. In two numerical examples the accuracy of this approximation is assessed, with the emphasis on its application to the computation of buffer asymptotics.

3.6.1 System description

Consider the multiplexer queue to which N stochastically independent on-off sources are fed. This superposition is taken to be homogeneous, i.e. all sources have the same characteristics. During an on-period, a source generates exactly one cell per slot, whereas no cells are generated during an off-period. The generating functions corresponding to the duration of the on- and the off-periods are denoted respectively by $A(z) = \sum_{k=1}^{\infty} a_k z^k$ and $B(z) = \sum_{k=1}^{\infty} b_k z^k$.

3.6.2 The matrix-analytical approach

Since this case study is not intended in the first place as an illustration of the theory of Section 3.2, the reader is referred to [63] for the FEA-derivation of the functional equation. The MAA starts with modelling a single on-off source as a DBMAP. As in [63] the states $A_1, A_2, \ldots$ and $B_1, B_2, \ldots$ are introduced. A source is said to be in state A_n if it is in the n-th slot of an on-period. Similarly, a source is said to be in state B_n if it is in the n-th slot of an off-period. Using these states an on-off source can be described as an infinite-dimensional Markov chain, which is defined by the following transitions:

- from A_n to A_{n+1} with probability $p_n^a = \dfrac{1-\sum_{i=1}^{n} a_i}{1-\sum_{i=1}^{n-1} a_i}$,
- from B_n to B_{n+1} with probability $p_n^b = \dfrac{1-\sum_{i=1}^{n} b_i}{1-\sum_{i=1}^{n-1} b_i}$,
- from B_n to A_1 with probability $1-p_n^b$ and
- from A_n to B_1 with probability $1-p_n^a$.

To be able to write down the matrices $\mathbf{D}$, $\mathbf{D}_l$ and $\mathbf{D}(z)$, we order the states in the following way: A_1, A_2, ... , B_1, B_2, ... Following the same reasoning as in Section 3.5.3 we obtain

$$\mathbf{D}(z) = \begin{pmatrix} 0 & p_1^a z & 0 & \dots & 1-p_1^a & 0 & 0 & \dots \\ 0 & 0 & p_2^a z & \dots & 1-p_2^a & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots & 1-p_3^a & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots \\ (1-p_1^b)z & 0 & 0 & \dots & 0 & p_1^b & 0 & \dots \\ (1-p_2^b)z & 0 & 0 & \dots & 0 & 0 & p_2^b & \dots \\ (1-p_3^b)z & 0 & 0 & \dots & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{3.31}$$

The DBMAP corresponding to the superposition is denoted by $\left(\mathbf{D}^N, \mathbf{D}_l^N\right)$. Hence $\mathbf{D}^N(z) = \bigotimes_{j=1}^N \mathbf{D}(z)$ as demonstrated in Section 2.2.1. Multiply both sides of the rewritten Pollachek-Kinchin equation

$$z\mathbf{X}(z) = \mathbf{X}(z)\mathbf{D}^N(z) + (z-1)\mathbf{x}_0\mathbf{D}^N(z) \tag{3.32}$$

by

$$\mathbf{w}^N(\underline{x},\underline{y}) = \bigotimes_{j=1}^{N} \mathbf{w}(\underline{x},\underline{y}) = \bigotimes_{j=1}^{N} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ 1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix}, \tag{3.33}$$

with $\underline{x} = (x_1, x_2, x_3, \dots)$ and $\underline{y} = (1, y_2, y_3, \dots)$. This results in the functional equation

$$\begin{aligned} z\mathbf{X}(z)\mathbf{w}^N(\underline{x},\underline{y}) = \mathbf{X}(z)\Big[x_1 z D\Big(\frac{y_2}{x_1 z}\Big)\Big]^N \mathbf{w}^N\big(\underline{G}(\underline{x},\underline{y},z), \underline{H}(\underline{x},\underline{y},z)\big) \\ + (z-1)\mathbf{x}_0\mathbf{w}^N\big(\underline{G}(\underline{x},\underline{y},z), \underline{H}(\underline{x},\underline{y},z)\big). \end{aligned} \tag{3.34}$$

Here

$$\underline{G}(\underline{x},\underline{y},z) = \big(G_1(\underline{x},\underline{y},z), G_2(\underline{x},\underline{y},z), G_3(\underline{x},\underline{y},z), \dots\big),$$
$$\underline{H}(\underline{x},\underline{y},z) = \big(1, H_2(\underline{x},\underline{y},z), H_3(\underline{x},\underline{y},z), \dots\big)$$

and

$$G_n(\underline{x},\underline{y},z) = \frac{C_n(x_{n+1}z)}{x_1 z D_1\left(\frac{y_2}{x_1 z}\right)},$$
$$H_n(\underline{x},\underline{y},z) = \frac{D_n\left(\frac{y_{n+1}}{x_1 z}\right)}{D_1\left(\frac{y_2}{x_1 z}\right)}.$$

The generating functions $C_i(z)$ and $D_i(z)$ are defined by $C_i(z) = 1 - p_n^a + p_n^a z$ and $D_i(z) = 1 - p_n^b + p_n^b z$.

3.6.3 An approximation for $\mathbf{x}_0$

Since the off-periods are generally distributed, it is impossible to write down $\mathbf{x}_0$ or $\mathbf{g}$ (see Section 2.6) by straightforward probabilistic reasoning. Hence $\mathbf{g}$ has to be calculated numerically as the invariant probability vector of $\mathbf{G}$. Since in general the state space of the DBMAP $(\mathbf{D}^N, \mathbf{D}_l^N)$ is infinite, the on-off sources have to be approximated first by ones with bounded on- and off-periods. But even then the matrices may remain very large, resulting in intractable computations.

The FEA encounters the same problem, although otherwise formulated, and proposes an approximation, see e.g. [65] or [66]. Let us translate this approximation to the DBMAP setting. Consider the steady-state vector $\boldsymbol{\pi}^N$ of $\mathbf{D}^N$. Clearly $\boldsymbol{\pi}^N = \bigotimes_{j=1}^N \boldsymbol{\pi}$, with $\boldsymbol{\pi}\mathbf{D} = \boldsymbol{\pi}$ and $\boldsymbol{\pi}\mathbf{e} = 1$. Only if a source is in one of the states B_k, it generated no cell during the previous time slot. Hence the system can only be empty if each source is in some B_k-state. The approximation now is based on the assumption that the state of the total arrival process will not differ much if observed when the system is empty, or when all sources are passive. This reasoning results in the approximation $\tilde{\mathbf{g}}$ for $\mathbf{g}$ given by

$$\tilde{\mathbf{g}} = \bigotimes_{j=1}^{N} \frac{1}{\sum_{k=1}^{\infty} \pi_{B_k}} \begin{pmatrix} 0 & 0 & 0 & \dots & \pi_{B_1} & \pi_{B_2} & \pi_{B_3} & \dots \end{pmatrix}. \tag{3.35}$$

The accuracy of this approximation will be examined numerically for two specific queueing systems.

Example 1: geometric on-periods, mixed-geometric off-periods

Here the N on-off sources have geometrically distributed on-periods and mixed-geometrically distributed off-periods:

$$A(z) = \frac{(1-\alpha)z}{1-\alpha z}, \tag{3.36}$$

and

$$B(z) = q\frac{(1-\beta_1)z}{1-\beta_1 z} + (1-q)\frac{(1-\beta_2)z}{1-\beta_2 z}. \tag{3.37}$$

Three states are needed to model such a source as a DBMAP: one on-state state and two off-states, one corresponding to β_1, the other to β_2. One obtains

$$\mathbf{D}(z) = \begin{pmatrix} \alpha z & q(1-\alpha) & (1-q)(1-\alpha) \\ (1-\beta_1)z & \beta_1 & 0 \\ (1-\beta_2)z & 0 & \beta_2 \end{pmatrix} \tag{3.38}$$

Since only 3 states are involved, it is possible to compute $\mathbf{g}$, and hence also $\mathbf{x}_0$, by following the methods given in [49] or [51].

Although the set of parameters $(N, \alpha, \beta_1, \beta_2, q)$ fully characterises the arrival process, in [63] a more intuitive set of parameters is used, namely (N, ρ, K, L, q). The parameter ρ denotes the mean arrival rate and equals $N/(1+(1-\alpha)B'(1))$. The so-called burstiness factor K is defined as the ratio of the mean on-period of a single source to the corresponding quantity of a Bernoulli arrival process with the same load $\sigma = \rho/N$, i.e. $K = (1-\sigma)/(1-\alpha) = \sigma B'(1)$. Note that the Bernoulli process is simply a sequence of i.i.d. Bernoulli random variables. The parameter L is defined as the ratio of the variance of an off-period to the variance of a geometrically distributed off-period with the same mean length. As stated in [63, pg. 38], the parameters β_1 and β_2 can be calculated from q, L and $B'(1)$ in the following way:

$$\frac{1}{1-\beta_1} = B'(1) + \sqrt{\frac{(1-q)(L-1)B'(1)[B'(1)-1]}{2q}},$$
$$\beta_2 = 1 - \frac{(1-q)(1-\beta_1)}{B'(1)(1-\beta_1)-q}.$$

Furthermore α can be obtained from the values for ρ, N and K.

In Figure 3.1 two different measures for the accuracy of $\tilde{\mathbf{g}}$ are plotted against the load ρ, while the other parameters are kept constant, i.e. $N = 4$, $K = 2$, $L = 2$ and $q = 0.3$. The first measure is the so-called $\mathbf{g}$-error, which can be seen as an absolute measure for the error. Its value is given by

$$|\mathbf{g} - \tilde{\mathbf{g}}|_1 = \sum_i |\mathbf{g}_i - \tilde{\mathbf{g}}_i|.$$

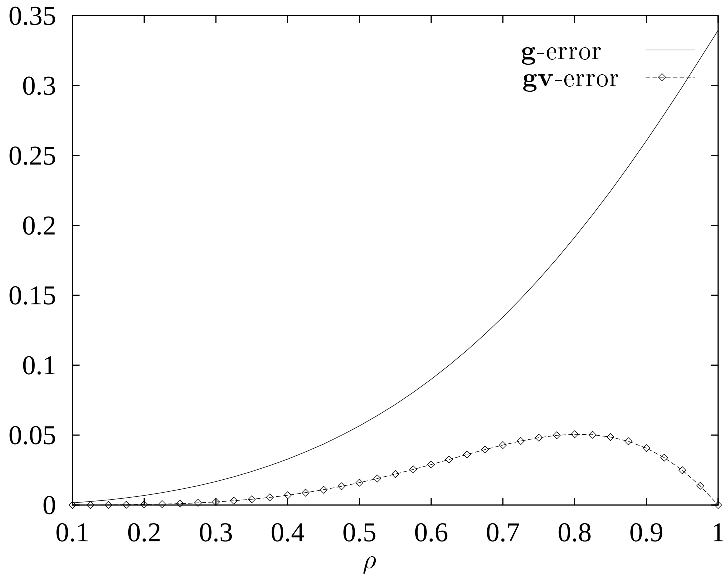

Figure 3.1: Error-plot example 1

The second error, the so-called **gv**-error, is related with the asymptotic behaviour. For its definition we recall the dominant pole approximation for the queueing system under study. Let $S(z) = \sum_{k=0}^{\infty} s_k z^k$ denote the generating function of the stationary buffer distribution. It is given by $S(z) = \mathbf{X}(z)\mathbf{e}$. Hence by applying Theorem 2.7.2 one obtains

$$s_n = \mathbf{x}_n\mathbf{e} \sim \frac{z_0 - 1}{\phi'(z_0) - 1}(1-\rho), \mathbf{g}\mathbf{v}^N(z_0) \tag{3.39}$$

with $\phi(z)$ the Perron-Frobenius eigenvalue of $\mathbf{D}^N(z)$, $\phi(z_0) = z_0$ and $\mathbf{v}^N(z_0)$ the normalised left eigenvector of $\mathbf{D}(z_0)$ corresponding to $\phi(z_0)$. From (3.39) it is clear that the value $|\mathbf{g}\mathbf{v}^N(z_0) - \tilde{\mathbf{g}}\mathbf{v}^N(z_0)|$, which is called the **gv**-error, is much more relevant here.

Figure 3.1 shows that the error $|\mathbf{g} - \tilde{\mathbf{g}}|_1$ increases with the load. The error $|\mathbf{g}\mathbf{v}(z_0) - \tilde{\mathbf{g}}\mathbf{v}(z_0)|$ however decreases again for a high enough load (note that z_0 is function of ρ). This phenomenon can be explained by noting that if $\rho \to 1$, then z_0 decreases towards 1, and hence $\mathbf{v}^N(z_0) \to \mathbf{e}$. So the **gv**-error vanishes because both $\tilde{\mathbf{g}}\mathbf{v}(z_0)$ and $\mathbf{g}\mathbf{v}(z_0)$ converge to 1 when $\rho \to 1$.

Example 2: Markovian on-off sources

The same two errors as in the previous example are plotted in Figure 3.2, now for the superposition of 5 Markovian on-off sources. The load is varied by varying the value for p. The other parameters are kept constant at $\alpha = 0.06$ and $\beta = 0.08$. Note that Figure 3.2 shows the same behaviour as Figure 3.1.

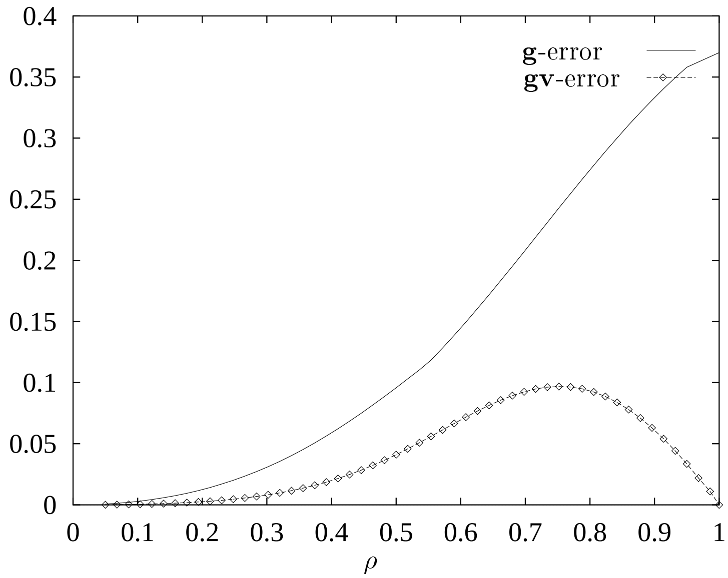

Figure 3.2: Error-plot example 2

Chapter 4

Tail Transitions

Not all DBMAP-D-1 queues have an asymptotic behaviour which fits within the framework of the dominant pole approximation. We will illustrate this by studying the asymptotic behaviour of two queueing systems to which the dominant pole approximation cannot be applied. These systems are neither pathological nor far-fetched.

The first system deals with two queues in tandem: the output process of a first queue serves, together with background traffic, as input for a second queue. The asymptotic behaviour of the tail probabilities of the second queue can be determined using Darboux's theorem. In some sense, we state and proof an extension of the dominant pole approximation as presented in Theorem 2.7.2. It turns out that the character of the asymptotic behaviour can essentially change when smoothly varying the parameters determining the system. These changes are called tail transitions.

The second system we study consists of a multiplexer queue with as input a mixture of LRD and SRD traffic. It will be shown that the tail probabilities can decay exponentially or according to a power-law, depending on the composition of the traffic mix.

The involved arrival processes belong to the same class of DBMAPs. This class is introduced below, together with an analysis of the related DBMAP-D-1 queue.

4.1 Preparation

The traffic stream we consider is the superposition of an on-off source and so-called background traffic. The on-off source has geometrically distributed off-periods and generally distributed on-periods. As such it belongs to the class of arrival processes introduced in Section 3.6. The generating function associated with the off-periods is given by

$$\frac{(1-\beta)z}{1-\beta z},$$

the generating function of the distribution of the length of the on-periods is denoted by $A(z) = \sum_{k=1}^{\infty} a_k z^k$. The background traffic generates X_n arrivals in slot n, with $\{X_n, n \geqslant 1\}$ a sequence of i.i.d. random variables. Their generating function is denoted by $\phi(z)$.

For the DBMAP description of the on-off source only one off-state is needed, because the off-periods are geometrically distributed. The on-states A_1, A_2, ... are defined as in Section 3.6. The off-state is taken to be the first state, hence $\mathbf{D}(z)$ is given by

$$\mathbf{D}(z) = \begin{pmatrix} \beta & (1-\beta)z & 0 & 0 & 0 & \dots \\ 1-p_1^a & 0 & p_1^a z & 0 & 0 & \dots \\ 1-p_2^a & 0 & 0 & p_2^a z & 0 & \dots \\ 1-p_3^a & 0 & 0 & 0 & p_3^a z & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

By the special structure of $\mathbf{D}(z)$ it is possible to obtain an implicit equation for its Perron-Frobenius eigenvalue $\lambda(z)$:

$$\lambda(z) = \beta + (1-\beta)A\Big(\frac{z}{\lambda(z)}\Big).$$

The DBMAP describing the superposition of the on-off source and the background traffic has $\phi(z)\mathbf{D}(z)$ as z-transform. Its mean arrival rate is given by

$$\rho = \phi'(1) + \frac{(1-\beta)A'(1)}{(1-\beta)A'(1)+1}.$$

This superposition will be used in Sections 4.2 and 4.3.2 as input for the multiplexer queue. Therefore we derive an explicit expression for the generating function

$$Q(z) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} q_k z^k,$$

with q_k is the steady-state probability of having k customers in this system. Clearly $Q(z) = \mathbf{X}(z)\mathbf{e}$, with $\mathbf{X}(z)$ given by

$$\mathbf{X}(z) = (z-1)\mathbf{x}_0\phi(z)\mathbf{D}(z)\big(z\mathbf{I} - \phi(z)\mathbf{D}(z)\big)^{-1},$$

which is the Pollachek-Kinchin equation for the DBMAP-D-1 queue having as input the DBMAP determined by $\phi(z)\mathbf{D}(z)$. Since there is only one off-state, the vector $\mathbf{x}_0$ is given by $\mathbf{x}_0 = \begin{pmatrix} 1-\rho & 0 & 0 & \ldots \end{pmatrix}$. The structure of $\mathbf{D}(z)$ makes it possible to compute the vector

$$\mathbf{w}(z) \stackrel{\text{def}}{=} \mathbf{x}_0 \phi(z)\mathbf{D}(z)\big(z\mathbf{I} - \phi(z)\mathbf{D}(z)\big)^{-1}, \tag{4.1}$$

with $\mathbf{w}(z) = \begin{pmatrix} w_0(z) & w_1(z) & \ldots \end{pmatrix}$. Some straightforward algebra leads to

$$w_0(z) = \frac{\beta\phi(z)z + (1-\beta)\phi(z)A\big(\phi(z)\big)z}{z^2 - \beta z\phi(z) - A\big(\phi(z)\big)(1-\beta)z\phi(z)},$$

$$w_1(z) = \frac{(1-\beta)z^2}{z^2 - \beta z\phi(z) - A\big(\phi(z)\big)(1-\beta)z\phi(z)},$$

and for $k \geqslant 2$,

$$w_k(z) = \left[1 - \sum_{i=1}^{k-1} a_i\right]\phi(z)^k w_1(z). \tag{4.2}$$

Since $Q(z) = \mathbf{X}(z)\mathbf{e} = (1-\rho)(z-1)\mathbf{w}(z)\mathbf{e}$, it follows that

$$Q(z) = (1-\rho)(z-1)\left[w_0(z) + w_1(z)\frac{1 - A\big(\phi(z)\big)}{1-\phi(z)}\right]. \tag{4.3}$$

4.2 Tandem Queues

Two queues are said to be in tandem if they are configured as in Figure 4.1. It is assumed that they both have an infinite waiting room and that their deterministic service time equals 1 slot. The number of cells arriving at the first queue in slot n is denoted by Y_n, where $\{Y_n, n \geqslant 1\}$ is a sequence of i.i.d. random variables. The generating function of the Y_n will be denoted by $\xi(z)$. The output stream of the first system, together with background traffic as defined in Section 4.1, is used as input for the second system. Let us demonstrate that this system is modelled by the DBMAP-D-1 queue introduced in Section 4.1.

Because the service time of the first queue equals one time slot, its output stream is an on-off source. The on-periods correspond to the busy-periods, the off-periods to the idle periods of the queue, which formally belongs to the class of GI-D-1 queueing systems, examined in e.g. [13]. It is shown there that the generating function of the distribution of the busy periods, which we denote by $A(z)$ in accordance with Section 4.1, satisfies the functional equation

$$A(z) = \frac{\xi(z[(1-\xi(0))A(z) + \xi(0)]) - \xi(0)}{1-\xi(0)}. \tag{4.4}$$

Since furthermore the idle period is geometrically distributed with parameter $\beta = \xi(0)$, it is clear that the theory developed in Section 4.1 applies to the second queue.

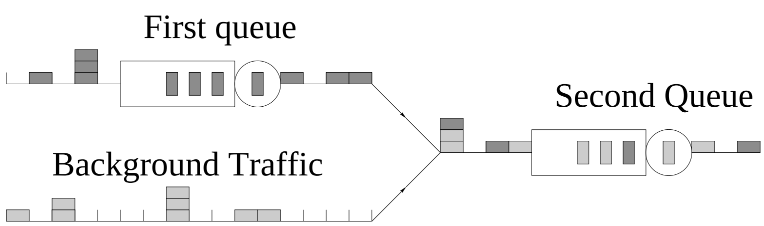

Figure 4.1: Tandem Queues

If one is able — given a concrete instance of ξ — to determine $A(z)$ from (4.4), then (4.3) can be used to describe the generating function associated with the second queue. It is of course not surprising that in general it is impossible to obtain $A(z)$ explicitly from (4.4). This is however not an obstacle to study the asymptotic behaviour of the second buffer as we will show in a Section 4.2.2. But first we examine in Section 4.2.1 one of these rare cases with an explicit formula for $A(z)$. This example, introduced and studied in [46], indicates which behaviour can be expected when dealing with arbitrary ξ and ϕ. For this general case a technique to obtain the asymptotic behaviour of the tail probabilities is presented in Section 4.2.2.

The reader should note that the approach presented here differs radically from [46]. Whereas in [46] the asymptotic behaviour is calculated by ad-hoc techniques, here Darboux's theorem is used.

4.2.1 The analytical case

As in [46] let

$$\xi(z) = \frac{1}{1+\kappa-\kappa z},$$

and

$$\phi(z) = 1 - \nu + \nu z,$$

with $0 < \kappa, \nu < 1$. Solving (4.4) leads to

$$A(z) = \frac{(1+\kappa)^2 - 2\kappa z - (1+\kappa)\sqrt{(1+\kappa)^2 - 4\kappa z}}{2\kappa^2 z}. \tag{4.5}$$

Furthermore $\beta = 1/(1+\kappa)$. The total arrival rate ρ at the second queue equals the sum of the rate at the first queue and the background rate, hence $\rho = \kappa + \nu$. For stability reasons it is assumed that $\kappa + \nu < 1$. It is also of interest to observe that the Perron-Frobenius eigenvalue $\lambda(z)$ of $\mathbf{D}(z)$ is exactly $\xi(z)$.

