On the average waiting time for regular routing to deterministic queues

Arie Hordijk * Dinard van der Laan †

January 10, 2003

Abstract

We provide an algorithm to compute a lower bound for the total average waiting time for the optimal routing to $N \geq 2$ deterministic queues. We also obtain an upper bound for the total average waiting time of the optimal routing policy. The upper and lower bounds coincide in case $N = 2$. Our analysis of deterministic queues is based on results from number theory, involving Farey intervals, best approximation points, convergents and continued fraction expansions. It gives explicit formulae for the average waiting time in case of regular routing to a deterministic queue.

1 Introduction

In this paper we consider the deterministic optimal routing to N parallel inhomogeneous queues. We assume that the arrival epochs are a deterministic sequence, $T = 0, 1, 2, \ldots$ say, and the service times at the queues are deterministic, S_i at queue i say. In stochastic models there is a difference between open-loop control in which there is no state information and closed-loop control in which the controller knows the loads at the queues. In the deterministic model there is no such difference in state information. But, as in the open-loop control we describe the routing policy as a sequence of the numbers of the queues to which the successive arriving customers are routed. For an overview of papers on open-loop control see [12]. An overview of results on closed-loop control can be found in [11] and [15].

The optimal routing for homogeneous queues, i.e. queues with $S_i = S$, for $i = 1, \ldots, N$, is well-known to be the round robin policy (see [17]). For inhomogeneous queues the computation of the optimal policy may be difficult (see [4] and [6] for approximating algorithms). In the seminal paper [8] Hajek showed for an exponential queue that the optimal routing of a given fraction of the customers to that queue is a regular or bracket sequence. For $N = 2$ it is proved in [1] and [3] that the optimal routing to each of the queues is a regular or bracket sequence for general stochastic interarrival and service time sequences. However, if $N \geq 3$ the optimal routing fails to be the composition of regular sequences, because in general regular sequences can not be combined to a feasible routing policy. Finding the optimal routing becomes a combinatorial hard problem for $N \geq 3$. In [16] the deterministic model is studied where the system is fully loaded, i.e. $\sum_{i=1}^{N} \frac{1}{S_i} = 1$, and a mathematical programming problem is designed to compute the optimal routing. In this paper we consider the case $\sum_{i=1}^{N} \frac{1}{S_i} \geq 1$. Now finding the optimal routing is split into two parts:

*Dept. of Mathematics, Leiden University, P.O.Box 9512, 2300 RA Leiden, The Netherlands. E-mail: hordijk@math.leidenuniv.nl . Telephone: (+31)715277146

†Dept. of Mathematics, Leiden University, P.O.Box 9512, 2300 RA Leiden, The Netherlands. E-mail: dvd-laan@math.leidenuniv.nl . Telephone: (+31)715277140

- Find best routing densities d_i to the queues $i = 1, \ldots, N$
- Construct a best allocation pattern for the given routing densities d_i

For $N = 2$ the second part is completely solved, since the allocation uses regular sequences as mentioned above. In [7] the problem of finding the best densities is studied, and a very interesting connection is made there to combinatorics on words. Moreover, a recursive formula for the average waiting time for a given density is obtained by using a (backward) continued fraction decomposition of the inverse of the service time, and an algorithm is derived to compute an optimal routing sequence. The present paper is in the same spirit but considers the more general case of $N \geq 2$ queues. We provide an algorithm to compute a lower bound for the total average waiting time for the optimal routing by finding best routing densities given that the routing to each of the queues is regular. In [12] we derived an upper bound for the total waiting time for a given allocation pattern as a function of its unbalance (which is a measure of the deviation from the regular routing). Combining these results we use a billiard sequence to provide an upper bound for the average waiting time of the optimal routing policy too. The upper and lower bounds coincide in case $N = 2$. This paper provides an analysis of deterministic queues based on results from number theory, involving Farey intervals, best approximation points, convergents and continued fraction expansions.

The paper is organized as follows. Section 2 gives a description of the queueing system together with the definitions of routing policy and regular sequences. In section 3 the Lindley recursion together with the extension of Little's relation between the average number of customers and the waiting time is given for the routing model. Its proof is given in the appendix. In section 4 Farey intervals are introduced. The period sequence of the upper bracket sequence of which the rational density belongs to some Farey interval is determined as a finite composition of the period sequences of the upper bracket sequences of the end-points densities of that Farey interval. This result gives the key to the proof that for upper bracket sequences the average number of customers $L(d)$ is a linear function of the density d on a Farey interval. This result is proved in section 5, where it is also shown that $L(d)$ is a convex function of the density. In section 6 best lower approximation points of $\frac{1}{S_i}$, where S_i is the service time for queue i, are introduced. They correspond to the so-called jump points in [7], where recursive formulae for computing the average waiting time $W(d_i)$ were obtained. The main result of section 6 is the explicit formulae for the average number of customers $L_{S_i}(d_i)$ (and of the average waiting time $W(d_i)$) at best lower approximation points d_i of $\frac{1}{S_i}$. Together with the fact that consecutive best lower approximation points constitute a Farey interval on which $L_{S_i}(d_i)$ is linear, this provides an efficient algorithm for computing $L_{S_i}(d_i)$ given that the best lower approximation points are known. Section 6 also contains an exposition, based on well-known results from number theory, how the ordered set of best approximation points can be computed by the use of the continued fraction expansion and the induced convergents. The efficient algorithm to compute this ordered set is illustrated by an example. In section 7 the above mentioned results are applied to find the minimal regular routing. More precisely we derive an algorithm for computing

$$\min_{d_1+\cdots+d_N=1} \sum_{i=1}^{N} L_{S_i}(d_i),$$

using the above mentioned convexity and linearity properties on Farey intervals. The algorithm is again illustrated by an example. We characterize the set of densities for which the minimum is achieved. Also, as a consequence of the characterization we find that there exist minimizing rational densities if $\sum_{i=1}^{N} \frac{1}{S_i} > 1$. Indeed, the algorithm computes such a set of rational densities. For $N = 2$ this implies that the regular policy with these densities is periodic, and hence an optimal periodic routing policy is found by the algorithm. For $N \geq 3$ the algorithm computes a lower bound for the optimal routing policy. Together with results of [12] which we summarize in section 7, we find a billiard sequence with rational densities (hence a periodic policy) which has an average waiting time not larger than the lower bound plus $\frac{N}{2} - 1$. The performance of this billiard sequence should be close to optimal if N is not too large. For $N = 2$ the lower bound equals the upper bound.

2 Description of the queueing system and notations

We consider a queueing system with $N \geq 2$ parallel FIFO (First In First Out) queues. Customers arrive at moments $T_1, T_2, \ldots$ with $T_1 = 0$ and $T_1 \leq T_2 \leq \ldots$ and the n -th arriving customer has to be routed to one of the queues at the moment T_n of its arrival. We consider deterministic routing policies using no state information. Then the routing policy is described by an infinite word $U = (U_1, U_2, \ldots)$ on the alphabet $\{1, 2, \ldots, N\}$ where U_n is the queue to which the n - th arriving customer is routed (see [12]). For every queue $i \in \{1, 2, \ldots, N\}$ we have a sequence $u^i = (u^i_1, u^i_2, \ldots)$ of zeros and ones where $u^i_n = 1$ if $U_n = i$ and $u^n_i = 0$ if $U_n \neq i$. Then u^i is called the routing, splitting or admission sequence for queue i (see for example [8], [23], [2], [12] and [16]).

For a sequence of nonnegative integers $u = (u_1, u_2, \ldots)$ we define the counting function $\kappa_u : \mathbb{Z}_{\geq 0} \to \mathbb{Z}$ by $\kappa_u(n) := \sum_{i=1}^{n} u_i$. Thus $\kappa_u(n)$ is the partial sum of the first n terms. Note that if u^i is a splitting sequence then $\kappa_{u^i}(n)$ is the number of customers that is routed to queue i among the first n arriving customers.

Put $\delta_n := T_{n+1} - T_n$ for $n = 1, 2, \ldots$. Then $\{\delta_n\}$ is the sequence of interarrival times. Further for queue $i \in \{1, 2, \ldots, N\}$ we have a sequence of service times $\{\sigma^i_j\}_{j=1,2,\ldots}$, where σ^i_j is the service time of the j -th customer that is routed to queue i.

In a deterministic queueing system the interarrival times are constant and the service times are constant for every queue. Thus there exist positive numbers $\delta, \sigma_1, \sigma_2, \ldots, \sigma_N$ such that $\delta = \delta_1 = \delta_2 = \ldots$ and $\sigma_i = \sigma^i_1 = \sigma^i_2 = \sigma^i_3 = \ldots$ for $i = 1, 2, \ldots, N$. By time scaling with a factor δ we assume without loss of generality that $\delta = 1$ and then $S_i := \frac{\sigma_i}{\delta}$ is the service time of server i for the new time scale. Thus such a deterministic queueing system with N parallel FIFO queues can be described by the vector $(S_1, S_2, \ldots, S_N)$ of service times when the interarrival times are assumed to be equal to the time unit. In the sequel we use the description $(S_1, S_2, \ldots, S_N)$ system when we mean a queueing system with deterministic interarrival and service times as described above.

2.1 Routing policies and words

A routing policy U is described by a sequence of symbols (letters) $(U_1, U_2, \ldots)$ and can be seen as an (infinite) word (sequence) on a finite alphabet. If we have N parallel queues then U is a so-called $\mathbb{N}$ -word on the alphabet $\{1, 2, \ldots, N\}$. We refer to the beginning of section 2 of [12] for the definitions of the density of a letter, period of a word, subword, prefix, suffix, concatenation of sequences, etc. See [18] and [10] for more on words and combinatorics.

We recall that the domain of a word I is a subinterval of $\mathbb{Z}$ and the length of a finite (i.e I is finite) word W is denoted by $\mid W \mid$. Moreover, a sequence (word) of zeros and ones is said to have density d if the letter 1 has density d. It follows that the density of letter 0 is $1 - d$ in that case.

Definition 2.1 *Let u be an infinite sequence (I - word) of zeros and ones for some appropriate infinite interval I. We say that u is regular with density d if for every $N \in \mathbb{N}$ we have for every subword u' of u of length N that $|u'|_1$, the number of ones in u', equals $\lfloor Nd \rfloor$ or $\lceil Nd \rceil$. In that case we also say that the set $H_1 \subseteq I$, corresponding to the places of the ones, is a regular set (of density d).*

The following theorem follows for $I = \mathbb{N}$ from the classification of balanced sequences in [20] and the fact that regular sequences are particular balanced sequences (see [25]).

Theorem 2.2 *Let u and I be as above. Then u is regular with density d if and only if there exists $\theta \in \mathbb{R}$ such that $u_t = \lfloor (t+1)d + \theta \rfloor - \lfloor td + \theta \rfloor$ for every $t \in I$ or there exists a $\theta' \in \mathbb{R}$ such that $u_t = \lceil (t+1)d + \theta' \rceil - \lceil td + \theta' \rceil$ for every $t \in I$.*

Regular sequences are also called bracket sequences and Theorem 2.2 shows the reason for this. Moreover, the θ in the bracket form is called the phase of the bracket sequence. In the sequel all infinite regular sequences we consider are $\mathbb{N}$ -words. Hence if I is infinite, then $I = \mathbb{N}$.

Definition 2.3 *The upper bracket sequence of density d is the $\mathbb{N}$-word $u = (u_1, u_2, \ldots, u_n)$ on the alphabet $\{0,1\}$ for which $u_t = \lceil td \rceil - \lceil (t-1)d \rceil$ for every $t \in \mathbb{N}$. The lower bracket sequence of density d is the $\mathbb{N}$-word $u = (u_1, u_2, \ldots, u_n)$ on the alphabet $\{0,1\}$ for which $u_t = \lfloor td \rfloor - \lfloor (t-1)d \rfloor$ for every $t \in \mathbb{N}$.*

If we make the assumption that letter 0 is smaller than letter 1 in the alphabet $\{0,1\}$ then the upper bracket sequence of density d is the lexicographic greatest regular $\mathbb{N}$ -word of density d and the lower bracket sequence is the lexicographic smallest regular sequence of density d. We have the following properties of these sequences (see [12]).

Lemma 2.4 *For the upper bracket sequence u of density d we have that $\kappa_u(n) = \lceil nd \rceil$ for every $n \in \mathbb{Z}_{\geq 0}$ and for the lower bracket sequence v of density density d we have that $\kappa_v(n) = \lfloor nd \rfloor$ for every $n \in \mathbb{Z}_{\geq 0}$.*

Lemma 2.5 *Let u and v be as in Lemma 2.4. Then for the support sets $H_1(u)$ and $H_1(v)$ of the ones of these bracket sequences we have that*

$$H_1(u) = \{\lfloor \frac{t}{d} \rfloor + 1\}_{t=0}^{\infty} \text{ and } H_1(v) = \{\lceil \frac{t}{d} \rceil\}_{t=1}^{\infty}.$$

3 Definitions and preliminary results

For $t \geq 0$ and $i \in \{1, 2, \ldots, N\}$ let $V_i(t)$ be the remaining workload measured in time units for server i at time t and let $B_i(t)$ be the number of customers waiting in the buffer of queue i at moment t. Then $\underline{V}(t) := (V_1(t), V_2(t), \ldots, V_N(t))$ is the vector of remaining workloads in the system at time t, $\underline{B}(t) := (B_1(t), B_2(t), \ldots, B_N(t))$ is the vector of buffer contents at time t and $B(t) := \sum_{i=1}^{N} B_i(t)$ is the total number of customers waiting in one of the buffers at time t. Let $G_i(t) = \{t' \in [0,t] : V_i(t') = 0\}$ and let $m_i(t)$ be the Lebesque measure of $G_i(t)$. Then $m_i(t)$ is the time that server i was idle up to time t. We say that a server i is non-idling in a time interval $[t_1, t_2]$ if $m_i(t_1) = m_i(t_2)$.

For an $(S_1, S_2, \ldots, S_N)$ system we assume that $V_i(0) = 0$ for $i = 1, 2, \ldots, N$ and for a given routing policy $U = (U_1, U_2, \ldots)$ and time $t > 0$ we can compute $\underline{V}(t)$ recursively by Lindley's formula which gives that

$$V_i(t) = \max(V_i(t') + u_{t'}^i \cdot S_i + (t' - t), 0) \text{ for } i = 1, 2, \ldots, N \text{ where } t' = \lceil t \rceil - 1. \quad (1)$$

Note that we define $V_i(t)$ such that the function $t \to V_i(t)$ is left continuous, but not right continuous in general. However, the right discontinuities of this function can only be at integer t. Moreover, it follows inductively that for every $i \in \{1, 2, \ldots, N\}$ and $t, t_0 \in \mathbb{R}$ with $t \geq t_0 \geq 0$ we have that

$$V_i(t) - V_i(t_0) = (\kappa_{u^i}(\lceil t \rceil) - \kappa_{u^i}(\lceil t_0 \rceil)) \cdot S_i + (m_i(t) - m_i(t_0)) - (t - t_0). \tag{2}$$

Note that in an $(S_1, S_2, \ldots, S_N)$ system it holds that $T_t = t - 1$ for $t = 1, 2, \ldots$ and thus $\kappa_{u^i}(\lceil t \rceil)$ is the number of customers routed to server i before time t. Further for an $(S_1, S_2, \ldots, S_N)$ system we have for every $t \geq 0$ that $B_i(t) = \lfloor \frac{V_i(t)}{S_i} \rfloor$ for $i = 1, 2, \ldots, N$ and thus $\underline{B}(t)$ and $B(t)$ can be easily obtained from $\underline{V}(t)$.

