Periodic routing to parallel queues with bounds on the average waiting time

A. Hordijk and D.A. van der Laan
Mathematical Institute, University Leiden

Abstract

In this paper we introduce the combinatorial notion of unbalance for a zero-one splitting sequence and for a routing pattern. Using this unbalance we derive an upper bound for the average expected waiting time of jobs which are routed to parallel queues according to a periodic routing rule. A billiard sequence is obtained with unbalance smaller than or equal to $\frac{N}{2} - 1$, where N is the number of different symbols in the sequence which corresponds to the number of parallel queues in the routing problem.

KEY WORDS: periodic routing, parallel queues, bounds on waiting time, unbalance, billiard sequences

1 Introduction

Load balancing in distributed multiprocessors and in communication networks is an important tool to improve their performance in mean delay of the jobs (for overviews of the literature see [11] and [22]). In many systems the load balancing protocol is an open-loop control, in which case it does not depend on the information in the system, as queue-sizes etc., but only depends on the system parameters. In this paper we focus on the performance under this type of control, which usually is called static routing control. In the publications ([26],[37], [20] and [25]) the static routing problem in a network is reduced to finding the best utilization of each service centre. Given this utilization of a particular service centre the problem then becomes to find the static routing to parallel queues with minimal average expected waiting time. For this routing problem probabilistic (also called Bernoulli) and deterministic routing policies have been analyzed. In case of probabilistic routing to homogeneous servers, the optimality of equal routing probabilities is proved in [14] and [27]. In general deterministic (generalized) round robin policies are superior to probabilistic routing for homogeneous (heterogeneous) servers. For homogeneous servers, the optimality of the round robin routing is proved in [29]. The problem of constructing the optimal generalized round robin policy for heterogeneous servers is an unsolved problem. In the papers [23] and [2] the optimal routing problem for exponential service times is transformed to a Markov decision process (see [1] for an overview of the applications of these processes in telecommunication).

In a recent approach the optimization procedure is splitted into two problems. First the fractions of jobs routed to the servers are computed via a best probabilistic routing. For networks these fractions are computed via the optimal rates (see [26],[37], [20] and [25]) , and for the single server with parallel queues also different approximation methods have been used (see [15]). Given the fractions of jobs that have to be routed to the queues, the problem then becomes to construct the best routing pattern. In the seminal paper of Hajek ([21]) it is shown for one queue with exponential service times that the regular splitting sequence with the density d is optimal for any initial phase ϕ, if a fraction d of the arriving customers have to be routed to that queue. A regular zero-one valued splitting sequence with rational density d and initial phase $\phi \in \mathbb{R}$ is:

$$\{b_k^d(\phi)\} = \lfloor (k+1)d + \phi \rfloor - \lfloor kd + \phi \rfloor ,$$

$k = 1, 2, \ldots$. The corresponding routing rule assigns the k-th arrival to the queue if $b_k^d(\phi) = 1$. Since the zeros of a regular splitting sequence with density d form a regular splitting sequence with density $(1-d)$, it follows that the regular splitting sequence with density d gives an optimal routing sequence to two queues with exponential service times for given fractions d and $(1-d)$. It is proved in [3] and [5] that a regular splitting policy is optimal in the class of all static routing policies (without fixing the routing fractions beforehand) for two queues with general stationary interarrival and service times. An algorithm for computing the optimal routing policy for two queues with deterministic interarrival and service times has been derived in [18].

The problem of finding the best optimal routing pattern with given splitting fractions is a complex optimization problem for $N \geq 3$ queues (see [9]). There are only few sets of fractions for which the regular splitting sequences generate an exact covering of the positive integers (see [4] and [36]), and only in these cases the optimal routing pattern is known. Regular sequences, which are also called Beatty, Sturmian, or bracket sequences have applications in many branches of sciences (see [35]).

For routing to $N \geq 3$ parallel queues many algorithms for finding good allocation patterns have been derived. In [32] the golden ratio policy is used for the optimal routing of jobs, in [24] for the dual problem of the static control of a multiple-access channel, and in [16] for optimal robot scheduling of web search engines. In these papers the asymptotic optimality of the golden ratio policy as the number of queues tends to infinity is studied. Other algorithms for generalized round robin routing sequences have been given in [7] and [9]. Methods to obtain good allocation patterns can be found in [15] and

[12]. Although simulations make it evident that the proposed algorithms behave well, bounds on their performance were only derived for exponential queues with no buffer space.

In this paper we concentrate on obtaining bounds for the average expected waiting time of the jobs for a routing pattern to parallel queues with unlimited buffer capacity and general sequences of interarrival and service times. Our analysis is mainly combinatorial. For a given pattern allocation we introduce its unbalance. All regular sequences have unbalance equal to zero, and for a splitting sequence to one queue the unbalance is roughly speaking its 'distance' to the regular sequence with the same density. For the routing pattern it is the sum of the unbalances of the splitting sequences to the queues. In [3] and [5] it is shown for routing policy ψ which routes a fraction d_i to queue i for $i = 1, 2, \ldots, N$ that the average expected waiting time $\overline{W}(\psi)$ is bounded from below by $\widetilde{R}$, where $\widetilde{R} = \sum_{i=1}^{N} d_i \cdot \overline{W}^i(\omega(d_i))$, with $\overline{W}^i(\omega(d_i))$ the average expected waiting time for jobs routed to queue i when the routing to queue i is according to a regular sequence with density d_i. The upper bound for $\overline{W}(\psi)$ which we derive in this paper is $\widetilde{R} + \delta \cdot \overline{O}(U)$, where δ is the mean interarrival time, U is the routing sequence corresponding to policy ψ and $\overline{O}(U)$ is the unbalance of U. Hence,

$$\widetilde{R} \leq \overline{W}(\psi) \leq \widetilde{R} + \delta \cdot \overline{O}(U).$$

The unbalance $\overline{O}(U)$ is a combinatorial notion and it does not depend on the service or the interarrival distribution. Hence the bounds are insensitive, i.e. they are valid for any stationary sequence of interarrival and service times. However, the lower bound $\widetilde{R}$ does depend on the distribution of the interarrival and service times. The upper bound is tight for the heavy traffic situation in which the interarrival and service times are deterministic and the traffic intensity is 1 for each of the queues. So the upper bound is accurate for high traffic, whereas $\overline{W}(\psi)$ will be close to the lower bound for low traffic.

Finding for given densities $d_i, i = 1, 2, \ldots, N$ the routing sequence U such that the unbalance $\overline{O}(U)$ is minimal can be transformed to a combinatorial optimization problem (see [28]). Greedy algorithms for obtaining good routing policies have been given in [7], [9], [33] and [28]. In this paper we consider greedy algorithms which in a special case generate billiard sequences. Suppose a billiard consisting of the N-dimensional cube $[0,1]^N$ (see [8], [10]), with sides numbered $1, 2, \ldots, N$ such that opposite sides have the same number. Give a billiard ball initial position $(x_1, x_2, \ldots, x_N)$ and initial velocity vector $(-d_1, -d_2, \ldots, -d_N)$, and take the sequence $U_b = (n_1, n_2, \ldots)$ indicating the order in which the sides are hit by the ball, breaking ties in an unique way and let ψ_b be the corresponding routing policy. Then ψ_b, which routes the k -th arrival to queue i if $n_k = i$, is a routing policy with densities d_i, $i = 1, 2, \ldots, N$. We show that by taking a special initial position the unbalance $\overline{O}(U_b)$ of this routing sequence is bounded by $\frac{N}{2} - 1$. Hence,

$$\widetilde{R} \leq \overline{W}(\psi_b) \leq \widetilde{R} + \delta(\frac{N}{2} - 1).$$

For $N = 2$ an optimal routing is constructed, and for $N \geq 3$ this billiard sequence is easily computed and has the above upper bound on its performance. In this paper we consider only rational densities $d_i = \frac{p_i}{q_i}$ with $p_i, q_i \in \mathbb{N}$, $\gcd(p_i, q_i) = 1$. In this case the billiard sequence is periodic with period $T = \text{lcm}(q_1, q_2, \ldots, q_N)$, and in fact it gives a routing pattern with period T.

The paper is organized as follows. In section 2 we start with an introduction of regular sequences together with properties which we need for our analysis. We define a partial order in the set $\mathcal{P}(T,k)$ of all sequences of zeros and ones of length T containing exactly k ones. The conjugacy class of the regular sequences in $\mathcal{P}(T,k)$ is shown to be the minimal element in an induced order. We define the notion unbalance for periodic words of zeros and ones by using the period cycle of the word, which is an element of $\mathcal{P}(T,k)$ if there are k ones in the cycle length T. In fact we will define a primal unbalance and a dual unbalance, the dual unbalance can be used for the dual model of optimal routing of the server in polling models. The unbalance is a measure for the irregularity of the word. The unbalance has a graphical representation (see figure 1), it is the area between the graph of its representation as element of $\mathcal{P}(T,k)$ and that of the regular one. This leads to another partial order which we called the graph order. The conjugacy class of the regular sequences in $\mathcal{P}(T,k)$ is shown to be the minimal element in this graph order. The results of section 2 are purely combinatorial and have possibly other applications than in this routing problem.

In section 3 we consider the routing to one queue and we consider all static deterministic policies which route a fraction p of all jobs to the queue. We restrict our analysis to periodic routing policies and thus to routing sequences with rational fraction p. We use the unbalance to bound the difference in average waiting times. We do this by a sample path comparison for fixed sequences of interarrival and service times. By using results on renovation and ergodicity we then obtain the above mentioned bounds for one queue. In section 4 we extend the bounds to N parallel queues.

In section 5 we extend the combinatorial analysis of the unbalance and derive its relation with the discrepancy function defined in section 2. We introduce greedy algorithms and the billiard sequence which is obtained by a special greedy algorithm. It is shown that there is a billiard sequence with minimal unbalance for given densities. For rational routing densities $d_i = \frac{p_i}{q_i}$ with $p_i, q_i \in \mathbb{N}$, $\gcd(p_i, q_i) = 1$ we show that the unbalance of the billiard routing pattern is bounded by $\frac{N}{2} - 1$ if an initial position $(x_1, \ldots, x_N)$ is chosen with $x_k = 0$ for k such that $q_k = \max_{i \in \{1,2,\ldots,N\}} q_i$ and $x_j = 1 - \frac{1}{q_j}$ for $j \neq k$. Many examples illustrate the notions and results in the sections.

Notations. For $t \in \mathbb{R}$ we denote by $\mathbb{R}_{>t}$ ($\mathbb{R}_{\geq t}$) the set of real numbers that are greater (or equal) than t. For the integers $\mathbb{Z}$ and the rational numbers $\mathbb{Q}$ we denote such subsets in a similar way.
For $x \in \mathbb{R}$ we denote by $\lfloor x \rfloor$ and $\lceil x \rceil$ the maximal integer not larger than x respectively the minimum integer not smaller than x.
If $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ then $a \pmod b$ denotes the integer $c \in \{0, 1, \ldots, b-1\}$ for which there exists some $d \in \mathbb{Z}$ such that $c + d \cdot b = a$. Further gcd denotes the greatest common divisor and lcm denotes the least common multiple.
For any random variable X we denote by $\mathbb{E}X$ the expectation of X.

2 Comparing routing sequences

In this section we will compare routing sequences in a queueing system. For a particular queue the routing sequence will be a sequence of ones and zeros, where an one means that the corresponding arriving customer is routed to that queue and a zero means that the customer is not routed to that queue. Instead of sequences we also speak of words. We first give some definitions.

If u and v are finite sequences, $u = (a_1, a_2, \ldots, a_n)$ and $v = (b_1, b_2, \ldots, b_m)$,
then $uv = (a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m)$ denotes the concatenation of the two sequences. For $k \in \mathbb{N}$ we denote by u^k the sequence $uu \ldots u$ in which the number of u's is k and we say that u^k is a power of u. A finite sequence is primitive if it is not a power of some other sequence. The following lemma is part of Proposition 1.3.1 of [30].

Lemma 2.1 *For every non-empty finite sequence u there exists an unique primitive sequence w such that $u = w^n$ for some $n \in \mathbb{N}$.*

Two finite sequences u, u' are said to be conjugate if there exist sequences v and w (v or w may be empty) such that $u = vw$ and $u' = wv$. This is an equivalence relation since u is conjugate to u' if and only if u' is a cyclic permutation of u. If u and u' are conjugate then we write $u \sim u'$ and the conjugacy class of all the cyclic permutations of a finite sequence u is denoted by $\widetilde{u}$.

Lemma 2.2 *Let u be a finite sequence, $u = w^n$ with w primitive. Then $|\widetilde{u}|$, the number of distinct conjugates of u, equals the length of sequence w.*

Proof. By Proposition 1.3.3 in [30] the number of distinct conjugates of u equals the number of distinct conjugates of w and by Proposition 1.3.2 in [30] the number of distinct conjugates of w equals the length of w. $\square$

If $\mathbf{A}$ is a finite set (a so-called alphabet) and $I \subseteq \mathbb{Z}$ is an interval in $\mathbb{Z}$ then a mapping $\Lambda : I \to \mathbf{A}$ is called an I - word or just word. Thus a word is a sequence of letters from the alphabet $\mathbf{A}$. A word is (in)finite if I is in(finite). A subword of Λ is the restriction of Λ to some interval $J \subseteq I$. For a finite (sub)word Λ we denote the length of Λ by $|\Lambda|$. Further for a letter $a \in A$ the number of a's in Λ is denoted by $|\Lambda|_a$. For an I -word Λ the support of letter a is the set $S_a = \{t \in I : \Lambda(t) = a\}$. For an I -word Λ we have that letter a has density d if for every non-decreasing sequence of finite subintervals I_i with union I we have that $\frac{|S_a \cap I_i|}{|I_i|}$ tends to d. An (infinite) I - word Λ is periodic if for some $q \in \mathbb{N}$ it holds that $\Lambda(t) = \Lambda(t+q)$ for every t for which $t, t+q \in I$. Then q is said to be a period of the word and any subword of length q is said to be a period cycle of the word. The period of Λ is defined as the minimum of all the periods of Λ. If q is the period of word Λ then a subword of length q is said to be a primitive period cycle of the word. It is obvious that the following holds.

Lemma 2.3 *If k is a period of some word Λ then for every $m \in \mathbb{N}$ we have that $k \cdot m$ is a period of Λ.*

Note that a sequence of zeros and ones corresponds to a word on the alphabet $\{0, 1\}$. Vice versa, to every letter in a word corresponds a sequence of ones and zeros via the support of the letter. In the sequel we identify a sequence (of integers) with a word and sometimes it is useful to choose an appropriate domain I for the word the sequence is identified with. So, we consider sequences of integers and words as the same thing and therefore we have the same notions and operations for words as for sequences. A sequence of zeros and ones is said to have density d if the letter 1 has density d in the corresponding word.

Definition 2.4 *Let u be an infinite sequence of zeros and ones and Λ be the corresponding word defined on an 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 Λ' of Λ of length n that $|\Lambda'|_1$ equals $\lfloor nd \rfloor$ or $\lceil nd \rceil$. In that case we also say that the corresponding set S_1 is a regular set.*

The following theorem follows for $I = \mathbb{N}$ from the classification of balanced sequences in [31] and the fact that regular sequences are particular balanced sequences (see [36]).

Theorem 2.5 *Let u and Λ be as in Definition 2.4. If u is regular with density d then the support set S_1 satisfies one of the following cases.*

- *(irrational case) d is irrational and there exists some $\varphi \in \mathbb{R}$ such that*

$$S_1 = I \cap \{\lfloor k \cdot \frac{1}{d} + \varphi \rfloor\}_{k \in \mathbb{Z}} \text{ or } S_1 = I \cap \{\lceil k \cdot \frac{1}{d} + \varphi \rceil\}_{k \in \mathbb{Z}}.$$

- *(rational case) d is rational and there exists some $\varphi \in \mathbb{Q}$ such that*

$$S_1 = I \cap \{\lfloor k \cdot \frac{1}{d} + \varphi \rfloor\}_{k \in \mathbb{Z}}.$$

Theorem 2.5 has the following corollaries.

Corollary 2.6 *Let u and Λ be as above. If u is regular with irrational density d then there exists a $\theta \in \mathbb{R}$ such that $\Lambda(t) = \lfloor (t+1)d + \theta \rfloor - \lfloor td + \theta \rfloor$ or $\Lambda(t) = \lceil (t+1)d + \theta \rceil - \lceil td + \theta \rceil$ for every $t \in I$. If u is regular with rational density d then there exists a $\theta \in \mathbb{Q}$ such that $\Lambda(t) = \lfloor (t+1)d + \theta \rfloor - \lfloor td + \theta \rfloor$ for every $t \in I$.*

Corollary 2.7 *If u is regular with rational density $d = \frac{p}{q}$ with $p, q \in \mathbb{N}$, $\gcd(p,q) = 1$ then u is periodic with period q.*

We define a finite sequence of zeros and ones to be regular if it is a period cycle of some infinite regular periodic sequence.

In [3] and [5] it is shown that a regular sequence is optimal when one has to route to a single queue and a certain fraction of all the arriving customers has to be routed to that queue. They also show that routing according to some regular sequence is optimal when one has to route to two parallel queues. For $N \geq 3$ queues one often wants to route to the queues according to some (optimized) fractions, where the sum of these fractions is 1. However, for such prescribed fractions it is not always possible to route in such a way that the routing sequence for every queue is regular. Therefore we want to compare irregular sequences and for finite and periodic sequences (words) we introduce the notion of unbalance. In fact we will define a primal unbalance and a dual unbalance. The unbalance measures the irregularity of the sequence and has some useful properties. One of the obvious properties will be that a sequence has unbalance 0 if and only if it is a periodic regular sequence or it is a period cycle of a periodic regular sequence. We first define the unbalance for finite sequences.

Notation. Let $\mathcal{P}(T,k)$ be all sequences of zeros and ones of length T containing exactly k ones and let $\mathcal{R}(T,k) \subseteq \mathcal{P}(T,k)$ be the subset of the regular sequences in $\mathcal{P}(T,k)$. Let $u \in \mathcal{P}(T,k)$ and Λ the corresponding word on the alphabet $\{0,1\}$. We assume without loss of generality that $I = \{1, 2, \ldots, T\}$ and we denote u as well as Λ by $(u_1, u_2, \ldots, u_T)$, where $u_t = \Lambda(t)$ for $t = 1, 2, \ldots, T$. Further we say for $l = 0, 1, \ldots, T-1$ that
$(u_{1+l}, u_{2+l}, \ldots, u_T, u_1, u_2, \ldots, u_l)$ is the l -th cyclic permutation of u.

From the rational case of Theorem 2.5 follows the following lemma.

Lemma 2.8 *Let $u \in \mathcal{R}(T,k)$. Then $S_1 = \{\lfloor i \cdot \frac{T}{k} + \varphi \rfloor\}_{i=1}^{k}$ for some $\varphi \in \{1 - \frac{T}{k}, 1 - \frac{T - \gcd(T,k)}{k}, \ldots, 1 - \frac{\gcd(T,k)}{k}\}$.*

It is clear that if $u \in \mathcal{R}(T,k)$ and $u' \sim u$ then $u' \in \mathcal{R}(T,k)$. In fact the following lemma states that $\mathcal{R}(T,k)$ forms a conjugacy class of $\mathcal{P}(T,k)$ and thus the regular sequences in $\mathcal{P}(T,k)$ are all cyclic permutations of each other.