With the expressions for $A(z)$ and $\phi(z)$, (4.3) becomes

$$Q(z) = (1-\rho)\left[(1-z) + \frac{z\left(\sqrt{(1-\kappa)^2 + 4\kappa\nu(1-z)} - (1-\kappa-2\nu)\right)}{2\nu(1-\nu-\kappa z)}\right].$$

The denominator of this expression can be responsible for a pole at

$$z_p = \frac{1-\nu}{\kappa}, \tag{4.6}$$

if the numerator differs from 0 at this point. Because the square root $\sqrt{}$ always designates the positive root the numerator equals 0 if and only if $\kappa + 2\nu \leqslant 1$. Remark that the stability condition $\kappa + \nu < 1$ implies $z_p > 1$. The square root $\sqrt{}$ introduces a branch point at

$$z_b = \frac{(1-\kappa)^2 + 4\kappa\nu}{4\kappa\nu}. \tag{4.7}$$

For each admissible choice of κ and ν, $1 < z_p \leqslant z_b$ with equality only if $\kappa + 2\nu = 1$. But as indicated above, z_p is only a pole if $\kappa + 2\nu > 1$. Hence three cases can be distinguished:

1. $\kappa + 2\nu > 1$: $Q(z)$ has a simple dominant pole at z_p,
2. $\kappa + 2\nu = 1$: $Q(z)$ has the combination of a pole and a branch point of order $\frac{1}{2}$ at $z_p = z_b$,
3. $\kappa + 2\nu < 1$: $Q(z)$ has only a branch point of order $\frac{1}{2}$ at z_b.

We now examine these cases in detail. As in [46] we look at the asymptotics for the tail probabilities

$$\mathbf{P}\{q > k\} = \sum_{j>k} q_j, \tag{4.8}$$

where q is the random variable associated with $Q(z) = \sum_j q_j z^j$. These values are the coefficients of the series expansion of

$$Q^*(z) \stackrel{\text{def}}{=} \frac{1 - Q(z)}{1 - z}. \tag{4.9}$$

We apply Darboux's theorem to the function $Q^*(z)$ for the cases mentioned above.

Case 1: $\kappa + 2\nu > 1$

The part of $Q^*(z)$ determining the asymptotic behaviour of its coefficients can be written as

$$-\frac{(1-\kappa-\nu)z}{2\nu(1-\nu)(1-z)}\left(1 - \frac{z}{z_p}\right)^{-1}\left(\sqrt{(1-\kappa)^2 + 4\kappa\nu(1-z)} - (1-\kappa-2\nu)\right). \tag{4.10}$$

Hence, by Theorem 2.7.1, we are able to conclude

$$\mathbf{P}\{q > k\} \sim \frac{\kappa + 2\nu - 1}{\nu} z_p^{-k}.$$

The reader should note that this case is also covered by the dominant pole approximation. Observe that $\xi(z_p)\phi(z_p) = z_p$, hence z_p plays the role of τ of Theorem 2.7.2.

Case 2: $\kappa + 2\nu = 1$

The part of the expression for $Q^*(z)$ introducing the singularity with weight $\frac{1}{2}$ at $z_p = z_b$ is

$$-\frac{1-\kappa-\nu}{1-z}z\frac{\sqrt{(1-\kappa)^2+4\kappa\nu(1-z)}}{2\nu(1-\nu-\kappa z)},$$

which can be rewritten as

$$-\frac{1-\kappa-\nu}{1-z}z\left(1-\frac{z}{z_p}\right)^{-\frac{1}{2}}\frac{\sqrt{(1-\kappa)^2+4\kappa\nu}}{2\nu(1-\nu)}.$$

Hence Theorem 2.7.1 implies

$$\mathbf{P}\{q > k\} \sim \frac{1}{\sqrt{\pi}}\sqrt{\frac{1+\kappa}{1-\kappa}}k^{-\frac{1}{2}}z_b^{-k}.$$

Note that the decay is not purely exponential, but that also a factor $k^{-\frac{1}{2}}$ is involved.

Case 3: $\kappa + 2\nu < 1$

The part of $Q^*(z)$ introducing the singularity with weight $-\frac{1}{2}$ at z_b is

$$-\frac{1-\kappa-\nu}{1-z}z\frac{\sqrt{(1-\kappa)^2+4\kappa\nu(1-z)}}{2\nu(1-\nu-\kappa z)} = -\frac{1-\kappa-\nu}{1-z}\frac{\sqrt{(1-\kappa)^2+4\kappa\nu}}{2\nu(1-\nu-\kappa z)}\left(1-\frac{z}{z_p}\right)^{\frac{1}{2}}.$$

This results in

$$\mathbf{P}\{q > k\} \sim \frac{1}{\sqrt{\pi}}\frac{(1-\kappa-\nu)((1-\kappa)^2+4\kappa\nu)}{(1-\kappa)^2(1-\kappa-2\nu)^2}\sqrt{(1-\kappa)^2+4\kappa\nu}\,k^{-\frac{3}{2}}z_b^{-k}. \tag{4.11}$$

Again two factors contribute to the decay: $k^{-\frac{3}{2}}$ and z_b^{-k}.

4.2.2 The general case

In this section we allow arbitrary distributions for the Y_n, which determine the arrivals at the first queue, and for the X_n, which generate the background traffic. Hence no conditions are imposed on the generating functions ξ and ϕ, with the exception that ϕ is taken such that it does not introduce branchpoints. Results similar to the ones presented above will be obtained. We start again with looking for poles and branchpoints.

Considering (4.3) it is clear that the common denominator of $w_1(z)$ and $w_2(z)$,

$$d(z) = z^2 - \beta z\phi(z) - A\big(\phi(z)\big)(1-\beta)z\phi(z),$$

can introduce singularities whenever it becomes 0. Observe that the zero point of $d(z)$, which lies outside the complex unit disk and has minimal modulus, is on the real axis. We denote this point by z_p. Branch points can be introduced by $A(z)$. Because $A(z)$ is a generating function, its branchpoints lie outside the complex unit disk. Since $A(z)$ satisfies the implicit equation $F\big(z, A(z)\big) = 0$, with

$$F(z,w) = \frac{\xi(z[(1-\xi(0))w + \xi(0)]) - \xi(0)}{1-\xi(0)} - w,$$

we can conclude from [8] and [45, Chapter 3] that in general $A(z)$ has an algebraic branchpoint of weight $\frac{1}{2}$ at the point z_b if

$$\begin{aligned}
\frac{\partial}{\partial w} F(z,w)|_{z=\phi(z_b), w=A(\phi(z_b))} &= 0,\\
\frac{\partial}{\partial z} F(z,w)|_{z=\phi(z_b), w=A(\phi(z_b))} &\neq 0,\\
\frac{\partial^2}{\partial w^2} F(z,w)|_{z=\phi(z_b), w=A(\phi(z_b))} &\neq 0.
\end{aligned}$$

The expression for $F(z,w)$ implies that the branchpoint of $A(z)$ with minimal modulus is real. From now on z_b will designate this special branchpoint.

We distinguish again between the following three cases: $z_p < z_b$, $z_p = z_b$ and $z_b < z_p$.

Case 1: $z_p < z_b$

This case belongs to the domain of the dominant pole approximation and is given for the sake of completeness. By applying Darboux's theorem we simply have

$$\mathbf{P}\{q > k\} \sim g z_p^{-k},$$

with

$$g = \lim_{z\to z_p} \left(1 - \frac{z}{z_p}\right) Q^*(z).$$

This limit can be determined directly by applying l'Hôpital's rule. The only difficulty is determination of the value $A'(\phi(z_b))$. By the implicit function theorem one has

$$A'(\phi(z_p)) = -\frac{\frac{\partial}{\partial z} F(z,w)|_{z=\phi(z_p), w=A(\phi(z_p))}}{\frac{\partial}{\partial w} F(z,w)|_{z=\phi(z_p), w=A(\phi(z_p))}}.$$

Case 2: $z_p = z_b$

We use again the fact that $A(z)$ is implicitly determined by F. In [8] a theorem linking implicit functions and Darboux's theorem is presented. Although this theorem cannot be applied here in a straightforward manner, it is still appropriate to quote it in the original form [8, Theorem 5].

Theorem 4.2.1 (Bender). *Assume that the power series $w(z) = \sum a_n z^n$ with non-negative coefficients satisfies $F(z,w) = 0$. Suppose there exist real numbers $r > 0$ and $s > a_0$ such that*

1. *for some $\delta > 0$, $F(z,w)$ is analytic whenever $|z| < r + \delta$ and $|w| < s + \delta$;*
2. *$F(r,s) = 0$ and $F_w(r,s) = 0$*
3. *$F_z(r,s) \neq 0$ and $F_{ww}(r,s) \neq 0$ and*
4. *if $|z| \leqslant r$, $|w| \leqslant s$, and $F(z,w) = F_w(z,w) = 0$ then $z = r$ and $w = s$.*

Then

$$a_n \sim \sqrt{\frac{rF_z}{2\pi F_{ww}}} n^{-\frac{3}{2}} r^{-n}, \tag{4.12}$$

where the partial derivatives F_z, F_w and F_{ww} are evaluated at $z = r$ and $w = s$.

The key point in the proof of this theorem is that — as far as the statement (4.12) is concerned — $w(z)$ can be approximated, in a neighbourhood of $z = r$ and $w = s$, by the corresponding solution $w^\star(z)$ of the equation

$$F(r,s) + (w-s)F_w + (z-r)F_z + \frac{1}{2}(w-s)^2 F_{ww} = 0. \tag{4.13}$$

Since $F(r,s) = F_w = 0$ one has

$$w^\star(z) \approxeq s \pm \sqrt{\frac{2(r-z)Fz}{F_{ww}}}. \tag{4.14}$$

Formula (4.12) follows by applying Darboux's theorem to the positive branch of $w^\star(z)$.

To obtain the asymptotic behaviour of the tail probabilities the approximation presented in (4.14) will be applied to $A(z)$. Let $r = \phi(z_b)$ and $s = A(\phi(z_b))$. Hence, analogous to (4.14),

$$A(\phi(z)) \approxeq A(\phi(z_b)) \pm \sqrt{\frac{2\phi'(z_b)F_z}{F_{ww}}(z_b - z)}.$$

Let furthermore $a = \sqrt{(2\phi'(z_b)F_z)/F_{ww}}$ and $f = (1-\beta)\phi(z_b)z_b$. Using the approximation for $A(z)$, which gives — as shown by the proof of Bender's

Theorem in [8] — the analytically correct results, and by proceeding as in Section 4.2.1, we end up with

$$\mathbf{P}\{q > k\} \sim \frac{c_2}{\sqrt{\pi}} k^{-\frac{1}{2}} z_b^{-k},$$

where

$$c_2 = \frac{1-\rho}{fa\sqrt{z_b}} \left(z_b^2 - d(z_b) + (1-\beta) z_b^2 \left[1 - A\big(\phi(z_b)\big)\right] \frac{\phi(z_b)}{1-\phi(z_b)} \right).$$

Case 3: $z_b < z_p$

The derivation of the tail probabilities is analogous to the case $z_b = z_p$. One obtains

$$\mathbf{P}\{q > k\} \sim \frac{c_3}{2\sqrt{\pi}} k^{-\frac{3}{2}} z_b^{-k},$$

with

$$c_3 = \frac{-(1-\rho)\sqrt{z_b}}{d(z_b)^2} \left(z_b^2 fa + ((1-\beta) z_b^2 \big(\left[1 - A\big(\phi(z_b)\big)\right] f - d(z_b)\big) a) \frac{\phi(r)}{1-\phi(r)} \right).$$

Numerical example

The theory developed in Section 4.2.2 is applied to the case with $\xi(z) = e^{\lambda(z-1)}$ and $\phi(z) = 1 - \nu + \nu z1$. Hence cells arrive at the first queue according to a Poisson distribution with parameter λ. The background traffic is Bernoulli with parameter ν. The following cases are examined:

1. $\lambda = .7$, $\nu = .20$ resulting in $z_p < z_b$,
2. $\lambda = .7$, $\nu = .13$ resulting in $z_p \approx z_b$,
3. $\lambda = .7$, $\nu = .10$ resulting in $z_b < z_b$.

For each of them the value $\log \mathbf{P}\{q = k\}$ obtained by simulation is plotted, together with the values obtained by the appropriate formula for the asymptotic behaviour. Note that by the definition of $Q^*(z)$, the asymptotic formula for $\mathbf{P}\{q = k\}$ can be derived from the one for $\mathbf{P}\{q > k\}$, by dividing it by $z_p - 1$ or $z_b - 1$.

As can be seen from Figure 4.2, the asymptotic regime sets in almost immediately for the cases 1 and 2. This is clearly not true for the third case. What happens there can be understood by taking a look at the asymptotic formulas obtained for the analytical case in Section 4.2.1. One observes in the denominator of the constant in (4.11) the factor $(1 - \kappa - 2\nu)^2$. If $k + 2\nu$ approaches

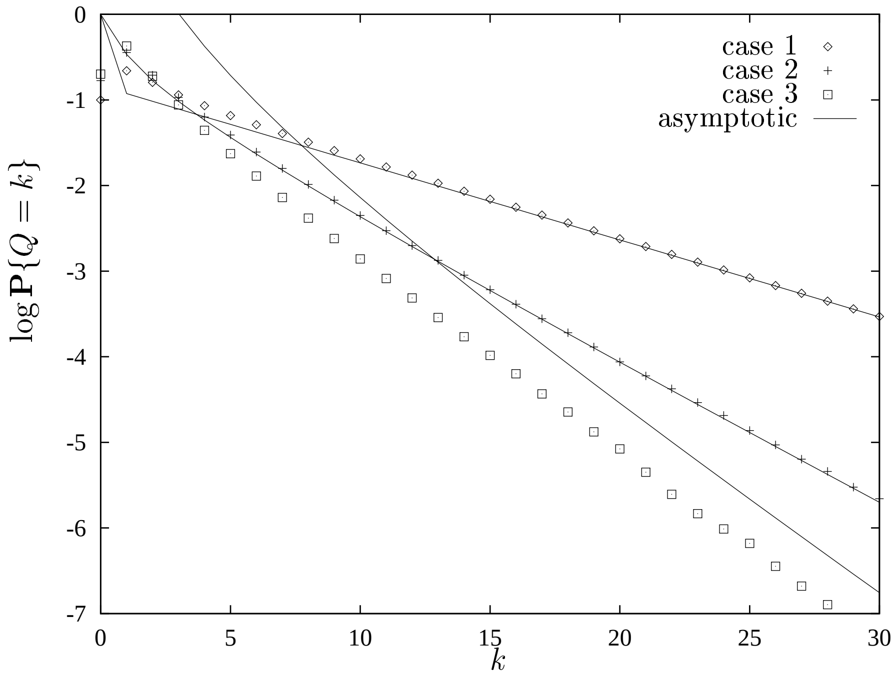

Figure 4.2: Numerical Example

1, this constant becomes very large. This is of course related to the fact that the system is approaching the transition point, where the asymptotic behaviour changes from $c_3 k^{-3/2} {z_b}^{-k}$ to $c_2 k^{-1/2} {z_b}^{-k}$. Since $k^{-3/2}$ decays faster than $k^{-1/2}$, the constant c_3 tries to compensate for this by diverging when $k+2\nu$ approaches 1. Consequently it takes high values for k to reach the asymptotic regime.

4.3 An LRD-SRD Traffic Mix

Tail transitions also happen when dealing with traffic mixes consisting of SRD and LRD components. It turns out that if such traffic is fed to the c-server queue introduced in Section 3.3.1, two radically different types of asymptotic behaviour may occur. Depending on the value of c and on the sizes of the SRD and the LRD components, the tail probabilities decay exponentially or according to a power-law.

This phenomenon is studied in [24] for a fluid-flow queue with input a mix of N stationary LRD and SRD on-off sources. The study presented here is limited to a c-server queue having as input the superposition of a single LRD on-off source and SRD background traffic. Despite being much more specific, this queueing system has all the relevant characteristics, and at the same time it is analytically tractable.

For the case $c = 1$, exact buffer asymptotics are obtained. When $c \geqslant 2$, we are still able to precisely describe the asymptotic behaviour of the tail probabilities, but the intrinsic complexity of the system prevents us from deriving closed form expressions. Nevertheless we obtain the point at which the tail transition occurs. Furthermore, the use of generating functions allows us to gather strong support for the conjecture made at the end of [24, Section 4.1].

4.3.1 The traffic mix: definitions and properties

The arrival process defined in Section 4.1 is reused. Hence we consider the superposition of background traffic and an on-off source with geometrically distributed off-periods. Let τ_A denote the random variable corresponding to the duration of the on-periods. By assuming a Pareto-like distribution for τ_A, i.e.

$$\mathbf{P}\{\tau_A = k\} \sim ak^{-s}, \tag{4.15}$$

with $2 < s < 3$ and $a > 0$, the on-off source generates LRD traffic. Indeed, if X_k denotes the number of arrivals generated by the on-off source in slot k, then the results presented in [57] imply that

$$\mathrm{Var}(X_1 + \ldots + X_n) \sim \frac{\mathbf{E}[\tau_B]^2}{(\mathbf{E}[\tau_A] + \mathbf{E}[\tau_B])^3} \frac{a}{(4-s)(3-s)(s-1)} n^{4-s}.$$

Hence by Proposition 1.3.10, the source generates LRD traffic with Hurst parameter $H = (4-s)/2$. The background traffic serves as the SRD component. Recall that its generating function is denoted by $\phi(z)$.

In Section 4.3.2 we consider the case where the total traffic stream is fed to the c-server queue with $c = 1$, while in Section 4.3.3 the traffic is offered to the c-server queue with $c > 1$.

4.3.2 The tail probabilities when $c = 1$: exact asymptotics

Since there is only 1 server we are actually dealing with the DBMAP-D-1 queue presented in Section 4.1. Hence the generating function associated with the stationary buffer distribution q is $Q(z)$, as defined by (4.3). Considering the equality

$$\mathbf{E}[q] = \sum_{k=0}^{\infty} \mathbf{P}\{q > k\},$$

it is clear that some information about the tail probabilities is contained in the mean queue length. By calculating $Q'(1)$ we obtain

$$\mathbf{E}[q] = \frac{A''(1)}{2} \frac{1}{\mathbf{E}[\tau_A] + \mathbf{E}[\tau_B]} \left(\frac{\phi'(1)^2}{1-\rho} + \phi'(1) \right) + \Omega,$$

with Ω denoting a term involving the finite quantities β, $\phi'(1)$, $\phi''(1)$ and $A'(1)$ (we suppose that all the moments of the background traffic exist). Since $\mathbf{E}[\tau_A^2] = \infty$, we have $A''(1) = \infty$, hence $\mathbf{E}[q] = \infty$. Consequently $Q'(z)$ diverges when z approaches 1. In Appendix 4.3.4 we show that the behaviour of $Q'(z)$ near 1 is described by

$$Q'(z) \sim \frac{L}{(1-z)^{3-s}} \text{ for } z \to 1-, \tag{4.16}$$

with

$$L = \frac{a\Gamma(3-s)}{s-1} \frac{1}{\mathbf{E}[\tau_A] + \mathbf{E}[\tau_B]} \left(\frac{\phi'(1)^{s-1}}{1-\rho} + \phi'(1)^{s-2} \right).$$

Note that $f(z) \sim g(z)$ for $z \to 1-$, with f, g real-valued functions, stands for $\lim_{z\to 1, z<1} f(z)/g(z) = 1$.

By using the Tauberian theorem for power series we are able to derive from (4.16) the asymptotic behaviour of the coefficients of the power series $Q'(z) = \sum_k k q_k z^{k-1}$. We recall this theorem from [26],

Theorem 4.3.1 (Tauberian theorem for power series). *Let $p_k \geqslant 0$ and suppose that*

$$P(z) = \sum_{k=0}^{\infty} p_k z^k$$

converges for $0 \leqslant z < 1$. If L is a constant and $0 \leqslant \sigma < \infty$, then the following relations are equivalent:

$$P(z) \sim \frac{L}{(1-z)^{\sigma}} \textit{ for } z \to 1-$$

and

$$p_0 + p_1 + \ldots + p_{n-1} \sim \frac{L}{\Gamma(\sigma+1)} n^{\sigma} \textit{ for } n \to \infty.$$

Applying this theorem to $Q'(z)$ results in

$$q_1 + 2q_2 + \ldots n q_n \sim \frac{L}{\Gamma(4-s)} n^{3-s}. \tag{4.17}$$

Using [43, 3.3 (c), pg. 59], the asymptotic behaviour of the tail probabilities can be derived from (4.17):

$$\sum_{j>k} q_j \sim \frac{L}{\Gamma(4-s)} \frac{3-s}{s-2} k^{2-s},$$

or

$$\mathbf{P}\{q > k\} \sim \frac{a}{(s-1)(s-2)} \frac{1}{\mathbf{E}[\tau_A] + \mathbf{E}[\tau_B]} \left(\frac{\phi'(1)^{s-1}}{1-\rho} + \phi'(1)^{s-2} \right) k^{2-s}. \tag{4.18}$$

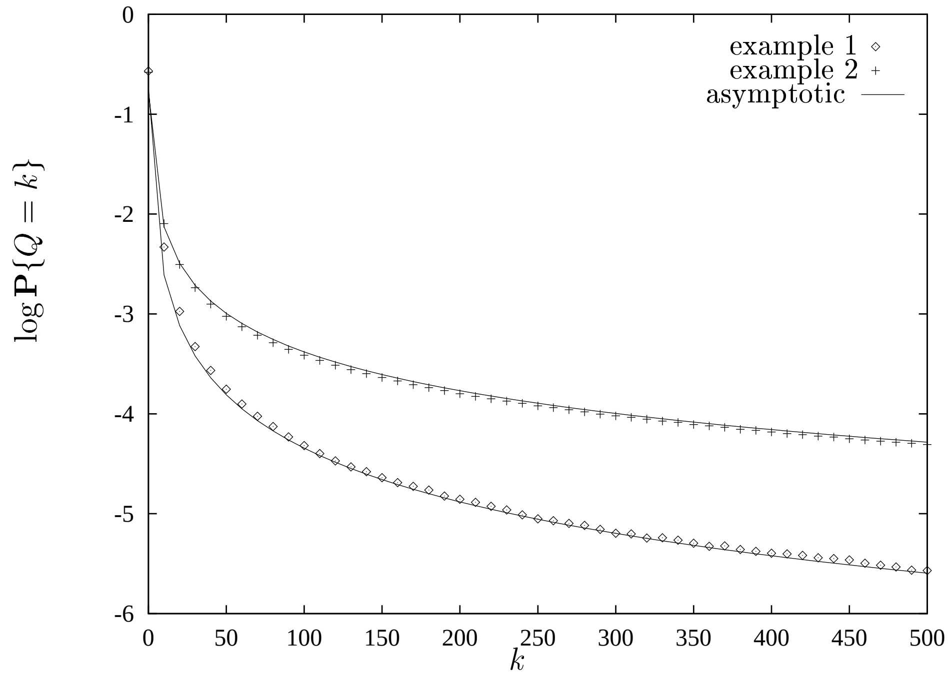

Figure 4.3: Numerical example with $c = 1$

Numerical example To observe how fast the asymptotic regime is reached we simulate the queue with

$$\mathbf{P}\{\tau_A = j\} = (j+1)^{-s} - j^{-s}, \tag{4.19}$$

and $\phi(z) = 1 - \nu + \nu z$. Here $a = s - 1$. For the first example $s = 2.8$, $\beta = 0.6$ and $\nu = 0.3$, and for the second example $s = 2.3$, $\beta = 0.8$ and $\nu = 0.3$. Hence the load of the system is about 0.68 for the first case and about 0.65 for the second case. As can be seen from Figure 4.3.2, the asymptotic regime is reached very quickly. The asymptotic behaviour of $\mathbf{P}\{q = k\}$ is given by

$$\mathbf{P}\{q = k\} \sim \frac{a}{(s-1)} \frac{1}{\mathbf{E}[\tau_A] + \mathbf{E}[\tau_B]} \left(\frac{\phi'(1)^{s-1}}{1-\rho} + \phi'(1)^{s-2} \right) k^{1-s}.$$

Note however that this formula is only valid because the $\mathbf{P}\{q = k\}$ are decreasing.