Definition 3.1 *Let $U = (U_1, U_2, \ldots)$ be a routing policy. Then we define*

$$L(U) := \limsup_{t' \to \infty} \frac{1}{t'} \cdot \int_{t=0}^{t'} B(t) d(t)$$

as the average number of customers waiting in one of the buffers for policy U. We define

$$W(U) := \limsup_{t' \to \infty} \frac{1}{t'} \sum_{t=1}^{t'} V_{U_t}(T_t)$$

as the average waiting time for the arriving customers for policy U.

We show in the appendix as an extension of Little's law for routing policies that

$$L(U) = \lambda \cdot W(U) \text{ if } W(U) < \infty, \tag{3}$$

where λ is the arrival rate of customers in the queueing system. In an $(S_1, S_2, \ldots, S_N)$ system we have by the time scaling that the arrival rate of customers λ is equal to 1 and thus it follows by (3) that

$$L(U) = W(U) \text{ if } W(U) < \infty. \tag{4}$$

The objective in this paper is optimal routing with respect to minimizing the average waiting time. Thus we look for routing policies U such that $L(U)$ and $W(U)$ are minimized.

For this problem it is useful to consider for a particular server $i \in \{1, 2, \ldots, N\}$ the average number of customers waiting in buffer i and the average waiting time of the customers that are routed to this server i. So, for $i = 1, 2, \ldots, N$ we define

$$L^i(U) := \limsup_{t' \to \infty} \frac{1}{t'} \cdot \int_{t=0}^{t'} B_i(t) d(t)$$

as the average number of customers waiting in buffer i for policy U. Recall that $H_i(U)$ is the support of letter i of the $\mathbb{N}$ -word U and for $t \in \mathbb{N}$ let $I_t = \{1, 2, \ldots, t\}$. Moreover, for $i = 1, 2, \ldots, N$ we define

$$W^i(U) := \limsup_{t' \to \infty} \frac{1}{\mid H_i(U) \cap I_{t'} \mid} \cdot \sum_{t \in H_i(U) \cap I_{t'}} V_i(T_t) = \limsup_{t' \to \infty} \frac{1}{\kappa_{u^i}(t')} \cdot \sum_{t=1}^{t'} u_t^i \cdot V_i(t-1) \tag{5}$$

as the average waiting time of customers routed to server i for policy U.

Note that $L^i(U)$ and $W^i(U)$ depend on u^i, the routing sequence for server i, but they do not depend on u^j if $j \neq i$. Therefore, if the routing to server i is according to some infinite sequence w of zeros and ones then we also use the notations $L^i(w)$ and $W^i(w)$ instead of $L^i(U)$ and $W^i(U)$ respectively. Moreover, if we consider a single queue of the system and it is not necessary to refer to the number of the server i then we just write $L(w)$ and $W(w)$ respectively. We define $\overline{d_i} := \limsup_{t\to\infty} \frac{\kappa_{u^i}(t)}{t}$ which we call the upper density of routing sequence u^i and we define $\underline{d_i} := \liminf_{t\to\infty} \frac{\kappa_{u^i}(t)}{t}$ which we call the lower density of u^i. If $\overline{d_i} = \underline{d_i}$ then the routing sequence u^i has a density $d_i := \lim_{t\to\infty} \frac{\kappa_{u^i}(t)}{t}$ and d_i is the arrival rate for server i in that case. In the appendix we show that

$$L^i(u^i) = d_i \cdot W^i(u^i) \text{ if } d_i \text{ exists and } W^i(u^i) < \infty. \tag{6}$$

By (4) and (6) we obtain the following proposition.

Proposition 3.2 *Let U be a routing policy applied in an $(S_1, S_2, \ldots, S_N)$ system such that routing sequence u^i has density d_i and $W^i(u^i) < \infty$ for $i = 1, 2, \ldots, N$. Then*

$$W(U) = L(U) = \sum_{i=1}^{N} L^i(u^i) = \sum_{i=1}^{N} d_i \cdot W^i(u^i).$$

Note that $\sum_{i=1}^{N} d_i = 1$ if the conditions of Proposition 3.2 are satisfied.

4 The upper bracket sequence and Farey intervals

In the sequel we denote by $w(d)$ the upper bracket sequence of density d. Thus for the counting function κ we have that $\kappa_{w(d)}(n) = \lceil n \cdot d \rceil$ for every $n \in \mathbb{Z}_{\geq 0}$. If d is rational, $d = \frac{p}{q}$ with $\gcd(p, q) = 1$ let $\omega(d) := (w(d)_1, w(d)_2, \ldots, w(d)_q)$ be the period word of $w(d)$. Note that we use the slightly different notations $w(d)$ and $\omega(d)$ for respectively the upper bracket sequence and the period word of this sequence.

Let d_1, d_2 be rational fractions with $0 \leq d_1 \leq d_2 \leq 1$, $d_i = \frac{p_i}{q_i}$ and $\gcd(p_i, q_i) = 1$ for $i = 1, 2$. Then $I = [d_1, d_2]$ is called a Farey interval if and only if $q_1 \cdot p_2 - p_1 \cdot q_2 = 1$.

From elementary number theory (see for example [9], chapter 3, page 23) we have the following facts about Farey intervals. If $[d_1, d_2]$ is a Farey interval and $d_0 = \frac{p_1+p_2}{q_1+q_2}$. Then $I' = [d_1, d_0]$ and $I'' = [d_0, d_2]$ are also Farey intervals. Moreover, all rational numbers in (d_1, d_2) have denominator greater than or equal to $q_1 + q_2$.

4.1 Factorisation of the upper bracket sequence

Recall that a factor of a word is a finite subsequence of consecutive letters of the word. Let X be a set of nonempty finite words over some alphabet $\mathbf{A}$. A finite X - factorisation of a finite word w is a finite sequence $(x_1, x_2, \ldots, x_n)$ such that $w = x_1 x_2 \ldots x_n$ and $x_i \in X$ for $i = 1, 2, \ldots, n$. An infinite X -factorisation of an $\mathbb{N}$ - word w is an infinite sequence $(x_1, x_2, \ldots)$ such that $w = x_1 x_2 \ldots$ and $x_i \in X$ for $i = 1, 2, \ldots$. Sometimes we just say factorisation instead of (in)finite X -factorisation. The set X is called a code ($\mathbb{N}$ -code) if every finite respectively infinite word over the alphabet $\mathbf{A}$ has at most one X - factorisation. It is easily seen that every $\mathbb{N}$ -code is also a code, but the converse is not true (see [18], chapter 6, page 177).

Lemma 4.1 *Let $d_1, d_2 \in \mathbb{Q} \cap [0,1]$ with $d_1 \neq d_2$. Then $\omega(d_1)$ and $\omega(d_2)$ are not powers of the same word.*

Proof. Suppose that there exist $k, l \in \mathbb{N}$ and a sequence of zeros and ones u such that $u^k = \omega(d_1)$ and $u^l = \omega(d_2)$. Then it follows that d_1, the density of $\omega(d_1)$, equals the density of u and also d_2, the density of $\omega(d_2)$, equals the density of u. Hence $d_1 = d_2$, which gives a contradiction. □

Theorem 4.2 *Let $d_1, d_2 \in \mathbb{Q} \cap [0,1]$ with $d_1 \neq d_2$. Then $X := \{\omega(d_1), \omega(d_2)\}$ is an $\mathbb{N}$-code for words over the alphabet $\{0,1\}$.*

Proof. By Lemma 4.1 and Corollary 6.2.5 in [18] (which follows immediately from the so-called Defect theorem) there exists neither a finite nor an infinite nontrivial relation between $\omega(d_1)$ and $\omega(d_2)$. Hence X is a code and also an $\mathbb{N}$-code.

Lemma 4.3 *Let $I = [d_1, d_2]$ be a Farey interval and $d_0 = \frac{p_1+p_2}{q_1+q_2}$ as above. Then $\omega(d_0) = \omega(d_2)\omega(d_1)$.*

Proof. We put $u := \omega(d_2)\omega(d_1)$. We have $\gcd(p_1 + p_2, q_1 + q_2) = 1$ because $q_1 + q_2$ is the smallest denominator in (d_1, d_2). Hence $| u |=| \omega(d_0) |= q_1 + q_2$ and it suffices to prove that $\kappa_u(n) = \kappa_{\omega(d_0)}(n)$ for $1 \leq n \leq q_1 + q_2$. For $1 \leq n \leq q_2$ we have that

$$\kappa_{\omega(d_0)}(n) = \lceil d_0 \cdot n \rceil \leq \lceil d_2 \cdot n \rceil = \kappa_u(n).$$

Suppose $\lceil d_0 \cdot n \rceil < \lceil d_2 \cdot n \rceil$ for some $1 \leq n \leq q_2$. Then there exist a positive integer k such that $\frac{n \cdot (p_1+p_2)}{q_1+q_2} \leq k < \frac{n \cdot p_2}{q_2}$. Hence $n \cdot q_2 \cdot (p_1 + p_2) \leq k \cdot q_2 \cdot (q_1 + q_2) < n \cdot p_2 \cdot (q_1 + q_2)$ and thus $0 \leq q_2 \cdot (k \cdot (q_1 + q_2) - n \cdot (p_1 + p_2)) < n$. Since $n \leq q_2$ it follows that $k \cdot (q_1 + q_2) - n \cdot (p_1 + p_2) = 0$ and thus $d_0 = \frac{p_1+p_2}{q_1+q_2} = \frac{k}{n}$ with $1 \leq n \leq q_2$, which gives a contradiction. So, $\kappa_u(n) = \kappa_{\omega(d_0)}(n)$ for $1 \leq n \leq q_2$. Analogously it follows that $\kappa_u(n) = \kappa_{\omega(d_0)}(n)$ for $q_2 + 1 \leq n \leq q_1 + q_2$. □

Theorem 4.4 *Let $I = [d_1, d_2]$ be a Farey interval and put $X := \{\omega(d_1), \omega(d_2)\}$. Then for every $d \in (d_1, d_2)$ there exists a unique infinite X-factorisation of the upper bracket sequence $w(d)$ of density d. Moreover, if d is rational then there exists a unique finite X-factorisation of the period word $\omega(d)$ of $w(d)$.*

Proof. The factorization is obtained by the following algorithm. Initialize by putting $x_0 = d_1 = \frac{p_0}{q_0}$, $y_0 = d_2 = \frac{r_0}{s_0}$, $u_0 = \omega(x_0)$, $v_0 = \omega(y_0)$ and $i := 0$. At step i we have that $x_i = \frac{p_i}{q_i}$ with $\gcd(p_i, q_i) = 1$ and $y_i = \frac{r_i}{s_i}$ with $\gcd(r_i, s_i) = 1$ and we do the following iteratively: $z_i := \frac{p_i+r_i}{q_i+s_i}$, $w_i := v_i u_i$. If $d = z_i$ then stop. Else if $d < z_i$ then $x_{i+1} := x_i$, $u_{i+1} := u_i$, $y_{i+1} := z_i$, $v_{i+1} := w_i$. Else if $d > z_i$ then $x_{i+1} := z_i$, $u_{i+1} := w_i$, $y_{i+1} := y_i$, $v_{i+1} := v_i$. If the algorithm has not stopped then let $i := i + 1$ and go to the next step.
From the algorithm it follows by induction that w_i is factored in $u_0 = \omega(d_1)$ and $v_0 = \omega(d_2)$ for every $i \in \mathbb{Z}_{\geq 0}$ and by Lemma 4.3 we have that $w_i = \omega(z_i)$. If d is rational then the algorithm stops for some $i = N$ and then $d = z_N$ and $\omega(d) = \omega(z_N) = w_N$ is factored in $\omega(d_1)$ and $\omega(d_2)$. If d is irrational then $\lim_{i\to\infty} z_i = d$ and $\lim_{i\to\infty} w_i = w(d)$. □

5 The average number of customers in a single queue

We consider a single queue $i \in \{1, 2, \ldots, N\}$ of an $(S_1, S_2, \ldots, S_N)$ system. Let $S = S_i$ be the service time in this queue. We suppose that the routing sequence for this queue has density $d \in [0,1]$. Then d

is the arrival rate of customers for this queue and the traffic intensity for this queue is $\rho := d \cdot S$. The stability condition for the queue is $\rho \leq 1$, or equivalently $d \leq \frac{1}{S}$. Suppose that the routing sequence u^i is a regular sequence of density $d \in [0,1]$. Then it is proved in [14] that $W^i(u^i)$ and $L^i(u^i)$ exist for every ρ and are independent of the initial phase θ. It follows from Loynes' stability theorem (see [5]) that the limit is finite for $\rho < 1$. For $\rho = 1$ this is proved in [16]. If $\rho > 1$ then the queue is not stable and $L^i(u^i) = W^i(u^i) = \infty$ in that case. Combining these results it follows that if u is a regular sequence of density $d \in [0,1]$ then

$$L^i(u) = L^i(w(d)) \text{ and } W^i(u) = W^i(w(d)). \tag{7}$$

Thus for the average number of customers waiting in a single queue all regular sequences of density d have the same performance as the upper bracket sequence $w(d)$ of density d. Moreover, this performance does only depend on the service time S of the queue and the density d. Therefore to simplify notation we write $L_S(d)$, or $L(d)$, if the value of S is clear from the context, instead of the more involved $L^i(w(d))$. Similarly we write $W_S(d)$ or $W(d)$ instead of $W^i(w(d))$. So, for a single queue with constant service time S we have that $L(d) = L_S(d)$ is the average number of customers waiting in the queue when the routing to the queue is according to a regular sequence of density d and $W(d) = W_S(d)$ is the average waiting time of these customers. Moreover, in [3] it is proved for rather general settings that routing according to a regular sequence of density d is optimal among all routing sequences of (lower) density at least d. This result implies the following lemma.

Lemma 5.1 *For every service time S and any routing sequence u^i of (lower) density at least d we have that $W^i(u^i) \geq W(d)$ and $L^i(u^i) \geq L(d)$.*

In the sequel if we consider a single queue then the remaining workload for the server (measured in time units) at time t is simply denoted by $V(t)$ instead of $V_i(t)$. Remember that the interarrival times of customers at the routing point are constant and equal to 1 by time scaling and that we assume that $V(0) = 0$. One of the objectives in this paper is to give an efficient algorithm for computing $L_S(d)$ and $W_S(d)$ for arbitrary $S > 0$ and $d \in [0,1]$. For this it is useful to consider for a fixed value of S the functions $d \to L(d)$ and $d \to W(d)$, where both these functions have domain $[0,1]$ and codomain $[0,\infty]$. In [7] an algorithm is given to compute $W_S(d)$ and some properties of the function $W(d)$ are obtained. The following lemma follows directly from the results in [7].