Lemma 2.9 *Let $u \in \mathcal{R}(T,k)$. Then $u' \in \mathcal{R}(T,k)$ if and only if $u \sim u'$.*

Proof. We still have to prove the "only if"-part. For $u \in \mathcal{R}(T,k)$ we have that $\widetilde{u} \subseteq \mathcal{R}(T,k)$. By Lemma 2.1 we have that $u = w^n$ for some primitive w and by Lemma 2.8 it follows that $n = \gcd(T,k)$. Hence the length of sequence w is $\frac{T}{\gcd(T,k)}$ and by Lemma 2.2 we have that $|\widetilde{u}| = \frac{T}{\gcd(T,k)}$. However, by Lemma 2.8 we have that $|\mathcal{R}(T,k)| \leq \frac{T}{\gcd(T,k)}$. Thus $\widetilde{u} = \mathcal{R}(T,k)$. □

Notation. For sequence $u \in \mathcal{P}(T,k)$ we define the counting function $\kappa_u : \{0,1,\dots,T\} \to \mathbb{Z}$ with $\kappa_u(n) = \sum_{t=1}^{n} u_t$. Thus $\kappa(n)$ counts the number of ones in the first n letters of the corresponding word Λ. We also define the discrepancy function $\chi_u : \{0,1,\dots,T\} \to \mathbb{Q}$ by $\chi_u(n) = n \cdot \frac{k}{T} - \kappa_u(n)$ for $n = 0,1,\dots,T$.

Now we define a partial order $\preceq$ on $\mathcal{P}(T,k)$. For $u,v \in \mathcal{P}(T,k)$ we say that $u \preceq v$ if $\kappa_u(n) \leq \kappa_v(n)$ for $n = 1,2,\dots,T$.

Lemma 2.10 *This partial order $\preceq$ on $\mathcal{P}(T,k)$ induces a total order on $\mathcal{R}(T,k)$.*

Proof. Let $u,u' \in \mathcal{R}(T,k)$. Then by Lemma 2.8 there exist $\varphi,\varphi' \in \{1-\frac{T}{k}, 1-\frac{T-\gcd(T,k)}{k}, \dots, 1-\frac{\gcd(T,k)}{k}\}$ such that the support set S_1 of u is $\{\lfloor i \cdot \frac{T}{k} + \varphi \rfloor\}_{i=1}^{k}$, while the support set S_1' of u' is $\{\lfloor i \cdot \frac{T}{k} + \varphi' \rfloor\}_{i=1}^{k}$. Hence, $u \preceq u'$ if $\varphi \geq \varphi'$ and $u' \preceq u$ if $\varphi' \geq \varphi$. Thus u and u' are ordered. □

Since $\mathcal{R}(T,k)$ is finite it follows from Lemma 2.10 that $\mathcal{R}(T,k)$ contains a greatest element for this order. This greatest element is the sequence in which " the ones are as much to the left as possible " (under the constraint that the sequence is regular of course). This greatest element of $\mathcal{R}(T,k)$ is important for our definition of the (primal) unbalance of a sequence $u \in \mathcal{P}(T,k)$. We denote this greatest element with $\overline{\omega}(T,k)$ or just $\overline{\omega}$ if no confusion is possible. By Lemma 2.8 and the proof of Lemma 2.10 we have the following lemma which can be used to determine $\overline{\omega}(T,k)$ quickly.

Lemma 2.11 *For the support set S_1 of $\overline{\omega}(T,k)$ we have that*

$$S_1 = \{\lfloor i \cdot \frac{T}{k} + 1 - \frac{T}{k} \rfloor\}_{i=1}^{k}.$$

Corollary 2.12 *For $\overline{\omega}(T,k)$ we have that*

$$\kappa_{\overline{\omega}(T,k)}(n) = \lceil n \cdot \frac{k}{T} \rceil \text{ for } n = 0,1,\dots,T.$$

We have seen in Lemma 2.9 that the regular sequences $\mathcal{R}(T,k)$ are a conjugacy class in $\mathcal{P}(T,k)$. We have the following theorem in which the partial order is used to give a characterising property (see Lemma 2.25) of the conjugacy class $\mathcal{R}(T,k)$ of regular sequences in the set of all the conjugacy classes of $\mathcal{P}(T,k)$.

Theorem 2.13 *Every conjugacy class $\widetilde{u}$ of $\mathcal{P}(T,k)$ contains an upper bound of $\mathcal{R}(T,k)$, i.e for every $u \in \mathcal{P}(T,k)$ there exists a $v \in \mathcal{P}(T,k)$ such that $v \sim u$ and $v \succeq w$ for every $w \in \mathcal{R}(T,k)$.*

Proof. Let $u \in \mathcal{P}(T,k)$ and let $l = \operatorname{argmax}_{n=0,1,\dots,T-1} \chi_u(n)$. We claim that $u' := (u_{1+l}, u_{2+l}, \dots, u_T, u_1, u_2, \dots, u_l)$, the l-th cyclic permutation of u, is an upper bound of $\mathcal{R}(T,k)$. Since $u' \sim u$ this proves the theorem.

To prove the claim it suffices to show that $u' \succeq \overline{\omega}$. By definition it follows that $\chi_u(l) \geq \chi_u((l+n) \pmod T)) = \chi_u(l) + \chi_{u'}(n)$ for $n = 0, 1, \dots, T-1$ and thus $\chi_{u'}(n) \leq 0$ for $n = 0, 1, \dots, T-1$. Since $\chi_{u'}(0) = \chi_{u'}(T) = 0$ it follows that

$$\max_{n=1,2,\dots,T} \chi_{u'}(n) = \max_{n=0,1,\dots,T-1} \chi_{u'}(n) = 0.$$

Hence $\kappa_{u'}(n) \geq n \cdot \frac{k}{T}$ and thus $\kappa_{u'}(n) \geq \lceil n \cdot \frac{k}{T} \rceil$ for $n = 1, 2, \dots, T$. So, by Corollary 2.12 we have indeed that $u' \succeq \overline{\omega}$. □

Remark. If $\gcd(k,T) = 1$ then it is easily seen for every $u \in \mathcal{P}(T,k)$ that χ_u is injective on the domain $\{0, 1, \dots, T-1\}$ and thus $\operatorname{argmax}_{n=0,1,\dots,T-1} \chi_u(n)$ is unique. A consequence is that for every $u \in \mathcal{P}(T,k)$ the upper bound of $\mathcal{R}(T,k)$ in the conjugacy class $\widetilde{u}$ is unique if $\gcd(k,T) = 1$. If $\gcd(k,T) > 1$ then the upper bound is in general not unique. For example, if $u = (1,0,0,0,1,0,0,1,0) \in \mathcal{P}(9,3)$ then $\overline{\omega} = (1,0,0,1,0,0,1,0,0)$ and the conjugacy class $\widetilde{u}$ contains two upper bounds of $\mathcal{R}(9,3)$. Namely, $(1,0,1,0,0,0,1,0,0)$ and $(1,0,0,1,0,1,0,0,0)$. Note that for $u \in \mathcal{P}(T,k)$ the proof of Theorem 2.13 yields an algorithm to determine the upper bound(s) of $\mathcal{R}(T,k)$ in the conjugacy class $\widetilde{u}$.

Theorem 2.13 has the following " dual " theorem which is obtained in a completely analogous way. This dual theorem is needed for the definition of the dual unbalance.

Theorem 2.14 *Every conjugacy class $\widetilde{u}$ of $\mathcal{P}(T,k)$ contains a lower bound of $\mathcal{R}(T,k)$.*

We are now ready to define the (primal) unbalance.

Definition 2.15 *Let $u \in \mathcal{P}(T,k)$, $\overline{\omega} = \overline{\omega}(T,k)$ and let u' be an upper bound with respect to the partial order $\preceq$ of $\mathcal{R}(T,k)$ in the conjugacy class $\widetilde{u}$. Then the primal unbalance $\overline{I}$ of u is defined by*

$$\overline{I}(u) = \frac{1}{T} \sum_{n=1}^{T} (\kappa_{u'}(n) - \kappa_{\overline{\omega}}(n)). \tag{1}$$

Theorem 2.16 *The primal unbalance $\overline{I}$ is well defined.*

Proof. The primal unbalance $\overline{I}$ is well defined if and only if for every pair u', u'' of upper bounds of $\mathcal{R}(T,k)$, $u' \sim u''$ we have that $\sum_{n=1}^{T} \kappa_{u'}(n) = \sum_{n=1}^{T} \kappa_{u''}(n)$. We can assume that $u' \neq u''$. Since $u' \sim u''$ there exists an $m \in \{1, 2, \dots, T-1\}$ such that $u' = vw$, $u'' = wv$ where $v = (v_1, v_2, \dots, v_m)$ and $w = (w_1, w_2, \dots, w_{T-m})$. Since $u' \succeq \overline{\omega}$ and $u'' \succeq \overline{\omega}$ we have that $\chi_{u'}(n) \leq 0$ and $\chi_{u''}(n) \leq 0$ for $n = 0, 1, \dots, T$. Moreover we have that $\chi_{u'}(0) = \chi_{u'}(T) = 0$ and $\chi_{u''}(0) = \chi_{u''}(T) = 0$. So,

$$\chi_{u'}(m) = m \cdot \frac{k}{T} - \kappa_{u'}(m) = (T - (T-m)) \cdot \frac{k}{T} - \sum_{i=1}^{m} v_i = (T \cdot \frac{k}{T} - \kappa_{u''}(T))$$

$$-((T-m) \cdot \frac{k}{T} - \kappa_{u''}(T-m)) = \chi_{u''}(T) - \chi_{u''}(T-m) = -\chi_{u''}(T-m)$$

and thus $\chi_{u'}(m) = \chi_{u''}(T-m) = 0$. Hence $\chi_{u'}(0) = \chi_{u''}(T-m)$ and by induction it follows that $\chi_{u'}(n') = \chi_{u''}(T-m+n')$ for $n' = 0, 1, \dots, m$. Further $\chi_{u'}(m) = \chi_{u''}(0)$ and by induction it follows that $\chi_{u'}(n') = \chi_{u''}(n'-m)$ for $n' = m, m+1, \dots, T$. Hence

$$\sum_{n=1}^{T} \chi_{u'}(n) = \sum_{n'=1}^{m} \chi_{u'}(n') + \sum_{n'=m+1}^{T} \chi_{u'}(n') = \sum_{n'=1}^{m} \chi_{u''}(T-m+n') + \sum_{n'=m+1}^{T} \chi_{u''}(n'-m)$$

$$= \sum_{n=1}^{T} \chi_{u''}(n) \text{ and thus } \sum_{n=1}^{T} \kappa_{u'}(n) = \sum_{n=1}^{T} \kappa_{u''}(n).$$

$\square$

Now that we have defined the primal unbalance we also define the dual unbalance. Therefore we use the dual order $\preceq'$ of the partial order $\preceq$, i.e $u \preceq' v$ if and only if $v \preceq u$. With this dual order $\preceq'$ we have in a completely analogous way the same results as for the order $\preceq$. Thus also for this order $\preceq'$ we have that $\mathcal{R}(T,k)$ has a greatest element, which we denote by $\underline{\omega}(T,k)$, or just $\underline{\omega}$ if no confusion is possible. Note that $\underline{\omega}$ is the smallest element of $\mathcal{R}(T,k)$ with respect to the primal order $\preceq$. With this dual order we can restate Theorem 2.14 as " Every conjugacy class $\widetilde{u}$ of $\mathcal{P}(T,k)$ contains an upper bound (with respect to the order $\preceq'$) of $\mathcal{R}(T,k)$." Hence, Theorem 2.14 says the same as Theorem 2.13 except that the primal order $\preceq$ is replaced with the dual order $\preceq'$. Now we also define the dual unbalance.

Definition 2.17 *Let $u \in \mathcal{P}(T,k)$, $\underline{\omega} = \underline{\omega}(T,k)$ and let u' be an upper bound with respect to the (dual) partial order $\preceq'$ of $\mathcal{R}(T,k)$ in the conjugacy class $\widetilde{u}$. Then the dual unbalance $\underline{I}$ of u is defined by*

$$\underline{I}(u) = \frac{1}{T}\sum_{n=1}^{T}(\kappa_{\underline{\omega}}(n) - \kappa_{u'}(n)). \tag{2}$$

The dual unbalance is just as the primal unbalance well defined and we can illustrate the primal unbalance $\overline{I}(u)$ and the dual unbalance $\underline{I}(u)$ of some sequence $u \in \mathcal{P}(T,k)$ graphically in the following way. We draw in the 2-dimensional plane a line segment l with slope $\frac{k}{T-k}$ from lattice point $(0,0)$ to lattice point $(T-k,k)$. Further a sequence $v \in \mathcal{P}(T,k)$ corresponds to a connected graph from $(0,0)$ to $(k, T-k)$ in the following way. The graph starts at $(0,0)$ and goes one length unit to the right or above if the next element of the sequence is a zero respectively an one. To determine $\overline{I}(u)$ and $\underline{I}(u)$ for some $u \in \mathcal{P}(T,k)$ we have to find $u' \in \mathcal{P}(T,k)$ respectively $u'' \in \mathcal{P}(T,k)$ such that u' and u'' are in the same conjugacy class as u, $\overline{\omega} \preceq u'$ and $u'' \preceq \underline{\omega}$. Then if we draw the graphs of $u', u'', \overline{\omega}$ and $\underline{\omega}$ as described above we have the following. We have that the graph of $\overline{\omega}$ never is below the line segment l, but remains as close as possible to l under this condition. Namely, the graph of $\overline{\omega}$ goes through the lattice points that are on l and the lattice points that are immediately above l. Similarly the graph of $\underline{\omega}$ is never above l and goes through the lattice points on l and the lattice points immediately below l. Further the graph of u' is never below $\overline{\omega}$ and thus never below l. It is easily seen that $T \cdot \overline{I}(u) = T \cdot \overline{I}(u')$ is the area below the graph of u' minus the area below the graph of $\overline{\omega}$. Similarly the graph of u'' is never above $\underline{\omega}$ and thus never above l. Moreover $T \cdot \underline{I}(u) = T \cdot \underline{I}(u'')$ is the area below the graph of $\underline{\omega}$ minus the area below the graph of u''.

For example we do this for the sequence $u = (1,0,0,0,1,0,0,1,0) \in \mathcal{P}(9,3)$ of the previous example. We take $u' = (1,0,1,0,0,0,1,0,0)$ as upper bound of $\mathcal{R}(9,3)$ in the conjugacy class of u and we take $u'' = (0,0,0,1,0,0,1,0,1)$ as lower bound of $\mathcal{R}(9,3)$ in the conjugacy class of u. In Figure 1 are the graphs of $\overline{\omega}, \underline{\omega}, u', u''$ and the line segment l. Note that we have drawn the lines of the graphs above l and below l thicker if the graphs of u' and $\overline{\omega}$ respectively the graphs of u'' and $\underline{\omega}$ coincide. It follows that $\overline{I}(u) = \frac{1}{9}$ and $\underline{I}(u) = \frac{2}{9}$.

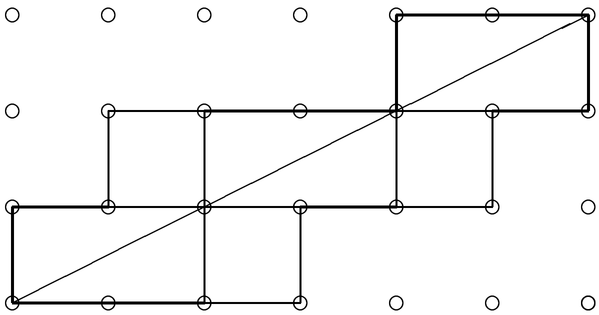

Figure 1

A direct consequence of the definitions of the primal and dual unbalance is the following lemma which states that conjugates have the same primal unbalance and the same dual unbalance.

Lemma 2.18 *If $u \sim u'$ then $\overline{I}(u) = \overline{I}(u')$ and $\underline{I}(u) = \underline{I}(u')$.*

For a finite word (sequence) $u = (u_1, u_2, \ldots, u_T)$ we denote by $\overleftarrow{u} := (u_T, u_{T-1}, \ldots, u_1)$ the reversed word. Further if for some sequences $u, v \in \mathcal{P}(T,k)$ we have that $\overleftarrow{u} \sim v$ then we say that u and v are mirrors of each other. Note that if $u \in \mathcal{P}(T,k)$ is regular then u is a mirror of itself.

Proposition 2.19 *If $u, v \in \mathcal{P}(T,k)$ are mirrors of each other then $\overline{I}(u) = \underline{I}(v)$ and $\underline{I}(u) = \overline{I}(v)$.*

Proof. Let $u' \sim u$ be an upper bound with respect to the partial order $\preceq$ of $\mathcal{R}(T,k)$ and put $v' = \overleftarrow{u'}$. Then

$$\overline{I}(u) = \frac{1}{T}\sum_{n=1}^{T}(\kappa_{u'}(n) - \kappa_{\overline{\omega}}(n)) = \frac{1}{T}\sum_{n=0}^{T}(\kappa_{u'}(n) - \lceil n \cdot \frac{k}{T} \rceil)$$

and $v' \sim v$. Moreover for $n' = 0, 1, \ldots, T$ we have that

$$\kappa_{v'}(T - n') = k - \kappa_{u'}(n') \leq k - \lceil n' \cdot \frac{k}{T} \rceil = \lfloor (T - n') \cdot \frac{k}{T} \rfloor$$

and thus $\kappa_{v'}(n) \leq \lfloor n \cdot \frac{k}{T} \rfloor$ for $n = 0, 1, 2, \ldots, T$. Hence $v' \sim v$ is an upper bound with respect to the partial order $\preceq'$ of $\mathcal{R}(T,k)$. Thus

$$\underline{I}(v) = \frac{1}{T}\sum_{n'=1}^{T}(\kappa_{\underline{\omega}}(n') - \kappa_{v'}(n')) = \frac{1}{T}\sum_{n'=0}^{T}(\lfloor n' \cdot \frac{k}{T} \rfloor - (k - \kappa_{u'}(T - n')) =$$

$$\frac{1}{T} \cdot \sum_{n'=0}^{T}(\kappa_{u'}(T - n') - \lceil (T - n') \cdot \frac{k}{T} \rceil) = \frac{1}{T} \cdot \sum_{n=0}^{T}(\kappa_{u'}(n) - \lceil n \frac{k}{T} \rceil) = \overline{I}(u).$$

Analogously it follows that $\underline{I}(u) = \overline{I}(v)$. □

For example if $u = (1,0,1,1,0,1,1,1,0,1,0) \in \mathcal{P}(11,7)$ then $v = (1,1,0,1,1,0,1,0,1,0,1) \in \mathcal{P}(11,7)$ is a mirror of u. The reader can check that $\overline{I}(u) = \underline{I}(v) = \frac{1}{11}$ and $\underline{I}(u) = \overline{I}(v) = \frac{2}{11}$.