4.3.3 The tail probabilities when $c > 1$: a qualitative analysis

Contrary to the case $c = 1$, no closed form formula describing the asymptotic behaviour can be obtained when $c > 1$. The main reason is that an explicit formula for $Q(z)$ like (4.18) does not exist for the DBMAP-D-c with $c > 1$.

Nevertheless it is possible to determine precisely the behaviour of the tail probabilities. It turns out that this behaviour depends on c and on the sizes of the LRD and the SRD shares of the traffic mix.

We start with deriving from the Pollachek-Kinchin equation for the DBMAP-D-c queue an expression for $Q(z)$, the generating function of the stationary buffer distribution q.

Calculating $Q(z)$ when $c > 1$

Recall that the Pollachek-Kinchin equation for the DBMAP-D-c queue (see Section 3.4.1) is given by:

$$\mathbf{X}(z) = \sum_{l=0}^{c-1} \mathbf{x}_l \mathbf{D}(z)(z^c - z^l)\big(z^c\mathbf{I} - \mathbf{D}(z)\big)^{-1}.$$

Whereas for $c = 1$ the vector $\mathbf{x}_0$ is known explicitly, here the vectors $\mathbf{x}_0, \ldots, \mathbf{x}_{c-1}$ can only be computed numerically, as is shown in [49, pg. 310–329]. This prevents us from deriving a closed form formula for $Q(z) = \mathbf{X}(z)\mathbf{e}$. Nevertheless we proceed as in Section 4.1. Define for $l = 0, 1, \ldots, c-1$,

$$\mathbf{w}^{(l)}(z) = \mathbf{x}_l \mathbf{D}(z)\big(z^2\mathbf{I} - \mathbf{D}(z)\big)^{-1},$$

with $\mathbf{w}^{(l)}(z) = \begin{pmatrix} w_0^{(l)} & w_1^{(l)} & \ldots \end{pmatrix}$. From this definition one can derive the equations

$$\begin{aligned}\big(z^c - \phi(z)\beta\big)w_0^{(l)}(z) &- z^{c-1}A\Big(\frac{\phi(z)}{z^{c-1}}\Big)w_1^{(l)}(z) \\ &= \phi(z)\sum_{j=1}^{\infty}(\mathbf{x}_l)_j\left[\sum_{j=k+1}^{\infty}\frac{a_k}{1-\sum_{i=1}^{j-1}a_i}\Big(\frac{\phi(z)}{z^{c-1}}\Big)^{k-j}\right] \\ &\quad + \phi(z)\beta(\mathbf{x}_l)_0 + \phi(z)\sum_{k=1}^{\infty}\big(1-p_k^a\big)(\mathbf{x}_l)_k,\end{aligned} \tag{4.20}$$

and

$$-(1-\beta)z\phi(z)w_0^{(l)}(z) + z^c w_1^{(l)}(z) = (1-\beta)z(\mathbf{x}_l)_0, \tag{4.21}$$

which determine $w_0^{(l)}(z)$ and $w_1^{(l)}(z)$. The $w_k^{(l)}(z)$, with $k \geqslant 2$, are given by

$$w_k^{(l)}(z) = \Big(\prod_{i=1}^{k-1}p_i^a\Big)\Big(\frac{\phi(z)}{z^{c-1}}\Big)^{k-1}w_1^{(l)}(z) + \sum_{j=1}^{k-1}\mathbf{x}_{i,j}\Big(\prod_{i=j}^{k-1}p_i^a\Big)\Big(\frac{\phi(z)}{z}\Big)^{k-j}.$$

Summarising this information we have

$$\begin{aligned}\mathbf{w}^{(l)}(z)\mathbf{e} = w_0^{(l)}(z) &+ w_1^{(l)}(z)\left[1 + \sum_{k=1}^{\infty}\Big(\sum_{j>k}a_j\Big)\Big(\frac{\phi(z)}{z^{c-1}}\Big)^k\right] \\ &+ \sum_{j=1}^{\infty}(\mathbf{x}_l)_j\sum_{k=j}^{\infty}\frac{1-\sum_{i=1}^{k}a_i}{1-\sum_{i=1}^{j-1}a_i}\Big(\frac{\phi(z)}{z^{c-1}}\Big)^{k-j+1}.\end{aligned} \tag{4.22}$$

The analysis of

$$Q(z) = \sum_{l=0}^{c-1} (z^c - z^l) \mathbf{w}^{(l)}(z) \mathbf{e}, \tag{4.23}$$

reveals two different types of asymptotic behaviour. If $\phi'(1) > c-1$, or otherwise stated, if the mean arrival rate of the background traffic exceeds $c-1$, then the tail probabilities decay according to a power-law, as is shown under Case 1. If the mean arrival rate of the background traffic is strictly less than $c-1$, then the tail probabilities decay approximately exponentially, as is demonstrated under Case 2. It turns out that the rate of this decay can be easily determined. For the case $\phi'(1) = c-1$, the transition point, no conclusions could be drawn.

Case 1: $\phi'(1) > c-1$

Since $\phi'(1) > c-1$ we have

$$\frac{d}{dz}\Big(\frac{\phi(z)}{z^{c-1}}\Big)\Big|_{z=1} > 0.$$

Hence there exists an open set $G \subset D(0,1)$, with $D(0,1)$ the closed complex unit disk, having the following properties:

$$\text{for each } z \in G\text{: } \Big|\frac{\phi(z)}{z^{c-1}}\Big| < 1; \tag{4.24}$$

$$\text{the interval } [\zeta, 1) \text{ belongs to } G \text{ for some } \zeta \in (0,1). \tag{4.25}$$

By (4.24) all series occurring in (4.20) and (4.22) are summable for $z \in G$. Hence on G the representation (4.23) of $Q(z)$ can be used. Since we focus on the influence of the LRD on-off source, $\phi(z)$ is taken to be analytical on some open set containing $D(0,1)$.

We obtain the asymptotic behaviour of the tail probabilities, as in Section 4.3.2, by examining the behaviour of $Q'(z)$. Therefore we list below the expressions which occur in (4.23) and involve the coefficients a_k of $A(z)$.

Note in the formulas for $w_0^{(l)}(z)$ and $w_1^{(l)}(z)$, to be derived from (4.20) and (4.21), the occurrence of

$$A\Big(\frac{\phi(z)}{z^{c-1}}\Big), \tag{4.26}$$

and

$$\sum_{j=1}^{\infty} (\mathbf{x}_l)_j \Bigg[\sum_{j=k+1}^{\infty} \frac{a_k}{1 - \sum_{i=1}^{j-1} a_i} \Big(\frac{\phi(z)}{z^{c-1}}\Big)^{k-j} \Bigg]. \tag{4.27}$$

Furthermore one notices in (4.22)

$$\sum_{k=1}^{\infty} \Big(\sum_{j>k} a_j\Big) \Big(\frac{\phi(z)}{z^{c-1}}\Big)^{k}, \tag{4.28}$$

and

$$\sum_{j=1}^{\infty} (\mathbf{x}_0)_j \sum_{k=j}^{\infty} \frac{1 - \sum_{i=1}^{k} a_i}{1 - \sum_{i=1}^{j-1} a_i} \left(\frac{\phi(z)}{z^{c-1}} \right)^{k-j+1}. \tag{4.29}$$

Let us analyse the behaviour of $Q'(z)$ for $z \to 1-$. Consider the term $(z^c - z^l)\mathbf{w}^{(1)}(z)\mathbf{e}$ of (4.23). Carefully examining the occurrences of (4.26) and (4.27) reveals that, as far as these expressions are concerned, the situation is essentially the same as with $Q'(z)$ in Section 4.3.2. The contributions of (4.28) and (4.29) can be coped with as follows. We restrict the outline to (4.29), as (4.28) can be dealt with in a similar way. In $Q'(z)$ the terms

$$f_l(z) \stackrel{\text{def}}{=} \frac{d}{dz} \left[\left(z^c - z^l \right) \sum_{j=1}^{\infty} (\mathbf{x}_l)_j \sum_{k=j}^{\infty} \frac{1 - \sum_{i=1}^{k} a_i}{1 - \sum_{i=1}^{j-1} a_i} \left(\frac{\phi(z)}{z^{c-1}} \right)^{k-j+1} \right],$$

show up for $l = 0, \dots, c-1$. Considering the proof of the statement (4.16), it is clear one has to determine the limits

$$\lim_{z \to 1-} (1-z)^{3-s} f_l(z),$$

which can be done by observing that

$$\lim_{z \to 1-} (1-z)^{3-s} \sum_{j=1}^{\infty} (\mathbf{x}_l)_j \sum_{k=j}^{\infty} \frac{1 - \sum_{i=1}^{k} a_i}{1 - \sum_{i=1}^{j-1} a_i} \left(\frac{\phi(z)}{z^{c-1}} \right)^{k-j+1} =$$
$$\sum_{j=1}^{\infty} (\mathbf{x}_l)_j \lim_{z \to 1-} (1-z)^{3-s} \sum_{k=j}^{\infty} \frac{1 - \sum_{i=1}^{k} a_i}{1 - \sum_{i=1}^{j-1} a_i} \left(\frac{\phi(z)}{z^{c-1}} \right)^{k-j+1}. \tag{4.30}$$

This equality is proved as follows. Note that $(\mathbf{x}_l)_j \leqslant \boldsymbol{\pi}_j$, with

$$\boldsymbol{\pi}_j = \frac{(1-\beta)\mathbf{P}\{\tau_A \geqslant j\}}{1 + (1-\beta)A'(1)}, \tag{4.31}$$

which is the steady-state probability that the on-off source is in state A_j. Together with the fact that

$$\lim_{z \to 1-} (1-z)^{3-s} \frac{(1-\beta)}{1 + (1-\beta)A'(1)} \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \left(\sum_{i=k+1}^{\infty} a_i \right) \left(\frac{\phi(z)}{z^{c-1}} \right)^{k-j+1} = M,$$

where $M > 0$, (4.31) implies that (4.30) is valid. Hence for some $C > 0$,

$$Q'(z) \sim \frac{C}{(1-z)^{3-s}} \text{ for } z \to 1-,$$

which implies

$$\mathbf{P}\{q > k\} \sim \frac{C(3-s)}{(s-2)\Gamma(4-s)} k^{2-s},$$

The constant C depends on $A(z)$, β, $\phi(z)$, c and the vectors $\mathbf{x}_0, \dots \mathbf{x}_{c-1}$. It should be kept in mind that these vectors $\mathbf{x}_0, \dots, \mathbf{x}_{c-1}$ can be calculated only numerically.

Case 2: $\phi'(1) < c - 1$

Contrary to the case $\phi'(1) > c-1$ there exists some $z_0 > 1$ such that $\phi(z_0) = z_0$. We suppose that $z_0 < \infty$, because $z_0 = \infty$ implies that the background traffic generates at most $c - 1$ arrivals in a slot. If this is the case no buffer is needed, because all arriving cells can be served instantaneously. For $z \in (1, z_0)$,

$$\frac{\phi(z)}{z^{c-1}} < 1.$$

Since $\phi(z)$ is a generating function, it follows that for the complex z with $1 < |z| < |z_0|$, one has $|\phi(z)| < |z^{c-1}|$. Furthermore, because $\phi(z) \neq z^n$ for each $n \geqslant 0$, there exist at most a finite number of complex points $z_i^\star$ with modulus 1 such that $|\phi(z_i^\star)| = |z_i^\star|$. Let G be the maximal open set such that for $z \in G$,

$$\left|\frac{\phi(z)}{z^{c-1}}\right| < 1.$$

Hence all complex z, with $1 < |z| < |z_0|$, belong to G. Furthermore $G \cap D(0,1)$ is non-empty. Notice that on G the function $Q(z)$, given by (4.23), is well-defined. Recall that q denotes the random variable associated with the stationary buffer distribution. Because $G \cap D(0,1)$ is non-empty, $Q(z)$ is a representation of the z-transform $\mathbf{E}[z^q]$ on G. Hence the power series $\sum_{k=0}^{\infty} q_k z^k$, associated with $\mathbf{E}[z^q]$, is holomorphic on the open disk with centre 0 and radius $|z_0|$. The point z_0 represents a non-removable singularity. Since on G,

$$\mathbf{E}[z^q] = \sum_{k=0}^{\infty} q_k z^k = Q(z),$$

it is possible to determine the asymptotic behaviour of $q_k = \mathbf{P}\{q = k\}$ by examining the behaviour of $Q'(z)$. Define for $y \in (0,1)$,

$$f(y) = Q'(z_0 y).$$

Analysing the behaviour of $f(y)$ for $y \to 1-$, by using similar arguments as for the $\phi'(1) > c - 1$ case, we obtain

$$q_1 + 2z_0 q_2 + \ldots + n z_0^{n-1} q_n \sim C n^{3-s},$$

with $C > 0$. If the sequence $(k z_0^{k-1} q_k)$ decreases, this implies

$$\mathbf{P}\{q = k\} \sim (3 - s) C z_0^{-k} k^{1-s}.$$

In [24] this type of asymptotic behaviour is conjectured for similar queueing systems.

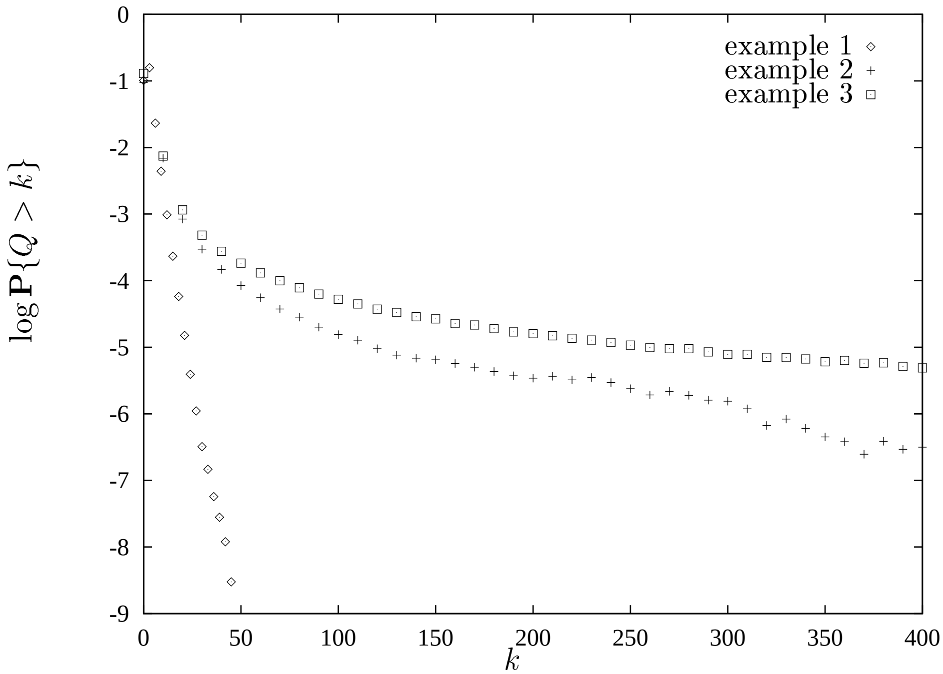

Figure 4.4: Numerical example with $c = 2$

Numerical example

The tail probabilities for three examples, each one corresponding to a case considered above, are shown in Figure 4.4. The on-periods are distributed as in (4.19), the background traffic is generated by Poisson variables with parameter λ. For the first example $s = 2.3$, $\beta = 0.4$ and $\lambda = 0.8$, which results in a mean arrival rate of 1.5. Clearly this example is covered by Case 2. The second example has $s = 2.6$, $\beta = 0.4$ and $\lambda = 1$. Hence the mean arrival rate is 1.6. Since $\lambda = 1$, this example is a transition point case. The third one has a power-law decay, here $s = 2.6$, $\beta = 0.8$ and $\lambda = 1.2$, hence mean arrival rate is 1.5.

4.3.4 Appendix

To validate the statement (4.16) we rewrite $Q(z)$:

$$Q(z) = (1-\rho)\big(f_0(z) + f_1(z)\big),$$

with

$$f_0(z) = (z-1)\frac{t(z)}{z^2 - t(z)}, \tag{4.32}$$

$$f_1(z) = (z-1)\frac{(1-\beta)z^2}{z^2-t(z)}\frac{1-A\big(\phi(z)\big)}{1-\phi(z)}, \tag{4.33}$$

$$t(z) = \beta z\phi(z) + (1-\beta)z\phi(z)A\big(\phi(z)\big). \tag{4.34}$$

Let us focus on the behaviour of $f_0'(z)$. The function $f_1'(z)$ can be analysed in a similar way. Combining the results for $f_0'(z)$ and $f_1'(z)$ we obtain (4.16). For $f_0'(z)$ we have

$$f_0'(z) = -\frac{\big(t(z)-z\big)^2}{\big(z^2-t(z)\big)^2} + z^2\frac{(z-1)^2\sum_{k=0}^{\infty}(k+1)T_{k+2}z^k}{\big(z^2-t(z)\big)^2},$$

with $T_k = \sum_{j\geqslant k} t_j$. Since the first term of $f_0'(z)$ does not diverge for $z \to 1-$, we only need to take the second term into account. First we have to determine the asymptotic behaviour of T_k. One way is applying Theorem 4.3.1 to $t''(z)$. First observe that

$$t''(z) \sim (1-\beta)z\phi(z)\frac{d^2}{dz^2}\Big[A(\phi(z))\Big] \text{for } z \to 1-.$$

Let $\sigma = 3 - s$. We have

$$\begin{aligned}
&\lim_{z\to 1-}(1-z)^\sigma \frac{d^2}{dz^2}\Big[A\big(\phi(z)\big)\Big] \\
&\quad= \lim_{z\to 1-}\Big(\frac{1-z}{1-\phi(z)}\Big)^\sigma (1-\phi(z))^\sigma A''\big(\phi(z)\big)\big(\phi'(z)\big)^2 \\
&\quad= \lim_{y\to 1-}(1-y)^\sigma A''(y)\big(\phi'(1)\big)^{2-\sigma}.
\end{aligned}$$

Since by Theorem 4.3.1,

$$A''(y) \sim \frac{1}{(1-y)^\sigma}a\Gamma(\sigma),$$

one has, by the same theorem,

$$\sum_{k=0}^{n} k(k-1)t_k \sim \frac{a\Gamma(\sigma)}{\Gamma(\sigma+1)}n^\sigma(\phi'(1))^{2-\sigma}(1-\beta),$$

hence

$$T_k \sim \frac{a}{\phi}'(1)^{s-1}(1-\beta)k^{1-s}.$$

One more application of Theorem 4.3.1 results in

$$f_0'(z) \sim \frac{1}{\big(2-t'(1)\big)^2}\frac{a\Gamma(3-s)}{(s-1)\Gamma(4-s)}(1-z)^{s-3}.$$

Furthermore observe that

$$(1-\rho) = \frac{2-t'(1)}{(1-\beta)A'(1)+1},$$

hence

$$f_0'(z) \sim \frac{1}{(1-\rho)^2\big((1-\beta)A'(1)+1\big)^2}\frac{a\Gamma(3-s)}{(s-1)\Gamma(4-s)}(1-z)^{s-3}.$$

Chapter 5

An LRD Arrival Process Based on Pseudo Self-Similar Traffic Models

In this chapter a specific class of LRD arrival processes is introduced, the definition of which is based on the theory of pseudo self-similar traffic models. Furthermore we investigate the asymptotic behaviour of the multiplexer queue having such an arrival process as input. To state our goals and to clarify the techniques we use, we begin this introduction with a short and informal overview of the field of the pseudo self-similar arrival processes.

According to [52, pg. 338] a pseudo self-similar traffic model is a SRD process, which mimics LRD behaviour over some range of time scales. Of course, when considering large enough time scales, the process reveals its true nature. Otherwise stated, over a certain interval of time the autocovariance of a pseudo self-similar process decays approximately as described in Definition 1.3.9, beyond this interval it starts to decay exponentially. In the literature most effort has gone hitherto into fitting LRD traffic to pseudo-self similar processes with as few states as possible. In e.g. [2] the authors fit a Markov modulated Poisson process with only eight states quite successfully to measured LRD LAN traffic.

The reason these pseudo self-similar traffic models are considered is straightforward. Although they behave approximately like LRD processes, are after all they are still genuine SRD processes. Hence one can reuse the huge amount of theory concerning queueing systems with SRD input to calculate, at least approximately, the performance characteristics of queueing systems with LRD input. In this chapter we originally aimed at obtaining theoretical results on asymptotic behaviour by following this philosophy. Inspired by the results presented in [52, pg. 339], we define a sequence of Markovian arrival processes Y^M, each one exhibiting LRD behaviour over a larger time scale than its predecessor. The limit of this sequence, denoted by Y^∞, is an LRD arrival process. The idea is to study the queue, having Y^∞ as input, as the limit of the queues having Y^M as input. It turns out that this approach only works out for the mean queue length. To obtain a lower bound for the tail probabilities we had to use large deviation techniques.

5.1 The Traffic Model Y^∞: Definitions

The most important properties of the arrival process Y^∞ are summarised below:

- it is defined by means of three parameters (a, b, p),
- under the condition that $a \geq b^2$ it is LRD,
- the Hurst parameter is given by
$$H = \frac{1}{2}\Big(\frac{\log b}{\log a - \log b}\Big),$$
- the Index of Dispersion for Counts (IDC) and its limit can be expressed in a closed form.

As indicated before Y^∞ is constructed as the limit of a sequence of Markovian processes. To define this sequence we use a collection $\{X^{(i)}, 1 \leqslant i < \infty\}$ of stochastically independent Markovian on-off sources. Let $1 < b < a$. The on-period of $X^{(i)}$ is geometrically distributed with mean duration $\left(\frac{a}{b}\right)^i$, the off-period is geometrically distributed with mean duration a^i. While on, the source $X^{(i)}$ generates a cell in a slot with probability p. According to Example 2.2, $X^{(i)}$ is a DBMAP determined by the matrices

$$\mathbf{D}_0^{(i)} = \begin{pmatrix} 1-(1/a)^i & (1/a)^i \\ (1-p)(b/a)^i & (1-p)\big(1-(b/a)^i\big) \end{pmatrix},$$

and

$$\mathbf{D}_1^{(i)} = \begin{pmatrix} 0 & 0 \\ p(b/a)^i & p\big(1-(b/a)^i\big) \end{pmatrix}.$$

Let $\mathbf{D}^{(i)} = \mathbf{D}_0^{(i)} + \mathbf{D}_1^{(i)}$. The stationary distribution $\boldsymbol{\pi}^{(i)}$ of $\mathbf{D}^{(i)}$ is given by

$$\boldsymbol{\pi}^{(i)} = \begin{pmatrix} \dfrac{b^i}{1+b^i} & \dfrac{1}{1+b^i} \end{pmatrix}.$$

The mean arrival rate $\rho^{(i)}$ of $X^{(i)}$ is given by

$$\rho^{(i)} = \frac{p}{1+b^i}.$$

When dealing with LRD the most important characteristic is of course the autocovariance. Recalling (2.2) we have

$$\mathrm{Cov}\Big(X_1^{(i)}, X_{1+k}^{(i)}\Big) = \boldsymbol{\pi}^{(i)}\mathbf{D}_1^{(i)}\Big(\big(\mathbf{D}^{(i)}\big)^{k-1} - \mathbf{e}\boldsymbol{\pi}^{(i)}\Big)\mathbf{D}_1^{(i)}\mathbf{e},$$

from Example 2.4 it follows that

$$\operatorname{Cov}\left(X_1^{(i)}, X_{1+k}^{(i)}\right) = \left[1 - \left(\frac{1}{a}\right)^i - \left(\frac{b}{a}\right)^i\right]^k p^2 \frac{b^i}{(1+b^i)^2}.$$

By the definition of $X^{(i)}$ it is clear that for increasing i the autocovariance is decaying more slowly. This enables us to obtain LRD. On the other hand, the mean arrival rate is decreasing fast, i.e. $\sum_{i=1}^{\infty} \rho^{(i)} < \infty$, which is essential for the construction of Y^∞.