Lemma 5.2 *If $d \leq \frac{1}{\lceil S \rceil}$ where S is the service time then $W(d) = 0$*

From Lemma 5.2 it follows that $W(d) = 0$ for every $d \in [0,1]$ if the service time $S \leq 1$, which is a trivial case. Therefore we assume in the sequel that $S > 1$. If $S > 1$ then the queue is stable if $d \in [0, \frac{1}{S}]$ and unstable if $d \in (\frac{1}{S}, 1]$. Thus by (6) we have the following lemma.

Lemma 5.3

$$L(d) = d \cdot W(d) \textit{ if } d \in [0, \frac{1}{S}].$$

Moreover, if $d \in (\frac{1}{S}, 1]$ then $L(d) = W(d) = \infty$.

So, there is a direct relation between $L(d)$ and $W(d)$ and if you have computed one of them then the other follows from Lemma 5.3. In [16] the value of $W(d)$ was computed for $d = \frac{1}{S}$, the most right point of the interval of stability and traffic intensity $\rho = 1$. The following result was obtained.

Lemma 5.4 *For service time $S > 1$ we have that $W(\frac{1}{S}) = \frac{1}{2}$ if S is irrational. If $S = \frac{p}{q}$ with $p, q \in \mathbb{N}$, $\gcd(p, q) = 1$ then $W(\frac{1}{S}) = \frac{1}{2} - \frac{1}{2q}$.*

The following properties of the function $W(d)$ are proved in [7].

Theorem 5.5 *For given service time $S > 1$ the function $W(d)$ is continuous and monotonically increasing on the interval $[0, \frac{1}{S}]$. Moreover, on the interval $[\frac{1}{\lceil S \rceil}, \frac{1}{S}]$ the function $W(d)$ is strictly increasing.*

It also follows from the more general results in [14] that the function $L(d)$ is continuous and monotonically increasing on the interval of stability. We prove some properties of the function $L(d)$ which do not hold for $W(d)$. These properties of the function $L(d)$ will be used in our algorithm for computing $L_S(d)$ and $W_S(d)$.

Let u be a finite sequence of zeros and ones of length k and suppose that k consecutively arriving customers are routed to the queue according to u. Suppose that the first of these k customers arrives at time t_0 and thus the last at time $t_0 + k - 1$. Then we say that u lasts from t_0 to $t_0 + k$ and thus $V(t_0)$ is the workload at the beginning of u and $V(t_0 + k)$ is the workload at the end of u. We say that u is workload non-increasing if for every initial workload $V(t_0)$ it holds that $V(t_0 + k) \leq V(t_0)$. Note that from Lindley's equation (1) it follows that for this it is sufficient that $V(t_0) = 0$ implies that $V(t_0 + k) = 0$. The following lemma is useful for proving properties of $L(d)$.

Lemma 5.6 *Let $d \in \mathbb{Q}$, $0 \leq d \leq 1$. Then $\omega(d)$, the period word of the upper bracket sequence of density d, is workload non-increasing if and only if $d \leq \frac{1}{S}$.*

Proof. Let $d = \frac{p}{q}$ where $p, q \in \mathbb{N}$ with $\gcd(p, q) = 1$. Then $q = |\omega(d)|$ and we may assume that $\omega(d)$ lasts from some $t_0 \in \mathbb{Z}_{\geq 0}$ to $t_1 = t_0 + q \in \mathbb{Z}_{\geq 0}$. Suppose that $d > \frac{1}{S}$. Then by (2) we have for $S = S_i$

$$V(t_1) = V(t_0) + \kappa_{\omega(d)}(q) \cdot S_i + (m_i(t_1) - m_i(t_0)) - q \geq V(t_0) + p \cdot S_i - q = V(t_0) + q(d \cdot S_i - 1) > V(t_0).$$

Hence $\omega(d)$ is not workload non-increasing if $d > \frac{1}{S}$. Conversely, suppose that $d \leq \frac{1}{S}$ and $V(t_1) > V(t_0)$. Let $t^* = \max : t \in [t_0, t_1]$ for which $V(t^*) \leq V(t_0)$. Since $V(t)$ is monotonically non-increasing in any interval $(a, a + 1]$ for which $a \in \mathbb{Z}_{\geq 0}$ it follows that $t^* \in \mathbb{Z}_{\geq 0}$. Moreover, it follows that the server is non-idling in the interval $[t^*, t_1]$. So, by (2) and Lemma 2.4 we have that

$$V(t_1) = V(t^*) + (\kappa_{\omega(d)}(q) - \kappa_{\omega(d)}(t^* - t_0)) \cdot S - (t_1 - t^*) \leq V(t_0) + (p - \lceil d(t^* - t_0) \rceil) \cdot S - (t_0 + q - t^*) \leq$$

$$V(t_0) + (p - d(t^* - t_0)) \cdot S - (t_0 + q - t^*) \leq V(t_0) + (p - d(t^* - t_0)) \cdot \frac{1}{d} - (t_0 + q - t^*) = V(t_0),$$

which contradicts $V(t_1) > V(t_0)$. Hence $\omega(d)$ is workload non-increasing if $d \leq \frac{1}{S}$. $\square$

Suppose that $d \in \mathbb{Q}$, $d \leq \frac{1}{S}$ and that the customers are routed to the queue according to the upper bracket sequence $w(d) = \omega(d)^\infty$. Then it follows from Lemma 5.6 that if the workload is 0 at the beginning of an $\omega(d)$ factor then it is also 0 at the end of the factor and thus at the beginning of the next factor. So, since the workload is 0 at $t = 0$, the beginning of the first $\omega(d)$ factor, it follows that the workload is 0 at the end of every $\omega(d)$ factor. Thus the workload process is renewed after every $\omega(d)$ factor. Hence we have the following corollary.

Corollary 5.7 *If $d \in \mathbb{Q}$, $d \leq \frac{1}{S}$ then $W(d)$ is equal to the average waiting time of customers routed to the queue during the first period $\omega(d)$. Similarly $L(d)$ is equal to the average number of customers in the queue during the first period $\omega(d)$.*

In [14] the following theorem is shown in a more general form. In fact, the convexity of $L(d)$ for any stationary sequence of generally distributed interarrival times and stationary generally distributed service times is proved, which implies the special deterministic case. We give a simple proof for the deterministic queueing system of this paper.

Theorem 5.8 *For given service time $S > 1$ the function $L(d)$ is convex on the interval of stability $[0, \frac{1}{S}]$.*

Proof. Since $L(d)$, the average number of customers in the queue under the bracket policy with density d, is a continuous function of d on the interval of stability $[0, \frac{1}{S}]$, it is sufficient to show the following mid-point convexity of $L(d)$ (see [19] Theorem 1.7). Let $d_1 = \frac{p_1}{q_1}$ and $d_2 = \frac{p_2}{q_2}$ where $p_i, q_i \in \mathbb{N}$ with $\gcd(p_i, q_i) = 1$ and period words $w_i := \omega(d_i)$, $i = 1, 2$. Define the mid-point $d := \frac{1}{2}(d_1 + d_2)$ and show that $L(d) \leq \frac{1}{2}(L(d_1) + L(d_2))$. In order to prove this we consider the admission sequence

$$w = (w_1^{q_2} w_2^{q_1})^\infty \tag{8}$$

A simple calculation gives that the density of w equals

$$\frac{p_1 q_2 + p_2 q_1}{2 q_1 q_2} = d.$$

It follows from Lemma 5.6 that the queue is empty after each w_i factor. Since $| w_1^{q_2} | = | w_2^{q_1} | = q_1 q_2$ it follows that

$$L(w) = \frac{1}{2}(L(d_1) + L(d_2)).$$

Hence, since the bracket policy is optimal (see [1]) we have

$$L(d) \leq L(w) = \frac{1}{2}(L(d_1) + L(d_2)).$$

□

The following theorem shows that the function $L(d)$ is not only convex, but also piecewise linear. This property will be of great use in our algorithm for computing $L_S(d)$ and $W_S(d)$.

Theorem 5.9 *Let $d_1, d_2 \in [0, 1]$ be rational numbers such that $I = [d_1, d_2]$ is a Farey interval and $d_1 < d_2 \leq \frac{1}{S}$, where S is the service time. Let $d \in I$, $d = \lambda \cdot d_1 + (1 - \lambda) \cdot d_2$, where $\lambda \in [0, 1]$. Then $L(d) = \lambda \cdot L(d_1) + (1 - \lambda) \cdot L(d_2)$.*

Proof. Arriving customers are routed to the queue according to routing sequence $w(d)$ and to compute $L(d)$ we assume that the workload is 0 at starting moment $t = 0$. According to Theorem 4.4 $w(d)$ can be factored in factors $\omega(d_1)$ and $\omega(d_2)$. From the assumption $d_1 < d_2 \leq \frac{1}{S}$ and Lemma 5.6 it follows that both the factors $\omega(d_1)$ and $\omega(d_2)$ are workload non-increasing. Thus if the workload in the queue is 0 at the beginning of such a factor then it is also 0 at the end of the factor and thus it is 0 at the beginning of the next factor. So, the queue is always empty at the beginning of an $\omega(d_1)$ or $\omega(d_2)$ factor. Hence the average number of customers in the queue during any $\omega(d_1)$ or $\omega(d_2)$ factor is $L(d_1)$ respectively $L(d_2)$. Hence, if p is the fraction of time the system is in an $\omega(d_1)$ factor and thus $(1 - p)$ is the fraction of time the system is in an $\omega(d_2)$ factor then

$$L(d) = p \cdot L(d_1) + (1 - p) \cdot L(d_2). \tag{9}$$

It is easily seen that $p = \lambda$ and thus (9) is valid with $p = \lambda$. This implies the theorem. $\square$

Remark. Thus we have that the function $L(d)$ is linear on Farey intervals contained in the stability interval. The linearity of the function $L(d)$ on Farey intervals can be generalised to some more generally distributed service times if the following condition holds. If $[d_1, d_2]$ with $0 \le d_1 < d_2 \le 1$ is the Farey interval then both the $\omega(d_1)$ and $\omega(d_2)$ factor should be workload non-increasing with probability one. If this condition holds then the proof of Theorem 5.9 remains valid.

6 Computation of the average waiting time

6.1 The value of $W(d)$ in best lower approximation points

Definition 6.1 *Let $x > 0$ be a given real number and let $r = \frac{p}{q}$ with $p \in \mathbb{Z}_{\ge 0}$, $q \in \mathbb{N}$ with $\gcd(p, q) = 1$ be a rational number in the interval $(0, x]$. Then r is a best lower approximation of x if there does not exist a rational number in the interval $(r, x])$ with denominator smaller than or equal to q.*

Note that x is a best lower approximation of x itself if and only if x is a rational number.

Lemma 6.2 *Let $S > 1$ be the service time and d be a best lower approximation of $\frac{1}{S}$. Then during an $\omega(d)$ factor the server is non-idling before the last letter of the factor. Moreover, if $d = \frac{p}{q}$ where $p, q \in \mathbb{N}$ with $\gcd(p, q) = 1$ and the arriving customers are routed to the queue according to the upper bracket sequence $w(d)$ then*

$$V(t) = \lceil \lceil t \rceil d \rceil S - t \text{ for every } t \in [0, q-1]. \tag{10}$$

Proof. Let $d = \frac{p}{q}$ where $p, q \in \mathbb{N}$ with $\gcd(p, q) = 1$. Then $q = |\omega(d)|$ and we may assume that $\omega(d)$ lasts from some $t_0 \in \mathbb{Z}_{\ge 0}$ to $t_1 = t_0 + q \in \mathbb{Z}_{\ge 0}$. We have that $0 < d \le \frac{1}{S} < 1$ and thus $q \ge 2$. Recall that $V(t)$ denotes the workload in the queue at moment $t \ge 0$. Let $k \in \{1, 2, \dots, q-1\}$. Suppose that $\lceil d \cdot k \rceil S - k \le 0$. Then it follows that $d = \frac{p}{q} \le \frac{\lceil d \cdot k \rceil}{k} \le \frac{1}{S}$. Since $\frac{\lceil d \cdot k \rceil}{k}$ is a rational number with denominator smaller than q it follows that $d < \frac{\lceil d \cdot k \rceil}{k} \le \frac{1}{S}$, which contradicts that d is a best lower approximation of $\frac{1}{S}$. Thus $\lceil d \cdot k \rceil S - k > 0$. Hence by (2) and Lemma 2.4 we have that

$$V(t_0 + k) \ge V(t_0) + \kappa_{\omega(d)}(k) \cdot S - k = V(t_0) + \lceil d \cdot k \rceil S - k > V(t_0).$$

So, we have that $V(t_0 + k) > V(t_0)$ for every $k \in \{1, 2, \dots, q-1\}$ and thus $V(t) > 0$ for every $t \in (t_0, t_0 + q - 1]$. Thus the server is non-idling in the time interval $[0, q-1]$. So, by (2) and Lemma 2.4 we have for every $t \in [0, q-1]$ that

$$V(t) = V(0) + \kappa_{w(d)}(\lceil t \rceil) \cdot S - t = \lceil \lceil t \rceil d \rceil S - t.$$

$\square$

Theorem 6.3 *Let $S > 1$ be the service time and d be a best lower approximation of $\frac{1}{S}$. Let $d = \frac{p}{q}$ where $p, q \in \mathbb{N}$ with $\gcd(p, q) = 1$. Then*

$$W(d) = \frac{p-1}{2p}(pS + 1 - q) \text{ and} \tag{11}$$

$$L(d) = \frac{p-1}{2q}(pS + 1 - q). \tag{12}$$

Proof. We have that $d \leq \frac{1}{S}$. Thus according to Lemma 5.7 it holds that $W(d)$ is equal to the average waiting time in the first period $\omega(d)$. From $|\omega(d)| = q$ and (5) it follows that

$$W(d) = \frac{1}{\kappa_{\omega(d)}(q)} \cdot \sum_{t=1}^{q} \omega(d)_t \cdot V(t-1) = \frac{1}{p} \cdot \sum_{t=0}^{q-1} \omega(d)_{t+1} \cdot V(t). \tag{13}$$

By Definition 2.3 and (10) we have for $t = 0, 1 \ldots, q-1$ that

$$\omega(d)_{t+1} \cdot V(t) = (\lceil (t+1)d \rceil - \lceil t \cdot d \rceil)(\lceil td \rceil S - t). \tag{14}$$