We also define the primal unbalance $\overline{I}$ and the dual unbalance $\underline{I}$ on the set of conjugacy classes by $\overline{I}(\widetilde{u}) = \overline{I}(u)$ and $\underline{I}(\widetilde{u}) = \underline{I}(u)$, where u is an arbitrary representative of $\widetilde{u}$. By Lemma 2.18 this is well defined.

Lemma 2.20 *If $u = v^m$ where v is a primitive sequence and $m \in \mathbb{N}$ then $\overline{I}(u) = \overline{I}(v)$ and $\underline{I}(u) = \underline{I}(v)$.*

Proof. We can assume that $v \in \mathcal{P}(T,k)$. Then we have that $u \in \mathcal{P}(mT, mk)$ and by Lemma 2.11 we have that $\overline{\omega}(mT, mk) = \overline{\omega}(T,k)^m$. Further if $w \sim v$ and w is an upper bound of $\mathcal{R}(T,k)$ then $w^m \sim u$ and w^m is an upper bound of $\mathcal{R}(mT, mk)$. Thus we have for every $l \in \{1, 2, \ldots, T\}$ and $b \in \{0, 1, \ldots, m-1\}$ that

$$\kappa_w(l) - \kappa_{\overline{\omega}(T,k)}(l) = \kappa_{w^m}(l + bT) - \kappa_{\overline{\omega}(mT,mk)}(l + bT).$$

Hence

$$\begin{aligned}\overline{I}(u) &= \frac{1}{mT}\sum_{n=1}^{mT}\{\kappa_{w^m}(n) - \kappa_{\overline{\omega}(mT,mk)}(n)\} = \frac{1}{mT}\cdot m \cdot \sum_{n=1}^{T}\{\kappa_w(n) - \kappa_{\overline{\omega}(T,k)}(n)\} \\ &= \frac{1}{T}\sum_{n=1}^{T}\{\kappa_w(n) - \kappa_{\overline{\omega}(T,k)}(n)\} = \overline{I}(v).\end{aligned}$$

It follows completely analogously that $\underline{I}(u) = \underline{I}(v)$. □

Lemma 2.21 *For every $u \in \mathcal{P}(T,k)$ we have that $T \cdot \overline{I}(u) \in \mathbb{Z}_{\geq 0}$ and $\overline{I}(u) = 0$ if and only if $u \in \mathcal{R}(T,k)$. The same is true for the dual unbalance $\underline{I}(u)$.*

Proof. Let u' as in the definition of the the primal unbalance. Then $u' \succeq \overline{\omega}$ and thus $\kappa_{u'}(n) - \kappa_{\overline{\omega}}(n) \in \mathbb{Z}_{\geq 0}$ for $n = 1, 2, \ldots, T$. Thus $T \cdot \overline{I}(u) = \sum_{n=1}^{T}(\kappa_{u'}(n) - \kappa_{\overline{\omega}}(n)) \in \mathbb{Z}_{\geq 0}$. If $u \in \mathcal{R}(T,k)$ then by Lemma 2.9 we have that $u' = \overline{\omega}$ and thus $\overline{I}(u) = 0$. Conversely, if $\overline{I}(u) = 0$ then it follows that $\kappa_{u'}(n) - \kappa_{\overline{\omega}}(n) = 0$ for $n = 1, 2, \ldots, T$ and thus $u' = \overline{\omega} \in \mathcal{R}(T,k)$. So, by Lemma 2.9 we have that $u \in \mathcal{R}(T,k)$. For the dual unbalance $\underline{I}(u)$ the proof is completely analogous. □

Lemma 2.22 *Let $u, u', v, v' \in \mathcal{P}(T,k)$. Then $u \sim u'$, $v \sim v'$, $u \succeq v \succeq \overline{\omega}$ and $v' \succeq u' \succeq \overline{\omega}$ implies $u = v$ and $u' = v'$.*

Proof. By Lemma 2.18, the given orders and the definition of $\overline{I}(u)$ we have that $\overline{I}(u) = \overline{I}(u')$, $\overline{I}(v) = \overline{I}(v')$, $\overline{I}(u) \geq \overline{I}(v)$ and $\overline{I}(v') \geq \overline{I}(u')$. Hence $\overline{I}(u) = \overline{I}(u') = \overline{I}(v') = \overline{I}(v)$ and thus $u = v$ and $u' = v'$. □

In the sequel we denote the set of conjugacy classes of $\mathcal{P}(T,k)$ by $\widetilde{\mathcal{P}}(T,k)$ and we denote the conjugacy class of regular sequences in $\mathcal{P}(T,k)$ by $\widetilde{\omega}(T,k)$ or just $\widetilde{\omega}$. We now define a partial order $\trianglelefteq_{\overline{g}}$ and a partial order $\trianglelefteq_{\underline{g}}$ on $\widetilde{\mathcal{P}}(T,k)$.

Definition 2.23 *Let $\widetilde{u}, \widetilde{v} \in \widetilde{\mathcal{P}}(T,k)$. Then $\widetilde{u} \trianglelefteq_{\overline{g}} \widetilde{v}$ if there exist $u', v' \in \mathcal{P}(T,k)$ such that $u' \in \widetilde{u}$, $v' \in \widetilde{v}$ and $\overline{\omega} \preceq u' \preceq v'$. Further $\widetilde{u} \trianglelefteq_{\underline{g}} \widetilde{v}$ if there exist $u'', v'' \in \mathcal{P}(T,k)$ such that $u'' \in \widetilde{u}$, $v'' \in \widetilde{v}$ and $\underline{\omega} \preceq' u'' \preceq' v''$.*

Suppose that $\widetilde{u} \trianglelefteq_{\overline{g}} \widetilde{v}$ for some $u, v \in \mathcal{P}(T,k)$ and let u', v' be as in Definition 2.23. Then the graph of u' is never above the graph of v' and they are both never below the graph of $\overline{\omega}$. Therefore we call $\trianglelefteq_{\overline{g}}$ the upper graph order and similarly we call $\trianglelefteq_{\underline{g}}$ the lower graph order.

Lemma 2.24 *$\trianglelefteq_{\overline{g}}$ and $\trianglelefteq_{\underline{g}}$ are indeed partial orders on $\widetilde{\mathcal{P}}(T,k)$.*

Proof. From Theorem 2.13 the reflexivity of $\trianglelefteq_{\overline{g}}$ follows and from Lemma 2.22 the antisymmetry. Further the transitivity of $\trianglelefteq_{\overline{g}}$ follows directly from the definition and thus $\trianglelefteq_{\overline{g}}$ is a partial order on $\widetilde{\mathcal{P}}(T,k)$. The proof that $\trianglelefteq_{\underline{g}}$ is a partial order on $\widetilde{\mathcal{P}}(T,k)$ is completely analogous. □

As a direct consequence of Theorem 2.13 and the " dual " Theorem 2.14 we have the following result.

Lemma 2.25 *The conjugacy class $\widetilde{\omega}$ is the smallest element of $\widetilde{\mathcal{P}}(T,k)$ with respect to the order $\trianglelefteq_{\overline{g}}$ and also with respect to the order $\trianglelefteq_{\underline{g}}$.*

We say that $\widetilde{u} \vartriangleleft_{\overline{g}} \widetilde{v}$ if $\widetilde{u} \trianglelefteq_{\overline{g}} \widetilde{v}$ and $\widetilde{u} \neq \widetilde{v}$ and similar for the other partial orders. The following theorem follows directly from the definitions.

Theorem 2.26 *Let $u,v \in \mathcal{P}(T,k)$. Then $\widetilde{u} \trianglelefteq_{\overline{g}} \widetilde{v}$ implies $\overline{I}(u) \leq \overline{I}(v)$ and $\widetilde{u} \trianglelefteq_{\underline{g}} \widetilde{v}$ implies $\underline{I}(u) \leq \underline{I}(v)$. Moreover, $\widetilde{u} \vartriangleleft_{\overline{g}} \widetilde{v}$ implies $\overline{I}(u) < \overline{I}(v)$ and $\widetilde{u} \vartriangleleft_{\underline{g}} \widetilde{v}$ implies $\underline{I}(u) < \underline{I}(v)$.*

The orders $\trianglelefteq_{\overline{g}}$ and $\trianglelefteq_{\underline{g}}$ are in general not the same on $\widetilde{\mathcal{P}}(T,k)$. For example if $u = (1,0,1,0,0,0,1,0,0) \in \mathcal{P}(9,3)$ and $v = (1,0,1,0,0,1,0,0,0) \in \mathcal{P}(9,3)$ then $\widetilde{u} \vartriangleleft_{\overline{g}} \widetilde{v}$ and $\widetilde{v} \vartriangleleft_{\underline{g}} \widetilde{u}$. Therefore it is useful to define the partial order $\trianglelefteq_{g}$ on $\widetilde{\mathcal{P}}(T,k)$ as "intersection" of the orders $\trianglelefteq_{\overline{g}}$ and $\trianglelefteq_{\underline{g}}$. We call $\trianglelefteq_{g}$ the (strong) graph order.

Definition 2.27 *Let $\widetilde{u},\widetilde{v} \in \widetilde{\mathcal{P}}(T,k)$. Then $\widetilde{u} \trianglelefteq_{g} \widetilde{v}$ if $\widetilde{u} \trianglelefteq_{\overline{g}} \widetilde{v}$ and $\widetilde{u} \trianglelefteq_{\underline{g}} \widetilde{v}$.*

Remark. The partial orders $\trianglelefteq_{\overline{g}}$, $\trianglelefteq_{\underline{g}}$ and $\trianglelefteq_{g}$ on $\widetilde{\mathcal{P}}(T,k)$ induce in a natural way corresponding pre-orders on $\mathcal{P}(T,k)$. For example for $u,v \in \mathcal{P}(T,k)$ we say that $u \trianglelefteq_{g} v$ if $\widetilde{u} \trianglelefteq_{g} \widetilde{v}$ and similar for the other orders. However, if the orders $\trianglelefteq_{\overline{g}}$, $\trianglelefteq_{\underline{g}}$ and $\trianglelefteq_{g}$ are used on $\mathcal{P}(T,k)$ in this way then they are not antisymmetric, since cyclic permutations of u are different words and therefore they are not partial orders on $\mathcal{P}(T,k)$.

Now that we have defined the unbalance and several orders for finite sequences we will extend these notions to infinite periodic sequences (words). First of all we extend in a natural way the partial order $\preceq$ and its dual order $\preceq'$ to the infinite sequences $u = (u_1, u_2, \dots)$ of zeros and ones corresponding to $\mathbb{N}$-words on the alphabet $\{0,1\}$. We define just as for finite sequences (words) the counting function $\kappa_u : \mathbb{Z}_{\geq 0} \to \mathbb{Z}$ by $\kappa_u(n) = \sum_{t=1}^{n} u_t$ and we say that $u \preceq v$ if $\kappa_u(n) \leq \kappa_v(n)$ for $n = 1,2,\dots$. For a sequence $u = (u_1, u_2, \dots)$ of zeros and ones of density d corresponding to an $\mathbb{N}$ - word we also define the discrepancy function $\chi_u : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ by $\chi_u(n) = n \cdot d - \kappa_u(n)$ for $n \in \mathbb{Z}_{\geq 0}$. Further we have the following equivalence relation $\sim$ for infinite periodic sequences. In fact $\sim$ is an equivalence relation for all infinite periodic integer sequences and the corresponding words on a finite alphabet A and not only for the sequences of zeros and ones.

Definition 2.28 *Let u and v be infinite periodic integer sequences. Then we say that u is equivalent to v, $u \sim v$ if there exists a finite sequence w such that w is a period cycle of both u and v.*

In our application of routing sequences in a queueing system we have in general that equivalent infinite periodic sequences have the same performance. Therefore in most cases we consider infinite periodic sequences modulo this equivalence relation, i.e if $u \sim v$ then we consider them as the same sequence. This is also necessary for the several partial orders that we define on infinite periodic sequences of zeros and ones to be antisymmetric. Further an infinite periodic sequence is now completely determined by a period cycle of the sequence. If $u \in \mathcal{P}(T,k)$ then $u^\infty = (u_1, u_2, \dots, u_T)^\infty$ is the infinite periodic sequence with period cycle u.

Let $\mathcal{P}$ be the set of infinite periodic sequences of zeros and ones (modulo the equivalence relation), $\mathcal{R} \subseteq \mathcal{P}$ the subset of regular periodic sequences. Further for $d \in \mathbb{Q}$, $0 \le d \le 1$ let $\mathcal{P}(d) \subseteq \mathcal{P}$ be the subset of sequences with density d and $\mathcal{R}(d) := \mathcal{R} \cap \mathcal{P}(d)$. Then we have the following lemma.

Lemma 2.29 *We have that*

$$\mathcal{P} = \cup_{d\in\mathbb{Q}, 0\le d\le 1} \mathcal{P}(d),$$

$$\mathcal{R} = \cup_{d\in\mathbb{Q}, 0\le d\le 1} \mathcal{R}(d)$$

and $|\mathcal{R}(d)| = 1$ *for every* $d \in \mathbb{Q}$, $0 \le d \le 1$.

Proof. If $u \in \mathcal{P}$ then there exist $k \in \mathbb{Z}_{\ge 0}$, $T \in \mathbb{N}$, $k \le T$ such that $u = v^\infty$ for some $v \in \mathcal{P}(T,k)$ and thus $u \in \mathcal{P}(\frac{k}{T}) = \mathcal{P}(d)$ for some $d \in Q$, $0 \le d \le 1$. Hence

$$\mathcal{P} = \cup_{d\in\mathbb{Q}, 0\le d\le 1} \mathcal{P}(d)$$

and thus

$$\mathcal{R} = \mathcal{R} \cap \mathcal{P} = \mathcal{R} \cap \{\cup_{d\in\mathbb{Q}, 0\le d\le 1} \mathcal{P}(d)\} = \cup_{d\in\mathbb{Q}, 0\le d\le 1} \mathcal{R}(d).$$

By Lemma 2.9 we have that $|\mathcal{R}(d)| = 1$ for every $d \in \mathbb{Q}$, $0 \le d \le 1$. □

We denote the unique element of $\mathcal{R}(d)$ by $\omega(d)$ for every $d \in \mathbb{Q}$, $0 \le d \le 1$. We now define the primal and dual unbalance for infinite periodic sequences.

Definition 2.30 *Let* $u \in \mathcal{P}$ *and let* $u' \in \mathcal{P}(T,k)$ *be a period cycle of* u. *Then we define the primal unbalance of* u *as* $\overline{I}(u) := \overline{I}(u')$ *and we define the dual unbalance of* u *as* $\underline{I}(u) := \underline{I}(u')$.

Lemma 2.31 *The primal unbalance* $\overline{I}$ *and dual unbalance* $\underline{I}$ *are well defined on* $\mathcal{P}$.

Proof. Suppose u' and u'' are both period cycles of u. Then we have to show that $\overline{I}(u') = \overline{I}(u'')$. There exists a primitive period cycle w of u such that $u' \sim w^m$ and $u'' \sim w^n$ for some $m, n \in \mathbb{N}$. Then by Lemma 2.18 and Lemma 2.20 we have that $\overline{I}(u') = \overline{I}(w) = \overline{I}(u'')$. Hence $\overline{I}$ is well defined on $\mathcal{P}$ and completely analogously it follows that $\underline{I}$ is well defined on $\mathcal{P}$. □

Remark. When the unbalance of an infinite periodic sequence has to be computed it is of course most practical to compute the unbalance of a primitive period cycle of the sequence and not an arbitrary period cycle.

Suppose that $u, v \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$, $0 \le d \le 1$. Then we say that u, v are mirrors of each other if there exist u', v' such that u' and v' are period cycles of u respectively v and u' and v' are mirrors of each other. By proposition 2.19 and the definitions of the primal and dual unbalance for infinite period sequences of zeros and ones we have the following corollary.

Corollary 2.32 *Suppose that $u, v \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$, $0 \le d \le 1$ and u and v are mirrors of each other. Then $\overline{I}(u) = \underline{I}(v)$ and $\underline{I}(u) = \overline{I}(v)$.*

Now that we have defined the primal and dual unbalance on $\mathcal{P}$ we also define for every $d \in \mathbb{Q}$, $0 \le d \le 1$ the partial orders $\trianglelefteq_{\overline{g}}$, $\trianglelefteq_{\underline{g}}$ and $\trianglelefteq_g$ on $\mathcal{P}(d)$. Let $u, v \in \mathcal{P}(d)$ and let f_1 be the period of u and f_2 the the period of v. Then by Lemma 2.3 we have that u and v have both period $T := \mathrm{lcm}(f_1, f_2)$. Moreover, if u' is a period cycle of length T of u and v' is a period cycle of length T of v then we have that $u' \in \mathcal{P}(T, d \cdot T)$ and $v' \in \mathcal{P}(T, d \cdot T)$.

Definition 2.33 *Let $u, v \in \mathcal{P}(d)$ for some $d \in Q, 0 \le d \le 1$ and let $u', v' \in \mathcal{P}(T, d \cdot T)$ be corresponding period cycles as above. Then we define the partial orders $\trianglelefteq_{\overline{g}}$, $\trianglelefteq_{\underline{g}}$ and $\trianglelefteq_g$ on $\mathcal{P}(d)$ by*

$$
\begin{aligned}
&u \trianglelefteq_{\overline{g}} v \text{ if } \widetilde{u'} \trianglelefteq_{\overline{g}} \widetilde{v'} \text{ in } \widetilde{\mathcal{P}}(T, d \cdot T),\\
&u \trianglelefteq_{\underline{g}} v \text{ if } \widetilde{u'} \trianglelefteq_{\underline{g}} \widetilde{v'} \text{ in } \widetilde{\mathcal{P}}(T, d \cdot T) \text{ and}\\
&u \trianglelefteq_{g} v \text{ if } \widetilde{u'} \trianglelefteq_{g} \widetilde{v'} \text{ in } \widetilde{\mathcal{P}}(T, d \cdot T).
\end{aligned}
$$

All the results we had for the partial orders $\trianglelefteq_{\overline{g}}$, $\trianglelefteq_{\underline{g}}$ and $\trianglelefteq_g$ on $\widetilde{\mathcal{P}}(T, k)$ we also have for the partial orders $\trianglelefteq_{\overline{g}}$, $\trianglelefteq_{\underline{g}}$ and $\trianglelefteq_g$ on $\mathcal{P}(d)$. They are indeed partial orders and we have the following results.