We now construct the sequence Y^M of Markovian processes by which the process Y^∞ is defined. For each integer $M > 0$ define

$$Y^M = \sum_{i=1}^{M} X^{(i)}.$$

As mentioned in Section 2.2.1 this superposition of DMAPs is again a DBMAP, determined by the matrices

$$\begin{aligned}
\mathbf{D}_0^M &= \mathbf{D}_0^{(M)} \otimes \mathbf{D}_0^{(M-1)} \otimes \cdots \otimes \mathbf{D}_0^{(1)}, \\
&\vdots \\
\mathbf{D}_i^M &= \sum_{k_M + \ldots + k_1 = i} \bigotimes_{j=M}^{1} \mathbf{D}_{k_j}^{(j)}, \\
&\vdots \\
\mathbf{D}_M^M &= \mathbf{D}_1^{(M)} \otimes \mathbf{D}_1^{(M-1)} \otimes \cdots \otimes \mathbf{D}_1^{(1)}.
\end{aligned}$$

Furthermore $\mathbf{D}^M = \mathbf{D}^{(M)} \otimes \ldots \otimes \mathbf{D}^{(1)}$, $\mathbf{D}^M(z) = \mathbf{D}^{(M)}(z) \otimes \ldots \otimes \mathbf{D}^{(1)}(z)$ and the stationary probability distribution $\boldsymbol{\pi}^M$ of $\mathbf{D}^M$ is given by

$$\boldsymbol{\pi}^M = \boldsymbol{\pi}^{(M)} \otimes \boldsymbol{\pi}^{(M-1)} \otimes \ldots \otimes \boldsymbol{\pi}^{(1)}.$$

The mean arrival ρ^M of Y^M is given by $\rho^M = \sum_{i=1}^{M} \rho^{(i)}$. The limit process of the sequence Y^M is, as indicated before, denoted by Y^∞. Clearly the mean arrival rate ρ^∞ of Y^∞ is given by

$$\rho^\infty = \sum_{i=1}^{\infty} \frac{p}{1+b^i}.$$

By the independence of the sources $X^{(i)}$, the covariance structure of Y^∞ satisfies

$$\operatorname{Cov}\left(Y_1^\infty, Y_{1+k}^\infty\right) = \sum_{i=1}^{\infty} \left[1 - \left(\frac{1}{a}\right)^i - \left(\frac{b}{a}\right)^i\right]^k p^2 \frac{b^i}{(1+b^i)^2}. \tag{5.1}$$

5.2 Second Order Characteristics of Y^∞

Because of the analytical tractability of (5.1), it is possible to determine the condition on the parameters a and b for the process Y^∞ to be LRD. For the LRD instances of Y^∞ one can also calculate the Hurst parameter. We conclude this section with examining the IDC of Y^∞.

5.2.1 Determination of the Hurst parameter

Proposition 5.2.1. *The arrival process Y^∞ is LRD if and only if $b^2 \leqslant a$.*

Proof. Following Definition 1.3.1, Y^∞ is LRD if and only if the series

$$\sum_{k=1}^{\infty} \mathrm{Cov}\left(Y_1^\infty, Y_{1+k}^\infty\right) \tag{5.2}$$

diverges. Using (5.1) we have

$$\begin{aligned}\sum_{k=1}^{\infty} \mathrm{Cov}\left(Y_1^\infty, Y_{1+k}^\infty\right) &= \sum_{k=1}^{\infty}\sum_{i=1}^{\infty}\left[1-\left(\frac{1}{a}\right)^i-\left(\frac{b}{a}\right)^i\right]^k \frac{p^2 b^i}{(1+b^i)^2} \\ &= \sum_{i=1}^{\infty}\frac{b^i}{(1+b^i)^2}\sum_{k=1}^{\infty}\left[1-\left(\frac{1}{a}\right)^i-\left(\frac{b}{a}\right)^i\right]^k \\ &= \sum_{i=1}^{\infty}\frac{a^i b^i - b^i - b^{2i}}{(1+b^i)^3}.\end{aligned}$$

This last series diverges if and only if $b^2 \leqslant a$. ☐

To determine the Hurst parameter H we examine the asymptotic behaviour of the autocovariance of Y^∞.

Theorem 5.2.2. *There exist some $C_1, C_2 > 0$ such that*

$$C_1 k^{-\beta} < \mathrm{Cov}\left(Y_1^\infty, Y_{1+k}^\infty\right) < C_2 k^{-\beta}, \tag{5.3}$$

with

$$\beta = \frac{\log b}{\log a - \log b}. \tag{5.4}$$

Proof. See Appendix 5.4. ☐

Although the inequalities stated in Theorem 5.2.2 are insufficient to conclude that

$$\mathrm{Cov}\left(Y_1^\infty, Y_{1+k}^\infty\right) \sim c k^{2H-2},$$

for some $c > 0$ and

$$H = \frac{1}{2}\left(2 - \frac{\log b}{\log a - \log b}\right),$$

this value for H will be called the Hurst parameter of the process Y^∞.

5.2.2 The IDC of Y^∞

In this section we investigate the covariance structure of the process Y^∞ by means of the Index of Dispersion for Counts (IDC), introduced in Section 2.4.1. Denote by $I^{(i)}(k)$ the IDC of the process $X^{(i)}$ and let $J^{(i)} = \lim_{k\to\infty} I^{(i)}(k)$. The IDC of the process Y^∞ is denoted by $I^\infty(k)$ and $J^\infty = \lim_{k\to\infty} I^\infty(k)$.
From (2.3) it follows that

$$J^{(i)} = \frac{\boldsymbol{\pi}^{(i)}\mathbf{D}_1^{(i)}\mathbf{e} - 3[\boldsymbol{\pi}^{(i)}\mathbf{D}_1^{(i)}\mathbf{e}]^2 + 2\boldsymbol{\pi}\mathbf{D}_1^{(i)}\mathbf{Z}^{(i)}\mathbf{D}_1^{(i)}\mathbf{e}}{\boldsymbol{\pi}^{(i)}\mathbf{D}_1^{(i)}\mathbf{e}}, \tag{5.5}$$

with $\mathbf{Z}^{(i)}$ the fundamental matrix of the Markov chain associated with $\mathbf{D}^{(i)}$:

$$\mathbf{Z}^{(i)} = \left(\mathbf{I} - \left(\mathbf{D}^{(i)} - \mathbf{e}\boldsymbol{\pi}^{(i)}\right)\right)^{-1}.$$

Using the expressions for $\mathbf{D}_0^{(i)}$ and $\mathbf{D}_1^{(i)}$, given in Section 5.1, it is easy to show that

$$\mathbf{Z}^{(i)} = \frac{1}{(1+b^i)^2}\begin{pmatrix} a^i + b^i(1+b^i) & 1 - a^i + b^i \\ b^i(1-a^i+b^i) & 1 + b^i + a^i b^i \end{pmatrix}.$$

Hence,

$$\boldsymbol{\pi}^{(i)}\mathbf{D}_1^{(i)}\mathbf{Z}^{(i)}\mathbf{D}_1^{(i)}\mathbf{e} = \frac{p^2}{(1+b^i)^3}\big(1 + b^i(a^i - b^i)\big),$$

and (5.5) becomes

$$J^{(i)} = 1 - 3\frac{p}{1+b^i} + 2\frac{p}{(1+b^i)^2}\big(1 + b^i(a^i - b^i)\big). \tag{5.6}$$

Now we compute J^∞, i.e. the limit of the IDC of the process Y^∞. Since $Y^\infty = \sum_{i=1}^\infty X^{(i)}$ it follows that

$$I^\infty(k) = \sum_{i=1}^\infty \frac{\rho^{(i)}}{\rho^\infty} I^{(i)}(k).$$

Hence,

$$I^\infty(k) = \frac{\sum_{i=1}^\infty \operatorname{Cov}(X_1^{(i)}, X_1^{(i)}) + \sum_{i=1}^\infty 2\sum_{j=1}^{k-1}\frac{k-j}{k}\operatorname{Cov}(X_1^{(i)}, X_{1+j}^{(i)})}{\rho^\infty}.$$

Taking the limit for $k \longrightarrow \infty$, we obtain

$$J^\infty = \frac{\rho^\infty - 3\displaystyle\sum_{i=1}^\infty (\rho^{(i)})^2 + 2p^2\sum_{i=1}^\infty \frac{1 + b^i(a^i - b^i)}{(1+b^i)^3}}{\rho^\infty}. \tag{5.7}$$

From (5.7) it follows that the limit of the IDC of the process Y^∞ is infinite if $b^2 \leqslant a$, which is the condition under which the process is LRD. This is in agreement with the criterion that a process is LRD if its IDC is diverging.

5.3 Queueing Behaviour

Originally we intended to use the philosophy behind the pseudo self-similar traffic models to obtain buffer asymptotics for the multiplexer queue with Y^∞ as input. Otherwise stated, deriving bounds for the tail probabilities of this queue, which is called the Y^∞-queue, should be done by studying the queues with the processes Y^M as input. If we could quantify the behaviour of these Y^M-queues as a function of M, then a description of the behaviour of the Y^∞-queue would emerge by letting $M \to \infty$. Unfortunately this approach only works out for the calculation of the mean queue length. Obtaining buffer asymptotics this way is impossible by the rapidly increasing complexity of the calculations.

The mean queue lengths of the Y^M-queues constitute a sequence diverging to ∞. Hence one can conclude that the mean queue length of the Y^∞-queue does not exists, or otherwise stated, equals ∞. This observation indicates a power-law decay of the tail probabilities, which is confirmed in Section 5.3.2 by the calculation of a large deviations lower bound.

Simulation results reveal that the buffer asymptotics can be described by formula (6.12) derived for a multiplexer queue with discrete-time LRD M/G/∞ input. Details can be found in Section 6.3.

The invariant probability vector of the Y^M-queue, which is a DBMAP-D-1 queue, is denoted by $\mathbf{x}^M$, with $\mathbf{X}^M(z) = \sum_{k=}^{\infty} \mathbf{x}_k^M z^k$ the associated z-transform. From now on we assume that the cells arrive back-to-back, i.e. $p = 1$, which makes it possible to derive a closed form for $\mathbf{x}_0^M$. The existence of the $\mathbf{x}^M$ is guaranteed by assuming $\rho^\infty < 1$.

5.3.1 Mean queue length

It is clear that the mean queue length of the Y^M-queue is given by $\left(\mathbf{X}^M\right)'\mathbf{e}$ with

$$\left(\mathbf{X}^M\right)' \stackrel{\text{def}}{=} \frac{d}{dz}\mathbf{X}^M(z)\Big|_{z=1}.$$

In [42] this quantity is calculated by deriving both sides of

$$\mathbf{X}^M(z)\left(z\mathbf{I} - \mathbf{D}^M(z)\right) = (z-1)\mathbf{x}_0^M \mathbf{D}^M(z).$$

This results in

$$\begin{aligned}\left(\mathbf{X}^M\right)'\mathbf{e} = &\frac{\boldsymbol{\pi}\left(\mathbf{D}^M\right)''\mathbf{e} + 2\mathbf{x}_0^M\left(\mathbf{D}^M\right)'\mathbf{e} - 2\rho^M}{2(1-\rho^M)} \\ &+ \frac{\left(2\mathbf{x}_0^M\mathbf{D}^M + 2\boldsymbol{\pi}^M\left(\mathbf{D}^M\right)'\right)\left(\mathbf{I} - \mathbf{D}^M + \mathbf{e}\boldsymbol{\pi}^M\right)^{-1}\left(\mathbf{D}^M\right)'\mathbf{e}}{2(1-\rho^M)}.\end{aligned}$$

We show that the r.h.s. diverges for $M \to \infty$. Since both $\boldsymbol{\pi}\left(\mathbf{D}^M\right)''\mathbf{e}$ and $2\mathbf{x}_0\left(\mathbf{D}^M\right)'\mathbf{e}$ are positive and finite, and since $\rho^M < \rho^\infty < 1$, it is sufficient to investigate the behaviour for $M \to \infty$ of the expression

$$2\mathbf{x}_0^M \mathbf{D}^M \left(\mathbf{I} - \mathbf{D}^M + \mathbf{e}\boldsymbol{\pi}^M\right)^{-1} \left(\mathbf{D}^M\right)' \mathbf{e} + \\ 2\boldsymbol{\pi}^M \left(\mathbf{D}^M\right)' \left(\mathbf{I} - \mathbf{D}^M + \mathbf{e}\boldsymbol{\pi}^M\right)^{-1} \left(\mathbf{D}^M\right)' \mathbf{e}. \quad (5.8)$$

We first focus on the second term. Since

$$\left(\mathbf{I} - \left(\mathbf{D}^M - \mathbf{e}\boldsymbol{\pi}^M\right)\right)^{-1} = \mathbf{I} + \sum_{k=1}^{\infty} \left(\left(\mathbf{D}^M\right)^k - \mathbf{e}\boldsymbol{\pi}^M\right), \quad (5.9)$$

and because $2\boldsymbol{\pi}^M \left(\mathbf{D}^M\right)' \left(\mathbf{D}^M\right)' \mathbf{e}$ is finite and positive, we only need to take

$$\boldsymbol{\pi} \left(\mathbf{D}^M\right)' \left[\sum_{k=1}^{\infty} \left(\left(\mathbf{D}^M\right)^k - \mathbf{e}\boldsymbol{\pi}\right)\right] \left(\mathbf{D}^M\right)' \mathbf{e}$$

into account. Since

$$\operatorname{Cov}\left(Y_1^{\infty}, Y_{1+k}^{\infty}\right) = \boldsymbol{\pi} \left(\mathbf{D}^M\right)' \left(\left(\mathbf{D}^M\right)^k - \mathbf{e}\boldsymbol{\pi}\right) \left(\mathbf{D}^M\right)' \mathbf{e},$$

we have

$$\boldsymbol{\pi} \left(\mathbf{D}^M\right)' \left[\sum_{k=1}^{\infty} \left(\left(\mathbf{D}^M\right)^k - \mathbf{e}\boldsymbol{\pi}\right)\right] \left(\mathbf{D}^M\right)' \mathbf{e} = \\ \sum_{k=2}^{\infty} \sum_{i=1}^{M} \left[1 - \left(\frac{1}{a}\right)^i - \left(\frac{b}{a}\right)^i\right]^k \frac{b^i}{(1+b^i)^2}. \quad (5.10)$$

The first term of the sum (5.8) can be reduced in the same way to

$$\mathbf{x}_0 \mathbf{D}^M \left[\sum_{k=1}^{\infty} \left(\left(\mathbf{D}^M\right)^k - \mathbf{e}\boldsymbol{\pi}\right)\right] \left(\mathbf{D}^M\right)' \mathbf{e}.$$

Let us now determine $\mathbf{x}_0^M$. Since the system can only be empty if all sources were in the off-state during the previous slot, one has

$$\mathbf{x}_0^M = (1 - \rho^M) \mathbf{u}^{(M)} \otimes \mathbf{u}^{(M-1)} \otimes \cdots \otimes \mathbf{u}^{(1)},$$

with $\mathbf{u}^{(i)}$ the first row of the matrix $\mathbf{D}^{(i)}$. Using the elementary properties of the Kronecker product $\otimes$ one obtains

$$\mathbf{x}_0 \mathbf{D}^M \left[\sum_{k=1}^{\infty} \left(\left(\mathbf{D}^M\right)^k - \mathbf{e}\boldsymbol{\pi}\right)\right] \left(\mathbf{D}^M\right)' \mathbf{e} = \\ -(1-\rho) \sum_{k=1}^{\infty} \sum_{i=1}^{M} \left[1 - \left(\frac{1}{a}\right)^i - \left(\frac{b}{a}\right)^i\right]^k \frac{1}{1+b^i}. \quad (5.11)$$

Combining (5.10) and (5.11) it follows that the mean queue length of the Y^∞-queue is ∞ if

$$\lim_{M\to\infty}\sum_{i=1}^{M}\sum_{k=0}^{\infty}\left[1-\left(\frac{1}{a}\right)^i-\left(\frac{b}{a}\right)^i\right]^k\left(\frac{b^i}{(1+b^i)^2}-(1-\rho)\frac{1}{1+b^i}\right)=\infty.$$

One can check this is the case if and only if $b^2 \leqslant a$, or otherwise stated, if and only if the arrival process Y^∞ is LRD. To simplify the calculations we could have used Viterbi's formula for the mean queue length (see [49, pg. 353–360]). But doing so also hides the influence of the autocovariance, which causes the diverging mean queue lengths. For a generalisation of this result we refer the reader to [28].

5.3.2 A large deviations lower bound

Calculating large deviations bounds for the tail probabilities of queueing systems is based on the notion of workload process (see e.g. [23]). This process expresses the difference between the amount of work that arrives at the queue in some time interval, and the amount of work that can be processed in this interval. Formally the workload process W_n is defined by

$$W_n = A_{[-n,0)} - n,$$

with $A_{[s,t)}$ the amount of work, here equal to the number of cells, that arrive during the slots $s, s+1, \dots, t-1$. For a DBMAP-D-1 queue the number of cells that can be processed in this interval equals $t-s$. The stationary buffer distribution Q can be expressed in terms of the workload process by Reich's formula (see e.g.[52]):

$$Q = \sup_{n>0} W_n. \tag{5.12}$$

The reader should note that this distribution Q takes into account only those cells waiting in the buffer, disregarding the one that is being served. A lower bound for $\mathbf{P}\{Q>k\}$ can be derived from (5.12) as follows:

$$\mathbf{P}\{Q>k\} = \mathbf{P}\{\sup_{n>0} W_n > k\} \geqslant \sup_{n>0} \mathbf{P}\{W_n > k\}. \tag{5.13}$$

Using this inequality, and inspired by the paper [41], we are able to derive the following lower bound.

Theorem 5.3.1 (Large deviations lower bound). *The buffer distribution Q as defined by Reich's formula (5.12) satisfies*

$$\liminf_{k\to\infty}\frac{\log \mathbf{P}\{Q>k\}}{\log k} \geqslant \frac{\log b}{\log a - \log b}.$$

Proof. First note that the Markov chain determined by $\mathbf{D}^{(i)}$ is reversible:

$$\boldsymbol{\pi}_p^{(i)}\big(\mathbf{D}^{(i)}\big)_{p,q} = \boldsymbol{\pi}_q^{(i)}\big(\mathbf{D}^{(i)}\big)_{q,p}.$$

Hence we can rewrite the workload process W_n associated with Y^∞ as follows:

$$\begin{aligned} W_n &= \sum_{j=1}^{n} Y_j^\infty - n \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{n} X_j^{(i)} - n. \end{aligned}$$

Choose some ε such that $0 < \varepsilon < \rho^\infty$. Define $\gamma = \rho^\infty - \varepsilon$. Take M_ε such that

$$\rho^{M_\varepsilon} = \sum_{i=1}^{M_\varepsilon} \rho^{(i)} > \rho^\infty - \frac{\varepsilon}{2}.$$

Define

$$\hat{Y}^{M_\varepsilon} = \sum_{i=M_\epsilon+1}^{\infty} X^{(i)},$$

hence $Y^{M_\varepsilon} + \hat{Y}^{M_\varepsilon} = Y^\infty$. The following series of inequalities is justified by inequality (5.13) and by the independence of the $X^{(i)}$:

$$\begin{aligned} &\liminf_{k\to\infty} \frac{\log \mathbf{P}\{Q > k\}}{\log k} \\ &\quad = \liminf_{k\to\infty} \frac{\log \mathbf{P}\{Q > \gamma k\}}{\log \gamma k} \\ &\quad \geqslant \liminf_{k\to\infty} \frac{\log \mathbf{P}\{\sum_{j=1}^k Y_j^\infty - k > \gamma k\}}{\log \gamma k} \\ &\quad \geqslant \liminf_{k\to\infty} \frac{\log \mathbf{P}\{\sum_{j=1}^k \hat{Y}_j^{M_\varepsilon} \geqslant k, \sum_{j=1}^k Y^{M_\varepsilon} > (\rho^\infty - \varepsilon)k\}}{\log \gamma k} \\ &\quad = \liminf_{k\to\infty} \frac{\log \mathbf{P}\{\sum_{j=1}^k \hat{Y}_j^{M_\varepsilon} \geqslant k\} + \log \mathbf{P}\{\sum_{j=1}^k Y^{M_\varepsilon} > (\rho^\infty - \varepsilon)k\}}{\log \gamma k} \\ &\quad \geqslant \liminf_{k\to\infty} \frac{\log \mathbf{P}\{\sum_{j=1}^k \hat{Y}_j^{M_\varepsilon} \geqslant k\}}{\log \gamma k} \\ &\qquad + \liminf_{k\to\infty} \frac{\log \mathbf{P}\{\sum_{j=1}^k Y^{M_\varepsilon} > (\rho^\infty - \varepsilon)k\}}{\log \gamma k}. \end{aligned}$$

We start with examining the second term of the last expression. Since Y^{M_ε} can be interpreted as a functional of a Markov chain with a finite state space, one has that with probability 1,

$$\frac{1}{k}\sum_{j=1}^{k} Y^{M_\varepsilon} \to \rho^{M_\varepsilon}.$$

Hence one can conclude that

$$\lim_{k\to\infty} \mathbf{P}\Big\{\sum_{j=1}^{k} Y^{M_\varepsilon} > (\rho^\infty - \varepsilon)k\Big\} = 1,$$

or

$$\liminf_{k\to\infty} \frac{\log \mathbf{P}\{\sum_{j=1}^{k} Y^{M_\varepsilon} > (\rho^\infty - \varepsilon)k\}}{\log \gamma k} = 0$$

We now turn to the other term. In order to deal with $\mathbf{P}\{\sum_{j=0}^{k} \hat{Y}_j^{M_\varepsilon} \geqslant k\}$, we introduce for each k the collection of stochastically independent random variables $Z_i^{(k)}$ with $i \geqslant M_\varepsilon + 1$. The distribution of $Z_i^{(k)}$ is defined via its generating function:

$$\mathbf{E}[z^{Z_i^{(k)}}] = 1 - \left[1 - \left(\frac{b}{a}\right)^i\right]^{k-1} \frac{1}{1+b^i} + \left[1 - \left(\frac{b}{a}\right)^i\right]^{k-1} \frac{1}{1+b^i} z^k.$$

The definition of $Z_i^{(k)}$ is based on the total number of arrivals generated by the source $X^{(i)}$ during k consecutive slots. Note that the z-transform of $\sum_{j=1}^{k} X^{(i)}$ equals

$$\boldsymbol{\pi}^{(i)}\Big(\mathbf{D}^{(i)}(z)\Big)^{k-1}\mathbf{e},$$

from this formula it follows that

$$Z_k^{(i)} = \begin{cases} 0 & \text{if } \sum_{j=1}^{k} X^{(i)} < k, \\ k & \text{if } \sum_{j=1}^{k} X^{(i)} = k. \end{cases}$$

Consequently

$$\mathbf{P}\Big\{\sum_{j=1}^{k} \hat{Y}_j^{(M_\varepsilon)} \geqslant k\Big\} \geqslant \mathbf{P}\Big\{\sum_{i=M_\varepsilon+1}^{\infty} Z_i^{(k)} \geqslant k\Big\}.$$