Let $A = \{0, \lfloor \frac{1}{d} \rfloor, \lfloor \frac{2}{d} \rfloor, \ldots\}$. Then it follows by Lemma 2.5 for every $t \in \mathbb{Z}_{\geq 0}$ that

$$w(d)_{t+1} = \lceil (t+1)d \rceil - \lceil t \cdot d \rceil = \begin{cases} 1 & \text{if } t \in A \\ 0 & \text{if } t \notin A \end{cases} \tag{15}$$

On combining (13),(14) and (15) we obtain

$$W(d) = \frac{1}{p} \cdot \sum_{t \in A \cap \{0,1,\ldots,q-1\}} (\lceil td \rceil S - t) = \frac{1}{p} \cdot \sum_{t'=0}^{p-1} (\lceil \lfloor \frac{t'}{d} \rfloor d \rceil S - \lfloor \frac{t'}{d} \rfloor) =$$

$$\frac{S}{p} \cdot \sum_{t'=0}^{p-1} (\lceil \lfloor \frac{t'}{d} \rfloor d \rceil) - \frac{1}{p} \cdot \sum_{t'=0}^{p-1} (\lfloor t' \cdot \frac{q}{p} \rfloor). \tag{16}$$

It is easily seen that if $0 < x < 1$ then $\lceil \lfloor \frac{k}{x} \rfloor x \rceil = k$ for $k = 0, 1, \ldots$. Thus

$$\frac{S}{p} \cdot \sum_{t'=0}^{p-1} (\lceil \lfloor \frac{t'}{d} \rfloor d \rceil) = \frac{S}{p} \cdot \sum_{t'=0}^{p-1} t' = \frac{S(p-1)}{2}.$$

Moreover, analogously to Theorem 100 in [9] (cf. [12], Lemma 5.7) it follows that

$$\sum_{t'=0}^{p-1} (\lfloor t' \cdot \frac{q}{p} \rfloor) = \frac{1}{2}(p-1)(q-1).$$

Hence by (16) we have that

$$W(d) = \frac{S(p-1)}{2} - \frac{1}{2p}(p-1)(q-1) = \frac{p-1}{2p}(pS + 1 - q).$$

By Lemma 5.3 it follows that

$$L(d) = d \cdot W(d) = \frac{p-1}{2q}(pS + 1 - q).$$

□

So, for given service time $S > 1$ the values of $W(d)$ and $L(d)$ are easily obtained for all best lower approximations of $\frac{1}{S}$. Before we give a procedure to obtain iteratively the best lower approximations of $\frac{1}{S}$, we first have the following lemma.

Lemma 6.4 *If I is a subinterval of positive Lebesque measure of the open interval $(0, 1)$, then there exists a unique rational number of lowest denominator in I.*

Proof. Since I has positive Lebesque measure and the rational numbers are dense in the real numbers, I contains rational numbers. Suppose there exist $a, b \in \mathbb{N}$ such that $[\frac{a}{b}, \frac{a+1}{b}] \subseteq I$ and there exist no rational numbers in the interval $(\frac{a}{b}, \frac{a+1}{b})$ with denominator smaller or equal than b. Then $\frac{a}{b}, \frac{a+1}{b}$ are consecutive elements of the Farey series of order b (see [9], page 23) and thus we have by Theorem 28 in [9] that $(a+1)b - ba = b = 1$. Hence $\frac{a}{b}$ is an integer which contradicts that $I \subseteq (0,1)$. □

Let r_1 be the rational number of lowest denominator in the interval $(0, \frac{1}{S}]$ and put i := 1. Do the following iteratively. If $r_i = \frac{1}{S}$ then stop. Else if $r_i < \frac{1}{S}$ then let r_{i+1} be the rational number of lowest denominator in the interval $(r_i, \frac{1}{S}]$ and let $i := i+1$. According to Lemma 6.4 the r_i are well defined.

Lemma 6.5 *The number d is a best lower approximation of $\frac{1}{S}$ if and only if $d = r_i$ for some $i \in \mathbb{N}$. Moreover, if S is rational then there exists some $k \in \mathbb{N}$ such that*

$$\frac{1}{\lceil S \rceil} = r_1 < r_2 < \ldots < r_k = \frac{1}{S}.$$

If S is irrational then

$$\frac{1}{\lceil S \rceil} = r_1 < r_2 < \ldots \text{ and } \lim_{i \to \infty} r_i = \frac{1}{S}.$$

Proof. Let $r_i = \frac{p_i}{q_i}$ with $p_i, q_i \in \mathbb{N}$ and $\gcd(p_i, q_i) = 1$. We first prove that $r_1 = \frac{1}{\lceil S \rceil}$. Suppose that $q_1 < \lceil S \rceil$. Then $r_1 = \frac{p_1}{q_1} \geq \frac{1}{q_1} > \frac{1}{S}$, which gives a contradiction. Since $\frac{1}{\lceil S \rceil} \leq \frac{1}{S}$ we have that $q_1 = \lceil S \rceil$ and by Lemma 6.4 it follows that $r_1 = \frac{1}{\lceil S \rceil}$.
By definition and Lemma 6.4 we have that the r_i are best lower approximations. Moreover, both the sequence $r_1, r_2, \ldots$ of best lower approximations and the sequence $q_1, q_2, \ldots$ of denominators are strictly increasing. Let $d = \frac{a_1}{a_2}$ be a best lower approximation of $\frac{1}{S}$ with $a_1, a_2 \in \mathbb{N}$ and $\gcd(a_1, a_2) = 1$. Then there exist some $l \in \mathbb{N}$ such that $q_i \geq a_2$ if and only if $i \geq l$. From the definition it follows that $q_l = a_2$ and thus we have by Lemma 6.4 that $r_l = \frac{a_1}{a_2} = d$. So, d is a best lower approximation of $\frac{1}{S}$ if and only if $d = r_i$ for some $i \in \mathbb{N}$. If S is rational then $\frac{1}{S}$ is a best lower approximation of itself and thus we have that $r_k = \frac{1}{S}$ for some $k \in \mathbb{N}$. If S is irrational then $\frac{1}{S}$ is irrational and it follows from the definition that $r_i < \frac{1}{S}$ for every $i \in \mathbb{N}$. Thus we have that $r := \lim_{i \to \infty} r_i \leq \frac{1}{S}$. Suppose that $r < \frac{1}{S}$. Then by definition there exist no rational numbers in the interval $(r, \frac{1}{S})$, which contradicts the fact that the rational numbers are dense in the real numbers. □

Lemma 6.6 *If r_i, r_{i+1} are consecutive best lower approximations of $\frac{1}{S}$ then $[r_i, r_{i+1}]$ is a Farey interval.*

Proof. Let $r_i = \frac{a_1}{b_1}$ with $a_1, b_1 \in \mathbb{N}$, $\gcd(a_1, b_1) = 1$ and let $r_{i+1} = \frac{a_2}{b_2}$ with $a_2, b_2 \in \mathbb{N}$, $\gcd(a_2, b_2) = 1$. By definition and Lemma 6.4 we have that $b_1 < b_2$ and that there exist no rational numbers in the interval (r_i, r_{i+1}) with denominator smaller or equal than b_2. Hence $\frac{a_1}{b_1}, \frac{a_2}{b_2}$ are consecutive members of the Farey series of order b_2 and thus by Theorem 28 in [9] we have that $b_1 \cdot a_2 - a_1 \cdot b_2 = 1$. Thus $[r_i, r_{i+1}]$ is a Farey interval. □

For every service time $S > 1$ we have by Theorem 5.8 that the function $L(d)$ is convex on the interval $[0, \frac{1}{S}]$. Moreover, it follows from Lemma 5.2 that $L(d)$ is linear on $[0, r_1]$ and from Theorem 5.9 and Lemma 6.6 that $L(d)$ is linear on every interval $[r_i, r_{i+1}]$. Thus it follows from Lemma 6.5 that $L(d)$ is piecewise linear on the interval $[0, \frac{1}{S}]$. Therefore the slope of $L(d)$ can only change in the best lower approximation points r_i. In [7] these points r_i are called jump points. According to Lemma 5.3 $L(d) = W(d) = \infty$ for $d > \frac{1}{S}$.

The value of $L_S(d)$ and $W_S(d)$ can be computed exactly for every given service time $S > 1$ and $d \in [0, 1]$. The following algorithm can be used to compute $L(d)$ and $W(d)$ for any d.

Algorithm. If $0 \le d \le \frac{1}{\lceil S \rceil}$ then $L(d) = W(d) = 0$. Else if $\frac{1}{S} < d \le 1$ then $L(d) = W(d) = \infty$. Else if $d = \frac{1}{S}$ then $W(d)$ is given by Lemma 5.4 and $L(d) = d \cdot W(d)$. Else we have that $\frac{1}{\lceil S \rceil} < d < \frac{1}{S}$. Then there exist consecutive best lower approximations points r_i and r_{i+1} such that $d \in [r_i, r_{i+1}]$ (in the following subsection we give a method based on the continued fraction expansion of $\frac{1}{S}$ to determine these points r_i, r_{i+1} in an efficient way). If $d = r_i$ or $d = r_{i+1}$ then $W(d)$ and $L(d)$ are given by Theorem 6.3. Else $r_i < d < r_{i+1}$. Then compute $L(r_i)$ and $L(r_{i+1})$ by applying Theorem 6.3 and put $\lambda = \frac{r_{i+1}-d}{r_{i+1}-r_i}$. Then $L(d) = \lambda \cdot L(r_1) + (1-\lambda) \cdot L(r_2)$ and we can compute $W(d)$ by $W(d) = \frac{L(d)}{d}$.

We illustrate the algorithm by the following example.
Example. We calculate $W_S(d)$ for $S = \frac{31}{19}$ and $d = \frac{\sqrt{6}}{4}$. The best lower approximations of $\frac{1}{S} = \frac{19}{31}$ are $r_1 = \frac{1}{2}$, $r_2 = \frac{3}{5}$, $r_3 = \frac{11}{18}$ and $r_4 = \frac{19}{31}$. Since $d \in [\frac{11}{18}, \frac{19}{31}]$ we obtain from Theorem 6.3 that $L(\frac{11}{18}) = \frac{5}{19}$ and $L(\frac{19}{31}) = \frac{9}{31}$. We have $\lambda = \frac{\frac{19}{31}-d}{\frac{19}{31}-\frac{11}{18}} = 342 - \frac{279\sqrt{6}}{2}$ and $L(d) = \lambda \cdot L(\frac{11}{18}) + (1-\lambda) \cdot L(\frac{19}{31}) = \frac{72\sqrt{6}}{19} - 9$. Hence $W(d) = \frac{L(d)}{d} = \frac{288}{19} - 6\sqrt{6}$.

6.2 Best lower approximations and the continued fraction expansion

In this subsection we show that the best lower approximations $r_1, r_2, \ldots$ of $\frac{1}{S}$ can be obtained by applying the continued fraction expansion of $\frac{1}{S}$. This gives for given $d \in (\frac{1}{\lceil S \rceil}, \frac{1}{S})$ an efficient algorithm for computing consecutive best lower approximation r_i, r_{i+1} of $\frac{1}{S}$ such that $d \in [r_i, r_{i+1}]$.

See for example [9] for definitions and properties of the continued fraction expansion of a real number. Put $\alpha := \frac{1}{S}$. The partial quotients $a_0, a_1, \ldots$ of the (simple) continued fraction expansion of α are recursively defined by:

$$\left\{ \begin{array}{ll} a_0 = \lfloor \alpha \rfloor; & \alpha_1 = \frac{1}{\alpha - a_0} \\ a_n = \lfloor \alpha_n \rfloor; & \alpha_{n+1} = \frac{1}{\alpha_n - a_n} \text{ for } n = 1, 2, \ldots \end{array} \right\}.$$

Then

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots \cfrac{1}{a_n + \cdots}}} := [a_0, a_1, \ldots, a_n, \ldots].$$

Since $S > 1$ we have that $0 < \alpha < 1$ and thus $a_0 = 0$. Moreover, note that $a_1, a_2, \ldots$ are positive integers. If α is rational then $\alpha_m - a_m = 0$ for some in $m \in \mathbb{N}$ and the process of computing the partial quotients stops for $n = m$. Then the continued fraction expansion of α is finite and $\alpha = [0, a_1, a_2, \ldots, a_m]$. If α is irrational then the continued fraction expansion of α is infinite.

We define p_n, q_n recursively by

$$\begin{array}{lll} p_0 = a_0, & p_1 = a_0 a_0 + 1, & p_n = a_n p_{n-1} + p_{n-2} (n \ge 2), \\ q_0 = 1, & q_1 = a_1, & q_n = a_n q_{n-1} + q_{n-2} (n \ge 2) \end{array}.$$

Then $\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n]$ and $\frac{p_n}{q_n}$ is called the nth convergent of $\alpha = [a_0, a_1, \ldots]$. If $\alpha = [a_0, a_1, \ldots, a_m]$ is rational then the mth convergent $\frac{p_m}{q_m} = \alpha$ is the last convergent of α and thus the number of convergents of α is finite. If α is irrational then the number of convergents of α is infinite. We put $x_n := \frac{p_n}{q_n}$. Then the convergents of α have the following properties (see [9]).

Lemma 6.7 **(i)** *The even convergents x_{2n} increase strictly with n, while the odd convergents x_{2n+1} decrease strictly.*

(ii) *Every odd convergent is greater than any even convergent.*

(iii) *Every even convergent is smaller than α and every odd convergent is greater than α (except that in case of rational α the last convergent is equal to α, whether this be even or odd).*

(iv) *If α is irrational then $\lim_{n\to\infty} x_n = \alpha$.*

Lemma 6.8 *If $n > 1$, $0 < q \le q_n$, and $\frac{p}{q} \ne \frac{p_n}{q_n}$, then*

$$|\frac{p_n}{q_n} - \alpha | < | \frac{p}{q} - \alpha | .$$

Lemma 6.7 and Lemma 6.8 have the following corollary.

Corollary 6.9 *The even convergents $\frac{p_{2n}}{q_{2n}}$, $n = 1, 2, \ldots$ are best lower approximations of α.*

So, the even convergents obtained by the continued fraction expansion of α are best lower approximations of α. However, in general the converse is not true. First of all if α is equal to the n-th convergent $\frac{p_n}{q_n}$ for some odd positive integer then this odd convergent is also a best approximation of α. Moreover, to obtain all best lower approximations of α from the continued fraction of α we also have to consider the so-called even intermediate convergents, which we define now.