Lemma 2.34 *For every $d \in \mathbb{Q}$, $0 \le d \le 1$, we have that the regular sequence $\omega(d)$ is the smallest element of $\mathcal{P}(d)$ with respect to the order $\trianglelefteq_{\overline{g}}$, the order $\trianglelefteq_{\underline{g}}$ and with respect to the order $\trianglelefteq_g$.*

Theorem 2.35 *Let $u, v \in \mathcal{P}(d)$. Then $u \trianglelefteq_{\overline{g}} v$ implies $\overline{I}(u) \le \overline{I}(v)$ and $u \trianglelefteq_{\underline{g}} v$ implies $\underline{I}(u) \le \underline{I}(v)$. Moreover, $u \triangleleft_{\overline{g}} v$ implies $\overline{I}(u) < \overline{I}(v)$ and $u \triangleleft_{\underline{g}} v$ implies $\underline{I}(u) < \underline{I}(v)$.*

We give an example to illustrate this . Let $u = (1,1,0,1,0,0)^\infty \in \mathcal{P}(\frac{1}{2})$ and $v = (1,1,0,1,0,1,0,1,0,1,0,0)^\infty \in \mathcal{P}(\frac{1}{2})$. Then we have that $u \triangleleft_{\overline{g}} v$, $v \triangleleft_{\underline{g}} u$ and (thus) u and v are not $\trianglelefteq_g$ ordered. Further $\overline{I}(u) = \frac{1}{3}$, $\overline{I}(v) = \frac{5}{12}$, $\underline{I}(u) = \frac{1}{6}$ and $\underline{I}(v) = \frac{1}{12}$.

3 Bounding the difference between sequences in expected average waiting time

Let $\{T_i\}_{i=1,2,\ldots}$ be a sequence of arrival times of customers, with the convention that $T_1 = 0$. Put $\delta_i := T_{i+1} - T_i$ for $i = 1, 2, \ldots$. Then $\{\delta_i\}$ is the sequence of interarrival times. Further a fraction of these arriving customers is routed to a server according to some routing sequence $u = (u_1, u_2, \ldots)$ of zeros and ones. For such a routing sequence we have the counting function $\kappa_u(n) = \sum_{t=1}^{n} u_t$ that we used to define a partial order $\preceq$ for such routing sequences. Further we define the the following related function $\nu_u(i) : \mathbb{Z}_{\ge 0} \to \mathbb{Z}$ with $\nu_u(j) = \min\{n \in \mathbb{Z}_{\ge 0} : \kappa_u(n) = j\}$ and we put $\tau_u(j) = \sum_{i=\max(\nu_u(j-1),1)}^{\nu_u(j)-1} \delta_i$ for $j = 1, 2, \ldots$. Then $\tau_u(j)$ is the time elapsed between the routing of the $j-1$-th and j-th customer to the server according to routing sequence u. If we put $\lambda_u(j) = \sum_{k=1}^{j} \tau_u(k)$ for $j = 1, 2, \ldots$ then $\lambda_u(j)$ is the time at which the $j-th$ customer is routed to the server according to routing sequence u. Further we have a sequence of service times $\{\sigma_j\}_{j=1,2,\ldots}$, where σ_j is the service time of the j-th customer that is routed to the server according to any routing sequence. For now we assume that the interarrival times $\{\delta_i\}_{i=1,2,\ldots}$ and service times $\{\sigma_j\}_{j=1,2,\ldots}$ are fixed sequences of non-negative real numbers. Later we will also consider the case that these are random variables. Further we define $W_u(j,w)$ to be the workload for the server at the moment the j-th customer is routed to the server according to routing sequence u, given that the initial workload at time $T_1 = 0$ is equal to w. In other words $W_u(j,w)$ is the waiting time for the j -th customer that is routed to the server. If the initial workload w is fixed (thus the system starts for all routing sequences u with the same initial workload w) then we write for short $W_u(j)$ in stead of $W_u(j,w)$. For example this is the case if it is assumed that the server is empty at $T_1 = 0$. Note that $W_u(j) + \lambda_u(j)$ is the moment that the server starts serving the j-th customer that is routed to the server. We have the following lemma.

Lemma 3.1 *Let $u = (u_1, u_2, \dots)$ and $v = (v_1, v_2, \dots)$ be routing sequences as above. If $u \preceq v$ and $W_v(j,w) + \lambda_v(j) \leq W_u(j,w') + \lambda_u(j)$ for some $j \in \mathbb{N}$ then for every $k \geq j$ we have that $W_v(k,w) + \lambda_v(k) \leq W_u(k,w') + \lambda_u(k)$.*

Proof. It suffices to show the assertion for $k = j+1$, because it then follows with induction. We have that $W_v(j+1,w) = \max(W_v(j,w)+\sigma_j-\tau_v(j+1),0)$ and $W_u(j+1,w') = \max(W_u(j,w')+\sigma_j-\tau_u(j+1),0)$. Further $\lambda_u(j+1) - \lambda_v(j+1) \geq 0$ since $u \preceq v$. Hence

$$W_v(j+1,w) \leq \max(W_u(j,w') + \lambda_u(j) - \lambda_v(j) + \sigma_j - \tau_v(j+1), 0) \leq$$

$$\max(W_u(j,w') + \sigma_j - \tau_u(j+1), 0) + \max(\lambda_u(j+1) - \lambda_v(j+1), 0) =$$

$$W_u(j+1,w') + \lambda_u(j+1) - \lambda_v(j+1).$$

$\square$

If $w = w'$ then the inequality holds for $j = 1$ and thus we have the following corollary.

Corollary 3.2 *Let u and v as above, $u \preceq v$. Then $W_v(j) + \lambda_v(j) \leq W_u(j) + \lambda_u(j)$ for every $j \in \mathbb{N}$.*

Note that for every $j \in \mathbb{N}$ we have that $\lambda_u(j) - \lambda_v(j) = \sum_{i=\nu_v(j)}^{\nu_u(j)-1} \delta_i$. This is a sum of interarrival times δ_i, where the number of terms in the sum is $\nu_u(j) - \nu_v(j)$. Therefore, if we put $N_{uv}(m) = \sum_{j=1}^{m}(\nu_u(j) - \nu_v(j))$ for $m = 1, 2, \dots$ then we have by the following lemma by Corollary 3.2.

Lemma 3.3 *Let u and v be routing sequences with $u \preceq v$. For every $m \in \mathbb{N}$ we have that*

$$\sum_{j=1}^{m} W_v(j) \leq \sum_{j=1}^{m} W_u(j) + \sum_{j=1}^{m} \sum_{i=\nu_v(j)}^{\nu_u(j)-1} \delta_i. \tag{3}$$

In the double sum the number of terms is $N_{uv}(m)$.

Since for every sequence $u = (u_1, u_2, \dots)$ the functions κ_u and ν_u are strongly related , we can also use the counting function κ instead of ν to find an expression for $N_{uv}(m)$. Namely, we have the following lemma.

Lemma 3.4 *Let $u = (u_1, u_2, \dots)$ and $v = (v_1, v_2, \dots)$ be sequences of zeros and ones. Then we have for every $m \in \mathbb{N}$ that*

$$\sum_{j=1}^{m}(\nu_u(j) - \nu_v(j)) = \sum_{n=1}^{\max(\nu_u(m),\nu_v(m))} \{\min(m, \kappa_v(n)) - \min(m, \kappa_u(n))\}.$$

Proof. We prove it by induction to m. For $m = 1$ we may assume that $\nu_u(1) \geq \nu_v(1)$. Then we have that

$$\nu_u(1) - \nu_v(1) = \sum_{n=\nu_v(1)}^{\nu_u(1)-1} 1 = \sum_{n=1}^{\nu_u(1)-1} \min(1, \kappa_v(n)) =$$

$$\sum_{n=1}^{\max(\nu_u(1),\nu_v(1))} \{\min(1, \kappa_v(n)) - \min(1, \kappa_u(n))\}.$$

Similarly we have for $m = 2, 3, \ldots$ that

$$\nu_u(m) - \nu_v(m) = \sum_{n=\min(\nu_u(m),\nu_v(m))}^{\max(\nu_u(m),\nu_v(m))-1} \{\min(\kappa_v(n), \kappa_u(n) + 1) - \min(\kappa_u(n), \kappa_v(n) + 1)\}.$$

So, by induction we have for $m = 2, 3, \ldots$ that

$$\sum_{j=1}^{m} (\nu_u(j) - \nu_v(j)) = \sum_{j=1}^{m-1} (\nu_u(j) - \nu_v(j)) + (\nu_u(m) - \nu_v(m)) =$$

$$\sum_{n=1}^{\max(\nu_u(m-1),\nu_v(m-1))} \{\min(m-1, \kappa_v(n)) - \min(m-1, \kappa_u(n))\} +$$

$$\sum_{n=\min(\nu_u(m),\nu_v(m))}^{\max(\nu_u(m),\nu_v(m))-1} \{\min(\kappa_v(n), \kappa_u(n) + 1) - \min(\kappa_u(n), \kappa_v(n) + 1)\} =$$

$$\sum_{n=1}^{\max(\nu_u(m),\nu_v(m))} \{\min(m, \kappa_v(n)) - \min(m, \kappa_u(n))\}.$$

$\square$

Corollary 3.5 *Let u and v be sequences as above and suppose that for some $l, m \in N$ we have that $\kappa_u(l) = \kappa_v(l) = m$. Then $N_{uv}(m) = \sum_{n=1}^{l} (\kappa_v(n) - \kappa_u(n))$.*

For periodic sequences of the same density we have the following lemma.

Lemma 3.6 *Suppose that $u = (u')^\infty$ and $v = (v')^\infty$ for some $u', v' \in \mathcal{P}(T, k)$ and $\overline{\omega}(T, k) \preceq u' \preceq v'$. Then if $m = sk$ for some $s \in \mathbb{N}$ we have that $N_{uv}(m) = s \cdot T \cdot (\overline{I}(v) - \overline{I}(u))$.*

Proof. Let $m = sk$ with $s/in/N$. Then $\kappa_u(sT) = \kappa_v(sT) = sk$. So, by Corollary 3.5 we have that $N_{uv}(m) = \sum_{n=1}^{sT} (\kappa_v(n) - \kappa_u(n))$. So, since $\{\kappa_v(n) - \kappa_u(n)\}$ is periodic with period T we have that

$$N_{uv}(m) = \sum_{n=1}^{sT} (\kappa_v(n) - \kappa_u(n)) = s \cdot \sum_{n=1}^{T} (\kappa_v(n) - \kappa_u(n)) =$$

$$s \cdot \sum_{n=1}^{T} \{(\kappa_v(n) - \kappa_{\overline{\omega}}(n)) - (\kappa_u(n) - \kappa_{\overline{\omega}}(n))\} = s \cdot T \cdot (\overline{I}(v) - \overline{I}(u)).$$

$\square$

Corollary 3.7 *Suppose that* $u = (u')^\infty$ *and* $v = (v')^\infty$ *for some* $u', v' \in \mathcal{P}(T,k)$ *and* $\overline{\omega}(T,k) \preceq u' \preceq v'$. *Then for every* $m \in \mathbb{N}$ *we have that*

$$\lfloor \frac{m}{k} \rfloor \cdot T \cdot (\overline{I}(v) - \overline{I}(u)) \leq N_{uv}(m) \leq \lceil \frac{m}{k} \rceil \cdot T \cdot (\overline{I}(v) - \overline{I}(u))$$

and $\lim_{m\to\infty} \frac{N_{uv}(m)}{m} = \frac{T}{k} \cdot (\overline{I}(v) - \overline{I}(u))$.

Proof. Because $u' \preceq v'$ we have that $u \preceq v$ and it follows that $N_{uv}(m)$ is monotonically non-decreasing in m. Therefore the first statement follows from Lemma 3.6 and the second statement follows from the first. $\square$.

By Lemma 3.3 and Corollary 3.7 we have the following theorem.

Theorem 3.8 *Let* u *and* v *be as in Corollary 3.7. Suppose that for* $\{\delta_i\}_{i=1,2,\ldots}$, *the sequence of inter-arrival times, we have for* $l = 1, 2, \ldots, T$ *that* $\limsup_{t\to\infty} \frac{1}{t} \cdot \sum_{i=0}^{t-1} \delta_{i\cdot T+l} := \gamma_l \leq \delta$. *Then*

$$\limsup_{m\to\infty} \frac{1}{m} \cdot (\sum_{j=1}^{m} W_v(j) - \sum_{j=1}^{m} W_u(j)) \leq \delta \cdot \frac{T}{k} \cdot (\overline{I}(v) - \overline{I}(u)).$$

Proof. By Lemma 3.3 together with some straightforward derivations,

$$\limsup_{m\to\infty} \frac{1}{m} \cdot (\sum_{j=1}^{m} W_v(j) - \sum_{j=1}^{m} W_u(j)) \leq \limsup_{m\to\infty} \frac{N_{uv}(m)}{m} \cdot \max_{l=1,\ldots,T} \gamma_l \leq \delta \cdot \limsup_{m\to\infty} \frac{N_{uv}(m)}{m}.$$

Hence by Corollary 3.7

$$\limsup_{m\to\infty} \frac{1}{m} \cdot (\sum_{j=1}^{m} W_v(j) - \sum_{j=1}^{m} W_u(j)) \leq \delta \cdot \frac{T}{k} \cdot (\overline{I}(v) - \overline{I}(u)).$$

$\square$

Let u be a routing sequences of zeros and ones. If the interarrival times $\{\delta_i\}$ and service times $\{\sigma_j\}$ are fixed sequences then we say that $\overline{W}(u)$ is the long-run average waiting time of customers routed to the server according to routing sequence u if $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^{m} W_u(j) = \overline{W}(u)$. So, if $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^{m} W_u(j)$ exists then we say that $\overline{W}(u)$ exists. If $\overline{W}(u)$ exists, but the value depends on the initial workload w then $\overline{W}(u)$ denotes the value for initial workload $w = 0$ (thus the server is assumed to be empty at $T_1 = 0$ in that case).
If the interarrival times $\{\delta_i\}$ and service times $\{\sigma_j\}$ are random variables then we say that $\overline{W}(u)$ is the almost sure long-run average waiting time of customers routed to the server according to routing sequence u if $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^{m} W_u(j) = \overline{W}(u)$ with probability one. So, if in case of random variables $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^{m} W_u(j)$ almost surely (a.s.) exists then we say that $\overline{W}(u)$ exists and again if the value of $\overline{W}(u)$ would depend on the initial workload w then $\overline{W}(u)$ denotes the value for $w = 0$.

Theorem 3.8 has the following corollary.

Corollary 3.9 *Let* $\{\delta_i\}$, u *and* v *be as in Theorem 3.8. If* $\overline{W}(u)$ *and* $\overline{W}(v)$ *exist then*

$$\overline{W}(v) - \overline{W}(u) \leq \delta \cdot \frac{T}{k} \cdot (\overline{I}(v) - \overline{I}(u)).$$

From ergodic theory we have the following theorem for stochastic interarrival and service times. In fact a more general result will be proved in the appendix (see Theorem 6.1).

Theorem 3.10 *Suppose that the interarrival times $\{\delta_i\}$ of customers arriving at the system are independent and identically distributed (i.i.d.) random variables with mean δ and the service times $\{\sigma_j\}$ of the considered server are i.i.d. random variables with mean σ which are independent of the interarrival times. Further let u' and u'' be routing sequences of zeros and ones that are both representatives of some $u \in \mathcal{P}(d)$ with $d \in \mathbb{Q}$ and $\frac{\sigma}{\delta} \cdot d < 1$. Then $\overline{W}(u')$ and $\overline{W}(u'')$ exist and are finite. Moreover $\overline{W}(u') = \overline{W}(u'')$ and the long-run average waiting time does not depend on the initial workload w.*

Notation. Let δ , σ and d be as in Theorem 3.10. Then we say that $\rho := \frac{\sigma}{\delta} \cdot d$ is the traffic intensity for the server.

By Theorem 3.10 all the routing sequences which are representatives of some given $u \in \mathcal{P}(d)$ have the same long-run average waiting time if the stability condition $\rho < 1$ is fulfilled. Therefore we denote this long-run average waiting time simply by $\overline{W}(u)$ for $u \in \mathcal{P}(d)$.

Theorem 3.11 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j\}$ be as in Theorem 3.10. Further let $u, v \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$ such that the traffic intensity $\rho < 1$ and $u \trianglelefteq_{\overline{g}} v$. Then*

$$\overline{W}(v) - \overline{W}(u) \leq \frac{\delta}{d}(\overline{I}(v) - \overline{I}(u)).$$

Proof. Since $u, v \in \mathcal{P}(d)$ with $d \in \mathbb{Q}$ and $u \trianglelefteq_{\overline{g}} v$ there exist $T, k \in \mathbb{N}$, $u', v' \in \mathcal{P}(T,k)$ and routing sequences u'', v'' such that $d = \frac{k}{T}$, $\overline{\omega}(T,k) \preceq u' \preceq v'$, $u'' = (u')^\infty$ is representative of u and $v'' = (v')^\infty$ is representative of v. By Theorem 3.10 we have that both $\overline{W}(u) = \overline{W}(u'')$ and $\overline{W}(v) = \overline{W}(v'')$ exist. Let the sequence $\{\delta_i(u'')\}_{i=1,2,\ldots}$ be a realisation of the sequence of interarrival times when routing sequence u'' is used and let the sequence $\{\delta_i(v'')\}_{i=1,2,\ldots}$ be a realisation of the interarrival times when routing sequence v'' is used. Because they are sequences of i.i.d. random variables we make the coupling that $\delta_i(u'') = \delta_i(v'') = \delta_i$ for $i = 1, 2, \ldots$ where the sequence $\{\delta_i\}$ is a realisation of the sequence of interarrival times. Let the sequence $\{\sigma_j(u'')\}_{j=1,2,\ldots}$ be a realisation of the sequence of service times when routing sequence u'' is used and let the sequence $\{\sigma_j(v'')\}_{j=1,2,\ldots}$ be a realisation of the service times when routing sequence v'' is used. Because they are sequences of i.i.d random variables we make the coupling that $\sigma_j(u'') = \sigma_j(v'') = \sigma_j$ for $j = 1, 2, \ldots$ where the sequence $\{\sigma_j\}$ is a realistation of the sequence of service times. Let $W^c_{u''}(j), W^c_{v''}(j)$ be the workload for the server at the moment the j-th customer is routed to the server according to routing sequence u'' respectively v'' with this coupling of the interarrival times and service times. Then by Theorem 3.10 $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^m W^c_{u''}(j) = \overline{W}(u'') = \overline{W}(u)$ with probability one and $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^m W^c_{v''}(j) = \overline{W}(v'') = \overline{W}(v)$ with probability one. Further for every $T \in \mathbb{N}$ and $l \in \mathbb{N}$ we have that $\lim_{t\to\infty} \frac{1}{t} \cdot \sum_{i=0}^{t-1} \delta_{i\cdot T+l} = \delta$ with probability one and thus we have analogously to Theorem 3.8 that

$$\limsup_{m\to\infty} \frac{1}{m} \cdot \{\sum_{j=1}^m W^c_{v''}(j) - \sum_{j=1}^m W^c_{u''}(j)\} \leq \delta \cdot \frac{T}{k} \cdot (\overline{I}(v) - \overline{I}(u)) = \frac{\delta}{d}(\overline{I}(v) - \overline{I}(u))$$

with probability one. Hence $\overline{W}(v) - \overline{W}(u) \leq \frac{\delta}{d}(\overline{I}(v) - \overline{I}(u))$. □

By Theorem 3.11 and the results in the previous section we have the following corollary.