Let $\alpha_i^k = \big(1 - (b/a)^i\big)^{k-1} 1/(1+b^i)$ and consider the z-transform of the sum $\sum_{i=M_\varepsilon+1}^{\infty} Z_i^{(k)}$,

$$\begin{aligned}\mathbf{E}\Big[z^{\sum_{i=M_\varepsilon+1}^{\infty} Z_i^{(k)}}\Big] &= \prod_{i>M_\epsilon} \mathbf{E}\big[z^{Z_i^{(k)}}\big] \\ &= \prod_{i>M_\epsilon+1} (1-\alpha_i^k) + \sum_{j>M_\epsilon}^{\infty} \prod_{\substack{i>M_\epsilon \\ i\neq j}} (1-\alpha_i^k)\alpha_j^k z^k + \ldots\end{aligned}$$

Clearly for arbitrary k and j

$$\prod_{\substack{i>M_\varepsilon \\ i\neq j}} (1-\alpha_i^k) > \prod_{i=1}^{\infty}\Big[1-\Big(\frac{b}{a}\Big)^i\Big]^{k-1}.$$

Denote the r.h.s. of this inequality by C, hence

$$\mathbf{P}\Big\{\sum_{i=M_\varepsilon+1}^{\infty} Z_i^{(k)} \geqslant k\Big\} \geqslant C \sum_{i=M_\varepsilon+1}\Big[1-\Big(\frac{b}{a}\Big)^i\Big]^{k-1}\frac{1}{1+b^i}.$$

From the calculations in the proof of Theorem 5.2.2 it follows that

$$\sum_{i=M_\varepsilon+1}^{\infty}\Big[1-\Big(\frac{b}{a}\Big)^i\Big]^{k-1}\frac{1}{1+b^i} > ck^{-\beta},$$

with $\beta = \log b/(\log a - \log b)$ and some $c > 0$. Hence

$$\liminf_{k\to\infty} \frac{\log \mathbf{P}\{\sum_{j=1}^k \hat{Y}_j^{M_\varepsilon} \geqslant k\}}{\log \gamma k} \geqslant -\beta,$$

which concludes the proof. □

5.4 Appendix

A detailed outline of the proof of Theorem 5.2.2 is presented. We have to show the existence of two positive constants C_1 and C_2 such that, for k large enough,

$$C_1 k^{-\beta} < \mathrm{Cov}\big(Y_1^\infty, Y_{1+k}^\infty\big) < C_2 k^{-\beta},$$

with

$$\beta = \frac{\log b}{\log a - \log b}.$$

We will assume $b^2 < a$, the case $b^2 = a$ can be handled in a similar way. The proof consists of two parts. In the first part the behaviour of an auxiliary function $g(k)$ is determined. In the second part it is proved that $\mathrm{Cov}\big(Y_1^\infty, Y_{1+k}^\infty\big)$ behaves in essentially the same way. Without losing generality, let $p = 1$. The function $g(k)$ is defined by:

$$g(k) = \sum_{i=1}^{\infty}\Big[1-\Big(\frac{b}{a}\Big)^i\Big]^k\Big(\frac{1}{b}\Big)^i. \tag{5.14}$$

Since integrals are easier to evaluate we squeeze $g(k)$ as follows:

$$\int_2^\infty \left[1-\left(\frac{b}{a}\right)^{(x-1)}\right]^k \left(\frac{1}{b}\right)^x dx < \sum_{i=2}^\infty \left[1-\left(\frac{b}{a}\right)^i\right]^k \left(\frac{1}{b}\right)^i$$
$$< \sum_{i=1}^\infty \left[1-\left(\frac{b}{a}\right)^i\right]^k \left(\frac{1}{b}\right)^i < \int_1^\infty \left[1-\left(\frac{b}{a}\right)^x\right]^k \left(\frac{1}{b}\right)^{x-1} dx$$

For our purposes it suffices to examine the behaviour of the integral

$$\int_1^\infty \left[1-\left(\frac{b}{a}\right)^x\right]^k \left(\frac{1}{b}\right)^x dx$$

when $k \to \infty$. We rewrite it as the sum

$$\int_1^{\gamma_k} \left[1-\left(\frac{b}{a}\right)^x\right]^k \left(\frac{1}{b}\right)^x dx + \int_{\gamma_k}^\infty \left[1-\left(\frac{b}{a}\right)^x\right]^k \left(\frac{1}{b}\right)^x dx \tag{5.15}$$

with

$$\gamma_k = \frac{\log k}{\log a - \log b}.$$

The reasoning for choosing this γ_k goes as follows. We want to replace $\big(1-(b/a)^x\big)^k$ by $1-k(b/a)^x$, but this approximation is only reasonable as long as $1-k(b/a)^x \geqslant 0$, so one must have $x \geqslant \gamma_k$ for k fixed. By the inequalities

$$1-k\left(\frac{b}{a}\right)^x < \left[1-\left(\frac{b}{a}\right)^x\right]^k,$$

and

$$\frac{1}{b}\int_{\gamma_k}^\infty \left[1-\left(\frac{b}{a}\right)^x\right]^k \left(\frac{1}{b}\right)^x dx < g(k),$$

we immediately have

$$\frac{1}{b}\left(\frac{1}{\log b} - \frac{1}{\log a}\right) k^{-\beta} < g(k).$$

For calculating an upper bound the approximation mentioned above is no longer sufficient. Better approximations, i.e. second-order ones, are needed for bounding the second integral of (5.15). For the first integral it suffices that

$$\left[1-\left(\frac{b}{a}\right)^x\right]^k < e^{-k(b/a)^x} < \frac{1}{k}\left(\frac{b}{a}\right)^{-x},$$

resulting in the upper bound

$$\int_1^{\gamma_k}\left[1-\left(\frac{b}{a}\right)^x\right]^k\left(\frac{1}{b}\right)^x dx < \int_1^{\gamma_k}\frac{1}{k}\left(\frac{a}{b^2}\right)^x dx$$
$$< \frac{1}{\log a - 2\log b}k^{-\beta}.$$

For the second integral of (5.15) we use the inequality

$$\left[1-\left(\frac{b}{a}\right)^x\right]^k < 1 - k\left(\frac{b}{a}\right)^x + \frac{k^2}{2}\left(\frac{b}{a}\right)^{2x}.$$

Hence

$$\int_{\gamma_k}^{\infty}\left[1-\left(\frac{b}{a}\right)^x\right]^k\left(\frac{1}{b}\right)^x dx < \int_{\gamma_k}^{\infty}\left[1-k\left(\frac{b}{a}\right)^x\right]\left(\frac{1}{b}\right)^x dx$$
$$+\int_{\gamma_k}^{\infty}\frac{k^2}{2}\left(\frac{b}{a}\right)^{2x}\left(\frac{1}{b}\right)^x dx$$
$$= \left(\frac{1}{\log b} - \frac{1}{\log a}\right)k^{-\beta} + \frac{1}{2\log a - \log b}k^{-\beta}.$$

Combining these results in

$$g(k) < \left(\frac{1}{\log a - 2\log b} + \frac{1}{\log b} - \frac{1}{\log a} + \frac{1}{2\log a - \log b}\right)bk^{-\beta}.$$

The fact that the coefficients of $k^{-\beta}$ are positive is a consequence of $b < a$ and $b^2 < a$.

We are left with justifying the replacement of the autocovariance by the function $g(k)$. The upper bound does not pose any problems since

$$\mathrm{Cov}\left(Y_1^{\infty}, Y_{k+1}^{\infty}\right) < g(k).$$

For the lower bound it can be shown that the difference between $g(k)$ and the autocovariance decays as $k^{-2\beta}$, by using the estimate

$$g(k) - \mathrm{Cov}\left(Y_1^{\infty}, Y_{k+1}^{\infty}\right) < \sum_{i=1}^{\infty}\left[1-\left(\frac{b}{a}\right)^i\right]^k\frac{3}{b^{2i}} + \sum_{i=1}^{\infty}\left[1-\left(\frac{b}{a}\right)^i\right]^k\frac{1}{a^i}.$$

Chapter 6

LRD M/G/∞ Arrival Processes and Buffer Asymptotics

The M/G/∞ process is, together with arrival processes derived from fractional Brownian motion, the most frequently used mathematical model to describe LRD traffic streams. In this chapter we derive the exact buffer asymptotics for the multiplexer queueing system having a discrete-time LRD M/G/∞ process as input.

In Section 6.1 we introduce the studied class of discrete-time LRD M/G/∞ processes, together with a short discussion of the applicability of these arrival processes to the modelling of today's teletraffic.

In section 6.2 the buffer asymptotics are derived. This section starts with an overview of the results obtained in the literature for related queueing systems with continuous-time or discrete-time LRD M/G/∞ input. The derivation of the buffer asymptotics begins with the modelling of the discrete-time LRD M/G/∞ process as a DBMAP. Hence we are dealing again with a DBMAP-D-1 queue, which allows us to obtain an analytically tractable expression for the generating function associated with the stationary buffer distribution. From this expression we derive, by using the Tauberian theorem for power series, the formula describing the asymptotic behaviour of the tail probabilities. In Section 6.2.4 this result is generalised to the superposition of an LRD M/G/∞ process and an SRD traffic stream, modelled by a finite state DBMAP. Simulation results indicating further generalisations are presented in Section 6.3.

The chapter concludes with a discussion of the practical relevance of the studied queueing model. At the same time we point out several possibilities to obtain, by the techniques presented here, buffer asymptotics for related but more generally applicable queueing systems.

6.1 M/G/∞ Arrival Processes and LRD

6.1.1 M/G/∞ arrival processes: definitions

The M/G/∞ arrival process has both a continuous-time and a discrete-time version. We start with recalling the definition of the discrete-time M/G/∞ input process as it can be found in e.g. [37] and [58]. The number of arrivals generated by the M/G/∞ process is defined by means of the following auxiliary M/G/∞ queueing system, which clarifies the origin of its name. Consider a system with infinitely many servers to which, during slot k, γ_k new customers arrive. The γ_k are i.i.d. Poisson variables with mean λ. Each of these customers is presented to an idle server and the service starts at slot $k+1$. The service times are i.i.d. random variables with distribution G. We now come to the M/G/∞ arrival process: the number of arrivals it generates during slot k equals the number of busy servers b_k in the M/G/∞ queueing system.

A more intuitive but equivalent description goes as follows. At the k-th time slot a number γ_k of new so-called trains are generated according to a Poisson distribution with mean λ. The random variables γ_k are taken to be independent. Each train consists of a number of back-to-back customers (i.e. the customers belonging to the same train arrive in consecutive time slots). The random variable representing the length of a train, in number of customers, has distribution G. Again b_k denotes the number of customers arriving during slot k. One can derive from this description that the mean arrival rate $\mathbf{E}[b_k]$ equals $\lambda\mu$, with μ the mean of G.

For future reference we also introduce the continuous-time, fluid-flow M/G/∞ arrival process. In [22] it is defined as a random marked Poisson process of intensity λ, the marks of which are i.i.d. random variables with common distribution G. The arrival times are those of customers, each of which starts service immediately for a time given by the corresponding mark. Let N_t be the number of customers in the system at time t. Then $A_t = \int_0^t dsN(s)$ denotes the total amount of fluid generated in the interval $[0,t]$. This definition allows for an interpretation similar to the one for the discrete-time case.

6.1.2 LRD M/G/∞ arrival processes

The M/G/∞ process was first mentioned in the context of LRD in [15]. To be able to identify which discrete-time M/G/∞ processes are LRD, we recall the covariance structure for the general case. Let σ denote a random variable with distribution G. As shown in detail in [37],

$$\mathrm{Cov}(b_j, b_{j+k}) = \lambda \sum_{n=k}^{\infty} \mathbf{P}\{\sigma > n\}, \tag{6.1}$$

with b_j the number of arrivals in slot j. From this formula it follows that a very wide range of covariance structures can be covered by just choosing the appropriate G. Since

$$\sum_{j=0}^{\infty} \mathrm{Cov}(b_k, b_{k+j}) = \frac{\lambda}{2}\mathbf{E}[\sigma(\sigma+1)],$$

it is clear that the process b_k is LRD if the second moment of G is infinite. This is e.g. true for Pareto-like instances of G. A Pareto-like distributed random variable, which will be denoted by τ_A, is characterised by

$$\mathbf{P}\{\tau_A = k\} \sim ak^{-s}, \tag{6.2}$$

with $2 < s < 3$ and $a > 0$. Note that as before $f(k) \sim g(k)$ stands for $\lim_{k\to\infty} f(k)/g(k) = 1$. Hence by (6.1),

$$\mathrm{Cov}(b_j, b_{j+k}) \sim \frac{\lambda a}{(s-1)(s-2)} k^{2-s}. \tag{6.3}$$

According to (1.4) the Hurst parameter of the process b_k is given by $H = (4-s)/2$. From now on the term discrete-time LRD M/G/∞ arrival process is reserved for the case with G being a Pareto-like distribution, represented by τ_A and determined by $2 < s < 3$ and $a > 0$. Furthermore such an arrival process will be denoted by $\mathbf{A}$, the generating function of τ_A by $A(z) = \sum_{k=1}^{\infty} a_k z^k$ and $B(z)$ designates the generating function $\exp\big(\lambda(z-1)\big)$. Hence the mean arrival rate ρ of $\mathbf{A}$ is given by $\rho = B'(1)A'(1) = \lambda B'(1)$.

6.1.3 LRD M/G/∞ processes and traffic measurements

Currently much research is devoted to the analysis of traffic measurements. The results of these studies can be used to judge the validity of mathematical traffic models. Measurements of Internet traffic have shown that in several cases the M/G/∞ process can be used as a first approximation, see e.g. [5], [50], [4] and [47]. Consider the following example.

In [50] it is demonstrated that requests for service arrive at an FTP-server according to a Poisson process. Furthermore the lengths of the downloaded files are approximately distributed like τ_A in (6.2). So as a first approximation an LRD M/G/∞ process seems appropriate. When however modelling an FTP-connection in more detail, one has to take into account that the rate by which data is transmitted during this connection has a — possibly complicated — stochastic structure of its own. Hence for more accurate results one should adapt the M/G/∞ process to incorporate this small-time scale behaviour. For WWW-sessions one gets more or less the same picture, but with an even more complicated "within-session" structures, see e.g. [16].

A further discussion of this topic is given in Section 6.4.

6.2 Buffer Asymptotics

As indicated in the introduction we will derive the exact buffer asymptotics for the multiplexer queue with a discrete-time LRD M/G/∞ process **A** as input. Hence we consider a single-server, infinite capacity queue with a deterministic service time equal to one time slot. For stability reasons it is assumed that the mean arrival rate $\rho < 1$. From now on this queueing system will be denoted by **QT** (referring to the train arrivals).

We start with presenting an overview of some related results concerning the buffer asymptotics for queues with LRD M/G/∞ input, as they can be found in the literature.

To actually determine the exact buffer asymptotics for the queueing system under study, we will derive in Section 6.2.2 an analytically tractable expression for the generating function $Q(z)$, associated with the stationary distribution q of the buffer occupancy. Applying the Tauberian theorem for power series (4.3.1), we obtain the asymptotic formula for the tail probabilities. In Section 6.2.4 we generalise this result for a queue having as input the superposition of an LRD M/G/∞ process and an SRD traffic stream, modelled by a finite state DBMAP.

6.2.1 Overview of related results

The buffer asymptotics for queues with LRD M/G/∞ input are rather extensively studied in the literature: see e.g. [12], [34], [33], [41] and [58] We limit our discussion to the two most relevant — with respect to our work — papers.

In [41] the authors apply large deviation techniques. The application of their bounds to the case studied here results in

$$\lim_{k\to\infty} \frac{\log \mathbf{P}\{q > k\}}{\log k} = 2 - s, \tag{6.4}$$

indicating tail probabilities which decay according to a power-law. The formulation of the statement reveals at the same time the drawback of the large deviations approach: (6.4) implies $\mathbf{P}\{q > k\} \approxeq f(k)k^{2-s}$, but the behaviour of $f(k)$ cannot be precisely determined. By using a Tauberian theorem we conclude that $f(k)$ is a constant and we determine its value.

For a fluid-flow queue with input a continuous-time LRD M/G/∞ arrival process, exact buffer asymptotics are conjectured in [34] and [35]. Here the distribution G is supposed to be regularly varying with non-integer exponent $\alpha > 0$:

$$\mathbf{P}\{\tau^{\text{on}} > x\} = \frac{l(x)}{x^{\alpha}},$$

with τ^{on} distributed according to G and $l(x)$ slowly varying, i.e. for $\delta > 1$ one has $\lim_{x\to\infty} l(\delta x)/l(x) = 1$. The authors state that for the steady-state amount of fluid q^{f} in the queue with service rate 1,

$$\lim_{x\to\infty} \frac{\mathbf{P}\{q^{\mathrm{f}} > x\}}{\int_{x/\rho}^{\infty} \mathbf{P}\{\tau^{\mathrm{on}} > u\}\, du} = \frac{\lambda}{1-\rho},$$

with $\rho = \lambda \mathbf{E}[\tau^{\mathrm{on}}]$. Let us translate this result to our setting: the integral $\int_{x/\rho}^{\infty} \mathbf{P}\{\tau^{\mathrm{on}} > u\} du$ becomes $\sum_{j=\lfloor k/\rho \rfloor}^{\infty} \mathbf{P}\{\tau_A > k\}$, with $\lfloor x \rfloor$ the integer part of x. Hence we obtain

$$\mathbf{P}\{q > k\} \sim \frac{\lambda a \rho^{2-s}}{(s-2)(s-1)(1-\rho)} k^{2-s}.$$

In Section 6.2.3 this formula will be proved. This implies at the same time that the conjecture in [35] is bound to be true.

6.2.2 The generating function $Q(z)$ of the buffer distribution of QT

The system **QT** will be modelled as a DBMAP-D-1 queue. Doing so we can use the Pollachek-Kinchin equation to derive an appropriate representation of $Q(z)$.

The system QT as a DBMAP-D-1 queue

We define the DBMAP associated with **A** by translating the FEA description of this arrival process as presented in [62] and [64]. We do not just follow the FEA, mainly because the DBMAP approach enables us to state and prove, in a more direct way, the generalisation presented in Section 6.2.4.

As shown in Section 3.2, a driving process θ_k and the link between this process and the number of arrivals, is needed. We recall the definition of the infinite-dimensional driving process $\theta_k = \left(\theta_k^{(i)}\right)_{i \geqslant 1}$ from [64]: $\theta_k^{(i)}$ is defined as the number of sequences that generate their i-th packet during slot k. Hence

$$\theta_k^{(1)} = e^{-\lambda} \frac{\lambda^k}{k!},$$

and

$$\theta_k^{(i+1)} = \sum_{j=1}^{\theta_{k-1}^{(i)}} c_{i,j} \text{ for } i = 1, \ldots, n-1,$$

with $c_{i,j}$ a collection of i.i.d. random variables with generating function

$$C_i(z) = 1 - p_i + p_i z,$$

where

$$p_i = \frac{1 - \sum_{j=1}^{i} a_j}{1 - \sum_{j=1}^{i-1} a_j}.$$

Hence p_i expresses the probability that a train which lasted already i slots, will last at least $i+1$ slots. Clearly the positive integers constitute the state space of each component $\theta_k^{(i)}$. The total number of arrivals is given by

$$e_k = \sum_{i=1}^{\infty} \theta_k^{(i)}.$$

We will denote the DBMAP associated with $\mathbf{A}$ by $(\mathbf{D}, \mathbf{D}_l)$. According to (3.4) the matrices $\mathbf{D}_l$ are defined by

$$\big[\mathbf{D}_l\big]_{(j_i)_{i\geqslant 1},(j_i')_{i\geqslant 1}} = \mathbf{P}\big\{e_k = l, \theta_k^{(i)} = j_i'; i \geqslant 1 | \theta_{k-1}^{(i)} = j_i; i \geqslant 1\big\}.$$

A representation of $Q(z)$ will be derived from the Pollachek-Kinchin equation

$$\mathbf{X}(z) = (z-1)\mathbf{x}_0 \mathbf{D}(z)\big(z\mathbf{I} - \mathbf{D}(z)\big)^{-1}.$$

It is therefore important to observe that the vector $\mathbf{x}_0$ has only one non-zero entry: the one corresponding to all the $\theta^{(i)} = 0$, its value being $p_0 = 1 - \rho = 1 - \lambda A'(1)$.