Definition 6.10 *Let $\alpha = [a_0, a_1, \ldots]$ if α is irrational or $\alpha = [a_0, a_1, \ldots, a_m]$ for some $m \in \mathbb{N}$ if α is rational. Then a rational number $\frac{p}{q}$ is an even intermediate convergent of α if and only if $p = p_{n-2} + c \cdot p_{n-1}$ and $q = q_{n-2} + c \cdot q_{n-1}$ for some positive even integer n (with n smaller or equal than m if α is rational) and $c \in \{1, 2, \ldots, a_n - 1\}$, where $\frac{p_n}{q_n}$ is the n-th convergent of α.*

Note that for every positive even integer $n \le m$ we have that

$$\frac{p_{n-2}}{q_{n-2}} < \frac{p_{n-2} + p_{n-1}}{q_{n-2} + q_{n-1}} < \frac{p_{n-2} + 2p_{n-1}}{q_{n-2} + 2q_{n-1}} < \ldots < \frac{p_{n-2} + (a_n - 1)p_{n-1}}{q_{n-2} + (a_n - 1)q_{n-1}} < \frac{p_{n-2} + (a_n)p_{n-1}}{q_{n-2} + (a_n)q_{n-1}} = \frac{p_n}{q_n} \le \alpha. \tag{17}$$

So, all even intermediate convergents of α are smaller than α. Moreover, it can be shown (see for example [21]) that every even intermediate convergent $\frac{p}{q}$ of α is a best lower approximations of α. Conversely, we have the following lemma, which follows directly from results in [21].

Lemma 6.11 *Let $r_i = \frac{p}{q}$ with $\gcd(p, q) = 1$ be a best lower approximation of $\alpha \in (0, 1)$. Then one of the following three possibilities holds for r_i.*

- *r_i is an even convergent $\frac{p_{2n}}{q_{2n}}$ with $n \ge 1$ of α.*
- *r_i is an even intermediate convergent of α.*

- *r_i is equal to α, which is an odd convergent of itself.*

Recall that we ordered the best lower approximations $r_1, r_2, \dots$ of $\alpha \in (0,1)$ such that $r_1 < r_2 < \dots$ and that $\frac{p_n}{q_n}$ is the n -th convergent of α. Using the preceding results it is possible to identify the ordered set $B := (r_1, r_2, \dots)$ of best lower approximations of α in terms of the convergents and the partial quotients of α. Doing this we have three slightly different cases. First note that $p_0 = 0$ and $q_0 = 1$.

1. If $\alpha = [0, a_1, a_2, \dots]$ is irrational then the ordered set B is infinite,

$$B = (\frac{p_1}{q_1+1}, \frac{2p_1}{2q_1+1}, \dots, \frac{a_2 p_1}{a_2 q_1 + 1} = \frac{p_2}{q_2}, \frac{p_3+p_2}{q_3+q_2}, \frac{2p_3+p_2}{2q_3+q_2}, \dots, \frac{p_4}{q_4}, \dots). \tag{18}$$

2. If α is rational and an even convergent of itself, $\alpha = [0, a_1, a_2, \dots, a_k] = \frac{p_k}{q_k}$ where k is even, then B is finite,

$$B = (\frac{p_1}{q_1+1}, \frac{2p_1}{2q_1+1}, \dots, \frac{a_2 p_1}{a_2 q_1 + 1} = \frac{p_2}{q_2}, \frac{p_3+p_2}{q_3+q_2}, \frac{2p_3+p_2}{2q_3+q_2}, \dots,$$

$$\frac{p_{k-2}}{q_{k-2}}, \frac{p_{k-1}+p_{k-2}}{q_{k-1}+q_{k-2}}, \frac{2p_{k-1}+p_{k-2}}{2q_{k-1}+q_{k-2}}, \dots, \frac{a_k p_{k-1}+p_{k-2}}{a_k q_{k-1}+q_{k-2}} = \frac{p_k}{q_k} = \alpha). \tag{19}$$

3. If α is rational and an odd convergent of itself, $\alpha = [0, a_1, a_2, \dots, a_k] = \frac{p_k}{q_k}$ where k is odd, then B is finite,

$$B = (\frac{p_1}{q_1+1}, \frac{2p_1}{2q_1+1}, \dots, \frac{a_2 p_1}{a_2 q_1 + 1} = \frac{p_2}{q_2}, \frac{p_3+p_2}{q_3+q_2}, \frac{2p_3+p_2}{2q_3+q_2}, \dots, \frac{p_{k-1}}{q_{k-1}}, \frac{p_k}{q_k} = \alpha). \tag{20}$$

Example. Suppose that $\alpha = \frac{339}{410}$. Then the continued fraction algorithm gives that $\alpha = [0, 1, 4, 1, 3, 2, 3, 2]$ and the convergents $\{\frac{p_n}{q_n}\}_{n=0}^{7}$ of α are $\frac{0}{1}, \frac{1}{1}, \frac{4}{5}, \frac{5}{6}, \frac{19}{23}, \frac{43}{52}, \frac{148}{179}, \frac{339}{410}$. By (20) it follows that the ordered set of best lower approximations of α is $B = (\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{9}{11}, \frac{14}{17}, \frac{19}{23}, \frac{62}{75}, \frac{105}{127}, \frac{148}{179}, \frac{339}{410})$.

We now give the algorithm based on continued fraction expansion for computing for any $d \in (\frac{1}{\lceil S \rceil}, \frac{1}{S})$ two consecutive best lower approximations r_i, r_{i+1} of $\alpha = \frac{1}{S}$ such that $d \in [r_i, r_{i+1}]$.

Algorithm. Apply the continued fraction algorithm to find consecutively the partial quotients $a_1, a_2, \dots$ and the corresponding convergents $\frac{p_1}{q_1}, \frac{p_2}{q_2}, \dots$ of α until we have found an even convergent $\frac{p_{2n}}{q_{2n}}$, that is greater or equal than d or we have that $\frac{p_{N-1}}{q_{N-1}} < d < \frac{p_N}{q_N} = \alpha$ for some odd integer N, $N > 1$. Note that according to Lemma 6.7 this first part of the algorithm stops after finite time. If we have the latter case then we put $r_i = \frac{p_{N-1}}{q_{N-1}}$ and $r_{i+1} = \frac{p_N}{q_N}$. Then $d \in (r_i, r_{i+1})$ and we are finished. So, suppose we have the former case. Then by Lemma 6.7 we have that $\frac{p_{2n-2}}{q_{2n-2}} < d \le \frac{p_{2n}}{q_{2n}}$ and by (17) there exists some unique integer k, $0 \le k < a_{2n}$ such that $\frac{k \cdot p_{2n-1} + p_{2n-2}}{k \cdot q_{2n-1} + q_{2n-2}} < d \le \frac{(k+1) \cdot p_{2n-1} + p_{2n-2}}{(k+1) \cdot q_{2n-1} + q_{2n-2}}$. It is easily seen that this holds for $k = \lfloor \frac{d \cdot q_{2n-2} - p_{2n-2}}{p_{2n-1} - d \cdot q_{2n-1}} \rfloor$. By putting $r_i = \frac{k \cdot p_{2n-1} + p_{2n-2}}{k \cdot q_{2n-1} + q_{2n-2}}$ and $r_{i+1} = \frac{(k+1) \cdot p_{2n-1} + p_{2n-2}}{(k+1) \cdot q_{2n-1} + q_{2n-2}}$ we have that $d \in (r_i, r_{i+1}]$ and we are finished.

We give an example to illustrate this algorithm.

Example.

Suppose $\alpha := \frac{1}{S} = \frac{7433}{9734} + \frac{\sqrt{18773}}{9734}$ and $d = 0.77761$. Then applying the continued fraction algorithm to α we consecutively find $a_0 = 0$, $\frac{p_0}{q_0} = \frac{0}{1}$, $a_1 = 1$, $\frac{p_1}{q_1} = 1$, $a_2 = 3$, $\frac{p_2}{q_2} = \frac{3}{4}$, $a_3 = 2$, $\frac{p_3}{q_3} = \frac{7}{9}$, $a_4 = 137$, $\frac{p_4}{q_4} = \frac{962}{1237}$ and we have that $d \leq \frac{p_4}{q_4}$. Then we compute that $k = \lfloor \frac{d \cdot q_2 - p_2}{p_3 - d \cdot q_3} \rfloor = 73$ and it follows that $r_i = \frac{k \cdot p_3 + p_2}{k \cdot q_3 + q_2} = \frac{514}{661}$ and $r_{i+1} = \frac{(k+1) \cdot p_3 + p_2}{(k+1) \cdot q_3 + q_2} = \frac{521}{670}$.

Remark. The number of operations needed for this algorithm based on the continued fraction expansion is linear in the number of digits of the denominator of r_{i+1}. So, this algorithm is very efficient in practice. Moreover, as illustrated in the last example of the previous subsection the exact value of $W_S(d)$ can be easily calculated after we have obtained the consecutive best lower approximations r_i, r_{i+1} of $\frac{1}{S}$ such that $d \in (r_i, r_{i+1}]$. So, this is an efficient method for calculating the exact value of $W_S(d)$ and in most cases it should be quicker than the recursive algorithm used in [7]. Indeed, if this recursive algorithm is applied then, when $d \in (r_i, r_{i+1}]$, all the best lower approximations $r_1 < r_2 < \ldots < r_i < r_{i+1}$ of $\frac{1}{S}$ and the corresponding value of $W_S(r_n)$ for $n = 1, 2, \ldots, i+1$ are explicitly calculated. So, if q is the denominator of r_{i+1} then this recursive algorithm for computing $W_S(d)$ has to do a number of order q calculations in worst case. In our algorithm not all the best lower approximations of $\frac{1}{S}$ are explicitly calculated. Indeed, only the convergents of $\frac{1}{S}$ and the last two best lower approximations r_i, r_{i+1} with corresponding values $W_S(r_i)$ and $W_S(r_{i+1})$ are calculated. Therefore our algorithm only has to do a number of order $\log(q)$ calculations.

7 Minimization over multiple queues

7.1 Properties of a minimal point

In this section we obtain results on the minimimal average waiting time in case of N parallel queues as introduced in section 2.

For a given $\overline{S} := (S_1, S_2, \ldots, S_N)$ queueing system with N parallel FIFO queues and $(d_1, d_2, \ldots, d_n) \in [0,1]^N$ we put

$$L_{\overline{S}}(d_1, d_2, \ldots, d_n) = L(d_1, d_2, \ldots, d_n) := \sum_{i=1}^{N} L_{S_i}(d_i).$$

Let U be a routing policy applied in an $\overline{S} = (S_1, S_2, \ldots, S_N)$ system such that routing sequence u^i has density d_i for $i = 1, 2, \ldots, N$. Then by Proposition 3.2 and Lemma 5.1 we have that

$$W(U) = \sum_{i=1}^{N} L^i(u^i) \geq \sum_{i=1}^{N} L_{S_i}(d_i) = L_{\overline{S}}(d_1, d_2, \ldots, d_n) \tag{21}$$

Put

$$H^N := \{(x_1, x_2, \ldots, x_N) \in \mathbb{R}^N : x_i \geq 0, i = 1, 2, \ldots, N \text{ and } \sum_{i=1}^{N} x_i = 1.$$

It is easily seen that if some routing policy U has densities $d_1, d_2, \ldots, d_N$ then $(d_1, d_2, \ldots, d_N) \in H^N$. Moreover, since H^N is a convex and compact subset of $\mathbb{R}^N$ and $L_{\overline{S}}(d_1, d_2, \ldots, d_n)$ is convex in $(d_1, d_2, \ldots, d_n)$ by Theorem 5.8 we have the following lemma.

Lemma 7.1 *For an* $\overline{S} := (S_1, S_2, \ldots, S_N)$ *queueing system put*

$$M(\overline{S}) := \{(d_1, d_2, \ldots, d_N) \in H : L_{\overline{S}}(d_1, d_2, \ldots, d_N) = \min_{H^N} L_{\overline{S}}(x_1, x_2, \ldots, x_N)\}. \tag{22}$$

Then $M(\overline{S})$ *is an nonempty convex and compact subset of* H^N.

We put

$$R(\overline{S}) = R((S_1, S_2, \ldots, S_N)) := L_{\overline{S}}(d_1, d_2, \ldots, d_n), \text{ where } (d_1, d_2, \ldots, d_n) \in M(\overline{S}).$$

Then $R(\overline{S})$ is a lower bound for the average waiting time in the $\overline{S}$ system, since Theorem 4.1 and Theorem 4.2 of [3] give the following result, where we do not require that the routing sequences u^i have densities d_i for $i = 1, 2, \ldots, N$.

Proposition 7.2 *For any routing policy* U *applied to the* $\overline{S} := (S_1, S_2, \ldots, S_N)$ *queueing system we have that*

$$W(U) \geq R(\overline{S}).$$

We also have upper bounds for the average waiting time in the $\overline{S} = (S_1, S_2, \ldots, S_N)$ system. By Theorem 4.7 in [12] we have the following proposition.

Proposition 7.3 *Let* U *be a periodic routing policy such that the routing sequence* u^i *has density* d_i *for* $i = 1, 2, \ldots, N$. *Then*

$$W(U) \leq L_{\overline{S}}(d_1, d_2, \ldots, d_N) + \overline{O}(U),$$

where $\overline{O}(U) \geq 0$ *is the so-called total primal unbalance of* U, *which is some kind of measure of the irregularity of the sequence (see [12] or [13] for the definition of unbalance).*

The total primal unbalance can always be bounded. Indeed, by Theorem 5.1 in [12] we have the following proposition.

Proposition 7.4 *Let rational densities* $d_1, d_2, \ldots, d_N \in \mathbb{Q}_{>0}$ *with* $\sum_{i=1}^N d_i = 1$ *be given. Then there exists a periodic routing policy* U *such that routing sequence* u^i *has density* d_i *for* $i = 1, 2, \ldots, N$ *and* $\overline{O}(U) \leq \frac{N}{2} - 1$.

We note that such a routing sequence policy U with given densities d_i and $\overline{O}(U) \leq \frac{N}{2} - 1$ can easily be obtained by constructing a so-called billiard sequence (see [12]). Moreover, the upper bound of Proposition 7.3 is tight. Note that by the definition of $L_{\overline{S}}(d_1, d_2, \ldots, d_N)$ we have that $W(U) = L_{\overline{S}}(d_1, d_2, \ldots, d_N)$ if for every $i \in \{1, 2 \ldots, N\}$ the routing sequence u^i is a regular sequence of density d_i. Indeed, from the definition of the total primal unbalance it follows that $\overline{O}(U) = 0$ if and only if u^i is a regular sequence for $i = 1, 2, \ldots, N$.
Suppose that $\sum_{i=1}^N \frac{1}{S_i} < 1$. Then for every $(x_1, x_2, \ldots, x_N) \in H^N$ there exists some i for which $x_i > \frac{1}{S_i}$ and thus we have by Lemma 5.3 that $R(\overline{S}) = \infty$. So, by Proposition 7.2 it follows that $W(U) = \infty$ for every routing policy U. Thus the queueing system is unstable in this case and optimal policies do not exist. Therefore we assume in the sequel that $\sum_{i=1}^N \frac{1}{S_i} \geq 1$. We shall show in Proposition 7.14 that the lower bound $R(\overline{S})$ is finite and in Theorem 7.15 that there exist policies U with $W(U) \leq \frac{N-1}{2}$.