Corollary 3.12 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j\}$ be as in Theorem 3.10. Further let $u \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$ such that $\rho < 1$ and let $\omega = \omega(d)$ be the regular sequence of density d. Then*

$$\overline{W}(u) \leq \overline{W}(\omega) + \frac{\delta}{d} \cdot \overline{I}(u).$$

We also have the dual results of Theorem 3.11 and Corollary 3.12. In these dual results the dual unbalance $\underline{I}$ is used instead of the primal unbalance $\overline{I}$. These can be proved completely analogously using the dual order instead of the primal order. The dual results of Theorem 3.11 and Corollary 3.12 are the following.

Theorem 3.13 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j\}$ be as in Theorem 3.10. Further let $u, v \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$ such that $\rho < 1$ and $u \trianglelefteq_{\underline{g}} v$. Then*

$$\overline{W}(v) - \overline{W}(u) \geq \frac{\delta}{d}(\underline{I}(u) - \underline{I}(v)).$$

Corollary 3.14 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j\}$ be as in Theorem 3.10. Further let $u \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$ such that $\rho < 1$ and let $\omega = \omega(d)$ be the regular sequence of density d. Then*

$$\overline{W}(u) \geq \overline{W}(\omega) - \frac{\delta}{d} \cdot \underline{I}(u).$$

Remark. Contrary to Corollary 3.12 this dual result Corollary 3.14 is not useful for this type of routing system, because if u and ω are as in these corollaries then it is known that $\overline{W}(u) \geq \overline{W}(\omega)$ (See [3] and [5]).

In Theorem 3.10 we assume that the interarrival times $\{\delta_i\}$ are i.i.d. random variables with mean δ and the service times $\{\sigma_j\}$ are i.i.d. random variables with mean σ. So, the results hold in particular if the interarrival and service times are deterministic, i.e the interarrival times are all equal to δ and the service times are all equal to σ. Further in this case the assumption that the traffic intensity $\rho < 1$ can be relaxed to $\rho \leq 1$. Since the long-run average waiting time depends on the initial workload w if $\rho = 1$ we assume that the queue is empty at $T_1 = 0$, i.e the initial workload $w = 0$ in this case. In [28] it is shown that Theorem 3.10 (except the statement that the long-run average waiting does not depend on the initial workload) in this deterministic case also holds for $\rho = 1$. Since $\rho < 1$ was only needed to prove Theorem 3.10 it follows that results following after Theorem 3.10 also hold for $\rho = 1$ in this deterministic case. In fact for routing to a deterministic queue and deterministic interarrival times with traffic intensity $\rho = 1$ we have instead of Theorem 3.11 and Corollary 3.12 the following stronger result.

Theorem 3.15 *Suppose that the interarrival times and service times are deterministic and equal to δ respectively σ. Further suppose that the queue is initially empty. Let $u \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$ such that $\delta = d\sigma$, hence $\rho = 1$, and let $\omega = \omega(d)$ be the regular sequence of density d. Then*

$$\overline{W}(u) = \overline{W}(\omega) + \frac{\delta}{d} \cdot \overline{I}(u).$$

Proof. Let $d = \frac{k}{T}$ with $\gcd(k,T) = 1$. Let $\overline{\omega} = (\overline{\omega}(T,k))^\infty$, which is a representative of ω. Let $u' \in \mathcal{P}(T,k)$ such that $\overline{\omega}(T,k) \preceq u'$ and $u'' := (u')^\infty$ is a representative of u. From [28] we have that $\overline{W}(\omega) = \overline{W}(\overline{\omega}) = \lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^m W_{\overline{\omega}}(j)$ and $\overline{W}(u) = \overline{W}(u'') = \lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^m W_{u''}(j)$. For both routing sequence $\overline{\omega}$ and u'' the server is always busy serving customers from $T_1 = 0$. Namely, let $t_0 \in \mathbb{R}_{\geq 0}$ be arbitrarily. Since customers arrive at times $0, \delta, 2\delta, \ldots$ the number of customers routed to the server before time t_0 according to routing sequence $\overline{\omega}$ is $\kappa_{\overline{\omega}}(\lceil \frac{t_0}{\delta} \rceil) \geq \lceil \frac{t_0}{\delta} d \rceil = \lceil \frac{t_0}{\sigma} \rceil$. Since the server could not start earlier with the serving of these customers than at time $T_1 = 0$ and each of them has a serving time of σ, the server is not finished with serving these customers before time $\sigma \cdot \lceil \frac{t_0}{\sigma} \rceil \geq t_0$. So, for routing sequence $\overline{\omega}$ the server is always busy from $T_1 = 0$ and this follows analogously for routing sequence u''. From this and the fact that the server is assumed to be empty at $T_1 = 0$ it follows that $W_{u''}(j) + \lambda_{u''}(j) = W_{\overline{\omega}}(j) + \lambda_{\overline{\omega}}(j) = (j-1)\sigma$ for every $j \in \mathbb{N}$. Therefore we have for every $m \in \mathbb{N}$ that (3) holds with equality for u'' and $\overline{\omega}$ with $\delta_i = \delta$ for $i = 1, 2, \ldots$. Thus

$$\sum_{j=1}^{m} W_{u''}(j) - \sum_{j=1}^{m} W_{\overline{\omega}}(j) = \sum_{j=1}^{m} \sum_{i=\nu_{u''}(j)}^{\nu_{\overline{\omega}}(j)-1} \delta = \delta \cdot N_{\overline{\omega}u''}(m)$$

for every $m \in \mathbb{N}$. Hence by Corollary 3.7 we have that

$$\overline{W}(u) - \overline{W}(\omega) = \delta \cdot \lim_{m\to\infty} \frac{N_{\overline{\omega}u''}(m)}{m} = \delta \cdot \frac{T}{k} \cdot \overline{I}(u'') = \frac{\delta}{d} \cdot \overline{I}(u).$$

□

According to Theorem 3.15 the inequalities of Theorem 3.11 and Corollary 3.12 hold with equality for routing to a deterministic queue with traffic intensity $\rho = 1$ if the interarrival times are deterministic. For non-deterministic interarrival and service times the inequalities of Theorem 3.11 and Corollary 3.12 are rather tight if ρ is close to 1, or in other words if the traffic is heavy.

4 Routing to parallel queues

In this section we derive upper bounds on the average waiting time for routing customers to parallel queues according to some (periodic) routing policy. We assume that we have $N \geq 2$ parallel servers. Then a routing policy ψ corresponds to an $\mathbb{N}$ - word (integer sequence) $U = (U_1, U_2, \ldots)$ on the alphabet $\{1, 2, \ldots, N\}$, where U_i is the server to which the i-th arriving customer is routed according to policy ψ. If the applied policy and thus also the corresponding word are both periodic we have for some $T \in \mathbb{N}$ that $U_1, U_2, \ldots, U_T$ is a primitive period cycle of the word. Then we say that T is the period of the applied policy ψ and we denote ψ and the corresponding word U as $(U_1, U_2, \ldots, U_T)^\infty$.

Definition 4.1 *Let ψ and U be as above. Then for $t \in \mathbb{N}$ we define $W(t) = W_\psi(t)$ as the waiting time of the t -th arriving customer if policy ψ is applied, which is the remaining workload for server U_t at the moment that the t-th customer arrives. If the interarrival times and service times of all servers are fixed sequences then we say that $\overline{W}(\psi) = \overline{W}(U)$ is the long-run average waiting time of the arriving customers routed according to policy ψ if $\lim_{\tau\to\infty} \frac{1}{\tau} \cdot \sum_{t=1}^{\tau} W_\psi(t) = \overline{W}(\psi)$ and the system is empty at $T_1 = 0$. If the interarrival times and service times of the several servers are random variables then we say that $\overline{W}(\psi) = \overline{W}(U)$ is the almost sure long-run average waiting time of the arriving customers routed according to policy ψ if $\lim_{\tau\to\infty} \frac{1}{\tau} \cdot \sum_{t=1}^{\tau} W_\psi(t) = \overline{W}(\psi)$ with probability one and the system is empty at $T_1 = 0$.*

Notation. If U is an $\mathbb{N}$ -word (not necessarily periodic) on the alphabet $\{1, 2, \ldots, N\}$ (corresponding to some routing policy) then for $i = 1, 2, \ldots, N$ we denote by $u^i = (u^i_1, u^i_2, \ldots)$ the corresponding routing sequence of zeros and ones for server i, i.e $u_t^i = \begin{cases} 1 & \text{if } U_t = i \\ 0 & \text{if } U_t \neq i \end{cases}$. For $i = 1, 2, \ldots, N$ and $t \in \mathbb{N}$ we put $A^t(i) := \{t' \in \mathbb{N} : u_{t'}^i = 1\} \cap \{1, 2, \ldots, t\}$. Let $d_1, d_2, \ldots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$. We denote by $\mathcal{S}(d_1, d_2, \ldots, d_N)$ all the infinite $\mathbb{N}$ -words $U = (U_1, U_2, \ldots)$ on the alphabet $\{1, 2, \ldots, N\}$ such that u^i has density d_i for $i = 1, 2, \ldots, N$. If U is periodic then we have for every $i \in \{1, 2, \ldots, N\}$ that $u^i \in \mathcal{P}(d)$ for some $d \in \mathbb{Q}$. Let $d_1, d_2, \ldots, d_N \in \mathbb{Q}_{>0}$ such that $\sum_{i=1}^N d_i = 1$. Then we denote by $\mathcal{Q}(d_1, d_2, \ldots, d_N)$ all the infinite periodic $\mathbb{N}$ - words U on the alphabet $\{1, 2, \ldots, N\}$ for which $u^i \in \mathcal{P}(d_i)$ for $i = 1, 2, \ldots, N$. Further for $T, k_1, k_2, \ldots, k_N \in \mathbb{N}$ with $\sum_{i=1}^N k_i = T$ we denote by $\mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N)$ all the infinite $\mathbb{N}$-words on the alphabet $\{1, 2, \ldots, N\}$ for which every subword of length T contains exactly k_i letters i for $i = 1, 2, \ldots, N$.

Note that $\mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N) \subseteq \mathcal{Q}(\frac{k_1}{T}, \frac{k_2}{T}, \ldots, \frac{k_N}{T})$ and $\mathcal{Q}(d_1, d_2, \ldots, d_N) \subseteq \mathcal{S}(d_1, d_2, \ldots, d_N)$. Let ψ be a periodic routing policy with period T and U the corresponding word on the alphabet $\{1, 2, \ldots, N\}$. Then there exist non-negative integers $k_1, k_2, \ldots, k_N$ with $\sum_{i=1}^N k_i = T$ such that from T consecutively arriving customers policy ψ routes exactly k_i of them to server i for $i = 1, 2, \ldots, N$. Hence $U \in \mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N) \subseteq \mathcal{Q}(\frac{k_1}{T}, \frac{k_2}{T}, \ldots, \frac{k_N}{T}) \subseteq \mathcal{S}(\frac{k_1}{T}, \frac{k_2}{T}, \ldots, \frac{k_N}{T})$. Vice versa if $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ with $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \ldots, N$ and $\sum_{i=1}^N d_i = 1$ then there exist $T, k_1, k_2, \ldots, k_N \in \mathbb{N}$ with $\sum_{i=1}^N k_i = T$ such that $U \in \mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N)$. Further for every server $i \in \{1, 2, \ldots, N\}$ we have a corresponding routing sequence $u^i = (u_1^i, u_2^i, \ldots, u_T^i)^\infty \in \mathcal{P}(d_i)$ where $d_i = \frac{k_i}{T}$. For example if $U = (1,2,1,2,1,3,1,2,1,1,2,3)^\infty$ then $U \in \mathcal{Q}(\{12\}, 6, 4, 2) \subseteq \mathcal{Q}(\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$. Further $u^1 = (1,0,1,0,1,0,1,0,1,1,0,0)^\infty \in \mathcal{P}(\frac{1}{2})$, $u^2 = (0,1,0,1,0,0,0,1,0,0,1,0)^\infty \in \mathcal{P}(\frac{1}{3})$ and $u^3 = (0,0,0,0,0,1)^\infty \in \mathcal{P}(\frac{1}{6})$.

Let $\{\delta_i\}$ be the sequence of interarrival times and let $\{\sigma_j^i\}$ be the sequence of service times of server i for $i = 1, 2, \ldots, N$, i.e σ_j^i is the service time of the j - th customer that is routed to server i. We define $\overline{W}^i(\psi) = \overline{W}^i(u^i)$ as the long-run average waiting time of customers routed to server i if policy ψ is applied in the same way as in the previous section. The only differences are that routing sequence u is replaced by routing sequence u^i and the sequence of service times $\{\sigma_j\}$ is replaced with the sequence of service times $\{\sigma_j^i\}$. We have the following theorem.

Theorem 4.2 *Suppose that the interarrival times $\{\delta_i\}$ of customers arriving at the system are i.i.d. random variables with mean δ and for every $i \in \{1, 2, \ldots, N\}$ the service times $\{\sigma_j^i\}$ are i.i.d. random variables with mean σ_i which are independent of the interarrival times. Let ψ be a routing policy that corresponds to some word $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ such that $\frac{\sigma_i}{\delta} \cdot d_i < 1$ for $i = 1, 2, \ldots, N$. Then $\overline{W}(\psi)$ exists and is finite. Moreover $\overline{W}(\psi) = \sum_{i=1}^n d_i \cdot \overline{W}^i(\psi)$.*

Proof. We have that $u^i \in \mathcal{P}(d_i)$ and $\frac{\sigma_i}{\delta} \cdot d_i < 1$ for $i = 1, 2, \ldots, N$. Hence by Theorem 3.10 we have for every $i \in \{1, 2, \ldots, N\}$ that $\overline{W}^i(\psi) = \overline{W}^i(u^i)$ exists and is finite. Hence

$$\lim_{\tau \to \infty} \frac{1}{\tau} \cdot \sum_{t=1}^{\tau} W_\psi(t) = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{i=1}^{N} \sum_{t \in A^\tau(i)} W_\psi(t) =$$

$$\lim_{\tau \to \infty} \sum_{i=1}^{N} d_i \sum_{t \in A^\tau(i)} \frac{W_\psi(t)}{|A^\tau(i)|} = \sum_{i=1}^{N} d_i \cdot \overline{W}^i(\psi).$$

□

Notation. Let $\psi, \delta, \sigma_i, d_i$ be as in Theorem 4.2 for $i = 1, 2, \ldots, N$. Then we say that $\rho_i := \frac{\sigma_i}{\delta} \cdot d_i$ is the traffic intensity for server i if policy ψ is applied.

Remark. If the interarrival times and the service times are deterministic then for Theorem 4.2 it suffices that $\rho_i \leq 1$ for $i = 1, 2, \dots, N$, since in that case $\overline{W}^i(\psi)$ exists and is finite.

Definition 4.3 *Let $U \in \mathcal{Q}(d_1, d_2, \dots, d_N)$ for some $d_i \in \mathbb{Q}_{>0}$ with $\sum_{i=1}^N d_i = 1$. Then we define the total primal unbalance $\overline{O}$ of U as*

$$\overline{O}(U) := \sum_{i=1}^{N} \overline{I}(u^i). \tag{4}$$

Further we define the total dual unbalance $\underline{O}$ of U as

$$\underline{O}(U) := \sum_{i=1}^{N} \underline{I}(u^i). \tag{5}$$

Further we define the partial orders $\trianglelefteq_{\overline{g}}, \trianglelefteq_{\underline{g}}$ and $\trianglelefteq_g$ on $\mathcal{Q}(d_1, d_2, \dots, d_N)$ in the following way. If $U, V \in \mathcal{Q}(d_1, d_2, \dots, d_N)$ then $U \trianglelefteq_{\overline{g}} V$ if $u^i \trianglelefteq_{\overline{g}} v^i$ for every $i \in \{1, 2, \dots, N\}$, $U \trianglelefteq_{\underline{g}} V$ if $u^i \trianglelefteq_{\underline{g}} v^i$ for every $i \in \{1, 2, \dots, N\}$ and $U \trianglelefteq_g V$ if $u^i \trianglelefteq_g v^i$ for every $i \in \{1, 2, \dots, N\}$.

If it is clear what is meant we just say unbalance and not total unbalance. Suppose that $U, V \in \mathcal{Q}(d_1, d_2, \dots, d_N)$ for $d_1, d_2, \dots, d_N \in \mathbb{Q}_{>0}$ with $\sum_{i=1}^N d_i = 1$. Then we say that U, V are mirrors of each other if there exist U', V' such that U' and V' are period cycles of U respectively V and $V' = \overleftarrow{U'}$.

Theorem 4.4 *Suppose that $U, V \in \mathcal{Q}(d_1, d_2, \dots, d_N)$ are mirrors of each other. Then $\overline{O}(U) = \underline{O}(V)$ and $\underline{O}(U) = \overline{O}(V)$.*

Proof. Since U and V are mirrors of each other it follows for every $i \in \{1, 2, \dots, N\}$ that $u^i, v^i \in \mathcal{P}(d_i)$ are mirrors of each other. Thus by Corollary 2.32 we have for $i = 1, 2, \dots, N$ that $\overline{I}(u^i) = \underline{I}(v^i)$ and thus

$$\overline{O}(U) = \sum_{i=1}^{N} \overline{I}(u^i) = \sum_{i=1}^{N} \underline{I}(v^i) = \underline{O}(V).$$

Analogously it follows that $\underline{O}(U) = \overline{O}(V)$. □

Example. Let $U = (1,2,3,1,4,2,1,3,1,2,4,3)^\infty \in \mathcal{Q}(\{12\}, 4, 3, 3, 2) \subseteq \mathcal{Q}(\frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \frac{1}{6})$. Then we have that $\overline{I}(u^1) = \frac{1}{12}$, $\overline{I}(u^2) = 0$, $\overline{I}(u^3) = \frac{1}{12}$, $\overline{I}(u^4) = 0$ and thus $\overline{O}(U) = \frac{1}{6}$. Further $\underline{I}(u^1) = \frac{1}{4}$, $\underline{I}(u^2) = 0$, $\underline{I}(u^3) = \frac{1}{6}$, $\underline{I}(u^4) = 0$ and thus $\underline{O}(U) = \frac{5}{12}$. Further for $V = (1,2,4,1,3,2,1,3,4,2,1,3)^\infty \in \mathcal{Q}(\{12\}, 4, 3, 3, 2) \subseteq \mathcal{Q}(\frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \frac{1}{6})$ which is a mirror of U the reader can verify that $\overline{O}(V) = \frac{5}{12}$ and $\underline{O}(V) = \frac{1}{6}$.

By Theorem 3.11 and Theorem 4.2 we have the following theorem.

Theorem 4.5 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j^i\}$ for $i = 1, 2, \dots, N$ be as in Theorem 4.2. Let $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \dots, N$ with $\sum_{i=1}^N d_i = 1$. Suppose that ψ, ψ' are routing policies with corresponding words $U, V \in \mathcal{Q}(d_1, d_2, \dots, d_N)$ respectively. Suppose that $\rho_i < 1$ for $i = 1, 2, \dots, N$ and $U \trianglelefteq_{\overline{g}} V$. Then*

$$\overline{W}(\psi') - \overline{W}(\psi) \leq \delta \cdot (\overline{O}(V) - \overline{O}(U)).$$

Proof. We have that

$$\overline{W}(\psi') - \overline{W}(\psi) = \sum_{i=1}^{N} d_i \cdot \overline{W}^i(v_i) - \sum_{i=1}^{N} d_i \cdot \overline{W}^i(u_i) = \sum_{i=1}^{N} d_i \cdot (\overline{W}^i(v_i) - \overline{W}^i(v_i)) \leq$$

$$\sum_{i=1}^{N} d_i \cdot (\frac{\delta}{d_i}) \cdot (\overline{I}(v^i) - \overline{I}(u^i)) = \delta \cdot (\overline{O}(V) - \overline{O}(U)).$$

□

Analogously we have by Theorem 3.13 and Theorem 4.2 the following theorem.