Determination of $Q(z)$

The generating function $Q(z) = \mathbf{X}(z)\mathbf{e}$, with $\mathbf{e}$ a column vector of 1's, satisfies locally — where the r.h.s. converges — the equality

$$Q(z) = \mathbf{X}(z)\mathbf{e} = (z-1)\mathbf{x}_0 \sum_{k=1}^{\infty} \frac{\big(\mathbf{D}^{(n)}(z)\big)^k}{z^k}\mathbf{e}. \tag{6.5}$$

By using the special structure of $\mathbf{D}(z)$, we are able to point out for which z the equality above holds, and which value $Q(z)$ takes at these points. Let us therefore examine the entries of $\big(\mathbf{D}(z)\big)^k\mathbf{e}$. For $k = 1$,

$$\Big[\mathbf{D}(z)\mathbf{e}\Big]_{(j_i)_{i\geqslant 1}} = B(z)\Big(C_1(z)^{j_1} C_2(z)^{j_2} \dots C_n(z)^{j_n} \dots\Big),$$

and for $k = 2$,

$$\Big[\big(\mathbf{D}(z)\big)^2\mathbf{e}\Big]_{(j_i)_{i\geqslant 1}} =$$
$$B(z)B\big(zC_1(z)\big)\Big[C_1\big(zC_2(z)\big)^{j_1} C_2\big(zC_3(z)\big)^{j_2} \dots C_n\big(zC_{n+1}(z)\big)^{j_n} \dots\Big].$$

Continuing this way it is clear that for $|z| \geqslant 1$:

$$\lim_{k\to\infty} \big(\mathbf{D}(z)\big)^k \mathbf{e} = \\ B(z)B\big(zC_1(z)\big)B\big[zC_1\big(zC_2(z)\big)\big]\dots B\Big[C_1\Big(zC_2\big(\dots zC_n(z)\big)\Big)\Big]\dots \mathbf{v}(z),$$

with the vector $\mathbf{v}(z)$ defined by

$$\big[\mathbf{v}(z)\big]_{(j_i)_{i\geqslant 1}} = \prod_{i=1}^{\infty} \chi_i(z)^{j_i},$$

and

$$\chi_j(z) = \frac{\sum_{k=j}^{\infty} a_k z^{k-j}}{1 - \sum_{k=1}^{j-1} a_k}.$$

The vector $\mathbf{v}(z)$ is the left eigenvector of $\mathbf{D}(z)$ with eigenvalue $B\big(z\chi_1(z)\big) = B\big(A(z)\big)$. To evaluate the expressions $C_1\Big(zC_2\big(\dots zC_{j-1}(z)\big)\Big)$ we introduce the sequence of random variables τ_{A_n} defined by

$$\mathbf{P}\{\tau_{A_n} = i\} = \begin{cases} \mathbf{P}\{\tau_A = i\} & \text{if } i < n, \\ \mathbf{P}\{\tau_A \geqslant n\} & \text{if } i = n, \\ 0 & \text{if } i > n. \end{cases}$$

The corresponding generating functions are denoted by $A_n(z)$, hence $A_1(z) = z$, $\dots$, $A_n(z) = a_1 z + a_2 z^2 + \dots + a_{n-1} z^{n-1} + (a_n + a_{n+1} + \dots) z^n$. It is clear that the sequence τ_{A_n} converges weakly to τ_A. Furthermore the A_n converge uniformly to A on $D(0,1)$, the closed unit disk of the complex plane. Hence also $B(A_n(z))$ converges uniformly to $B(A(z))$. Also note that

$$C_1\Big(zC_2\big(\dots zC_{j-1}(z)\big)\Big) = A_j(z).$$

Using this equality, (6.5) becomes

$$Q(z) = p_0(z-1)\sum_{i=1}^{\infty}\prod_{k=1}^{i} \frac{B(A_k(z))}{z}. \tag{6.6}$$

We now identify some open subset of $D(0,1)$ on which the r.h.s. of this equation converges. Since $B(A(-1)) < 1$, there exists some open set $G \subset D(0,1)$ and some $\gamma > 0$ such that for $z \in G$:

$$\left|\frac{B(A(z))}{z}\right| < 1 - 2\gamma.$$

By the uniform convergence of $B(A_n(z))$ to $B(A(z))$, it is clear that there exists some n_0 such that for $z \in G$ and $n \geqslant n_0$,

$$\left| \frac{B(A_n(z))}{z} \right| < 1 - \gamma.$$

Hence it follows that the r.h.s. of (6.6) converges for $z \in G$. The problem with the expression (6.6) is however that it diverges for z in the neighbourhood of 1. Indeed, for $z \in [0,1)$, $B(A_j(z)) > z$. Since we want to apply the Tauberian theorem for power series, we need an expression for Q which can be evaluated for z near 1. We solve this problem by deriving from (6.6) a formula for Q which is holomorphic on the open complex unit disk $U(0,1)$. This formula will be derived by a limit process. For each $n \geq 1$ and $z \in G$, (6.6) can be rewritten as

$$\begin{aligned} Q(z) &= p_0(z-1) \sum_{i=1}^{n} \prod_{k=1}^{i} \frac{B(A_k(z))}{z} \\ &\quad + p_0(z-1) \prod_{k=1}^{n} \frac{B(A_k(z))}{z} \sum_{i=1}^{\infty} \prod_{j=1}^{i} \frac{B(A_{n+j}(z))}{z}. \end{aligned}$$

Define for $z \in G$,

$$\begin{aligned} Q_n(z) &= p_0(z-1) \sum_{i=1}^{n} \prod_{k=1}^{i} \frac{B(A_k(z))}{z} \\ &\quad + p_0(z-1) \prod_{k=1}^{n} \frac{B(A_k(z))}{z} \sum_{i=1}^{\infty} \prod_{j=1}^{i} \frac{B(A_n(z))}{z} \\ &= p_0(z-1) \sum_{i=1}^{n} \prod_{k=1}^{i} \frac{B(A_k(z))}{z} + \prod_{k=1}^{n} \frac{B(A_k(z))}{z} V_n(z), \end{aligned} \tag{6.7}$$

with

$$V_n(z) = (z-1)p_0 \frac{B(A_n(z))}{z - B(A_n(z))}.$$

It is of interest to observe that the functions $Q_n(z)$ and $V_n(z)$ are also associated with queueing systems. Let $p_0^{(n)} = 1 - \lambda \mathbf{E}[\tau_{A_n}] = 1 - \lambda A_n'(1)$. The function $\left(p_0^{(n)}/p_0\right) Q_n(z)$ represents, on G, the generating function of the buffer distribution of the DBMAP-D-1 queue having as input the M/G/∞ process with distribution τ_{A_n}. The function $p_0^{(n)}/p_0 V_n(z)$ is the generating function of the buffer distribution of the GI-G-1 queue with a deterministic service time of 1 slot and with its arrivals distributed according to the random variable with generating function $B(A_n(z))$. For a detailed introduction to GI-G-1 queues we refer the reader to [13]. Clearly $V_n(z)$ is holomorphic on $U(0,1)$. Furthermore, one can see that the V_n converge uniformly on the compact subsets of $U(0,1)$ to V with

$$V(z) = p_0(z-1)\frac{B(A(z))}{z - B(A(z))}.$$

The function V is the generating function of the GI-G-1 queue with a deterministic service time of 1 slot and its arrivals distributed according to $B(A(z))$. This queueing system will be denoted by **QB** since the cells arrive in batches. Its buffer asymptotics will be derived in Section 6.2.3.

Let us return to Q. Notice that the Q_n converge uniformly to Q on G, a consequence of

$$Q(z) - Q_n(z) = \\ p_o(z-1)\prod_{k=1}^{n}\frac{B(A_k(z))}{z}\sum_{i=1}^{\infty}\left(\frac{B(A_n(z))}{z}\right)^i\left(1-\prod_{j=1}^{i}\frac{B(A_{n+j}(z))}{B(A_n(z))}\right),$$

and the inequality

$$\left|\sum_{j=1}^{i} A_{n+j}(z) - A_n(z)\right| \leqslant 2A'(1). \tag{6.8}$$

We proceed as follows. First the formula (6.7) for $Q_n(z)$ is rewritten, in order obtain an expression which does not diverge for z approaching 1 if $n \to \infty$. Then we will show that these new expressions for the $Q_n(z)$ converge uniformly on the compact subsets of $U(0,1)$ to a function $Q^\star(z)$. But since $Q^\star$ is equal to Q on G, it follows that $Q^\star = Q$ on $D(0,1)$. The actual derivation of this representation $Q^\star$ of Q is the subject of Theorem 6.2.7, which constitutes the main result of this section.

So let us transform the formula for $Q_n(z)$ given by (6.7). Using the equality

$$\frac{B(A_n(z))}{z - B(A_n(z))} = -\sum_{k=0}^{\infty}\frac{z^k}{B(A_n(z))^k},$$

which holds on some open disk contained in $D(0,1)$, it follows that

$$\begin{aligned} Q_n(z) = & p_0(z-1)\sum_{i=1}^{n-2}\left[\left(\prod_{k=1}^{i}B(A_k(z))\right)\frac{1}{z^i}\left(1-\prod_{k=i+1}^{n-1}\frac{B(A_k(z))}{B(A_n(z))}\right)\right] \\ & + \left[\prod_{k=1}^{n-1}\frac{B(A_k(z))}{B(A_n(z))}\right]V_n(z). \end{aligned} \tag{6.9}$$

The r.h.s. represents a holomorphic function on $U(0,1)$.

To keep the notation simple, we introduce the following definitions,

$$\begin{aligned}
\varphi_k^{(n)}(z) &= A_k(z) - A_n(z) + A_{k+1}(z) - A_n(z) + \ldots + A_{n-1}(z) - A_n(z) \\
&= \sum_{j=k}^{n-1} A_j(z) - A_n(z), \\
\varphi_k(z) &= A_k(z) - A(z) + A_{k+1}(z) - A(z) + \ldots = \sum_{j=k}^{\infty} \big[A_j(z) - A(z)\big], \\
\theta_k^{(n)}(z) &= \frac{B(A_k(z)) \ldots B(A_{n-1}(z))}{B(A_n(z))^{n-k}} = \exp\big[\lambda \varphi_k^{(n)}(z)\big], \\
\theta_k(z) &= \exp\big[\lambda \varphi_k(z)\big].
\end{aligned}$$

The formula for $Q_n(z)$ becomes, by making use of the definitions given above,

$$Q_n(z) = p_0^{(n)}(z-1) \sum_{i=1}^{n-2} \left[\left(\prod_{k=1}^{i} B(A_k(z)) \right) \frac{1}{z^i} \big(1 - \theta_{i+1}^{(n)}(z)\big) + \theta_1^{(n)}(z) V_n(z). \right.$$

We need the following lemmas before we are able to derive in Theorem 6.2.7 the representation of Q valid on $U(0,1)$.

Lemma 6.2.1. *For each $z \in D(0,1)$ the inequality $|B(A_j(z))| \leqslant \exp\big[\lambda(|z|-1)\big]$ holds.*

Proof. A straightforward calculation using the inequality $|A_j(z)| \leqslant A_j(|z|) \leqslant |z|$ on D(0, 1) leads to the result. □

Lemma 6.2.2. *The sequence $\dfrac{1}{z^{k-1}} \big[\varphi_k^{(n)}(z) - \varphi_k(z)\big]_n$ converges uniformly to 0 on $D(0,1)$, for each $k \geqslant 1$.*

Proof. For $z \in D(0,1)$ and $n > k$,

$$
\begin{aligned}
&\left|\frac{1}{z^{k-1}}\varphi_k(z) - \varphi_k^{(n)}(z)\right| \\
&\quad = \left|\frac{1}{z^{k-1}}(n-k)(A_n(z) - A(z)) + \frac{1}{z^{k-1}} \sum_{j=n+1}^{\infty} \left[A_j(z) - A(z)\right]\right| \\
&\quad = \left|(n-k)\left[\sum_{j=n+1}^{\infty} a_j\right] z^{n-k+1} - \sum_{j=n+1}^{\infty} a_k z^{j-k+1}\right. \\
&\qquad + \left[\sum_{j=n+2}^{\infty} a_j\right] z^{n-k+2} - \sum_{j=n+2}^{\infty} a_j z^{j-k+1} \\
&\qquad \left. + \left[\sum_{j=n+3}^{\infty} a_j\right] z^{n-k+3} - \sum_{j=n+3}^{\infty} a_j z^{j-k+1} + \dots \right| \\
&\quad \leqslant 2\left[(n+1) \sum_{j=n+1}^{\infty} a_j + \sum_{j=n+2}^{\infty} a_j + \sum_{j=n+3}^{\infty} a_j + \dots\right] \\
&\quad = 2 \sum_{j=n+1} j a_j .
\end{aligned}
$$

Hence the uniform convergence is a consequence of $A'(1) = \sum_{j=1}^{\infty} j a_j < \infty$. □

Corollary 6.2.3. *The sequence* $\left[\theta_k^{(n)}\right]_n$ *converges uniformly to* θ_k *on* $D(0,1)$*, for each* $k \geqslant 1$*.*

Lemma 6.2.4. *For each* $z \in D(0,1)$*, the inequalities*

$$\left|\frac{\varphi_k^{(n)}(z)}{z^{k-1}}\right| \leqslant 2 \sum_{m>k} m a_m \leqslant 2A'(1)$$

and

$$\left|\frac{\varphi_k(z)}{z^{k-1}}\right| \leqslant 2 \sum_{m>k} m a_m \leqslant 2A'(1)$$

hold for $1 \leqslant k \leqslant n-1$ *and* $1 \leqslant k < \infty$ *respectively.*

Proof. Analogous to the proof of Lemma 6.2.2. □

Lemma 6.2.5. *For each complex* z *the inequality* $\left|\exp(z) - 1\right| \leqslant |z| \exp(|z|)$ *holds.*

Proof. This is straightforward. □

Lemma 6.2.6. *The sequence* $\frac{1}{z^{k-1}}\left[\theta_k^{(n)}(z) - \theta_k(z)\right]_n$ *converges uniformly to 0 on* $D(0,1)$*, for each* $k \geqslant 1$*.*

Proof. This is a consequence of Lemma 6.2.5 and Lemma 6.2.2. □

Now we are able to proof the main result of this section.

Theorem 6.2.7. *The generating function Q is given on $U(0,1)$ by*

$$Q(z) = p_0(z-1)\sum_{i=1}^{\infty}\left[\left(\prod_{k=1}^{i} B\big(A_k(z)\big)\right)\frac{1}{z^i}\Big(1-\theta_{i+1}(z)\Big)\right]+\theta_1(z)V(z).$$

Proof. We show that the sequence Q_n converges uniformly to Q on the compact subsets of $U(0,1)$. The uniform convergence on compact subsets of $\theta_1^{(n)}V_n$ to $\theta_1 V$ is a direct consequence of the uniform convergence of $\theta_1^{(n)}$ to θ_1 on $D(0,1)$ (see Corollary 6.2.3), and of the uniform convergence on compact subsets of V_n to V. This observation allows us to restrict our attention to the convergence of

$$T_n(z) = p_0(z-1)\sum_{i=1}^{n-2}\left[\left(\prod_{k=1}^{i} B(A_k(z))\right)\frac{1}{z^i}\Big(1-\theta_{i+1}^{(n)}(z)\Big)\right]$$

to

$$T(z) = p_0(z-1)\sum_{i=1}^{\infty}\left[\left(\prod_{k=1}^{i} B(A_k(z))\right)\frac{1}{z^i}\Big(1-\theta_{i+1}(z)\Big)\right].$$

Let $\varepsilon > 0$. We will show the uniform convergence of T_n to T on the closed disk $D(0,\gamma)$ with radius $\gamma < 1$. Take $z \in D(0,\gamma)$. By Lemma 6.2.4 and Lemma 6.2.5 we have, for $i = 1,\dots,n-2$ and $1 \leqslant i < \infty$ respectively, that

$$\left|\frac{1}{z^i}(1-\theta_{i+1}^{(n)}(z))\right| \leqslant 2\lambda A'(1)e^{2\lambda A'(1))},$$

and

$$\left|\frac{1}{z^i}(1-\theta_{i+1}(z))\right| \leqslant 2\lambda A'(1)e^{2\lambda A'(1))}.$$

Hence, by choosing N large enough, one obtains, by using Lemma 6.2.1,

$$\left|p_0(z-1)\sum_{i=N}^{n-2}\left[\left(\prod_{k=1}^{i} B\big(A_k(z)\big)\right)\frac{1}{z^i}\Big(1-\theta_{i+1}^{(n)}(z)\Big)\right]\right| < \frac{\varepsilon}{4},$$

for each $n > N+2$ and

$$\left|p_0(z-1)\sum_{i=N}^{\infty}\left[\left(\prod_{k=1}^{i} B\big(A_k(z)\big)\right)\frac{1}{z^i}\Big(1-\theta_{i+1}(z)\Big)\right]\right| < \frac{\varepsilon}{4}.$$

Fix N. As a consequence of Lemma 6.2.6 there exists some $n_0 > N+2$ such that for $n \geqslant n_0$

$$\left| p_0(z-1)\sum_{i=1}^{N-1}\left[\left(\prod_{k=1}^{i} B\big(A_k(z)\big)\right)\frac{1}{z^i}\Big(\theta_{i+1}(z)-\theta_{i+1}^{(n)}(z)\Big)\right]\right| < \frac{\varepsilon}{2}.$$

Hence, for $z \in D(0,\gamma)$ and $n \geqslant n_0$, $|T(z)-T_n(z)| < \varepsilon$, which concludes the proof. $\square$

To simplify forthcoming calculations we rewrite once more the formula for Q. The following definitions are used:

$$\begin{aligned}
\varphi_k^\star(z) &= A_1(z)-A(z)+\ldots+A_k(z)-A(z),\\
\theta_k^\star(z) &= \exp\big[\lambda\varphi_k^\star(z)\big].
\end{aligned}$$

This enables us to obtain the following result:

$$\begin{aligned}
Q(z) &= p_0(z-1)\sum_{i=1}^{\infty}\Big[B(A(z))^i\theta_i^\star(z)\big(1-\theta_i(z)\big)\Big]+\theta_1(z)V(z)\\
&= p_0\frac{z-1}{1-\frac{B(A(z))}{z}}\left(1-\frac{B(A(z))}{z}\right)\sum_{i=1}^{\infty}B(A(z))^i\theta_i^\star(z)\big(1-\theta_i(z)\big)\\
&\quad+\theta_1(z)V(z)\\
&= p_0\frac{z-1}{z-B(A(z))}B(A(z))\theta_1^\star(z)\big(1-\theta_2(z)\big)\\
&\quad+p_0\frac{z-1}{1-\frac{B(A(z))}{z}}\times\\
&\quad\sum_{i=2}^{\infty}\frac{B(A(z))^i}{z^i}\big[\theta_i^\star(z)\big(1-\theta_{i+1}(z)\big)-\theta_{i-1}^\star(1-\theta_i(z))\big]\\
&\quad+\theta_1(z)V(z)\\
&= V(z)\theta_1^\star(z)+V(z)\sum_{i=1}^{\infty}\theta_i^\star(z)\frac{B(A(z))^i}{z^i}\big[\exp\big(A_{i+1}(z)-A(z)\big)-1\big].
\end{aligned} \tag{6.10}$$

By

$$A_{i+1}(z)-A(z) = z^i\left(\sum_{j=i+1}^{\infty}a_j-\sum_{j=i+1}^{\infty}a_jz^{j-i}\right),$$

it is clear that the last formula represents again a holomorphic function on $U(0,1)$, and so it is a valid representation of $Q(z)$.

6.2.3 The buffer asymptotics

From the last expression in the chain of equalities (6.10) it is possible to derive the buffer asymptotics. Recall that $Q(z)$ is the generating function associated with the stationary distribution q of the buffer occupancy of the system **QT**. Hence $Q(z) = \sum_{k=0}^{\infty} q_k z^k$ with $q_k = \mathbf{P}\{q = k\}$. Note that, using the formula

$$\mathbf{E}[q] = B'(1)\left[A'(1) - \frac{A''(1)}{2}\right] + \frac{B''(1)\left[A'(1)\right]^2 + B'(1)A''(1)}{2(1-\rho)},$$

derived in [62], one has

$$\lim_{z \to 1-} Q'(z) = \infty.$$

Identifying the way in which $Q'(z)$ diverges when $z \to 1-$, is the key step in determining the asymptotic behaviour of the buffer distribution.

Lemma 6.2.8. *The behaviour of $Q'(z)$ is given by*

$$Q'(z) \sim \frac{1}{(1-z)^{3-s}} \frac{\lambda a \Gamma(3-s+1)\rho^{s-2}}{(1-\rho)(s-1)(3-s)} \quad \text{for } z \to 1-$$

Proof. See Appendix 6.5. □

By Lemma 6.2.8 we are able to derive the asymptotic behaviour of the tail probabilities.

Theorem 6.2.9. *The asymptotic behaviour of the tail probabilities of the stationary distribution of the buffer occupancy of the queueing system* **QT** *is given by*

$$\mathbf{P}\{q > k\} \sim \frac{\lambda a \rho^{s-2}}{(s-2)(s-1)(1-\rho)} k^{2-s}. \tag{6.11}$$

Proof. Lemma 6.2.8 and the Tauberian theorem for power series 4.3.1 leads to

$$\sum_{j=1}^{k} j q_j \sim \frac{\lambda a \rho^{s-2}}{(1-\rho)(s-1)(3-s)} k^{3-s}.$$

An application of [43, 3.3 (c), pg. 59] concludes the proof. □

From Theorem 6.2.9 we conclude that the buffer distribution has a power-law decay with exponent $2-s$, where s the parameter of the Pareto-like distribution describing the train lengths. Besides the slope of the decay, we also obtain the constant which determines the asymptotic behaviour completely. As indicated before this asymptotic behaviour was postulated, for the related fluid flow queue, in [34] and [35].

In the same way it is possible to determine the asymptotic behaviour of the GI-D-1 queueing system **QB**. Here the stationary distribution of the buffer occupancy is denoted by v. The generating function $V(z) = \sum_{k=0}^{\infty} v_k z^k$ of v is given by (6.8).

Lemma 6.2.10. . *The behaviour of $V'(z)$ is given by*

$$V'(z) \sim \frac{1}{(1-z)^{3-s}} \frac{\lambda a \Gamma(3-s+1)}{(1-\rho)(s-1)(3-s)} \text{ for } z \to 1-.$$

Proof. See Appendix 6.5. □

Theorem 6.2.11. *The asymptotic behaviour of the tail probabilities of the stationary distribution of the buffer occupancy of the queueing system **QB** is given by*

$$\sum_{j>k} v_j \sim \frac{\lambda a}{(s-2)(s-1)(1-\rho)} k^{2-s}.$$

Proof. Analogous to the proof of Theorem 6.2.9. □

Remark

Theorem 6.2.9 can also be stated in a form which more or less hides that formula (6.11) is derived for a M/G/∞ arrival process. Equation (6.3) describing the correlation structure of the M/G/∞ process b_k with a Pareto-like distribution can be rewritten as

$$\mathrm{Cov}(b_k, b_{k+j}) \sim \gamma k^{-\sigma}$$

with

$$\gamma = \frac{a\lambda}{(s-2)(s-1)},$$

and $\sigma = s - 2$. With these definitions formula (6.11) for the tail probabilities becomes

$$\mathbf{P}\{q > k\} \sim \gamma \frac{\rho^{\sigma}}{1-\rho} k^{-\sigma}. \tag{6.12}$$

It turns out that the formulas obtained for the superpositions studied in the Section 6.2.4 are covered by (6.12). Simulation results presented in Section 6.3 indicate that (6.12) also applies to the queueing system with as input the process Y^{∞} studied in Chapter 5.

6.2.4 Superposition of an LRD M/G/∞ and an SRD arrival processes

In this section we study what happens with the behaviour of the tail probabilities if we feed the superposition of an LRD M/G/∞ and an SRD arrival process to the multiplexer queue. First we focus on the case that the second arrival process is also a discrete-time M/G/∞ process. Later on we generalise the result obtained for this superposition to the case with the SRD traffic stream modelled by a DBMAP with a finite state space.

Superposition of an LRD and an SRD M/G/∞ process

Observe that the superposition of two M/G/∞ processes, with train arrival rates λ_i, $i = 1, 2$ and train length distributions τ_{A_i}, $i = 1, 2$, is again a M/G/∞ process with $\lambda = \lambda_1 + \lambda_2$ and τ_A determined by $A(z) = \lambda_1/(\lambda_1 + \lambda_2)A_1(z) + \lambda_2/(\lambda_1 + \lambda_2)A_2(z)$. The arrival rates of the individual processes are denoted by ρ_i, hence $\rho_i = \lambda_i A_i'(1)$. For stability it is assumed that the total arrival rate $\rho = \rho_1 + \rho_2$ is less than 1.

Consider an arbitrary distribution with finite variance and generating function $A_1(z)$. It is clear from Section 6.1.2 that the M/G/∞ process associated with this distribution is SRD. The LRD traffic stream is modelled by an M/G/∞ process constructed from λ_2 and τ_{A_2} with $\mathbf{P}\{\tau_{A_2} = k\} \sim ak^{-s}$ with $s < 2 < 3$ and $a > 0$. It is clear that

$$\mathbf{P}\{\tau_A = k\} \sim \frac{\lambda_2}{\lambda_1 + \lambda_2} ak^{-s}.$$

Hence from Theorem 6.2.9 it follows that

$$\mathbf{P}\{q > k\} \sim \frac{\lambda_2 a(\rho_1 + \rho_2)^{s-2}}{(s-2)(s-1)\big(1 - (\rho_1 + \rho_2)\big)} k^{2-s}, \tag{6.13}$$

with q representing the stationary buffer distribution of the DBMAP-D-1 queue with as input the superposition. The contribution of the SRD arrivals is only reflected in the total load. Note that (6.13) can also be obtained from (6.3). This result can be easily generalised to the case with both M/G/∞ processes LRD.

Superposition of two LRD M/G/∞ processes

Here both the train length distributions are Pareto-like: $\mathbf{P}\{\tau_{A_i} = k\} \sim a_i k^{-s_i}$, $i = 1, 2$. Now

$$\mathbf{P}\{\tau_A = k\} \sim \frac{\lambda}{\lambda_1 + \lambda_2} ak^{-s},$$

with $s = \min(s_1, s_2)$, a and λ the a_i and the λ_i corresponding to this s. Again (6.12) describes the asymptotic behaviour. As expected the “worst” behaving traffic stream determines the decay of the tail probabilities.

Superposition of an LRD M/G/∞ process and an SRD DBMAP

Consider an SRD traffic stream modelled by the DBMAP $(\mathbf{D}^*, \mathbf{D}_l^*)$. We assume that this DBMAP has a finite state space and that its z-transform $\mathbf{D}^*(z)$ is diagonisable in a neighbourhood of 1. As in Section 6.2.2, the DBMAP modelling the LRD discrete-time M/G/∞ process is denoted by $(\mathbf{D}, \mathbf{D}_l)$. The DBMAP representing the superposition is denoted by $(\mathbf{D}^S, \mathbf{D}_l^S)$, hence $\mathbf{D}^S(z) = \mathbf{D}^*(z) \otimes \mathbf{D}(z)$. It is assumed that the total arrival rate $\rho^S = \rho^* + \rho < 1$. The stationary buffer distribution of the DBMAP-D-1 queue with input $(\mathbf{D}^S, \mathbf{D}_l^S)$ will be denoted by q, its generating function by $Q(z)$.