Moreover, we characterize the set $M(\overline{S})$ and we show that this set contains some special rational point if $\sum_{i=1}^{N}\frac{1}{S_i} > 1$. From this it follows that there exist a periodic policy for which the average waiting time is bounded by $R((S_1, S_2, \ldots, S_N)) + \frac{N}{2} - 1$. A corollary of this result is that there exists a periodic policy which is optimal if $N = 2$, which was proved in [7]. Finally we give an algorithm to obtain a rational point $x \in M(\overline{S})$ and the corresponding lower bound $R(\overline{S})$. If $N = 2$ then the rational point obtained by the algorithm yields an optimal routing policy as in [7]. If $N > 2$ then in general it is not possible to obtain an optimal policy from the rational point, but it can be used to construct periodic policies with performance close to the lower bound $R(\overline{S})$.

For given service time $S > 1$ and $d \in [0, 1]$ we define

Definition 7.5

$$l_S(d) = \begin{cases} 0 & \text{if } d \in [0, \frac{1}{\lceil S \rceil}] \\ \lim_{\varepsilon \downarrow 0} \frac{L_S(d) - L_S(d-\varepsilon)}{\varepsilon} & \text{if } d \in (\frac{1}{\lceil S \rceil}, \frac{1}{S}] \\ \infty & \text{if } d \in (\frac{1}{S}, 1] \end{cases} \tag{23}$$

and

$$r_S(d) = \begin{cases} 0 & \text{if } d \in [0, \frac{1}{\lceil S \rceil}) \\ \lim_{\varepsilon \downarrow 0} \frac{L_S(d+\varepsilon) - L_S(d)}{\varepsilon} & \text{if } d \in [\frac{1}{\lceil S \rceil}, \frac{1}{S}) \\ \infty & \text{if } d \in [\frac{1}{S}, 1] \end{cases} . \tag{24}$$

From Theorem 5.8 it follows that $l_S(d)$ and $r_S(d)$ are monotonically non-decreasing in d. Moreover, they assume only countably many values, since they are constant between consecutive best lower approximations of $\frac{1}{S}$.

The following lemma follows directly from Theorem 6.3 and Lemma 6.5.

Lemma 7.6 *If $S > 1$ is rational then $l_S(\frac{1}{S}) < \infty$. If $S > 1$ is irrational then $l_S(\frac{1}{S}) = \infty$.*

The following proposition gives an necessary and sufficient condition for elements of $M(\overline{S})$ and thus the set $M(\overline{S})$ is completely characterized by this.

Proposition 7.7 *Let $\overline{S} = (S_1, S_2, \ldots, S_N)$ and $x = (x_1, x_2, \ldots, x_N) \in H^N$. Then $x \in M(\overline{S})$ if and only if for every pair $i, j \in \{1, 2, \ldots, N\}$ it holds that either $l_{S_i}(x_i) \leq r_{S_j}(x_j) < \infty$ or $r_{S_j}(x_j) = \infty$.*

Proof. We first prove the " only if " part. Suppose that $x \in M(\overline{S})$ and there exist $i, j \in \{1, 2, \ldots, N\}$ such that $r_{S_j}(x_j) < \infty$ and $l_{S_i}(x_i) > r_{S_j}(x_j)$. Since $\sum_{i=1}^{N}\frac{1}{S_i} \geq 1$ and thus $L_{\overline{S}}(x_1, x_2, \ldots, x_N) = R(\overline{S}) < \infty$ it follows that $0 \leq x_n \leq \frac{1}{S_n}$ for $n = 1, 2, \ldots, N$. Moreover, it follows from (23) and (24) that $x_i > \frac{1}{\lceil S_i \rceil} > 0$ and $x_j < \frac{1}{S_j}$. Put $y(\epsilon) := (y_1, y_2, \ldots, y_N)$, where $y_i := x_i - \varepsilon$, $y_j := x_j + \varepsilon$ and $y_l := x_l$ for $l \neq i, j$. Then $L_{\overline{S}}(x) - L_{\overline{S}}(y(\varepsilon)) = \sum_{n=1}^{N}(L_{S_n}(x_n) - L_{S_n}(y_n)) = (L_{S_i}(x_i) - L_{S_i}(x_i - \varepsilon)) - (L_{S_j}(x_j + \varepsilon) - L_{S_j}(x_j))$ and thus

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \cdot (L_{\overline{S}}(x) - L_{\overline{S}}(y(\varepsilon))) = l_{S_i}(x_i) - r_{S_j}(x_j) > 0.$$

So, if $\varepsilon' > 0$ is an arbitrarily small positive number then $y(\varepsilon') \in H^N$ and $L_{\overline{S}}(x) - L_{\overline{S}}(y(\varepsilon')) > 0$, which contradicts the fact that $x \in M(\overline{S})$.
The " if " part follows from the convexity of the function. $\square$

By Lemma 7.6 and Proposition 7.7 we have the following corollary.

Corollary 7.8 *Let $x = (x_1, x_2, \ldots, x_N) \in M(\overline{S})$ where $\overline{S} = (S_1, S_2, \ldots, S_N)$ with $\sum_{i=1}^N \frac{1}{S_i} > 1$. Then for every $j \in \{1, 2, \ldots, N\}$ for which S_j is irrational we have that $x_j < \frac{1}{S_j}$.*

Proof. Since $\sum_{i=1}^N \frac{1}{S_i} \geq 1$ we have that $R(\overline{S}) < \infty$ and thus $x_i \leq \frac{1}{S_i}$ for $i = 1, 2, \ldots, N$. Suppose that S_j is irrational and $x_j = \frac{1}{S_j}$. Then by Lemma 7.6 we have that $l_{S_j}(x_j) = \infty$. Since $\sum_{i=1}^N \frac{1}{S_i} > 1$ and $\sum_{i=1}^N x_i = 1$ there exists some $k \neq j$ for which $x_k < \frac{1}{S_k}$. Then $r_{S_k}(x_k) < \infty = l_{S_j}(x_j)$, which according to Proposition 7.7 contradicts the fact that $x = (x_1, x_2, \ldots, x_N) \in M(\overline{S})$ $\square$.

Theorem 7.9 *Let $\overline{S} = (S_1, S_2, \ldots, S_N)$ be such that $\sum_{i=1}^N \frac{1}{S_i} > 1$. Then there exists some $x = (x_1, x_2, \ldots, x_N) \in M(\overline{S})$ and some $j \in \{1, 2, \ldots, N\}$ such that for every $i \neq j$ it holds that x_i is either a best lower approximation of $\frac{1}{S_i}$ or $x_i = 0$.*

Proof. Suppose that $x = (x_1, x_2, \ldots, x_N) \in M(\overline{S})$ and there exist $k, l \in \{1, 2, \ldots, N\}$, $k < l$ such that $x_k, x_l \neq 0$ and neither x_k is a best lower approximation of $\frac{1}{S_k}$ nor x_l is a best lower approximation of $\frac{1}{S_l}$. Put $r_0 = 0$ and let $\frac{1}{\lceil S_k \rceil} = r_1 < r_2 < \ldots$ be the best lower approximations of $\frac{1}{S_k}$. Similarly, put $s_0 = 0$ and let $\frac{1}{\lceil S_l \rceil} = s_1 < s_2 < \ldots$ be the best lower approximations of $\frac{1}{S_l}$. By Corollary 7.8 we have that $0 < x_k < \frac{1}{S_k}$ and $0 < x_l < \frac{1}{S_l}$. Hence there exist $n, m \in \mathbb{Z}_{\geq 0}$ such that $r_n < x_k < r_{n+1}$ and $s_m < x_l < s_{m+1}$. By Theorem 5.9 we have that the functions $L_{S_k}(d)$ and $L_{S_l}(d)$ are linear in d on the intervals $[r_n, r_{n+1}]$ and $[s_m, s_{m+1}]$ respectively. Let $x(\varepsilon) = (x_1, x_2, \ldots, x_{k-1}, x_k + \varepsilon, x_{k+1}, \ldots, x_{l-1}, x_l + \varepsilon, x_{l+1}, \ldots, x_N)$ and put $A = \min(r_{n+1} - x_k, x_l - s_m)$ and $B = \max(r_n - x_k, x_l - x_{m+1})$. Then there exists some $C \in \mathbb{R}$ such that

$$L_{\overline{S}}(x(\varepsilon)) = R(\overline{S}) + C \cdot \varepsilon$$

for every $\varepsilon \in [B, A]$. Since $B < 0$, $A > 0$ and $L_{\overline{S}}(x(\varepsilon)) \geq R(\overline{S})$ for every $\varepsilon \in \mathbb{R}$ it follows that $C = 0$. Hence $x(\varepsilon) \in M(\overline{S})$ for every $\varepsilon \in [B, A]$. In particular we have that $x(A) \in M(\overline{S})$ and thus either $(x_1, x_2, \ldots, x_{k-1}, r_{n+1}, x_{k+1}, \ldots, x_{l-1}, x_l - A, x_{l+1}, \ldots, x_N) \in M(\overline{S})$ or $(x_1, x_2, \ldots, x_{k-1}, x_k + A, x_{k+1}, \ldots, x_{l-1}, s_m, x_{l+1}, \ldots, x_N) \in M(\overline{S})$. The theorem follows by induction. $\square$

Corollary 7.10 *Let $\overline{S} = (S_1, S_2, \ldots, S_N)$ be such that $\sum_{i=1}^N \frac{1}{S_i} > 1$. Then the set $M(\overline{S})$ contains a point with rational coordinates.*

Remark. In fact, if $R(\overline{S}) > 0$ then in most cases the set $M(\overline{S})$ consists of only one point, which according to Corollary 7.10 and Theorem 7.9 has rational coordinates and all coordinates except at most one are best lower approximations of the inverse of the service time of the corresponding server.

By Proposition 7.3, Proposition 7.4 and Corollary 7.10 we have the following theorem.

Theorem 7.11 *For every $\overline{S} = (S_1, S_2, \ldots, S_N)$ system with $\sum_{i=1}^N \frac{1}{S_i} > 1$ there exists a periodic policy U such that $W(U) \leq R(\overline{S}) + \frac{N}{2} - 1$.*

Proof. By Corollary 7.10 there exists a point $(d_1, d_2, \ldots, d_N) \in M(\overline{S})$ with $d_i \in \mathbb{Q}_{\geq 0}$ for $i = 1, 2, \ldots, N$. By Proposition 7.4 there exists a periodic policy U such that for $i = 1, 2, \ldots, N$ the routing sequence u^i has density d_i for $i = 1, 2, \ldots, N$ and $\overline{O}(U) \leq \frac{N}{2} - 1$. Then by Proposition 7.3 we have that

$$W(U) \leq L_{\overline{S}}(d_1, d_2, \ldots, d_N) + \overline{O}(U) \leq R(\overline{S}) + \frac{N}{2} - 1.$$

□

From the sequel it becomes clear how such a policy U can be obtained. By Proposition 7.2 and Theorem 7.11 we obtain that for $N = 2$ there exist optimal periodic policies if $\frac{1}{S_1} + \frac{1}{S_2} > 1$ (see [7]).

Corollary 7.12 *For an $\overline{S} = (S_1, S_2)$ system with $\frac{1}{S_1} + \frac{1}{S_2} > 1$ there exists an optimal periodic policy U with $W(U) = R(\overline{S})$.*

Such an optimal periodic policy can be obtained as follows. Find a rational point $(d_1, d_2) \in M(\overline{S})$. Let u^1 be the upper bracket sequence of density d_1 (or any other regular sequence of zeros and ones of density d_1). Then we construct the routing policy U by taking u^1 and replacing every zero by a two. Since d_1 is rational we have that u^1 and thus also U is periodic. Moreover, u^1 is a regular sequence of density d_1 and u^2 is a regular sequence of density $1 - d_1 = d_2$, since u^2 is the complement (every zero is an one and every one is a zero) of u^1. Hence $\overline{O}(U) = 0$ and $W(U) = L_{\overline{S}}(d_1, d_2) = R(\overline{S})$.

In general if we have a routing policy U such that u^i is a regular sequence of density d_i for $i = 1, 2, \ldots, N$ and $(d_1, d_2, \ldots, d_N) \in M(\overline{S})$ then $W(U) = L_{\overline{S}}(d_1, d_2, \ldots, d_N) = R(\overline{S})$ and it follows that this routing policy U is optimal. However, if $N > 2$ then generally there does not exist a routing policy U such that u^i is a regular sequence of density d_i for $i = 1, 2, \ldots, N$ (see [2]). So, in that case it is possible that $W(U) > R(\overline{S})$ for every policy U.

If $\sum_{i=1}^{N} \frac{1}{S_i} = 1$ then $M(\overline{S})$ consists of the single point $(\frac{1}{S_1}, \frac{1}{S_2}, \ldots, \frac{1}{S_N})$, which in general is not a rational point. However, from [16] we have the following result.

Proposition 7.13 *For every $\overline{S} = (S_1, S_2, \ldots, S_N)$ system with $\sum_{i=1}^{N} \frac{1}{S_i} = 1$ there exists a policy U such that $W(U) \leq R(\overline{S}) + \frac{N}{2} - 1$.*

We have the following proposition.

Proposition 7.14 *For every $\overline{S} = (S_1, S_2, \ldots, S_N)$ system with $\sum_{i=1}^{N} \frac{1}{S_i} \geq 1$ we have that $R(\overline{S}) \leq \frac{1}{2}$.*

Proof. Let $x = (x_1, x_2, \ldots, x_N) \in M(\overline{S})$. Since $\sum_{i=1}^{N} \frac{1}{S_i} \geq 1$ we have that $0 \leq x_i \leq \frac{1}{S_i}$ for $i = 1, 2, \ldots, N$. So, by Lemma 5.4 and Theorem 5.5 we have that $W_{S_i}(x_i) \leq \frac{1}{2}$ for $i = 1, 2, \ldots, N$. Hence

$$R(\overline{S}) = \sum_{i=1}^{N} L_{S_i}(x_i) = \sum_{i=1}^{N} x_i \cdot W_{S_i}(x_i) \leq \sum_{i=1}^{N} x_i \cdot \frac{1}{2} = \frac{1}{2}.$$

□

By Theorem 7.11, Proposition 7.13 and Proposition 7.14 we have the following theorem.

Theorem 7.15 *For every $\overline{S} = (S_1, S_2, \ldots, S_N)$ system with $\sum_{i=1}^{N} \frac{1}{S_i} \geq 1$ there exists a policy U such that $W(U) \leq \frac{N-1}{2}$. Moreover, if $\sum_{i=1}^{N} \frac{1}{S_i} > 1$ then there exists a periodic policy U such that $W(U) \leq \frac{N-1}{2}$.*

7.2 Algorithms for determining a minimal point

In this subsection we give an algorithm for determining for a given $\overline{S} = (S_1, S_2, \ldots, S_N)$ system with $\sum_{i=1}^{N} \frac{1}{S_i} > 1$ a rational point $(x_1, x_2, \ldots, x_N) \in M(\overline{S})$ (recall that in most cases the set $M(\overline{S})$ consists of only one point and then this rational point is obviously unique) and the corresponding value of $R(\overline{S})$.