Theorem 4.6 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j^i\}$ for $i = 1, 2, \ldots, N$ be as in Theorem 4.2. Let $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \ldots, N$ with $\sum_{i=1}^{N} d_i = 1$. Suppose that ψ, ψ' are routing policies with corresponding words $U, V \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ respectively. Suppose that $\rho_i < 1$ for $i = 1, 2, \ldots, N$ and $U \trianglelefteq_g V$. Then*

$$\overline{W}(\psi') - \overline{W}(\psi) \geq \delta \cdot (\underline{O}(U) - \underline{O}(V)).$$

Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j^i\}$ for $i = 1, 2, \ldots, N$ be as in Theorem 4.2. Further let $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \ldots, N$ with $\sum_{i=1}^{N} d_i = 1$. Then, if $\overline{W}^i(\omega(d_i))$ exists and is finite for $i = 1, 2, \ldots, N$ then we put

$$\widetilde{R} = \widetilde{R}(d_1, d_2, \ldots, d_N) := \sum_{i=1}^{N} d_i \cdot \overline{W}^i(\omega(d_i)). \tag{6}$$

Remark. It is easily seen that if the interarrival times and service times are random variables then for the existence of $\widetilde{R}$ it suffices that $\rho_i < 1$ for $i = 1, 2, \ldots, N$, while if the interarrival times and service times are deterministic then it suffices that $\rho_i \leq 1$ for $i = 1, 2, \ldots, N$. Further the definition of $\widetilde{R}$ and these results for the existence of $\widetilde{R}$ can be extended to the case of irrational d_i. Namely, let $d_i \in \mathbb{R}_{>0}$ such that $\rho_i < 1$. Then in [3] and [5] it is shown that $W^i(\omega)$ exists and is the same for every regular sequence ω of zeros and ones of density d_i. In [28] this is also shown for the case of $\rho_i = 1$ and deterministic interarrival and service times. Moreover in Lemma 4.8 $W^i(\omega)$ is given explicitly for that case. So, if $d_i \in \mathbb{R}_{>0}$ for $i = 1, 2, \ldots, N$ with $\sum_{i=1}^{N} d_i = 1$ and $\rho_i < 1$, or $\rho_i = 1$ in case of deterministic interarrival and service times, then $\widetilde{R}$ is well defined as follows:

$$\widetilde{R} = \widetilde{R}(d_1, d_2, \ldots, d_N) := \sum_{i=1}^{N} d_i \cdot \overline{W}^i(\omega^i), \tag{7}$$

where ω^i is any regular sequence of zeros and ones of density d_i for $i = 1, 2, \ldots, N$.

In general $\widetilde{R}$ depends on the distribution of the interarrival times and service times and in some cases it is possible to compute $\widetilde{R}$ explicitly. We will do this in some examples.

By Corollary 3.12 and Theorem 4.2 we have the following theorem.

Theorem 4.7 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j^i\}$ for $i = 1, 2, \ldots, N$ be as in Theorem 4.2. Let $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \ldots, N$ with $\sum_{i=1}^{N} d_i = 1$. Suppose that a routing policy ψ is applied with corresponding word $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ and $\rho_i < 1$ for $i = 1, 2, \ldots, N$. Then*

$$\overline{W}(\psi) - \widetilde{R} \leq \delta \cdot \overline{O}(U).$$

From Theorem 7.12 in [28] we have an explicit formula for $\overline{W}^i(\omega^i)$ in case of $\rho_i = 1$, deterministic interarrival times and the server i has deterministic service times. In [28] the formula is given for interarrival times δ equal to 1 , but that is no restriction since the time-unit can always be chosen such that $\delta = 1$. So, for general δ we can just multiply by δ and we get the following lemma.

Lemma 4.8 *Suppose that the service times $\{\sigma_j^i\}$ of some server i are deterministic equal to σ_i and the the interarrival times $\{\delta_i\}$ are deterministic and equal to δ. Let ω^i be a regular sequence of zeros and ones with density $d_i \in \mathbb{R}_{>0}$ and suppose that the traffic intensity $\rho_i = d_i \cdot \frac{\sigma_i}{\delta}$ for server i is equal to 1. If d_i is rational, $d = \frac{p_i}{q_i}$ with $p_i, q_i \in \mathbb{N}$ and $\gcd(p_i, q_i) = 1$ then*

$$\overline{W}(\omega^i) = \delta \cdot (\frac{1}{2} - \frac{1}{2p_i}). \tag{8}$$

If d_i is irrational then $\overline{W}(\omega^i) = \frac{\delta}{2}$.

For the same case of deterministic interarrival and service times but with $\rho_i < 1$ an algorithm to calculate $\overline{W}(\omega^i)$ for a regular sequence ω^i is given in [18]. Thus in case of deterministic interarrival and service times $\widetilde{R}(d_1, d_2, \dots, d_N)$ can be determined exactly for any $d_1, d_2, \dots, d_N \in R_{>0}$ with $\sum_{i=1}^N d_i = 1$ if $\rho_i \leq 1$ for $i = 1, 2, \dots, N$. In the example at the end of this section $\widetilde{R}$ is determined in a special case with Poisson arrivals and exponentially distributed service times. For general bounds on $\widetilde{R}$ see [34].

We have the following theorem which emphasizes the sharpness of the bound of Theorem 4.7.

Theorem 4.9 *Let the interarrival times $\{\delta_i\}$ be deterministic equal to δ and the service times $\{\sigma_j^i\}$ be deterministic equal to σ_i for $i = 1, 2, \dots, N$. Let $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \dots, N$ with $\sum_{i=1}^N d_i = 1$. Let $p_i, q_i \in \mathbb{N}$ be such that $d_i = \frac{p_i}{q_i}$ with $\gcd(p_i, q_i) = 1$ for $i = 1, 2, \dots, N$. Suppose that a routing policy ψ is applied with corresponding word $U \in \mathcal{Q}(d_1, d_2, \dots, d_N)$ and $\rho_i = 1$ for $i = 1, 2, \dots, N$. Then*

$$\overline{W}(\psi) = \widetilde{R} + \delta \cdot \overline{O}(U) = \delta \cdot \left(\frac{1}{2} - \sum_{i=1}^{N} \frac{1}{2q_i} + \overline{O}(U) \right).$$

Proof. According to the remark following Theorem 4.2 we can apply Theorem 4.2 and by Theorem 3.15 it follows that

$$\overline{W}(\psi) = \widetilde{R} + \delta \cdot \overline{O}(U).$$

By (8) we have that

$$\widetilde{R} = \sum_{i=1}^{N} d_i \cdot \overline{W}(\omega(d_i)) = \sum_{i=1}^{N} \frac{p_i}{q_i} \cdot \delta \cdot (\frac{1}{2} - \frac{1}{2p_i}) = \delta \cdot (\frac{1}{2} - \sum_{i=1}^{N} \frac{1}{2q_i}). \tag{9}$$

Example. We consider a queueing system with 3 parallel servers where the interarrival times are deterministic and equal to $\delta = 3$. The arriving jobs are routed to the servers according to the policy ψ that corresponds to the word

$$U = (1, 2, 1, 2, 1, 3, 1, 2, 1, 3)^\infty \in \mathcal{Q}(\{10\}, 5, 3, 2) \subseteq \mathcal{Q}(\frac{1}{2}, \frac{3}{10}, \frac{1}{5}).$$

All the service times are deterministic and for server 1 they are equal to $\sigma_1 = 6$, for server 2 equal to $\sigma_2 = 10$ and for server 3 equal to $\sigma_3 = 15$. Hence $\rho_i = 1$ for $i = 1, 2, 3$. Further $\overline{I}(u^1) = 0$, $\overline{I}(u^2) = \frac{1}{10}$, $\overline{I}(u^3) = \frac{1}{10}$ and thus $\overline{O}(U) = \frac{1}{5}$. So, according to Theorem 4.9 we have that

$$\overline{W}(\psi) = 3 \cdot (\frac{1}{2} - (\frac{1}{4} + \frac{1}{20} + \frac{1}{10}) + \frac{1}{5}) = \frac{9}{10}$$

which can be checked by direct calculation.

Theorem 4.10 *Let the interarrival times $\{\delta_i\}$ and the service times $\{\sigma_j^i\}$ for $i = 1, 2, \ldots, N$ be as in Theorem 4.2. Let $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \ldots, N$ with $\sum_{i=1}^N d_i = 1$. Suppose that a routing policy ψ is applied with corresponding word $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ and $\rho_i < 1$ for $i = 1, 2, \ldots, N$. Then*

$$\widetilde{R} \leq \overline{W}(\psi) \leq \widetilde{R} + \delta \cdot \overline{O}(U).$$

Proof. From [3] and [5] we have that $\widetilde{R} \leq \overline{W}(\psi)$ and according to Theorem 4.7 we have that $\overline{W}(\psi) \leq \widetilde{R} + \delta \cdot \overline{O}(U)$. □

Note. Suppose that we have a queueing system where the arrivals are according to a Poisson process with parameter λ. Suppose that the service times of server i are exponentially distributed with parameter μ_i. If d_i, the fraction of jobs that is routed to server i, equals $\frac{1}{q_i}$ for some $q_i \in \mathbb{N}$, then $\overline{W}^i(\omega(d_i))$ can be calculated in the following way. For the routing sequence $\omega(d_i)$ we have that among every q_i arriving jobs exactly one job is routed to server i. So, the interarrival times at server i consist of q_i Poisson arrivals with parameter λ. Hence the interarrival times for the queue of server i are Erlang distributed, namely according to an $E_\lambda^{q_i}$ distribution. Thus $\overline{W}^i(\omega(d_i))$ is the same as the average waiting time for a $E_\lambda^{q_i}/M/1$ queue, where the parameter of the service times is μ_i. So (see [19]) if $x_i \in (0, 1)$ is a solution of the equation

$$x = (\frac{\lambda}{\lambda + \mu_i - \mu_i \cdot x})^{q_i}, \tag{10}$$

then $\overline{W}^i(\omega(d_i)) = \frac{x_i}{\mu_i \cdot (1 - x_i)}$.

In the following example we have calculated $\overline{W}^i(\omega(d_i))$ for $i = 1, 2, 3, 4$ in this way. Further we have explicitly calculated the lower bound $\widetilde{R}$ and upper bound $\widetilde{R} + \delta \cdot \overline{O}(U)$ of Theorem 4.10 for $\overline{W}(\psi)$, where U is the word corresponding to the applied routing policy ψ.

Example. We consider a queueing system with 4 parallel servers where the arrivals are according to a Poisson process with parameter $\lambda = 11$. Hence for the mean interarrival time δ we have that $\delta = \frac{1}{11}$. The arriving jobs are routed to the servers according to the policy ψ that corresponds to the word

$$U = (1, 2, 3, 1, 4, 2, 1, 3, 1, 2, 4, 3)^\infty \in \mathcal{Q}(\{12\}, 4, 3, 3, 2) \subseteq \mathcal{Q}(\frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \frac{1}{6})$$

which we considered in a previous example. For every $i \in \{1, 2, 3, 4\}$ the service times are exponentially distributed with parameter μ_i and $\mu_1 = 4$, $\mu_2 = \mu_3 = 3$ and $\mu_4 = 2$. Then we find that $\overline{W}^1(\omega(\frac{1}{3})) = 1.7792$ (rounded to 4 decimals) and thus by Corollary 3.12 we have that $\overline{W}^1(u^1) \leq 1.7792 + 3 \cdot \frac{1}{11} \cdot \frac{1}{12} = 1.8019$. Similarly we have that $\overline{W}^2(\omega(\frac{1}{4})) = \overline{W}^2(u^2) = 2.2105$, $\overline{W}^3(\omega(\frac{1}{4})) = 2.2105$, $\overline{W}^3(u^3) \leq 2.2408$ and $\overline{W}^4(\omega(\frac{1}{6})) = \overline{W}^4(u^4) = 3.0732$. Hence

$$\widetilde{R} = \frac{1}{3} \cdot \overline{W}^1(\omega(\frac{1}{3})) + \frac{1}{4} \cdot \overline{W}^2(\omega(\frac{1}{4})) + \frac{1}{4} \cdot \overline{W}^3(\omega(\frac{1}{4})) + \frac{1}{6} \cdot \overline{W}^4(\omega(\frac{1}{6})) = 2.2105.$$

and $\widetilde{R} + \delta \cdot \overline{O}(U) = 2.2105 + \frac{1}{11} \cdot \frac{1}{6} = 2.2257$. So, by Theorem 4.10 we have that $2.2105 \leq \overline{W}(\psi) \leq 2.2257$.

5 Billiard sequences and routing sequences

In this section we give some properties of sequences (words) and their implications for the corresponding routing policies. In particular we have some properties for so-called billiard sequences (see [8]and [10]). For every set of rational densities we obtain a billiard sequence with unbalance smaller or equal than $\frac{N}{2}-1$, where N is the number of parallel queues. We derive a relation between the discrepancy function and the unbalance and several properties can be obtained from that. We show that for given rational densities there exists a periodic billiard sequence which has minimal unbalance among all sequences with those densities.

Theorem 5.1 *Let $d_1, d_2, \ldots, d_N \in \mathbb{Q}_{>0}$ with $\sum_{i=1}^{N} d_i = 1$ for some $N \in \mathbb{Z}_{\geq 2}$. Then there exists a word $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ such that $\overline{O}(U) \leq \frac{N}{2} - 1$.*

Theorem 5.1 follows from Theorem 5.5 which we formulate and prove later. The dual of Theorem 5.1, which is a consequence of Theorem 5.1 and Theorem 4.4, is the following theorem.

Theorem 5.2 *Let $d_1, d_2, \ldots, d_N \in \mathbb{Q}_{>0}$ with $\sum_{i=1}^{N} d_i = 1$ for some $N \in \mathbb{Z}_{\geq 2}$. Then there exists a word $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ such that $\underline{O}(U) \leq \frac{N}{2} - 1$.*

Theorem 5.1 and Theorem 4.7 have the following corollary.

Corollary 5.3 *Let the interarrival times $\{\delta_i\}$, the service times $\{\sigma_j^i\}$ and d_i for $i = 1, 2, \ldots, N$ be as in Theorem 4.7. Suppose that $\rho_i < 1$ for $i = 1, 2, \ldots, N$. Then there exists a routing policy ψ corresponding to a word $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ such that*

$$\overline{W}(\psi) \leq \widetilde{R}(d_1, d_2, \ldots, d_N) + \delta \cdot (\frac{N}{2} - 1).$$

Theorem 5.5 shows how to construct a word $U \in \mathcal{Q}(d_1, d_2, \ldots, d_N)$ such that $\overline{O}(U) \leq \frac{N}{2} - 1$. For the construction we use the following algorithm which we call the SG (Special Greedy) algorithm.

SG algorithm. Let $d_1, d_2, \ldots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^{N} d_i = 1$ and $x_1, x_2, \ldots, x_N \in \mathbb{R}_{\geq 0}$ be given. Then $U = (U_1, U_2, \ldots)$ is determined inductively in the following way. Suppose that $U_1, U_2, \ldots, U_{n-1}$ have been determined and thus $\kappa_{u^i}(n-1)$ is known for $i = 1, 2, \ldots, N$. Choose U_n such that $U_n = i \in \{1, 2, \ldots, N\}$ for which $\frac{x_i + \kappa_{u^i}(n-1)}{d_i}$ is minimal.

If $0 \leq x_i \leq 1$ for $i = 1, 2, \ldots, N$ then a word U constructed by the SG algorithm is known in the literature as a billiard word or sequence (see for example [8]). Namely, if a billiard ball travels in an N -dimensional cube with elastic reflecton on the sides the cube and a sequence of integers from $\{1, 2, \ldots, N\}$ (the integers correspond to the sides of the cube and opposite sides correspond to the same integer) is constructed by writing down an integer if the ball reflects against a side of the cube corresponding to that integer then such a sequence is called a billiard sequence. It is easily seen that the word constructed by the SG algorithm is a billiard sequence obtained by starting in position $(x_1, x_2, \ldots, x_N)$ and having initial velocity vector $(-d_1, -d_2, \ldots, -d_N)$ in the unit cube $[0, 1]^N$. In the SG algorithm the choice of U_n is not totally determined if $\frac{x_i + \kappa_{u^i}(n-1)}{d_i}$ is minimal for several $i \in \{1, 2, \ldots, N\}$. We assume that in such cases a consistent choice is made. For example always the smallest i is chosen for which $\frac{x_i + \kappa_{u^i}(n-1)}{d_i}$ is minimal. If this modified SG algorithm is used then the word obtained by the algorithm is determined by $x_1, x_2, \ldots, x_N \in \mathbb{R}_{\geq 0}$. For billiard sequences this consistent choice means that if the billiard ball hits multiple sides at the same time then the order in which the letters corresponding to those sides appear in the billiard sequence is also prescribed. We will call such a billiard sequence a consistent billiard sequence. More precisely, we define a consistent billiard sequence on the alphabet $\{0, 1, \ldots, N\}$ to be a sequence ($\mathbb{N}$ - word) that can be otained by applying the SG algorithm in a consistent way to some $0 \leq x_i < 1$ for $i = 1, 2, \ldots, N$. Note that $x_i < 1$ instead of $x_i \leq 1$ for $i = 1, 2, \ldots, N$. The reason for this becomes clear in the proof of Lemma 5.4. A billiard sequence that is not consistent can be aperiodic even if the densities $d_1, d_2, \ldots, d_N \in \mathbb{Q}_{>0}$. However, for consistent billiard sequences we have the following lemma.