We start again with the rewritten Pollachek-Kinchin equation:

$$Q(z) = \mathbf{X}^S(z)\mathbf{e} = \mathbf{x}_0^S(z-1)\sum_{k=1}^{\infty} \frac{\left(\mathbf{D}^S(z)\right)^k}{z^k}.$$

As before this expression is only considered for those z for which the r.h.s. converges. Note that $\mathbf{x}_0^S = (1-\rho^S)\mathbf{y} \otimes \mathbf{g}$ for some $\mathbf{y}$, with $\mathbf{ye} = 1$, and with $(1-\rho)\mathbf{g} = \mathbf{x}_0$. The vectors $\mathbf{g}$ and $\mathbf{x}_0$ are associated with the DBMAP-D-1 queue having $(\mathbf{D}, \mathbf{D}_l)$ as input. The special form of $\mathbf{x}_0^S$ is caused by the fact that $\mathbf{g}$ has only one non zero entry. Consider the spectral decomposition of $\mathbf{D}^*(z)$:

$$\mathbf{D}^*(z) = \sum_{j=1}^{m} \lambda_j(z)\mathbf{B}_j(z),$$

with $\lambda_1(z)$ representing the Perron-Frobenius eigenvalue for z real and positive. Hence $Q(z)$ is given by

$$Q(z) = (1-\rho^S)(z-1)\sum_{j=1}^{m}\sum_{k=1}^{\infty}\Big(\lambda_j(z)^k \mathbf{y}\mathbf{B}_j(z)\mathbf{e}\Big)\left(\mathbf{g}\frac{\left(\mathbf{D}(z)\right)^k}{z^k}\mathbf{e}\right). \tag{6.14}$$

Since $\lambda_1(1) = 1$ and $\lambda_j(1) < 1$ for $j \geqslant 2$, we examine the contributions for $j = 1$ and $j \geqslant 2$ to the r.h.s. of (6.14) separately. Let us start with $j \geqslant 2$. Define $c_j(z) = \mathbf{y}\mathbf{B}_j(z)\mathbf{e}$. Because $\lambda_j(1) < 1$, the expression

$$(1-\rho^S)(z-1)c_j(z)\sum_{i=1}^{\infty}\prod_{k=1}^{i}\frac{B(A_k(z))\lambda_j(z)}{z}$$

converges in a neighbourhood of 1. At the same time the derivative to z of this expression is bounded in a neighbourhood of 1. Hence, by the techniques of the proof of Theorem 6.2.9, it is clear that the terms with $j \neq 1$ do not contribute to the asymptotic formula for the tail probabilities. Let us therefore focus on $j = 1$. First of all note that

$$\frac{d}{dz}\lambda_1(z)B(A(z))\Big|_{z=1} = \rho^* + \rho = \rho^S < 1. \tag{6.15}$$

This observation allows us to repeat the calculations of Section 6.2.2. The term with $j = 1$ in (6.14) equals

$$(1-\rho^S)(z-1)c_1(z)\sum_{i=1}^{\infty}\left[\left(\prod_{k=1}^{i} B(A_k(z))\right)\left(\frac{\lambda_1(z)}{z}\right)^i\left(1-\theta_{i+1}(z)\right)\right]+ \\ c_1(z)\theta_1(z)(1-\rho^S)\frac{\lambda_1(z)B(A(z))}{z-\lambda_1(z)B(A(z))}.$$

Since (6.11) describing the tail probabilities of the system **QT** is essentially obtained from (6.20) and (6.21), which can be found in the proof of Theorem 6.2.9, we translate (6.21) to the case considered here:

$$\lim_{z\to 1-}(1-z)^{3-s}\lambda\left[\sum_{k=1}^{\infty}\left(\sum_{j=k+1}^{\infty} a_k\right)\left(\lambda_1(z)B(A(z))\right)^k\right]' \\ = \frac{\lambda a\left(\rho^*+\rho\right)^{s-2}\Gamma(3-s+1)}{(s-1)(3-s)}.$$

Note furthermore that $c_1(1) = \mathbf{y}\mathbf{e}\boldsymbol{\pi}^*\mathbf{e} = 1$. Hence we conclude that

$$\mathbf{P}\{q > k\} \sim \frac{\lambda a(\rho^*+\rho)^{s-2}}{(s-2)(s-1)\left(1-(\rho^*+\rho)\right)}.$$

6.3 Simulation Results

Simulation results concerning three different queueing systems are presented. The first system fits within the scope of Theorem 6.2.9, the other two point out possible generalisations for this theorem.

Simulation of a Pareto queue with $2 < s < 3$

This first numerical example is a simulation of the so-called Pareto queue, which is an instance of the system **QT** with

$$\mathbf{P}\{\tau_A = j\} = j^{1-s} - (j+1)^{1-s}.$$

Hence $a = s - 1$ and the mean arrival rate is $\rho = \lambda\sum_{n=1}^{\infty} n^{-s}$. The first case shown in Figure 6.1 has $s = 2.8$ and $\lambda = 0.4$, resulting in a load of about 0.75. For the second one $s = 2.5$ and $\lambda = 0.2$, resulting in a load of 0.52. The approximation is based on (6.11). Although this formula is only asymptotically valid, it seems to be a good approximation over the whole range.

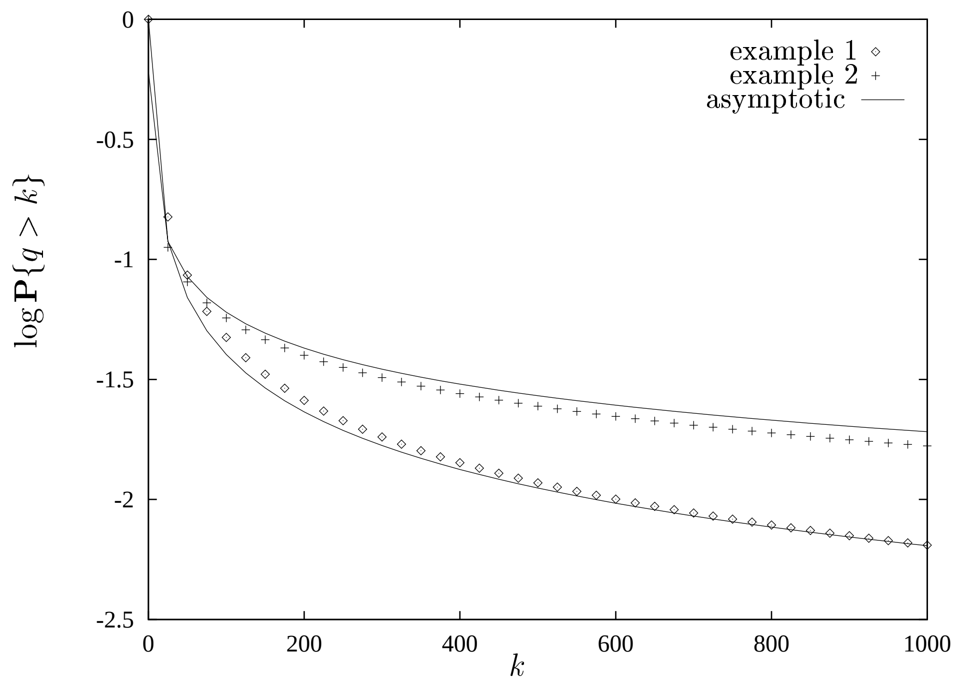

Figure 6.1: Pareto queue with $2 < s < 3$

Simulation of a Pareto queue with $s < 3$

Although Theorem 6.2.9 is proved for $2 < s < 3$, formula (6.11) also seems valid for $s > 3$. This is illustrated by the two cases shown in Figure 6.2. Here s is respectively 3.2 and 3.8, λ is respectively 0.6 and 0.5. Both cases have a load of 0.75.

Process Y^∞ of Chapter 5 revisited

Formula (6.12) seems also to be applicable to the DBMAP-D-1 queue with as input the process Y^∞ studied in Chapter 5. Let us first demonstrate how the M/G/∞ process and Y^∞ can be related.

Assume that $p = 1$. An active period of the source $X^{(i)}$ corresponds to the generation of a train of back-to-back cells. These trains are generated by the source $X^{(i)}$ at a rate of

$$\lambda^{(i)} = \frac{1}{\mathbf{E}[\tau_{\text{on}}^{(i)}] + \mathbf{E}[\tau_{\text{off}}^{(i)}]} = \frac{b^i}{a^i(1+b^i)},$$

with $\tau_{\text{on}}^{(i)}$ and $\tau_{\text{off}}^{(i)}$ corresponding to the on- and off-periods of $X^{(i)}$. Let $\lambda =$

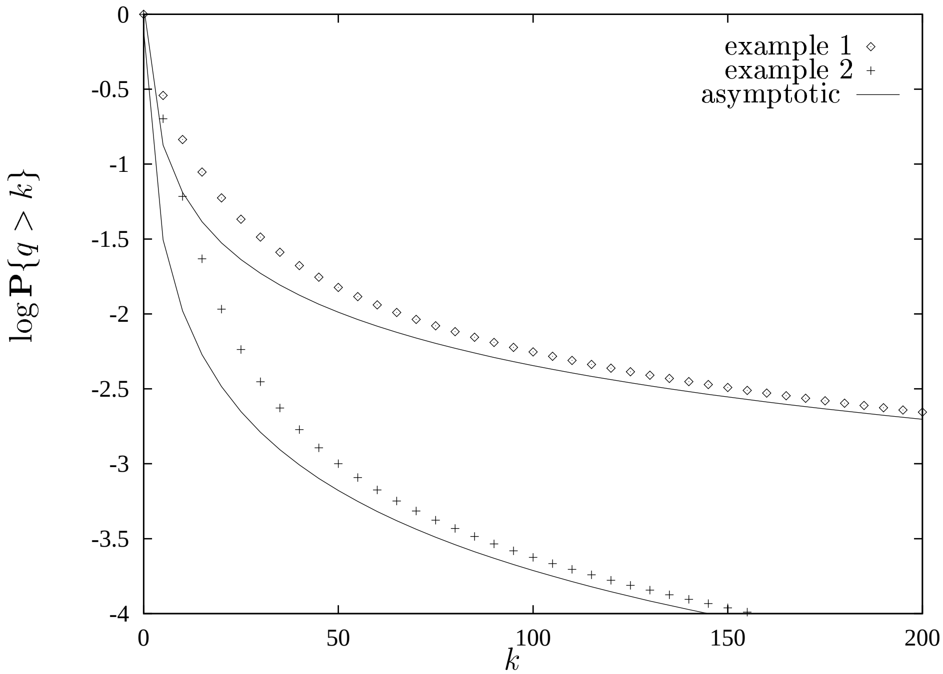

Figure 6.2: Pareto queue with $s > 3$

$\sum_{i=1}^{\infty} \lambda^{(i)}$. If X denotes the distribution of the Y^∞ train lengths then

$$\begin{aligned}
\mathbf{P}\{X = k\} &= \sum_{i=1}^{\infty} \mathbf{P}\{X = k | X \text{ is generated by source } i\} \\
&= \sum_{i=1}^{\infty} \left(\frac{b}{a}\right)^i \left[1 - \left(\frac{b}{a}\right)^i\right]^k \frac{\lambda^{(i)}}{\lambda}.
\end{aligned}$$

With X playing the role of τ_A and train arrival rate λ the associated M/G/∞ process has the same mean arrival rate as Y^∞:

$$\begin{aligned}
\lambda \mathbf{E}[X] &= \lambda \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} k \left[1 - \left(\frac{b}{a}\right)^i\right]^k \frac{\lambda^{(i)}}{\lambda} \\
&= \sum_{i=1}^{\infty} \frac{1}{1 + b^i}.
\end{aligned}$$

The outline presented above supports the application of (6.12) to the Y^∞-queue. Clearly $\sigma = \beta = \log b/(\log a - \log b)$. The value of γ has to be determined numerically. The first example shown in Figure 6.3 has $a = 5$, $b = 2$; for the second example $a = 8$, $b = 2$. In both cases the load equals 0.76.

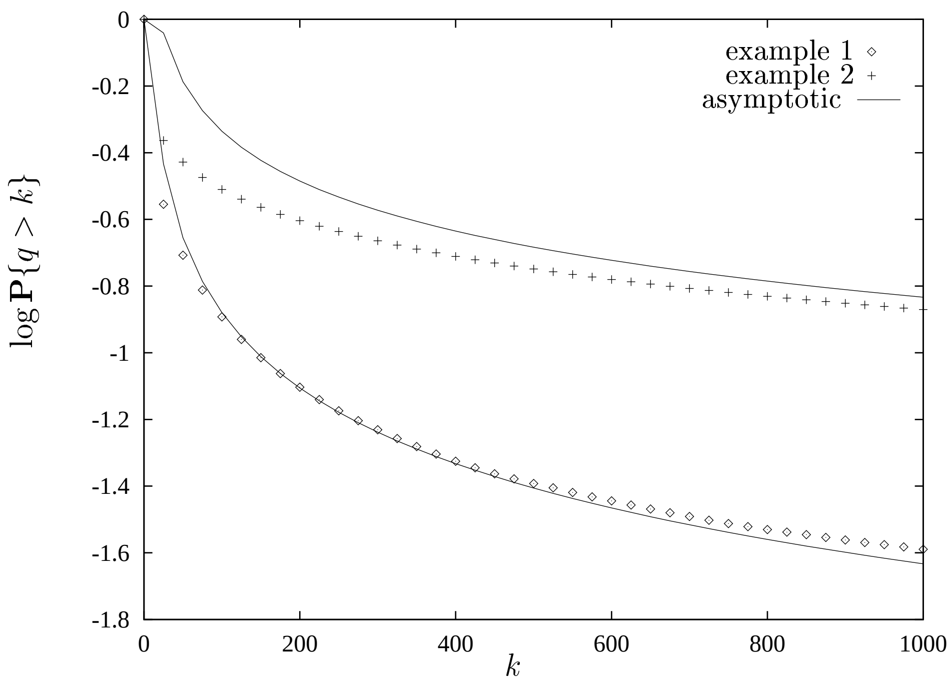

Figure 6.3: Simulation of the Y^∞-queue

6.4 Discussion and Open Problems

From a practical point of view the studied queueing system **QT** has a rather limited scope. In this discussion we first identify the two main factors limiting the applicability of this model. Then we indicate some possibilities to adapt the system **QT**, in order to broaden its scope. The consequences for the buffer asymptotics are also touched upon.

6.4.1 Factors limiting the applicability of QT

The two main factors limiting the scope of **QT** are:

1. the cells arrive back-to-back,
2. the service rate is limited to 1.

In Section 6.1.2 it was indicated that although the M/G/∞ process can be used as a first approximation for some important real traffic streams, it should, in order to achieve higher accuracy, also incorporate the modelling of the small-time scale behaviour. Having the cells arriving back-to-back clearly does not fit within this picture. Therefore an adaptation of the discrete-time LRD M/G/∞ process, introducing a — yet simple — stochastic substructure, is presented in the next section. Although it is only a slight modification, we demonstrate that it could bear a large influence on the tail probabilities.

The second limiting factor can be resolved by allowing c servers, with $c > 1$. Doing so we end up with a DBMAP-D-c queue, to which the techniques introduced in this chapter can be applied. By referring to results obtained by large deviation techniques, we point out which behaviour can be expected for the tail probabilities.

6.4.2 Thinning the M/G/∞ process

Instead of having a train generating its stream of cells back-to-back, cells are generated only with probability p during an activity period. For this class of arrival processes the mean arrival rate becomes $\lambda p\mathbf{E}[\tau_A]$.

The influence of this modification on the tail probabilities is demonstrated by considering a sequence of such processes $\mathbf{A}^N$. Consider some λ and τ_A. For the process $\mathbf{A}^N$ we let $p = 1/N$ and we take $N\lambda$ as the train arrival rate. Hence the mean arrival rate stays the same but the autocovariance ultimately vanishes when $N \to \infty$, which can be verified by the formulas presented in Section 2.3.1.

The queues having as input these processes $\mathbf{A}^N$ converge to the GI-D-1 queue, with arrivals distributed according to a Poisson distribution with mean λ. Of course this queue exhibits exponentially decaying tail probabilities. Since for $N = 1$ one has the Pareto-like tail probabilities of formula (6.11), a tail transition should occur one way or another. Attacking this problem by the techniques used introduced in this chapter is an open problem. One of the difficulties to overcome is the fact that there is no closed form expression for the vector $\mathbf{x}_0$ when $N > 1$.

6.4.3 Increasing the number of servers

Consider the DBMAP-D-c queue with $c > 1$ and having the discrete-time LRD M/G/∞ process as input. As shown in Section 3.4.1 we have the Pollachek-Kinchin equation

$$\mathbf{X}(z) = \sum_{j=0}^{c-1} \mathbf{x}_j \mathbf{D}(z)(z^c - z^j)\big(z^c\mathbf{I} - \mathbf{D}(z)\big)^{-1},$$

to analyse the asymptotic behaviour of the tail probabilities. As in Section 6.4.2, the fact that there are no closed form formulas for the vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{c-1}$ complicates the computations, but it may be possible to make a qualitative analysis like the one in Section 4.3.3. Preliminary attempts indicated that this problem is far from trivial.

As shown in the literature a change or transition in the asymptotic behaviour may occur if the number of servers increases. We clarify this by recalling several bounds obtained by large deviation techniques. They are written down in the notation of this chapter. It is assumed that $c - \lambda\mathbf{E}[\tau_A] > 0$, otherwise the stationary buffer distribution q would not exist.

In [41] the authors obtained the lower bound

$$\liminf_{k\to\infty} \frac{\log \mathbf{P}\{q > k\}}{\log k} \geqslant -(s-2)(\lfloor c - \lambda \mathbf{E}[\tau_A]\rfloor + 1), \tag{6.16}$$

with $\lfloor x \rfloor$ denoting the integer part of x. Furthermore they proved the upper bound

$$\limsup_{k\to\infty} \frac{\log \mathbf{P}\{q > k\}}{\log k} \leqslant -(s-2). \tag{6.17}$$

In [22] another upper bound is stated:

$$\limsup_{k\to\infty} \frac{\log \mathbf{P}\{q > k\}}{\log k} \leqslant 1 - (s-2)(c - \lambda \mathbf{E}[\tau_A]). \tag{6.18}$$

As indicated in [41] the upper bound (6.18) gets below the upper bound (6.17) if

$$c - \lambda \mathbf{E}[\tau_A] > \frac{s-1}{s-2}.$$

Hence when $c - \lambda \mathbf{E}[\tau_A]$ becomes large enough, the nature of the decay of the tail probabilities changes. When $c = 1$ this clearly never happens since $2 < s < 3$, which is in correspondence with the results of Section 6.2.3.

6.5 Appendix

We first give the proof of Lemma 6.2.10, because this lemma will be used in the proof of Lemma 6.2.8.

Proof of lemma 6.2.10. Differentiating $V(z)$ gives

$$p_0 \frac{-\left[1 - B(A(z))\right]^2 + 1 - B(A(z)) - (1-z)zB(A(z))\lambda A'(z)}{\left[z - B(A(z))\right]^2}.$$

By using $B(A(z)) = \sum_{k=0}^{\infty} [\lambda(A(z)-1)]^k / k!$ we obtain

$$\lim_{z\to 1-} (1-z)^{s-3} V'(z) = \lim_{z\to 1-} (1-z)^{s-3} p_0 \lambda \frac{1 - A(z) - (1-z)A'(z)}{\left[z - B(A(z))\right]^2}.$$

Furthermore,

$$\begin{aligned} 1 - A(z) - (1-z)A'(z) &= (1-z)\left[\sum_{k=1}^{\infty} a_k \sum_{j=0}^{k-1} z^j - \sum_{k=1}^{\infty} k a_k z^{k-1}\right] \\ &= (1-z)^2 \sum_{k=1}^{\infty} k \left[\sum_{j=k+1}^{\infty} a_j\right] z^{j-1}. \end{aligned}$$

Since for $k \to \infty$,

$$k \sum_{j=k+1}^{\infty} a_j \sim \frac{c}{s-1} k^{2-s},$$

and

$$\lim_{z \to 1-} \frac{(1-z)^2}{\big[z - B(A(z))\big]^2} = \frac{1}{p_0^2},$$

we obtain by applying the Theorem 4.3.1

$$\lim_{z \to 1-} (1-z)^{s-3} V'(z) = \frac{\lambda a \Gamma(3-s+1)}{(1-\rho)(s-1)(3-s)}.$$

This last equality concludes the proof. □

Proof of lemma 6.2.8. To keep the proof readable, the expression (6.10) for Q is repeated:

$$\begin{aligned} Q(z) &= V(z) \exp\big[\lambda\big(z - A(z)\big)\big] \\ &\quad + V(z) \sum_{i=1}^{\infty} \theta_i^{\star}(z) \frac{B(A(z))^i}{z^i} \big[\exp\big(A_{i+1}(z) - A(z)\big) - 1\big]. \end{aligned} \tag{6.19}$$

Furthermore we introduce the following definitions:

$$\begin{aligned} \Psi_i(z) &= B(A(z))^i \theta_i^{\star}(z) \frac{1}{z^i} \big[\exp\big(A_{i+1}(z) - A(z)\big) - 1\big], \\ \Psi(z) &= \sum_{i=1}^{\infty} \Psi_i(z). \end{aligned}$$

Hence, we can rewrite (6.19) as

$$Q(z) = V(z) \exp\big[\lambda\big(z - A(z)\big)\big] + V(z)\Psi(z)$$

and

$$\begin{aligned} Q'(z) &= V'(z) \exp\big[\lambda\big(z - A(z)\big)\big] + V(z) \frac{d}{dz} \exp\big[\lambda\big(z - A(z)\big)\big] \\ &\quad + V'(z)\Psi(z) + V(z)\Psi'(z). \end{aligned}$$

First of all we study the behaviour of $\Psi(z)$. Each term Ψ_i is non-negative on $[0,1]$ since $A_j(z) - A(z) \geqslant 0$ for $j \geqslant 1$ and for $z \in [0,1]$. Furthermore, by Lemma 6.2.5,

$$\begin{aligned} \Psi_i(z) &\leqslant \frac{1}{z^i} \lambda\big(A_{i+1}(z) - A(z)\big) \exp\big[A_{i+1}(z) - A(z)\big] \\ &\leqslant \frac{1}{z^i} \lambda\big(A_{i+1}(z) - A(z)\big) e^{\lambda} \\ &= \left[\left(\sum_{k=n+2}^{\infty} a_k\right) z - \sum_{k=n+2}^{\infty} a_k z^{k-n}\right] e^{\lambda}. \end{aligned}$$

This results in

$$0 \leqslant \Psi(z) \leqslant e^{\lambda}\left[\left(\sum_{k=3}^{\infty}(k-2)a_k\right)z - \sum_{k=3}^{\infty}(k-2)a_k z^k\right].$$

Hence,

$$\lim_{z\to 1-} \Psi(z) = 0.$$

Using this result, together with similar observations, we obtain

$$\lim_{z\to 1-}(1-z)^{3-s}Q'(z) = \lim_{z\to 1-}(1-z)^{3-s}V'(z) + \lim_{z\to 1-}(1-z)^{3-s}\Psi'(z).$$