Put

$$F(\overline{S}) := \{(x_1, x_2, \ldots, x_N) \in H^N : L_{\overline{S}}(x_1, x_2, \ldots, x_N) < \infty.\}$$

Then $M(\overline{S}) \subseteq F(\overline{S})$. Suppose that $x = (x_1, x_2, \ldots, x_N) \in F(\overline{S}) \setminus M(\overline{S})$. Then according to the characterization of $M(\overline{S})$ given by Theorem 7.7 we have that there exist $i, j \in \{1, 2, \ldots, N\}$, $i \neq j$ such that $l_{S_i}(x_i) > r_{S_j}(x_j)$ and $r_{S_j}(x_j) < \infty$. Hence there exists some $\varepsilon > 0$ such that $L_{\overline{S}}(x_1, x_2, \ldots, x_i - \varepsilon, \ldots, x_j + \varepsilon, \ldots, x_N) < L_{\overline{S}}(x_1, x_2, \ldots, x_N) < \infty$. So, if we have a point $x = (x_1, x_2, \ldots, x_N) \in H^N$ and we have for all $i, j \in \{1, 2, \ldots, N\}$ with $i \neq j$ that $L_{\overline{S}}(x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N) \geq L_{\overline{S}}(x_1, x_2, \ldots, x_N)$, where h is a small positive number, then the (Euclidean) distance between $x = (x_1, x_2, \ldots, x_N)$ and the set $M(\overline{S})$ is small, since this distance is of order h. For a given $\overline{S} = (S_1, S_2, \ldots, S_N)$ system with $\sum_{i=1}^{N} \frac{1}{S_i} > 1$ we can use this property to obtain a point (arbitrarily) close to the set $M(\overline{S})$. We now give the main steps of the algorithm for obtaining this point. Later we give more details on the implementation of every step.

Sketch of Algorithm.

step 1. Pick an initial point $x := (x_1, x_2, \ldots, x_N) \in F(\overline{S})$ and choose an appropriate initial step size $h > 0$ and an " acceptable error " $\varepsilon > 0$.

step 2. Find $i, j \in \{1, 2, \ldots, N\}$ with $i \neq j$ such that $L_{\overline{S}}(x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N) < L_{\overline{S}}(x_1, x_2, \ldots, x_N)$ or determine that there do not exist such i, j. In the former case go to step 3, in the latter case go to step 4.

step 3. Put $x := x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N)$ and return to step 2.

step 4. Reduce the value of the step size h by a reasonable amount (for example $h := \frac{h}{2}$) and determine whether the new value of h is smaller then ε or not. In the former case go to step 5, in the latter case return to step 2.

step 5. The algorithm terminates by giving as output the point x.

We will show that the point x obtained in this way can be used to obtain a rational point $x' \in M(\overline{S})$ if the distance between the point x and the set $M(\overline{S})$ is small enough. We first give some more details on our implementation of several steps of the algorithm above. In step 1 we suggest to put $x_i = \frac{\frac{1}{S_i}}{\sum_{j=1}^{N} \frac{1}{S_j}}$ for $i = 1, 2, \ldots, N$. It is easily seen that then $x \in F(\overline{S})$. Then for the initial step size h we choose $h = 2^{-(k+1)}$, where $k = \lceil \frac{\log(\max_{i=1,2\ldots N}(\frac{1}{S_i} - x_i))}{\log(\frac{1}{2})} \rceil$. We suggest putting $\varepsilon := \frac{h}{100}$, since then the distance between the obtained point x and the set $M(\overline{S})$ will usually be small enough to obtain a rational point in $M(\overline{S})$. However, if it turns out that this distance is not small enough to find the rational point then a smaller value for ϵ should be chosen, while in many other cases it is possible to use a greater value

for ϵ.
For step 2 the most obvious way to find such i, j is just systematically checking for $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, N$, $i \neq j$ whether $L_{\overline{S}}(x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N) < L_{\overline{S}}(x_1, x_2, \ldots, x_N)$ until you find such i, j or you have tried all pairs i, j with $i \neq j$ and none suffices. In this way the value of $L_{\overline{S}}(x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N)$ has to be computed $O(N^2)$ times. In practice this is fast enough, since N is usually a rather small number and $L_{\overline{S}}(x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N)$ can be computed quickly by the algorithm described in the previous section. However, it is not really necessary to check all pairs i, j since for example you could decide to go to step 4 and decrease the step size if you have determined that with high probability there do not exist such i, j. Moreover, some time can be gained if the pairs i, j are checked in a smart order. Therefore, we suggest a slightly more sophisticated method for implementing step 2 of the algorithm in case that N is not a small number. Start with computing the left derivatives $l_{S_i}(x_i)$ and the right derivatives $r_{S_i}(x_i)$ for $i = 1, 2, \ldots, N$. Then calculate the difference $l_{S_i}(x_i) - r_{S_j}(x_j)$ for all the pairs $i, j \in \{1, 2, \ldots, N\}$ with $i \neq j$. Check whether $L_{\overline{S}}(x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N) < L_{\overline{S}}(x_1, x_2, \ldots, x_N)$ for the pairs $i, j \in \{1, 2, \ldots, N\}$ for which $l_{S_i}(x_i) - r_{S_j}(x_j) > 0$ and run through these pairs in order of decreasing value of $l_{S_i}(x_i) - r_{S_j}(x_j)$. In the computation of $L_{\overline{S}}(x_1, x_2, \ldots, x_i - h, \ldots, x_j + h, \ldots, x_N)$ the (intermediate) results for computing the left and right derivatives $l_{S_i}(x_i)$ and $r_{S_i}(x_i)$ should be used to increase the efficiency of this method.
In step 4 we suggest to reduce the step size by a factor 2.

We illustrate the algorithm to obtain a point x close to the set $M(\overline{S})$ in the following example.

Example. Let $N = 3$ and $\overline{S} = (\frac{93}{57}, \frac{106}{39}, \frac{183}{20})$. Putting $x_i = \frac{\frac{1}{S_i}}{\sum_{j=1}^{N} \frac{1}{S_j}}$ for $i = 1, 2, 3$ we obtain the initial point $x = (\frac{368562}{655529}, \frac{221247}{655529}, \frac{65270}{655529}) = (0.562236, 0.337509, 0.100255) \in F(\overline{S})$ (rounded to 6 decimals) and corresponding value $L(x_1, x_2, x_3) = 0.128134$. The initial step size $h := 2^{-6} = 0.015625$ and the error ε is chosen to be $\frac{h}{100} = \frac{1}{6400} = 0.00015625$. For this step size h we do not find a pair i, j with $i \neq j$ that suffices. So, $h := 2^{-7}$ and we find that $i = 2$, $j = 1$ suffices. Then $x_1 := x_1 + h = 0.570048$, $x_2 := x_2 - h = 0.329697$, and we obtain the new point $(x_1, x_2, x_3) = (0.570048, 0.329697, 0.100255) \in F(\overline{S})$ and corresponding value $L(x_1, x_2, x_3) = 0.125732$. We do not find a new pair i, j which suffices and thus $h := 2^{-8}$. Next we find that $i = 1$, $j = 2$ suffices and we obtain the new point $(x_1, x_2, x_3) = (0.566142, 0.333603, 0.100255) \in F(\overline{S})$ and corresponding value $L(x_1, x_2, x_3) = 0.119799$. Continuing the procedure in this way we obtain consecutively $(x_1, x_2, x_3) := (0.562236, 0.333603, 0.104161)$, $L(x_1, x_2, x_3) = 0.118669$, $h := 2^{-9}$, $h := 2^{-10}$, $(x_1, x_2, x_3) := (0.561259, 0.333603, 0.105138)$, $L(x_1, x_2, x_3) = 0.118386$, $h := 2^{-11}$, $(x_1, x_2, x_3) := (0.561748, 0.3331155, 0.105138)$, $L(x_1, x_2, x_3) = 0.118202$, $h := 2^{-12}$, $(x_1, x_2, x_3) := (0.561504, 0.333359, 0.103138)$, $L(x_1, x_2, x_3) = 0.117865$, $h := 2^{-13}$. Since the new step size $h := 2^{-13}$ is smaller than the chosen error $\varepsilon = \frac{1}{6400}$, the algorithm stops. The distance between the last obtained point $x = (x_1, x_2, x_3) = (0.561504, 0.333359, 0.103138) \in F(\overline{S})$ and the set $M(\overline{S})$ is of order ε and thus $L(x_1, x_2, x_3) = 0.117865$ is an approximation of $R(\overline{S})$, the lower bound for the average waiting time in the $\overline{S}$ system.

So, as illustrated above we have an algorithm to obtain a point (arbitrarily) close to the set $M(\overline{S})$ and thus an approximation of $R(\overline{S})$. We now discuss a method to obtain a rational point in the set $M(\overline{S})$ from this and if we have found this rational point then we can calculate the exact value of $R(\overline{S})$ by the algorithm of the previous section.

Recall that by Theorem 7.9 there exists a point $(x_1, x_2, \ldots, x_N) \in M(\overline{S})$ and some $j \in \{1, 2, \ldots, N\}$ such that x_i is 0 or a best lower approximation of $\frac{1}{S_i}$ for every $i \neq j$. Hence x_i is a rational number of relatively low denominator for every $i \neq j$ and $x_j = 1 - \sum_{i \neq j} x_i$ is also a rational number. Let $(x_1^*, x_2^*, \ldots, x_N^*)$ be a point close to the set $M(\overline{S})$ obtained by the algorithm demonstrated above and let $x_1^i, x_2^i, \ldots$ be the convergents of x_i^*. Then for every $i \neq j$ it is very likely that $x_i = x_{k_i}^i$, where

k_i is a relatively small positive integer since the denominator of x_i is relatively small. Let ε be the error used in the algorithm to determine $(x_1^*, x_2^*, \ldots, x_N^*)$. Then for every $i \in \{1, 2, \ldots, N\}$ we use the continued fraction algorithm to determine the convergents $x_1^i, x_2^i, \ldots$ of x_i^* until we have that $\mid x_{k_i}^i - x_i^* \mid < \varepsilon$ for some $k_i \in \mathbb{N}$. We put $y_i := x_{k_i}^i$ for $i = 1, 2, \ldots, N$. Then y_i is a rational number for $i = 1, 2, \ldots, N$. Let y_j be the one with the greatest denominator among them. Put $y_j = 1 - \sum_{i \neq j} y_i$. Then $\sum_{i=1}^N y_i = 1$ and if ε was chosen small enough then $(y_1, y_2, \ldots, y_N)$ is likely to be a rational point in the set $M(\overline{S})$. In fact, if the set $M(\overline{S})$ consists of the single point $(x_1, x_2, \ldots, x_N)$ as is usually the case, then we should have that $y_i = x_i$ for $i = 1, 2, \ldots, N$. In that case we have for every $i \neq j$ that y_i is either 0 or a best lower approximation of $\frac{1}{S_i}$. In general we can verify whether the obtained rational point $(y_1, y_2, \ldots, y_N)$ is an element of $M(\overline{S})$ or not by the condition of Proposition 7.7. If $(y_1, y_2, \ldots, y_N) \in M(\overline{S})$ then the exact value of $R(\overline{S})$ can be calculated by $R(\overline{S}) = L(y_1, y_2, \ldots, y_N)$. If it turns out that $(y_1, y_2, \ldots, y_N) \notin M(\overline{S})$ then the algorithm could be modified. First of all it is possible to choose a smaller value for the error ε and try again.

It is also possible to exhaustively search for a rational point $(x_1, x_2, \ldots, x_N) \in M(\overline{S})$ for which x_i is either 0 or a best lower approximation of $\frac{1}{S_i}$ for every $i \neq j$ by using the enumerations (see (18),(19) and (20)) of the best lower approximations of $\frac{1}{S_i}$ for $i = 1, 2, \ldots, N$. For $N = 2$ this enumeration of possible values for the x_i and the fact that the function $L(x_1, x_2)$ is convex in x_1 and x_2 is used in [7] to find a rational point in $M(\overline{S})$. However, for larger values of N this is more complicated and takes much more time. On the other hand, if you first apply the above algorithm to obtain a point $(x_1^*, x_2^*, \ldots, x_N^*)$ close to the set $M(\overline{S})$ then you approximately know where you should start the search. This should reduce the time it takes to find the rational point by an exhaustive search.

We illustrate the algorithm to find a rational point $(y_1, y_2, \ldots, y_N) \in M(\overline{S})$ from a point $(x_1^*, x_2^*, \ldots, x_N^*)$ close to the set $M(\overline{S})$ by applying the continued fraction algorithm.

Example. We continue the previous example in which we obtained the point $(x_1^*, x_2^*, x_3^*) =$ $(0.561504, 0.333359, 0.103138)$ for the $\overline{S} = (\frac{93}{57}, \frac{106}{39}, \frac{183}{20})$ system by taking $\varepsilon = \frac{1}{6400}$. We apply the continued fraction algorithm to $x_1^* = 0.561504$ (rounded to 6 decimals) and we find that $x_1^* = [0, 1, 1, 3, 1, 1, 3, \ldots]$ with convergents $\frac{0}{1}, \frac{1}{1}, \frac{1}{2}, \frac{4}{7}, \frac{5}{9}, \frac{9}{16}, \frac{32}{57}, \ldots$. Since for the 6-th convergent $\frac{32}{57}$ we have that $\mid \frac{32}{57} - x_1^* \mid < \varepsilon$ we stop the continued fraction expansion of x_1^* and put $y_1 = \frac{32}{57}$. Similarly, we obtain $y_2 = \frac{1}{3}$ and $y_3 = \frac{2}{19}$. According to the algorithm we should put $y_1 := 1 - y_2 - y_3$, but the value of y_1 does not change by this, since we already have $y_1 + y_2 + y_3 = 1$. So, we have obtained the rational point $(y_1, y_2, y_3) = (\frac{32}{57}, \frac{1}{3}, \frac{2}{19})$ and next we verify that $(y_1, y_2, y_3) \in M(\overline{S})$ by the conditon of Proposition 7.7. So, we calculate the left derivatives $l_{S_i}(y_i)$ and the right derivatives $r_{S_i}(y_i)$ for $i = 1, 2, 3$. We obtain $l_{S_1}(y_1) = \frac{34}{19}$, $l_{S_2}(y_2) = 0$, $l_{S_3}(y_3) = \frac{3}{2}$, $r_{S_1}(y_1) = \frac{34}{19}$, $r_{S_2}(y_2) = \frac{51}{13}$ and $r_{S_3}(y_3) = \frac{87}{20}$. Since $l_{S_i}(y_i) \leq r_{S_j}(x_j)$ for every $i \neq j$ we have indeed that $(y_1, y_2, y_3) \in M(\overline{S})$. Moreover, y_2 and y_3 are best lower approximations of $\frac{1}{S_2}$ and $\frac{1}{S_3}$ respectively, as expected. This follows directly from the fact that $l_{S_2}(y_2) < r_{S_2}(y_2)$ and $l_{S_3}(x_3) < r_{S_3}(x_3)$. We have $R(\overline{S}) = L(y_1, y_2, y_3) = \frac{2551}{21660}$. So, for the $\overline{S} = (\frac{93}{57}, \frac{106}{39}, \frac{183}{20})$ system we have that $W(U) \geq \frac{2551}{21660}$ for every routing policy U. Moreover, it follows from Proposition 7.4 and Theorem 7.11 that it is easy to obtain (by constructing a billiard sequence) a periodic policy U' with densities $(d_1, d_2, d_3) = (y_1, y_2, y_3) = (\frac{32}{57}, \frac{1}{3}, \frac{2}{19})$ (and thus a period of 57) for which $W(U') \leq R(\overline{S}) + \frac{N}{2} - 1 = \frac{13381}{21660}$.