Lemma 5.4 *Let $d_1, d_2, \ldots, d_N \in \mathbb{Q}_{>0}$ with $\sum_{i=1}^{N} d_i = 1$. Let $p_i, q_i, k_i \in \mathbb{N}$ for $i = 1, 2, \ldots, N$ and $T \in \mathbb{N}$ be such that $d_i = \frac{p_i}{q_i} = \frac{k_i}{T}$ with $\gcd(p_i, q_i) = 1$ for $i = 1, 2, \ldots, N$ and $\gcd(k_1, k_2, \ldots, k_N) = 1$. Let U be a consistent billiard sequence with these densities. Then $U \in \mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N) \subseteq \mathcal{Q}(d_1, d_2, \ldots, d_N)$.*

Proof. The consistent billiard sequence U can be constructed by starting in $(x_1, x_2, \ldots, x_N)$ with $0 \leq x_i < 1$ for $i = 1, 2, \ldots, N$ and initial velocity vector $(-d_1, -d_2, \ldots, -d_N)$. We assume that the velocity of the billard ball is such that for every $i \in \{1, 2, \ldots, N\}$ it takes $1/d_i$ time units to travel between the opposite sides of the cube corresponding to i. Then for every $t \geq 0$ the billiard ball has reflected $\lceil t \cdot d_i - x_i \rceil$ times on sides corresponding to letter i in the time interval $[0, t]$. Hence in the time interval $[0, T]$ the billiard ball has reflected $\lceil T \cdot d_i - x_i \rceil = k_i$ times on sides corresponding to i for $i = 1, 2, \ldots, N$. Since $\sum_{i=1}^{N} k_i = T$ it follows that $\kappa_{u^i}(T) = k_i$ for every $i \in \{1, 2, \ldots, N\}$ and $\frac{x_i + \kappa_{u^i}(T)}{d_i} = \frac{x_i}{d_i} + T$ for every $i \in \{1, 2, \ldots, N\}$. Since U is obtained by applying the SG algorithm in a consistent way on $x_1, x_2, \ldots, x_N$ it follows that $U_{T+t} = U_t$ for every $t \in \{1, 2, \ldots, T\}$ and by induction we have that U is periodic with period T. Hence $U \in \mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N) \subseteq \mathcal{Q}(d_1, d_2, \ldots, d_N)$. $\square$

Theorem 5.5 *Let d_i, p_i, q_i, k_i and T be as in Lemma 5.4, $N \geq 2$ and let $k \in \{1, 2, \ldots, N\}$ be such that $q_k = \max_{i \in \{1,2,\ldots,N\}} q_i$. Put $x_k^* = 0$ and $x_j^* = 1 - \frac{1}{q_j}$ for every $j \in \{1, 2, \ldots, N\}$ for which $j \neq k$. Construct a consistent billiard sequence U starting in $(x_1^*, x_2^*, \ldots, x_N^*)$ with initial velocity vector $(-d_1, -d_2, \ldots, -d_N)$ in the unit cube $[0, 1]^N$. Then $U \in \mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N) \subseteq \mathcal{Q}(d_1, d_2, \ldots, d_N)$. Moreover $\overline{O}(U) = 0$ if $N = 2$ and $\overline{O}(U) < \frac{N}{2} - 1$ if $N \geq 3$.*

We prove this theorem later. In the theorem we have chosen a particular starting point for the billiard sequence, but we think that this is not necessary. We have the following conjecture.

Conjecture 5.6 *For every consistent periodic billiard sequence U on N letters it holds that $\overline{O}(U) \leq \frac{N}{2} - 1$ and $\underline{O}(U) \leq \frac{N}{2} - 1$ if $N \geq 2$.*

Billiard sequences have more nice properties. For example, in [33] there is a notion for (routing) sequences to be called m - balanced. Regular sequences are 1 - balanced for that notion. From the results in that paper it follows that billiard sequences on an alphabet of N letters are $N - 1$ - balanced.

We need a number of lemmas and theorems in order to prove Theorem 5.5. These results are also of independent interest. A relation between the discrepancy function and the unbalance is derived.

Lemma 5.7 *Let $p, q \in \mathbb{N}$. Then*

$$\sum_{n=1}^{q} \lceil n \cdot \frac{p}{q} \rceil = \frac{1}{2}(pq + p + q - \gcd(p,q)).$$

Proof. Put $p' = \frac{p}{\gcd(p,q)}$ and $q' = \frac{q}{\gcd(p,q)}$. Recall that for $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ we denote by $a \pmod b$ the integer $a' \in \{0, 1, \ldots, b-1\}$ such that $a = a' + kb$ for some $k \in \mathbb{Z}$. Then we have that

$$\sum_{n=1}^{q} (\lceil n \cdot \frac{p}{q} \rceil - n \cdot \frac{p}{q}) = \frac{1}{q} \cdot \sum_{n=1}^{q} (-np) \pmod q =$$

$$\frac{\gcd(p,q)}{q} \cdot \sum_{n=1}^{q'} \gcd(p,q) \cdot (-np') \pmod{q'} = \frac{(\gcd(p,q))^2}{q} \sum_{i=0}^{q'-1} i =$$

$$\frac{(\gcd(p,q))^2}{q} \cdot \frac{q'}{2} \cdot (q' - 1).$$

Hence

$$\sum_{n=1}^{q} \lceil n \cdot \frac{p}{q} \rceil = \sum_{n=1}^{q} n \cdot \frac{p}{q} + \sum_{n=1}^{q} (\lceil n \cdot \frac{p}{q} \rceil - n \cdot \frac{p}{q}) =$$

$$\frac{p}{2} \cdot (q+1) + \frac{1}{2} \gcd(p,q) \cdot (\frac{q}{\gcd(p,q)} - 1) = \frac{1}{2}(pq + p + q - \gcd(p,q)).$$

□

Recall the notation introduced in the previous section and the definition of the discrepancy function χ from section 2.

Definition 5.8 *Let $U \in \mathcal{S}(d_1, d_2, \ldots, d_N)$ for some $d_1, d_2, \ldots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^{N} d_i = 1$. Then we define*

$$c_i = c_i(U) = sup_{n \in \mathbb{Z}_{\geq 0}} \chi_{u^i}(n)$$

for $i = 1, 2, \ldots, N$.

Lemma 5.9 *If $U \in \mathcal{Q}(\{T\}, k_1, k_2, \ldots, k_N)$ for some $T, k_1, k_2, \ldots, k_N \in \mathbb{N}$ then*

$$c_i = \max_{n \in \{0,1,\ldots,T-1\}} \chi_{u^i}(n) \text{ for } i = 1, 2, \ldots, N.$$

Proof. For $i \in \{1, 2, \ldots, N\}$ we have that $d_i = \frac{k_i}{T}$ is the density of sequence u^i. So, for every $n \in \mathbb{Z}_{\geq 0}$ we have that

$$\chi_{u^i}(n + T) = \chi_{u^i}(n) + d_i \cdot T - (\kappa_{u^i}(n+T) - \kappa_{u^i}(n)) = \chi_{u^i}(n).$$

□

Theorem 5.10 *Let $d_i \in \mathbb{Q}_{>0}$ for $i = 1, 2, \dots, N$ such that $\sum_{i=1}^N d_i = 1$. Let $p_i, q_i \in \mathbb{N}$ such that $d_i = \frac{p_i}{q_i}$ with $\gcd(p_i, q_i) = 1$ for $i = 1, 2, \dots, N$ and let $U \in \mathcal{Q}(d_1, d_2, \dots, d_N)$. Then*

$$\overline{O}(U) = \sum_{i=1}^{N} c_i(U) - \frac{N}{2} + \sum_{i=1}^{N} \frac{1}{2q_i}.$$

Proof. There exist $T, k_1, k_2, \dots, k_N \in \mathbb{N}$ such that $U \in \mathcal{Q}(\{T\}, k_1, k_2, \dots, k_N) \subseteq \mathcal{Q}(d_1, d_2, \dots, d_N)$. For $i \in \{1, 2, \dots, N\}$ put $v_i = (u^i(1), u^i(2), \dots, u^i(T))$. Then $v_i \in \mathcal{P}(T, k_i)$ is a period cycle of u^i. Let $v_i' \sim v_i$ be an upper bound with respect to the partial order $\preceq$ of $\mathcal{R}(T, k_i)$. We have that v_i' is the l-th cyclic permutation of v_i for some $l \in \{0, 1, \dots, T-1\}$. It follows that (see the proof of Theorem 2.13) $\chi_{v_i}(l) = c_i = c_i + \chi_{v_i'}(0)$ and thus $\chi_{v_i}((l+n) \pmod{T}) = \chi_{v_i'}(n) + c_i$ for $n = 0, 1, \dots, T$. Hence by Definition of the primal unbalance

$$c_i = \frac{1}{T}\sum_{n=1}^{T}(\chi_{v_i}(n) - \chi_{v_i'}(n)) = \frac{1}{T}\sum_{n=1}^{T}(\kappa_{v_i'}(n) - \kappa_{v_i}(n)) = \tag{11}$$

$$\overline{I}(u^i) + \frac{1}{T}\sum_{n=1}^{T}\kappa_{\overline{\omega}(T,k_i)}(n) - \frac{1}{T}\sum_{n=1}^{T}\kappa_{u^i}(n).$$

By Corollary 2.12 and Lemma 5.7 we have that

$$\sum_{i=1}^{N}\frac{1}{T}\sum_{n=1}^{T}\kappa_{\overline{\omega}(T,k_i)}(n) = \sum_{i=1}^{N}(\frac{1}{T}\cdot\sum_{n=1}^{T}\lceil n\cdot\frac{k_i}{T}\rceil) =$$

$$\frac{1}{2}\cdot\sum_{i=1}^{N}(k_i + \frac{k_i}{T} + 1 - \frac{\gcd(k_i, T)}{T}) = \frac{1}{2}\cdot(T + 1 + N - \sum_{i=1}^{N}\frac{1}{q_i}).$$

So, by (11) we have that

$$\sum_{i=1}^{N} c_i = \overline{O}(U) - \frac{1}{T}\sum_{n=1}^{T} n + \frac{1}{2}\cdot(T + 1 + N - \sum_{i=1}^{N}\frac{1}{q_i}) = \overline{O}(U) + \frac{N}{2} - \sum_{i=1}^{N}\frac{1}{2q_i}.$$

□

We need the following definitions. Let $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$. Put

$$\widetilde{S} = \widetilde{S}(d_1, d_2, \dots, d_N) := \inf_{U \in \mathcal{S}(d_1, d_2, \dots, d_N)} \sum_{i=1}^{N} c_i(U).$$

Moreover, define

$$D = D(d_1, d_2, \dots, d_N) =$$

$$\{(x_1, x_2, \dots, x_N) \in \mathbb{R}^N : n \geq \sum_{i=1}^{N} \max(\lceil n\cdot d_i - x_i\rceil, 0) \text{ for every } n \in \mathbb{Z}_{\geq 0}\}$$

and

$$\widetilde{D} = \widetilde{D}(d_1, d_2, \dots, d_N) := \inf_{(x_1, x_2, \dots, x_N) \in D} \sum_{i=1}^{N} x_i.$$

It is easily seen that for every $U \in \mathcal{S}(d_1, d_2, \dots, d_N)$ it holds that

$$(c_1(U), c_2(U), \dots, c_N(U)) \in D(d_1, d_2, \dots, d_N). \tag{12}$$

Consider the following algorithm to construct a sequence U with given densities. It is similar to the SG algorithm, but the choice of U_n is slightly less restricted. Therefore it is called the GG (General Greedy) algorithm.

GG algorithm. Let $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$ and $x_1, x_2, \dots, x_N \in \mathbb{R}_{\geq 0}$ be given. Then $U = (U_1, U_2, \dots) \in \mathcal{S}(d_1, d_2, \dots, d_N)$ is determined inductively in the following way. Suppose that $U_1, U_2, \dots, U_{n-1}$ have been determined and thus $\kappa_{u^i}(n-1)$ is known for $i = 1, 2, \dots, N$. Choose U_n such that $U_n = i \in \{1, 2, \dots, N\}$ for which $\lfloor \frac{x_i + \kappa_{u^i}(n-1)}{d_i} \rfloor$ is minimal.

It is easily seen that for a word U constructed by the GG algorithm it holds that $U \in \mathcal{S}(d_1, d_2, \dots, d_N)$ and note that the SG algorithm is a special GG algorithm. The following results (until Theorem 5.19) are from [28] in which a slightly different notation is used. We restate the results using the notation of this paper. Note that some of the results are for arbitrary real densities and not only for rational densities.

Lemma 5.11 *Let $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$. Let $U \in \mathcal{S}(d_1, d_2, \dots, d_N)$ be constructed by applying the GG algorithm with $x_1, x_2, \dots, x_N \in \mathbb{R}_{\geq 0}$ such that $(x_1, x_2, \dots, x_N) \in D(d_1, d_2, \dots, d_N)$. Then $c_i(U) \leq x_i$ for $i = 1, 2, \dots, N$.*

A direct consequence of (12) and Lemma 5.11 is that

$$\widetilde{S} = \widetilde{D} \tag{13}$$

Theorem 5.12 *For every set of fractions $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$ there exist $U^* \in \mathcal{S}(d_1, d_2, \dots, d_N)$ and corresponding*
$(x_1^*, x_2^*, \dots, x_N^*) = (c_1(U^*), c_2(U^*), \dots, c_N(U^*)) \in D$ such that

$$\sum_{i=1}^N c_i(U^*) = \widetilde{S} \text{ and } \sum_{i=1}^N x_i^* = \widetilde{D}.$$

Thus according to Theorem 5.12 there are U and x_i which attain the respective infima.

Lemma 5.13 *Let $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$ and suppose that*

$$\sum_{i=1}^N x_i = N - 1 \text{ and } 0 \leq x_i \leq 1 + d_i \text{ for } i = 1, 2, \dots, N. \tag{14}$$

Then $(x_1, x_2, \dots, x_n) \in D$.

Particular $x_1, x_2, \dots, x_N$ that fulfill (14) are for example $x_i = 0$ for some $i \in \{1, 2, \dots, N\}$ and $x_j = 1$ for all $j \neq i$. Another example is $x_i = 1 - d_i$ for $i = 1, 2, \dots, N$. Combining (13), Theorem 5.12 and Lemma 5.13 gives the following corollary.

Corollary 5.14 *For every set of fractions $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$ we have that*

$$\widetilde{S} = \min_{U \in \mathcal{S}(d_1, d_2, \dots, d_N)} \sum_{i=1}^N c_i \le N - 1.$$

We think that for any billiard sequence U it holds that $\sum_{i=1}^N c_i(U) \le N-1$. The following theorem asserts that it is possible that $\widetilde{S} = N-1$ and thus Corollary 5.14 is sharp.

Theorem 5.15 *Let $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$. If $d_1, d_2, \dots, d_N$ are linearly independent over $\mathbb{Z}$ then $\widetilde{S}(d_1, d_2, \dots, d_N) = N-1$.*

Put

$$\mathcal{S}_0(d_1, d_2, \dots, d_N) = \{U \in \mathcal{S}(d_1, d_2, \dots, d_N) \text{ for which} \sum_{i=1}^N c_i = \widetilde{S}\}$$

and

$$D_0 = D_0(d_1, d_2, \dots, d_N) = \{(x_1, x_2, \dots, x_N) \in D(d_1, d_2, \dots, d_N) \text{ for which } \sum_{i=1}^N x_i = \widetilde{D}\}.$$

By Theorem 5.12 we have that $\mathcal{S}_0(d_1, d_2, \dots, d_N) \neq \emptyset$ and $D_0(d_1, d_2, \dots, d_N) \neq \emptyset$. Moreover, if $U \in \mathcal{S}_0(d_1, d_2, \dots, d_N)$ then $(c_1(U), c_2(U), \dots, c_N(U)) \in D_0(d_1, d_2, \dots, d_N)$ and vice versa if $(x_1, x_2, \dots, x_N) \in D_0(d_1, d_2, \dots, d_N)$ then we have for any word U constructed by applying the GG algorithm with these $x_1, x_2, \dots, x_N$ that $U \in \mathcal{S}_0(d_1, d_2, \dots, d_N)$.

Theorem 5.16 *Let $d_1, d_2, \dots, d_N \in \mathbb{R}_{>0}$ with $\sum_{i=1}^N d_i = 1$. Then for every $\varepsilon > 0$ and $i \in \{1, 2, \dots, N\}$ there exist $(x_1, x_2, \dots, x_N) \in D_0(d_1, d_2, \dots, d_N)$ with $x_i < \varepsilon$ and $x_j < 1$ for all $j \neq i$.*

In the foregoing results it was not required for the densities d_i to be rational. In the remaining of this section the densities d_i are rational with $\sum_{i=1}^N d_i = 1$. Further $p_i, q_i, k_i \in \mathbb{N}$ for $i = 1, 2, \dots, N$ and $T \in \mathbb{N}$ are such that $d_i = \frac{p_i}{q_i} = \frac{k_i}{T}$ with $\gcd(p_i, q_i) = 1$ and $\gcd(k_1, k_2, \dots, k_N) = 1$ for $i = 1, 2, \dots, N$.

Lemma 5.17 *Let $U \in \mathcal{S}(d_1, d_2, \dots, d_N)$ be such that $c_i(U) < \infty$ for $i \in \{1, 2, \dots, N\}$. Then $q_i \cdot c_i(U) \in \mathbb{Z}_{\ge 0}$.*

Proof. For every $n \in \mathbb{Z}_{\ge 0}$ we have that $q_i \cdot \chi_{u^i}(n) = p_i \cdot n - q_i \cdot \kappa_{u^i}(n) \in \mathbb{Z}$. □

Corollary 5.18 *Suppose that $(x_1, x_2, \dots, x_N) \in D_0(x_1, x_2, \dots, x_N)$. Then $q_i \cdot x_i \in \mathbb{Z}_{\ge 0}$ for $i = 1, 2, \dots, N$.*

From this corollary we have that $x_j < 1$ implies $x_j \le 1 - \frac{1}{q_j}$ for every j. Combining this with Theorem 5.16, we obtain the following result.

Theorem 5.19 *For every $i \in \{1, 2, \dots, N\}$ there exists some $(x_1, x_2, \dots, x_N) \in D_0(d_1, d_2, \dots, d_N)$ with $x_i = 0$ and $x_j \le 1 - \frac{1}{q_j}$ for every $j \neq i$. Moreover*

$$\widetilde{S} \le N - 1 - \sum_{i=1}^N \frac{1}{q_i} + \min_{i \in \{1, 2, \dots, N\}} \frac{1}{q_i}.$$

Proof of Theorem 5.5. By Lemma 5.4 we have that $U \in \mathcal{Q}(\{T\}, k_1, k_2, \dots, k_N) \subseteq \mathcal{Q}(d_1, d_2, \dots, d_N)$. According to Theorem 5.19 there exist $(x_1, x_2, \dots, x_N) \in D_0(d_1, d_2, \dots, d_N) \subseteq D(d_1, d_2, \dots, d_N)$ with $x_i \leq x_i^*$ for $i = 1, 2, \dots, N$. It follows directly from the definition of $D(d_1, d_2, \dots, d_N)$ that if $(x_1, x_2, \dots, x_N) \in D(d_1, d_2, \dots, d_N)$ and $x_i \leq x_i'$ for $i = 1, 2, \dots, N$ then $(x_1', x_2', \dots, x_N') \in D(d_1, d_2, \dots, d_N)$. Hence $(x_1^*, x_2^*, \dots, x_N^*) \in D(d_1, d_2, \dots, d_N)$. Since constructing a consistent billiard sequence is a special form of applying a GG algorithm we have by Lemma 5.11 that $c_i(U) \leq x_i^*$ for $i = 1, 2, \dots, N$. Thus

$$\sum_{i=1}^{N} c_i(U) \leq \sum_{i=1}^{N} x_i^* = N - 1 - \sum_{i=1}^{N} \frac{1}{q_i} + \min_{i \in \{1,2,\dots,N\}} \frac{1}{q_i}.$$

So, by Theorem 5.10 we have that

$$\overline{O}(U) = \sum_{i=1}^{N} c_i(U) - \frac{N}{2} + \sum_{i=1}^{N} \frac{1}{2q_i} \leq \frac{N}{2} - 1 + \min_{i \in \{1,2,\dots,N\}} \frac{1}{q_i} - \sum_{i=1}^{N} \frac{1}{2q_i}. \tag{15}$$

If $N = 2$ then $q_1 = q_2$ and thus $\sum_{i=1}^{N} \frac{1}{2q_i} = \min_{i \in \{1,2,\dots,N\}} \frac{1}{q_i}$. Combining this with (15) gives $\overline{O}(U) \leq 0$ and thus $\overline{O}(U) = 0$ if $N = 2$. If $N > 2$ then

$$\sum_{i=1}^{N} \frac{1}{2q_i} \geq \sum_{i=1}^{N} \min_{i \in \{1,2,\dots,N\}} \frac{1}{2q_i} > \min_{i \in \{1,2,\dots,N\}} \frac{1}{q_i}$$

and thus we have by (15) that $\overline{O}(U) < \frac{N}{2} - 1$ if $N > 2$. $\square$

We now show that for every set of rational densities there exist a consistent and thus periodic billiard sequence which has minimal unbalance among all (periodic) sequences with those densities.