Furthermore,

$$\Psi'(z) = \sum_{i=1}^{\infty}\Psi_i'(z),$$

because of the uniform convergence of $\left[\sum_{i=1}^{k}\Psi_i\right]_k$ to Ψ on compact subsets of $U(0,1)$. We will now focus on the behaviour of $\Psi'(z)$ since we already know the behaviour of $V'(z)$. Define:

$$\xi_i(z) = \frac{B(A(z))^i}{z^i}\lambda\big(A_{i+1}(z) - A(z)\big),$$
$$\omega_i(z) = 1 + \sum_{k=1}^{\infty}\frac{\big[\lambda\big(A_{i+1}(z) - A(z)\big)\big]^k}{(k+1)!}.$$

Hence,

$$\Psi(z) = \sum_{i=1}^{\infty}\theta_i^{\star}(z)\xi_i(z)\omega_i(z),$$

and

$$\Psi'(z) = \sum_{i=1}^{\infty}\big[\theta_i^{\star}(z)\big]'\xi_i(z)\omega_i(z) + \sum_{i=1}^{\infty}\theta_i^{\star}(z)\xi_i'(z)\omega_i(z) + \sum_{i=1}^{\infty}\theta_i^{\star}(z)\xi_i(z)\omega_i'(z).$$

The inequalities

$$\left[\frac{d}{dz}\theta_i^{\star}(z)\right]^+ \leqslant \lambda\theta_i^{\star}(z)\sum_{k=2}^{\infty}(k-1)\left(\sum_{j=k}^{\infty}a_j\right)z^{k-2}$$

and

$$\left[\frac{d}{dz}\theta_i^\star(z)\right]^- \leqslant \lambda\theta_i^\star(z)\sum_{k=2}^{\infty} k a_k z^{k-1},$$

where $[f(z)]^+ = \max\{0, f(z)\}$ and $[f(z)]^- = -\min\{0, f(z)\}$, imply

$$\lim_{z\to 1-}(1-z)^{3-s}\sum_{i=1}^{\infty}\big[\theta_i^\star(z)\big]'\xi_i(z)\omega_i(z) = 0.$$

Making use of similar techniques it also follows that

$$\lim_{z\to 1-}(1-z)^{3-s}\sum_{i=1}^{\infty}\theta_i^\star(z)\xi_i(z)\omega_i'(z) = 0.$$

Proceeding in the same way one ultimately obtains

$$\lim_{z\to 1-}(1-z)^{3-s}\Psi'(z) = \lim_{z\to 1-}(1-z)^{3-s}\left[\frac{1}{z}\sum_{i=1}^{\infty}\xi_i(z)\right]'.$$

Observe that $\frac{1}{z}\sum_{i=1}^{\infty}\xi_i(z)$ can be rewritten in a much more tractable way:

$$\frac{1}{z}\sum_{i=1}\xi_i(z) = \lambda\left[\sum_{k=1}^{\infty}\left(\sum_{j=k+1}^{\infty} a_k\right)B(A(z))^k + \frac{A(z)-1}{z} + a_1 - \frac{A[B(A(z))]-A(z)}{B(A(z))-z}\right]. \tag{6.20}$$

First of all note that

$$\lim_{z\to 1-}(1-z)^{3-s}\lambda\left[\sum_{k=1}^{\infty}\left(\sum_{j=k+1}^{\infty} a_k\right)B(A(z))^k\right]' = \frac{\lambda a\rho^{s-2}\Gamma(3-s+1)}{(s-1)(3-s)}. \tag{6.21}$$

Furthermore

$$\frac{A\big[B(A(z))\big]-A(z)}{B(A(z))-z} = \frac{A\big[B(A(z))\big]-1}{B(A(z))-z} + \frac{1-A(z)}{B(A(z))-z}$$

and

$$\frac{A\big[B(A(z))\big]-1}{B(A(z))-z} = v(z)t(z),$$

with

$$v(z) = \frac{1 - B(A(z))}{B(A(z)) - z}$$

and

$$t(z) = -\sum_{k=0}^{\infty} \left(\sum_{j>k} a_j \right) B(A(z))^k.$$

Clearly $t(1) = -A'(1)$ and

$$\lim_{z \to 1-} v(z) = \frac{\rho}{1-\rho}.$$

Furthermore,

$$\frac{d}{dz} v(z) = \frac{1 - B(A(z)) - (1-z)\big[B(A(z))\big]'}{\big[B(A(z)) - z\big]^2},$$

and since

$$\frac{1 - A(z)}{B(A(z)) - z} = v(z) \frac{1 - A(z)}{1 - B(A(z))} = v(z) \frac{1}{\lambda + \frac{\lambda^2}{2!}\big(A(z) - 1\big) + \ldots},$$

it follows that

$$\begin{aligned}
&\lim_{z \to 1-} (1-z)^{3-s} \frac{d}{dz} \left[\lambda \frac{A[B(A(z))] - A(z)}{B(A(z)) - z} \right] \\
&\qquad = \frac{\lambda a \rho^{s-2} \Gamma(3-s+1)}{(s-1)(3-s)} + \frac{\rho}{1-\rho} \frac{\lambda a \rho^{s-2} \Gamma(3-s+1)}{(s-1)(3-s)} \\
&\qquad\quad - \frac{\lambda a \Gamma(3-s+1)}{(1-\rho)(s-1)(3-s)}.
\end{aligned}$$

By taking a look at (6.5) one can conclude that

$$\lim_{z \to 1-} (1-z)^{3-s} Q'(z) = \frac{\lambda a \rho^{s-2} \Gamma(3-s+1)}{(1-\rho)(s-1)(3-s)},$$

which finishes the proof. $\square$

Conclusion

Delivering QoS guarantees together with an efficient use of network resources is one of the aims when designing today's multiservice communication networks. Mathematical contributions to the solution of this problem typically belong to one of the following domains:

1. the construction of arrival processes accurately modelling network traffic,
2. the development and study of methods to solve the queueing systems which model the various network elements,
3. the study of specific queueing systems to obtain applicable results.

The study presented in this thesis mainly contributes to the last two domains.

The demonstration in Chapter 3 of the equivalence of the FEA and the MAA can be identified as a first contribution, mainly because the examination of this equivalence provides new insights from which both methods benefit. Among other things, the treatment of the boundary probabilities, presented at the end of Chapter 3, is of interest here. Furthermore it should be noted that throughout this thesis both methods are used as complementary to one another.

A second main contribution concerns tail transitions. This study focuses primarily on the application of new techniques, rather than on analysing queueing systems of high practical relevance. The application of Darboux's theorem to the queues in tandem reveals a complex asymptotic behaviour, which cannot be captured by the application of large deviation techniques as presented in e.g. [23]. This illustrates the power of the careful analysis of the generating functions associated with a queueing system.

Although the model studied in Section 4.3 is straightforward, the results concerning the multi-server queue having as input the superposition of an LRD and an SRD traffic stream put the practical importance of LRD into a wider perspective. It is shown that although the total arrival process is LRD, the tail probabilities may still decay exponentially. With respect to the design of buffers, this a much better situation than with Weibullian or power-law tails, which are usually associated with LRD input.

The third contribution of this thesis is found in both Chapter 5 and Chapter 6, where the aim is to determine, as accurately as possible, the tail probabilities for specific types of queueing systems. The invoked techniques differ to a large extent. By its nature the system studied in Chapter 5 has to be analysed by large deviation techniques, resulting in a rather coarse lower bound for the tail probabilities. The analysis of the queue with LRD M/G/∞ input is much more successful, mainly because it is possible to obtain an analytically tractable expression for the generating function associated with the stationary buffer distribution. By a thorough analysis of this generating function, which involves the extensive use of the Tauberian theorem for power series, the exact asymptotic behaviour could be obtained. Unfortunately the studied model is rather limited with respect to practical applications. It is however indicated that the techniques developed in Chapter 6 are also applicable to extensions of this model which are more useful in practice.

Nederlandse Samenvatting

Dit werk behandelt het asymptotisch gedrag van wachtrijen. Onder asymptotisch gedrag verstaat men het staartgedrag van de stationaire distributies geassocieerd met de bezetting van buffers in wachtrijsystemen. Hoewel deze studie eerder wiskundig van aard is, zijn er nauwe banden met de huidige netwerkpraktijk. Het feit dat het in deze thesis gehanteerde wiskundige model afgeleid is van een reëel netwerkelement, met name de statistische multiplexer, onderstreept dit.

In Hoofdstuk 1 wordt het belang en de achtergrond van de studie van het asymptotisch gedrag geschetst. Vooral ATM (*Asynchronous Transfer Mode*) en QoS (*Quality of Service*) komen aan bod. Vervolgens wordt het mathematische multiplexer model ingevoerd. Zoals blijkt is het aankomstenproces van groot belang. Daarom wordt er tevens ingegaan op de karakteristieken van het huidige netwerkverkeer. Veel aandacht wordt besteed aan LRD (*Long Range Dependence*), daar uit talrijke metingen blijkt dat dit soort verkeer overvloedig voorkomt in hedendaagse datanetwerken.

Alle bestudeerde aankomstenprocessen worden gemodelleerd als een DBMAP (*Discrete-time Batch Markovian Arrival Process*). De meeste wachtrijen zijn van het DBMAP-D-1 type, afgezien van de DBMAP-D-c wachtrij die hoofzakelijk aan bod komt in Sectie 4.3.3. De DBMAPs en de wachtrijen van het DBMAP-D-1 type worden zorgvuldig gedefinieerd in Hoofdstuk 2. Wat de DBMAP betreft, wordt er vooral dieper ingegaan op de autocorrelatie. Deze autocorrelatie is na de gemiddelde aankomstensnelheid de belangrijkste karakteristiek van een aankomstenproces. De DBMAP-D-1 wachtrij behoort tot het domein van de M/G/1-type Markov ketens. De theorie hieraan verbonden wordt bondig gepresenteerd. Nadruk wordt vooral gelegd op de Pollachek-Kinchin vergelijking.

In Hoofdstuk 2 wordt tevens het asymptotisch gedrag van een grote klasse van DBMAP-D-1 wachtrijen bestudeerd aan de hand van de klassieke dominante pool benadering. Voor deze klasse van wachtrijen wordt het asymptotisch gedrag eveneens bepaald met behulp van de theorie der grote afwijkingen. Daarbij komen enkele interessante verbanden tussen beide methodes aan het licht.

Hoofdstuk 3 handelt niet rechtstreeks over asymptotisch gedrag. In dit hoofdstuk wordt aangetoond dat het hanteren van de methode der functionele vergelijking, of het gebruiken van matrix-analytische technieken om wachtrijen op te lossen, in wezen equivalent is. Dit wordt met name gedetailleerd aangetoond voor de DBMAP-D-1 wachtrij. Voor de DBMAP-G-1 wachtrij en de DBMAP-D-c wachtrij wordt enkel aangegeven hoe de equivalentie kan worden ingezien. Dit hoofdstuk besluit met twee diepgaand behandelde voorbeelden.

Staarttransities vormen het onderwerp van Hoofdstuk 4. Een staarttransitie is een abrupte verandering in het asymptotisch gedrag van een wachtrij, als gevolg van een continue verandering van de parameters die de wachtrij vastleggen. Twee wachtrijsystemen waarin deze staarttransities op een natuurlijke wijze voorkomen, worden bestudeerd. Het is van belang op te merken dat het hier om twee sterk verschillende wachtrijen gaat. De eerst wachtrij modelleert twee wachtrijen die in tandem zijn opgesteld. De technieken die hier gehanteerd worden om het asymptotisch gedrag te onderzoeken zijn een veralgemening van deze gebruikt voor de dominante pool benadering. De tweede wachtrij heeft als aankomstenproces een superpositie van een LRD en een Markoviaanse verkeersstroom. Afhankelijk van de relatieve grootte van beide componenten gedraagt de wachtrij zich asymptotisch radicaal verschillend.

In Hoofdstuk 5 wordt getracht, aan de hand van de theorie van de pseudo fractale aankomstenprocessen, het asymptotisch gedrag te bepalen van een wachtrij met LRD aankomsten. Daartoe wordt een LRD aankomstenproces gedefinieerd als de limiet van een rij van Markoviaanse processen. De correlatiestructuur van het LRD proces wordt gedetailleerd bepaald. Een ondergrens voor het asymptotisch gedrag van de wachtrij wordt bekomen door middel van de theorie der grote afwijkingen.

Hoofdstuk 6 behandelt een wachtrij met aankomstenstroom een LRD M/G/∞ proces. Er wordt middels de Pollachek-Kinchin vergelijking een handelbare formule afgeleid voor de stationaire distributie van de wachtrij. Met behulp van een Tauber stelling is het mogelijk het gedrag van deze genererende functie te analyseren. De informatie die alzo bekomen wordt, maakt een exacte beschrijving van het asymptotisch gedrag van de wachtrij mogelijk. Deze methodologie wordt vervolgens ook toegepast op een wachtrij met als aankomstenproces een superpositie van een Markoviaans proces en een LRD M/G/∞ proces. De resultaten worden tevens geëvalueerd vanuit een praktisch oogpunt. Dit leidt tot een beschouwing over het toepassen van de gehanteerde technieken voor de studie van meer algemene wachtrijen.

Bibliography

[1] J. Abate, G. L. Choudhury, and W. Whitt. Asymptotics for steady-state tail probabilities in structured markov queueing models. *Communications in Statistics - Stochastic Models*, 10(1), 1994.

[2] A. T. Andersen and B. F. Nielsen. An application of superpositions of two state Markovian sources to the modelling of self-similar behaviour. In *Proceedings of IEEE Infocom'97*, 1997.

[3] D. Anick, D. Mitra, and M. M. Sondhi. Stochastic theory of a data-handling system with multiple sources. *The Bell System technical Journal*, 61(8):1871–1894, 1982.

[4] J. Aracil, R. Edell, and P. Varaiya. An empirical Internet traffic study. In *35th Annual Allerton Conference on Communications, Control and Computing*, October 1997.

[5] J. Aracil, D. Morató, and M. Izal. On the combined effect of self-similarity and flow control in quality of service for transactional Internet services. In *Performance of Information and Communication Systems*, pages 111–122. IFIP, 1998.

[6] A. Baiocchi. Analysis of the loss probability of the MAP/G/1/K queue. part I: asymptotic theory. *Stochastic Models*, 10:867–893, 1994.

[7] A. Baiocchi and N. Blefari Melazzi. Analysis of the loss probability of the MAP/G/1/K queue. part II: approximations and numerical results. *Stochastic Models*, 10:895–925, 1994.

[8] E. A. Bender. Asymptotic methods in enumeration. *SIAM Review*, 16(4):485–515, October 1974.

[9] C. Blondia. A discrete-time batch markovian arrival process as B-ISDN traffic model. *Belgian Journal of Operations Research, Statistics and Computer Science*, 32, 1993.

[10] C. Blondia and F. Geerts. The correlation structure of the output of an ATM multiplexer. In *Proceedings of the fifth IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*. IFIP, 1997.

[11] A. A. Borovkov, G. Fayolle, and D. A. Korshunov. Transient phenomena for Markov chains and applications. *Adv. Appl. Prob.*, 24:322–342, 1992.

[12] O. J. Boxma. Fluid queues and regular variation. Technical report, CWI, 1996.

[13] H. Bruneel and B. G. Kim. *Discrete-Time Models for Communication Systems Including ATM.* Kluwer Academic Publishers, 1993.

[14] E. Çinlar. *Introduction to stochastic processes.* Prentice-Hall, 1975.

[15] D. R. Cox. *Statistics: an Appraisal*, chapter Long-range dependence: a review, pages 55–74. Iowa State University Press, 1984.

[16] M. E. Crovella and A. Bestavros. Explaining world wide web traffic self-similarity. Technical Report TR-95-015, Boston University, 111 Cummington St, Boston, MA 02215, October 1995.

[17] T. Daniëls and C. Blondia. Asymptotic behaviour of a discrete-time queue with long range dependent m/g/∞ input. submitted.

[18] T. Daniëls and C. Blondia. On the equivalence of the matrix-analytical and the functional equation approach to solve discrete-time queueing systems. submitted.

[19] T. Daniëls and C. Blondia. A discrete-time ATM traffic model with long range dependence characteristics. In *Performance and Management of Complex Communication Networks.* IFIP, 1997.

[20] T. Daniëls and C. Blondia. Asymptotic behavior of a discrete-time queue with long range dependent input. In *Proceedings of INFOCOM'99*, 1999.

[21] M. de Prycker. *Asynchronous Transfer Mode.* Ellis Horwood, 1991.

[22] N. G. Duffield. On the relevance of long-tailed durations for the statistical multiplexing of large aggregations. In *Proceedings of the 34th Annual Allerton Conference on Communication, Control and Computing*, October 1996.

[23] N. G. Duffield and N. O'Connell. Large deviations and overflow probabilities for the general single server queue, with applications. *Mathematical Proceedings of the Cambridge Philosophical Society*, (118):363–374, 1995.

[24] V. Dumas and A. Simonian. Asymptotic bounds for the fluid queue fed by sub-exponential on/off sources. Technical report, Mathématics Appliquées de Bordeaux, 1998.

[25] E. Falkenberg. On the asymptotic behaviour of the stationary distribution of markov chains of the m/g/1-type. *Communications in Statistics - Stochastic Models*, 10(1):75–97, 1994.

[26] W. Feller. *An Introduction to Probability and Its Applications, Volume II.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., November 1966.

[27] P. Ferguson and G. Huston. *Quality of Service.* John Wiley & Sons, Inc., 1998.

[28] F. Geerts and C. Blondia. Superposition of markov sources and long range dependence. In *Broadband Communications: the future of telecommunications*, pages 550–561, 1998.

[29] P. W. Glyn and W. Whitt. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *Journal of Applied Probability*, 31A:131–159, 1994.

[30] W. J. Goralski. *Introduction to ATM networking.* McGraw-Hill, Inc., 1995.

[31] A. Graham. *Kronecker Products and Matrix Calculus: with Applications.* Ellis Horwood Limited, 1981.

[32] R. Gusella. Characterizing the variability of arrival processes with indices of dispersion. *IEEE Journal on Selected Areas in Communications*, 9(2), February 1991.

[33] D. Heath, S. Resnick, and G. Samorodnitsky. Heavy tails and long range dependence in on/off processes and associated fluid models. Technical report, Cornell University, 1997.

[34] P. R. Jelenković and A. A. Lazar. Multiplexing on-off sources with subexponential on periods: Part ii. In V. Ramaswami and P. Wirth, editors, *Teletraffic contributions for the information Age*, volume 2b of *Teletraffic Science and Engineering*, pages 965–974, Washington D.C., june 1997. ITC, Elsevier.

[35] P. R. Jelenković and A. A. Lazar. Asymptotic results for multiplexing on-off sources with subexponential on period. *Advances in Applied Probability,*, 31(2), june 1999.

[36] J. G. Kemeny and J. L. Snell. *Finite Markov Chains.* Van Nostrand Reinhold Company, 1960.

[37] M. M. Krunz and A. M. Makowski. Modeling video traffic using m/g/∞ input processes: a compromise between Markovian and LRD models. *IEEE Journal on Selected Areas in Communications*, 16(5):733–748, june 1998.

[38] O. Kyas. *ATM networks.* International Thomson Publishing, 1995.

[39] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on networking*, 2:1–15, february 1994.

[40] S. Q. Li and C. L. Hwang. On input state space reduction and buffer noneffective region. In *Proc. IEEE Infocom'94*, pages 1018–1028, 1994.

[41] Z. Liu, P. Nain, D. Towsley, and Z.-L. Zhang. Asymptotic behavior of a multiplexer fed by a long-range dependent process. Technical Report 3230, INRIA, August 1997.

[42] D. M. Lucantoni. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 1991.

[43] B. M. Makarov, M. G. Goluzina, A. A. Lodkin, and A. N. Podkorytov. *Selected Problems in Real Analysis*, volume 107 of *Translations of Mathematical Monographs*. American Mathematical Society, 1992.

[44] B. B. Mandelbrot. Long-run linearity, locally gaussian processes, H-spectra and infinite variances. *International Economic Review*, 11:82–113, 1969.

[45] A. Markushevich. *Theory of functions of a complex variable*, volume II. Prentice-Hall, Inc., 1965.

[46] J. A. Morrison. Two discrete-time queues in tandem. *IEEE Transactions on Communications*, 27(3):563–573, 1979.

[47] M. Nabe, M. Murata, and H. Miyahara. Analysis and modelling of World Wide Web traffic for capacity dimensioning for Internet access. In *SPIE Video, Voice and Data Communications Conference*, November 1997.

[48] M. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779, 1979.

[49] M. F. Neuts. *Structured stochastic matrices of the M/G/1-type and their applications.* Marcel Dekker, 1989.

[50] V. Paxson and S. Floyd. Wide-area traffic: the failure of poison modeling. In *Proc. of the ACM Sigcomm'94*, pages 395–409, London, UK, 1994.

[51] V. Ramaswami. Nonlinear matrix equations in applied probability—solution techniques and open problems. *SIAM Review*, 30(2), June 1988.

[52] J. Roberts, U. Mocci, and J. Virtamo. *Broadband network teletraffic.* Springer, 1996.

[53] E. Seneta. *Non-negative matrices and Markov Chains.* Springer-Verlag New York, 1981.

[54] K. Spaey and C. Blondia. Circulant matching method for multiplexing ATM traffic applied to videosources. In U. Körner and A. A. Nilsson, editors, *Performance of Information and Communication Systems*, pages 234–245. IFIP, 1998.

[55] K. Sriram. Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE Journal on Selected Areas in Communications*, SAC-4(6), September 1986.

[56] B. Steyaert, Y. Xiong, and H. Bruneel. An efficient solution technique for discrete-time queues fed by heterogenuous traffic. *International Journal of Communication Systems*, 10:73–86, 1997.

[57] M. S. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar traffic modeling. *Computer Communication Review*, (27):5–23, 1997.

[58] K. P. Tsoukatos and A. M. Makowski. Heavy traffic analysis for a multiplexer driven by $M/GI/\infty$ input processes. In V. Ramaswami and P. Wirth, editors, *Teletraffic contributions for the information Age*, volume 2a of *Teletraffic Science and Engineering*, pages 497–506, Washington D.C., june 1997. ITC, Elsevier.

[59] B. Tsybakov and N. D. Georganas. On self-similar traffic in ATM queues: definitions, overflow probability bound and cell delay distribution. *IEEE/ACM Transactions on networking*, June 1997.

[60] P. Van Mieghem. The asymptotic behaviour of queueing systems: large deviations theory and dominant pole approximation. *Queueing Systems*, (23):27–55, 1996.

[61] S. Wittevrongel. *Prestatie-Analyse van Discrete-Tijd-Buffersystemen met Grillige Aankomstenstromen*. PhD thesis, University of Ghent, 1998.

[62] S. Wittevrongel and H. Bruneel. Queue length and delay for statistical multiplexers with variable-length messages. In *Proceedings of GLOBECOM'94*, pages 1080–1084, San Fransisco, november 1994.

[63] S. Wittevrongel and H. Bruneel. Deriving the tail distribution of the buffer contents in a statistical multiplexer with general heterogeneous on/off sources. In T. Hasegawa, H. Takagi, and Y. Takahashi, editors, *Performance and Management of Complex Communication Networks*. Chapman & Hall, November 1997.

[64] S. Wittevrongel and H. Bruneel. Correlation effects in ATM queues due to data format conversions. *Performance Evaluation*, (32):35–56, 1998.

[65] Y. Xiong and H. Bruneel. Buffer contents and delay for statistical multiplexers with finite number of inputs and train arrivals. In *Proceedings INFOCOM'92*, pages 2036–2044, Firenze, May 1992.

[66] Y. Xiong and H. Bruneel. On the asymptotic behavior of discrete-time single-server queueing systems with general markov-modulated arrival processes. In *Proceedings of ITC 14*, 1994.