8 Appendix: An extension of Little's Law for routing policies

Stidham proved in [24] the Little relation under the assumption that W, the limiting average of the waiting times, and L, the limiting average of the number of customers, do exist. In our optimal routing problem we do not want to restrict the set of policies to those for which these limits exist. Even if the limiting density does exist then the limiting averages L and W may fail to exist (see Example 8.3). Here we show that Little's relation can be replaced with the inequalities of Theorem 8.2 if L and W are defined in terms of limsups as in section 3.

The results in this appendix are valid for any input-output queueing system and thus we have a more general setting than before. However, we recall that the objective is to show the validity of (3), (4) and(6) for the particular deterministic queueing system considered in this paper. For convenience we use notation similar to that used in the paper.

We suppose that customers arrive at (or are admitted by some routing policy to) the queueing system at moments $t_n \in \mathbb{R}$, $n = 1, 2\ldots$ where $0 \le t_1 \le t_2 \le \ldots$. Let W_n be the waiting time in the system of the n -th arriving customer and let $B(t)$ be the number of customers waiting in the system at time t. Then we define $L := \limsup_{T\to\infty} \frac{1}{T} \cdot \int_{t=0}^{T} B(t)d(t)$ as the time average number of customers in the system and $W := \limsup_{n\to\infty} \frac{1}{T} \cdot \sum_{n=1}^{T} W_n$ as the average waiting time of the arriving customers. Moreover, we put $A(t) :=\mid n \in \mathbb{N} : t_n \le t \mid$, the number of arrivals before time t and $D(t) :=\mid n \in N : t_n + W_n \le t \mid$, the number of departures before time t. From the definition it is clear that

$$A(t) = B(t) + D(t) \text{ for every } t \in \mathbb{Z}_{\ge 0}.$$

We put $\overline{\lambda} := \limsup_{t\to\infty} \frac{A(t)}{t}$ as the upper arrival rate and $\underline{\lambda} := \liminf_{t\to\infty} \frac{A(t)}{t}$ as the lower arrival rate. We assume that $\overline{\lambda} < \infty$. If $\underline{\lambda} = \overline{\lambda}$ then $\lambda := \lim_{t\to\infty} \frac{A(t)}{t}$ is the arrival rate.

Lemma 8.1 *If $W < \infty$ then*

$$\liminf_{t\to\infty} \frac{D(t)}{t} = \liminf_{t\to\infty} \frac{A(t)}{t} \text{ and } \limsup_{t\to\infty} \frac{D(t)}{t} = \limsup_{t\to\infty} \frac{A(t)}{t}. \tag{25}$$

Moreover,

$$\limsup_{t\to\infty} \frac{1}{A(t)} \sum_{n=1}^{A(t)} W_n = \limsup_{t\to\infty} \frac{1}{A(t)} \sum_{n:t_n+W_n\le t} W_n. \tag{26}$$

Proof. Since $W < \infty$ we have $\lim_{n\to\infty} \frac{W_n}{n} = 0$. Thus $\lim_{n\to\infty} \frac{W_n}{t_n} = 0$, since $\overline{\lambda} < \infty$. Hence, for every $\varepsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that $W_n < t_n\varepsilon$ for every $n > n_0$. Let $t_0 \in R_{\ge 0}$ be chosen so large that $A(t_0) > n_0$. Then

$$\begin{aligned} &A(t_0) \ge D(t_0) =\mid n : t_n + W_n \le t_0 \mid \ge\mid n \le n_0 : t_n + W_n \le t_0 \mid + \mid n > n_0 : t_n(1+\varepsilon) \le t_0 \mid= \\ &\mid n \le n_0 : t_n + W_n \le t_0 \mid - \mid n \le n_0 : t_n \cdot (1+\varepsilon) \le t_0 \mid + \mid n : t_n \le t_0/(1+\varepsilon) \mid . \end{aligned} \tag{27}$$

Dividing (27) by t_0 and taking liminfs for $t_0 \to \infty$ we obtain

$$\liminf_{t_0\to\infty} \frac{A(t_0)}{t_0} \ge \liminf_{t_0\to\infty} \frac{D(t_0)}{t_0} \ge \liminf_{t_0\to\infty} \frac{A(t_0)/(1+\varepsilon)}{t_0}.$$

Since ε was arbitrary, it follows that $\liminf_{t\to\infty} \frac{D(t)}{t} = \liminf_{t\to\infty} \frac{A(t)}{t}$. Analogously it follows that $\limsup_{t\to\infty} \frac{D(t)}{t} = \limsup_{t\to\infty} \frac{A(t)}{t}$. Moreover, since $A(t) \to \infty$ as $t \to \infty$ it also follows from (27) that

$$\limsup_{t\to\infty} \frac{1}{A(t)} \sum_{n=1}^{A(t)} W_n \ge \limsup_{t\to\infty} \frac{1}{A(t)} \sum_{n:t_n+W_n\le t} W_n \ge \frac{1}{1+\varepsilon} \limsup_{t\to\infty} \frac{1}{A(t)} \sum_{n=1}^{A(t)} W_n.$$

This implies (26) since ϵ was arbitrary. $\square$

Remark. The proof of this lemma is similar to the proof of Little's Law by Stidham in [24]. The main difference is that the existence of certain limits is not assumed, but that instead of that liminfs and limsups are used.

By Lemma 8.1 we can prove that the following extension of Little's Law holds.

Theorem 8.2 *Suppose that $W < \infty$ for some input-output queueing system as described above. Then, for the upper arrival rate $\overline{\lambda}$ we have that*

$$L \leq \overline{\lambda} \cdot W. \tag{28}$$

Moreover, for the lower arrival rate $\underline{\lambda}$ we have that

$$L \geq \underline{\lambda} \cdot W \tag{29}$$

and if the arrival rate λ exists then

$$L = \lambda \cdot W \tag{30}$$

Proof. Using a standard argument in the derivation of Little's Law, more precisely by considering the total number of waiting time units up to time t (see [22] page 102 or [24]), we find the inequalities

$$\sum_{n \in \mathbb{N}: t_n + W_n \leq t} W_n \leq \int_{s=0}^{t} B(s) ds \leq \sum_{n=1}^{A(t)} W_n. \tag{31}$$

By the second inequality we have

$$\overline{\lambda} \cdot W = \limsup_{t \to \infty} \frac{A(t)}{t} \cdot \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} W_k = \limsup_{t \to \infty} \frac{A(t)}{t} \cdot \limsup_{t \to \infty} \frac{1}{A(t)} \sum_{k=1}^{A(t)} W_k \geq$$

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{k=1}^{A(t)} W_k \geq \limsup_{t \to \infty} \frac{1}{t} \int_{s=0}^{t} B(s) ds = L.$$

Similarly, (26) and the first inequality imply

$$\underline{\lambda} \cdot W = \liminf_{t \to \infty} \frac{A(t)}{t} \cdot \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} W_k = \liminf_{t \to \infty} \frac{A(t)}{t} \cdot \limsup_{t \to \infty} \frac{1}{A(t)} \sum_{k=1}^{A(t)} W_k =$$

$$\liminf_{t \to \infty} \frac{A(t)}{t} \cdot \limsup_{t \to \infty} \frac{1}{A(t)} \sum_{n: t_n + W_n \leq t} W_n \leq \limsup_{t \to \infty} \frac{1}{t} \sum_{n: t_n + W_n \leq t} W_n \leq \limsup_{t \to \infty} \frac{1}{t} \int_{s=0}^{t} B(s) ds = L.$$

By combining the first and second assertion we have $L = \lambda \cdot W$ if λ exists. $\square$

We now apply Theorem 8.2 for routing to a deterministic queueing system as described in section 3. If we consider the $(S_1, S_2, \ldots, S_N)$ system in total, then λ, is equal to 1 and independent of the routing policy U. Hence (4) follows immediately from (30). Moreover, if we consider a single queue i of the system then the arrival rate λ is equal to the density d_i (if this exists) of the routing sequence u^i corresponding to this queue. Thus (30) also implies (6).

In section 3 we defined the average waiting time W and the average buffer content L by limsups instead of limits. The following example shows that this is necessary, since in general the limits do not exist (even if you assume that the routing sequence has a density and thus the arrival rate exists).

Example 8.3 *Consider a single deterministic queue and assume that the density d of the routing sequence exists and is rational, $d = \frac{p}{q}$ with $\gcd(p,q) = 1$ say. Suppose that the service time S is such that*

$$\frac{1}{\lceil S \rceil} < d < \frac{1}{S}.$$

Let w be the upperbracket sequence with period word $\omega = (\omega_1, \dots, \omega_q)$ and with average workload $L(d)$. We construct a policy $u = (u_1, u_2, \dots)$ which consists of blocks of type ω^k alternated with blocks of type ϖ^l where $\varpi = (\varpi_1, \dots, \varpi_q)$, with $\varpi_k = 1 (1 \le k \le p)$ and $\varpi_k = 0 (p+1 \le k \le q)$.

Choose (p,q) and S such that

$$L(\varpi^\infty) > L(d).$$

Choose a sequence $\{\varepsilon_k > 0\}_{k=1}^\infty$ with $\lim_{k\to\infty} \varepsilon_k = 0$ and define inductively unbounded sequences t_k, t'_k. Let $t_k = t'_{k-1} + m_k q$ and define on the interval $t_k < t < t'_k$ the u_t as ω^{m_k}. Take m_k sufficiently large such that

$$\left| \frac{1}{t_k} \int_{s=0}^{t_k} B(s)ds - L(d) \right| < \varepsilon_k.$$

Let $t'_k = t_k + n_k q$ and define on the interval $t_k < t < t'_k$ the u_t as ϖ^{n_k} and take n_k sufficiently large such that

$$\left| \frac{1}{t'_k} \int_{s=0}^{t'_k} B(s)ds - L(\varpi^\infty) \right| < \varepsilon_k.$$

Then, since the fraction of ones in each ω and ϖ is equal to d, we have

$$\lim_{t\to\infty} \frac{\kappa_u(t)}{t} = d.$$

Moreover, $\liminf_{t\to\infty} \frac{1}{t} \int_{s=0}^t B(s)ds = L(d) < L(\varpi^\infty) = \limsup_{t\to\infty} \frac{1}{t} \int_{s=0}^t B(s)ds$.

The conclusion is that the existence of the limiting density does not imply the existence of the limiting average waiting time nor the existence of the limiting average buffer content.

References

[1] Altman, E., B. Gaujal, A. Hordijk. 2000. Admission control in stochastic event graphs. *IEEE Trans. Automat. Control* **45** 854–867.

[2] Altman, E., B. Gaujal, A. Hordijk. 2000. Balanced sequences and optimal routing. *J. Assoc. Comput. Mach.* **47** 752–775.

[3] Altman, E., B. Gaujal, A. Hordijk. 2000. Multimodularity, convexity and optimization properties. *Math. Oper. Res.* **25** 324–347.

[4] Arian, Y., Y. Levy. 1992. Algorithms for generalized round Robin routing. *Oper. Res. Lett.* **12** 313–319.

[5] Baccelli, F., P. Brémaud. 1994. *Elements of queueing theory.* Springer-Verlag.

[6] Combé, M.B., O.J. Boxma. 1994. Optimization of static traffic allocation policies. *Theoretical Computer Science* **125** 17–43.

[7] Gaujal, B., E. Hyon. 2000. Optimal routing policy in two deterministic queues. Technical Report INRIA RR-3997. To appear in Calculateurs Parallèles.

[8] Hajek, B. 1985. Extremal splittings of point processes. *Math. Oper. Res.* **10** 543–556.

[9] Hardy, G.H., E.M. Wright. 1960. *An introduction to the theory of numbers.* Fourth edition, Oxford.

[10] Heinis, A. 2001. *Arithmetics and combinatorics of words of low complexity.* Ph.D.-thesis, Leiden University.

[11] Hordijk, A., G. Koole. 1992. On the assignment of customers to parallel queues. *Prob. Engin. Inform. Sci.* **6** 495–511.

[12] Hordijk, A., D.A. van der Laan. 2000. Periodic routing to parallel queues with bounds on the average waiting time. Report MI no. 2000-44, Leiden University. Available on www.math.leidenuniv.nl/reports/2000-44.shtml.

[13] Hordijk, A., D.A. van der Laan. 2001. Bounds for deterministic periodic routing sequences. Integer programming and combinatorial optimization. K. Aardal, B. Gerards, ed., *Proceedings of the 8th international IPCO conference*, Springer, 236–250.

[14] Hordijk, A., D.A. van der Laan. 2002. Note on the convexity of the stationary waiting time as function of the density. Leiden University. Submitted to Probab. Engrg. Inform. Sci.

[15] Koole, G., P. Sparaggis, D. Towsley. 1999. Minimizing response times and queue lengths in systems of parallel queues. *J. Appl. Probab.* **36**), 1185–1193.

[16] van der Laan, D. A. 2000. Routing jobs to servers with deterministic service times. Report MI no. 2000-20, Leiden University. Submitted to Math. Oper. Res. Available on www.math.leidenuniv.nl/reports/2000-20.shtml.

[17] Liu, Z., R. Righter. 1998. Optimal load balancing on distributed homogeneous unreliable processors. *Oper. Res.* **46** 563–573.

[18] Lothaire, M. 2002. *Algebraic Combinatorics on Words.* Cambridge University Press, Cambridge.

[19] Moran, P.A.P. 1968. *An Introduction to Probability Theory.* Claridon Press Oxford.

[20] Morse, M., G.A. Hedlund. 1940. Symbolic dynamics II. *Amer. J. Math.* **60** 1–42.

[21] Perron, O. 1954. *Die Lehre von den Kettenbrüchen.* Third edition, Stuttgart.

[22] Ross, S.M. 1970. *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco.

[23] Shirakawa, H., M. Mori, M. Kijima. 1989. Evaluation of Regular Splitting Queues. *Comm. Statist. Stochastic Models* **5** 219–234.

[24] Stidham, S., Jr. 1974. A last word on $L = \lambda W$. *Oper. Res.* **22** 417–421.

[25] Tijdeman, R. 2000. Fraenkel's conjecture for six sequences. *Discrete Math.* **222** 223–234.