Theorem 5.20 *For every set of rational densities d_i there exists some consistent billiard sequence $U^* \in \mathcal{Q}(\{T\}, k_1, k_2, \dots, k_N) \subseteq \mathcal{Q}(d_1, d_2, \dots, d_N)$ such that*

$$\sum_{i=1}^{N} c_i(U^*) = \widetilde{S} \text{ and } \overline{O}(U^*) = \min_{U \in \mathcal{Q}(d_1, d_2, \dots, d_N)} \overline{O}(U).$$

Proof. By Theorem 5.16 there exist $(x_1, x_2, \dots, x_N) \in D_0(d_1, d_2, \dots, d_N)$ with $0 \leq x_i < 1$ for $i = 1, 2, \dots, N$. Let U^* be obtained by applying the SG algorithm in a consistent way on such $x_1, x_2, \dots, x_N$. Then U^* is a consistent billiard sequence and by Lemma 5.4 we have that $U^* \in \mathcal{Q}(\{T\}, k_1, k_2, \dots, k_N) \subseteq \mathcal{Q}(d_1, d_2, \dots, d_N)$. Moreover, since $(x_1, x_2, \dots, x_N) \in D_0(d_1, d_2, \dots, d_N)$ we have that $U^* \in \mathcal{S}_0(d_1, d_2, \dots, d_N)$ and thus $\sum_{i=1}^{N} c_i(U^*) = \widetilde{S}$. By Theorem 5.10 it then follows that $\overline{O}(U^*) = \min_{U \in \mathcal{Q}(d_1, d_2, \dots, d_N)} \overline{O}(U)$. $\square$

Such a billiard sequence with minimal unbalance can be found in an algorithmic way. Indeed, in case of rational densities d_i it is shown in [28] that an $(x_1, x_2, \dots, x_N) \in D_0(d_1, d_2, \dots, d_N)$ with $0 \leq x_i < 1$ for $i = 1, 2, \dots, N$ can be found by solving an integer linear programming problem (ILP). Thus for given rational densities it is possible to obtain a consistent billiard sequence U with minimal unbalance $\overline{O}$ by first solving an ILP and then applying the SG algorithm in a consistent way. Then the routing policy ψ corresponding to U is periodic and the upper bound from Theorem 4.7 for $\overline{W}(\psi)$ is minimized by ψ for those densities. Note that from Theorem 4.4 it follows that if V is a mirror of such a routing sequence U that then V has minimal dual unbalance $\underline{O}$. The following example illustrates some of the results of this section.

Example. Let $N=5$, $d_1=\frac{3}{8}$, $d_2=\frac{7}{24}$, $d_3=\frac{1}{6}$, $d_4=\frac{1}{8}$ and $d_5=\frac{1}{24}$. We apply the SG algorithm in a consistent way (in case of a tie the letter of smallest index is chosen) with $x_1=\frac{7}{8}$, $x_2=0$, $x_3=\frac{5}{6}$, $x_4=\frac{7}{8}$ and $x_5=\frac{23}{24}$. Then we obtain the word $U=(2,1,2,1,3,2,4,1,2,1,3,1,2,4,1,3,2,1,2,1,3,4,5,1)^{\infty}$ and thus $U\in\mathcal{Q}(\{24\},9,7,4,3,1)\subseteq\mathcal{Q}(\frac{3}{8},\frac{7}{24},\frac{1}{6},\frac{1}{8},\frac{1}{24})$. Using Lemma 5.9 it can be checked that $c_1(U)=\frac{5}{8}$, $c_2(U)=0$, $c_3(U)=\frac{2}{3}$, $c_4(U)=\frac{3}{4}$, $c_5(U)=\frac{11}{12}$ and thus $\sum_{i=1}^{N}c_i(U)=\frac{71}{24}$. Note that $c_i(U)\le x_i$ for every $i\in\{1,2,3,4,5\}$ as expected by Lemma 5.11. By Theorem 5.10 or direct computation it follows that $\overline{O}(U)=\frac{17}{24}$ and thus $\overline{O}(U)<\frac{N}{2}-1$ as expected by Theorem 5.5. For the word $V=(1,5,4,3,1,2,1,2,3,1,4,2,1,3,1,2,1,4,2,3,1,2,1,2)^{\infty}$ which is a mirror of U we have by Theorem 4.4 that $\underline{O}(V)=\overline{O}(U)=\frac{17}{24}<\frac{N}{2}-1$.

By Lemma 5.11 we have that if the SG algorithm is applied with $x_i=c_i(U)$ for $i=1,2,\ldots,5$ then a word $U'\in\mathcal{Q}(\frac{3}{8},\frac{7}{24},\frac{1}{6},\frac{1}{8},\frac{1}{24})$ is obtained with $c_i(U')\le c_i(U)$ for $i=1,2,\ldots,5$ and thus $\overline{O}(U')\le\overline{O}(U)$. Hence by using the SG algorithm iteratively in this way a sequence of periodic words with the given densities is found with non-increasing total primal unbalance. We do this for this example starting with U and we get the following sequence of words. First we obtain $U'=(2,1,2,3,1,4,2,1,1,3,2,1,2,4,1,3,2,1,1,2,3,4,5,1)^{\infty}$ with $c_1(U')=\frac{5}{8}$, $c_2(U')=0$, $c_3(U')=\frac{1}{2}$, $c_4(U')=\frac{5}{8}$, $c_5(U')=\frac{11}{12}$ and $\overline{O}(U')=\frac{5}{12}$. By applying the SG algorithm to $x_i=c_i(U')$ we obtain the word $U''=(2,1,3,2,1,4,2,1,3,1,2,1,4,2,1,3,2,1,1,2,3,4,5,1)^{\infty}$ with $c_1(U'')=\frac{5}{8}$, $c_2(U'')=0$, $c_3(U'')=\frac{1}{2}$, $c_4(U'')=\frac{5}{8}$, $c_5(U'')=\frac{11}{12}$ and $\overline{O}(U'')=\frac{5}{12}$. Since $c_i(U'')=c_i(U')$ for $i=1,2,\ldots,5$ applying the SG algorithm to $x_i=c_i(U'')$ gives the word U'' again. Thus the iterative process stops with U'', (see [28] for more on this iterative process) but the word U'' does not have minimal unbalance for these densities. Namely, we have found that $(0,\frac{7}{24},\frac{1}{2},\frac{3}{4},\frac{23}{24})\in D_0(\frac{3}{8},\frac{7}{24},\frac{1}{6},\frac{1}{8},\frac{1}{24})$, and thus the word $W=(1,2,1,3,2,1,4,2,1,3,1,2,1,4,2,3,1,2,1,3,1,2,4,5)^{\infty}$ obtained by applying the SG algorithm to this solution with $\overline{O}(W)=\frac{1}{4}$ has minimal primal unbalance for these densities.

Conclusion

We investigated in this paper the static routing to parallel queues with different service times. We introduced a combinatorial notion which we called unbalance. For a periodic routing pattern we derived an upper bound on the average waiting time as function of its unbalance. From a combinatorial point of view it would be of interest to extend the notion of unbalance to non-periodic routing sequences which have irrational routing fractions. For applications it would be important to extend the bounds derived in this paper to the following models:

1. optimal routing of the server in polling systems;
2. parallel queues with two or more servers for each queue;
3. sojourn times in queueing networks.

Acknowledgement

The authors would like to thank Bruno Gaujal for his hospitality during their visits to LORIA, Nancy. During one of the stimulating discussions with him the graph order was found. The authors thank Robert Tijdeman for his helpful comments.

6 Appendix

In Theorem 3.10 and subsequent results we assumed that $\{\delta_i\}$ ($\{\sigma_j\}$) were i.i.d. random variables. Below we show that instead of i.i.d. random variables we may assume that they are stationary ergodic sequences. This greater generality is very useful for applications to traffic processes in telecommunication networks. To the authors' knowledge there does not exist a proof for the i.i.d. case in the literature. But, it is a special case of the more general result of Theorem 6.1 (see also the remark following the theorem). Recall that $W_u(j,w)$ denotes the workload for the server at the moment that the j-th customer is routed to the server (according to routing sequence u), given that the initial workload at time $T_1 = 0$ is equal to w.

Theorem 6.1 *Let u' and u'' be routing sequences of zeros and ones that are both representatives of some $u \in \mathcal{P}(d)$ where $d = \frac{k}{T} \in \mathbb{Q}$ and the period cycle of u is an element of $\mathcal{P}(T,k)$. Suppose that the interarrival times $\{\delta_i\}$ of customers arriving at the system are a stationary ergodic sequence with $\delta = \mathbb{E}\delta_1$. Moreover, assume that the sequence $\{\delta_{nT+l}\}_{n=0}^{\infty}$ is ergodic for each $l \in \{1,2,\dots,T\}$. Suppose that the service times σ_j of the customers routed to the server form a stationary ergodic sequence with $\sigma = \mathbb{E}\sigma_1$, and are stochastically independent of the interarrival times. Moreover, assume that the sequences $\{\sigma_{nk+l}\}_{n=0}^{\infty}$ is ergodic for each $l \in \{1,2,\dots,k\}$. Suppose that $\frac{\sigma}{\delta} \cdot d < 1$. Then $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{n=0}^{m-1} \delta_{nT+l} = \delta$ almost surely (a.s.) for $l = 1,2,\dots,T$ and there exists a deterministic value $\overline{W}(u)$ such that*

$$\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=0}^{m-1} \overline{W}_{u'}(j,w') = \lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=0}^{m-1} W_{u''}(j,w'') = \overline{W}(u), \ \ a.s.$$

for any initial workloads w' and w''.

Proof. Since $\{\delta_{nT+l}\}_{n=0}^{\infty}$ is a stationary ergodic sequence it follows directly from Birkhoff's theorem that ([13] Theorem 6.28) $\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{n=0}^{m-1} \delta_{nT+l} = \mathbb{E}\delta_l$ a.s. Since δ_i is a stationary sequence we have that $\mathbb{E}\delta_l = \delta$ for $l = 1,2,\dots,T$. It is shown in [6] that there exist k sequences $\{W_u^*(nk+j)\}_{n=0}^{\infty}$, $j = 1,2,\dots,k$, which are all stationary ergodic, such that these sequences couple in finite time with the workload sequences. Indeed, for any w there exists a finite random variable M such that almost surely,

$$W_u(nk+j,w) = W_u^*(nk+j) \text{ for all } n \geq M. \tag{16}$$

Define,

$$\overline{W}(u) = \frac{1}{k} \cdot \sum_{j=1}^{k} \mathbb{E}W_u^*(j).$$

From Birkhoff's ergodic theorem it follows that almost surely

$$\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{n=0}^{m-1} W_u^*(nk+j) = \mathbb{E}W_u^*(j) \text{ for } j = 1,2,\dots,k.$$

Hence,

$$\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^{m} W_u^*(j) = \frac{1}{k} \cdot \sum_{j=1}^{k} \mathbb{E}W_u^*(j) = \overline{W}(u) \text{ a.s.}$$

With the coupling result (16) it then follows that

$$\lim_{m\to\infty} \frac{1}{m} \cdot \sum_{j=1}^{m} \overline{W}_u(j,w) = \overline{W}(u) \text{ a.s.}$$

For u' a representative of u there is a $l \in \{1, 2, \dots, T\}$ such that $u_n = u'_{l+n}$, $n = 1, 2, \dots$. If we couple the service times in the routing sequences u and u' such that $\sigma_j = \sigma'_{l+j}$ then $W_u(n, w) = W_{u'}(l + n, w')$ a.s. for $n = 1, 2, \dots$, where $w := W_{u'}(l, w')$. Hence

$$\lim_{m \to \infty} \frac{1}{m} \cdot \sum_{j=1}^{M} W_{u'}(j, w') = \overline{W}(u) \text{ a.s.}$$

□

Remark. If $\{\delta_i\}$ and $\{\sigma_j\}$ are independent sequences of i.i.d. random variables with mean δ respectively σ, then all the assumptions of Theorem 6.1 are satisfied. Indeed, every subsequence of i.i.d. random variables is stationary ergodic (see [13] Corollary 6.33). In modern telecommunication systems the arrival process is composed of heterogeneous traffic streams and in general more complex than a renewal process. Often a Markov Modulated Poisson process is considered as the appropriate process (see [17]). This process also satisfies the assumptions of Theorem 6.1.

References

[1] E. Altman, Applications of Markov decision processes in telecommunication: a survey, In *Markov Decision Processes*, Eds. A. Shwartz and E. Feinberg, 2000.

[2] E. Altman, S. Bhulai, B. Gaujal and A. Hordijk, Open-loop routing to M parallel servers with no buffers, *J. Appl. Probab.*, **37**, 668-684 (2000).

[3] E. Altman, B. Gaujal and A. Hordijk, Admission control in stochastic event graphs, *IEEE Trans. Automat. Control*, **45**, 854-867 (2000).

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

[5] E.Altman, B. Gaujal and A. Hordijk, Multimodularity, convexity and optimization properties, *Math. Oper. Res.*, **25**, 324-347 (2000).

[6] E. Altman and A. Hordijk, Applications of Borovkov's renovation theory to non-stationary stochastic recursive sequences and their control, *Adv. in Appl. Probab.*, **29**, 388-413 (1997).

[7] Y. Arian and Y. Levy, Algorithms for generalized round robin routing, *Oper. Res. Lett.*, **12**, 313-319 (1992).

[8] P. Arnoux, C. Mauduit, I. Shiokawa and J. Tamura, Complexity of sequences defined by billiards in the cube, *Bull. Soc. Math. France*, **122**, 1-12 (1994).

[9] A. Bar-Noy, R. Bhatia, J. Naor and B. Schieber, Minimizing service and operation costs of periodic scheduling, In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 11-20, ACM, New York, 1998.

[10] Y. Baryshnikov, Complexity of trajectories in rectangular billiards, *Comm. Math. Phys.*, **175** , 43-56 (1995).

[11] R. K. Boel and J. H. van Schuppen, Distributed routing for load balancing, *Proceedings of the IEEE*, **77**, 210-221 (1989).

[12] S. Borst and K. G. Kamakrishnan, Optimization of template-driven scheduling mechanisms, *J. Sched.*, **2**, 19-33 (1999).

[13] J. Breiman, *Probability*, Addison-Wesley, 1968.

[14] C. S. Chang, X. Chao and M. Pinedo, A note on queues with Bernoulli routing, In *29th Conference on Decision and Control*, 1990.

[15] M. B. Combé and O. J. Boxma (1994), Optimization of static traffic allocation policies, *Theoret. Comput. Sci.*, **125**, 17-43 (1994).

[16] E. G. Coffman, Jr., Z. Liu and R. R. Weber, Optimal robot scheduling for web search engines, *J. Sched.*, **1**, 15-29 (1998).

[17] W. Fischer and K. S. Meier-Hellstern, The Markov-modulated Poisson process (MMPP) cook book, *Performance Evaluation*, **18**, 149-171 (1992).

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

[19] D. Gross and C. M. Harris, *Fundamentals of queueing theory*, Wiley Series in Probability and Mathematical Statistics, 1974.

[20] B. Hajek, The proof of a folk theorem on queueing delay with applications to routing in networks, *J. Assoc. Comput. Mach.*, **30**, 834-851 (1983).

[21] B. Hajek, Extremal splittings of point processes, *Math. Oper. Res.*, **10**, 543-556 (1985).

[22] R. Hariharan, V. G. Kulkarni and S. Stidham, A survey of research relevant to virtual circuit routing in telecommunication networks, Technical Report WC/OC/TR 90-13, University of North Carolina at Chapel Hill, 1990.

[23] A. Hordijk, G. M. Koole and J. A. Loeve, Analysis of a customer assignment model with no state information, *Probab. Engrg. Inform. Sci.*, **8**, 419-429 (1994).

[24] A. Itai and Z. Rosberg, A golden ratio control policy for a multiple-access channel, *IEEE Trans. Automat. Control*, **29**, 712-718 (1984).

[25] H. Kameda and Y. Zhang, Uniqueness of the solution for optimal static routing in open BCMP queueing networks, Math. Comput. Modelling, **22**, 119-130 (1995).

[26] L. Kleinrock, *Queueing systems. Volume II: Computer applications*, Wiley, New York, 1976.

[27] G. M. Koole, On the static assignment to parallel servers, *IEEE Trans. Automat. Control*, **44**, 1588-1592 (1999).

[28] D. A. van der Laan, Routing jobs to servers with deterministic service times, MI Report no. 2000-20, Leiden University, 2000.

[29] Z. Liu and R. Righter, Optimal load balancing on distributed homogeneous unreliable processors, *Oper. Res.*, **46**, 563-573 (1998).

[30] M. Lothaire, *Combinatorics on Words*, Cambridge University Press, Cambridge, Reissued in 1997, 1983.

[31] M. Morse and G.A. Hedlund, Symbolic dynamics II, *Amer. J. Math.*, **62**, 1-42 (1940).

[32] Z. Rosberg and D. Towsley, Customer routing to parallel servers with different rates, *IEEE Trans. Automat. Control*, **30**, 1140-1143 (1985).

[33] S. Sano and N. Miyoshi, m-Balanced sequences and their applications to some network scheduling problems, *submitted* (2000).

[34] H. Shirakawa, M. Mori, and M. Kijima, Evaluation of Regular Splitting Queues, *Comm. Statist. Stochastic Models*, **5**, 219-234 (1989).

[35] R. Tijdeman, Exact covers of balanced sequences and Fraenkel's conjecture, *Algebraic number theory and diophantine analysis*, Eds.F. Halter-Koch and R.F. Tichy, Walter de Gruyter, 467-483 (2000).

[36] R. Tijdeman, Fraenkel's conjecture for six sequences, *Discrete Math.*, **222**, 223-234 (2000).

[37] T. P. Yum, The design and analysis of a semidynamic deterministic routing rule, *IEEE Trans. Comm.*, **29**, 498-504 (1981).

A. Hordijk: Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands; email: hordijk@math.leidenuniv.nl

D.A van der Laan: Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands; email: dvdlaan@math.leidenuniv.nl