

# On the static assignment to parallel servers

Ger Koole

Vrije Universiteit

Faculty of Mathematics and Computer Science

De Boelelaan 1081a, 1081 HV Amsterdam

The Netherlands

Email: koole@cs.vu.nl, Url: www.cs.vu.nl/~koole

Published in *IEEE Transactions on Automatic Control* **44**:1588-1592, 1999

## Abstract

We study the static assignment to  $M$  parallel, exponential, heterogeneous servers. Blocked customers are lost, while the objective is to minimize the average number of blocked customers. The problem is formulated as a stochastic control problem with partial observation, and an equivalent full observation problem is formulated. Numerical experiments are conducted and the structure of the optimal policies is studied.

*Keywords:* partially observed Markov Decision Processes, static assignments, round robin, regular sequences, dynamic programming

## 1 Introduction

In this paper we study the following assignment problem. Customers arrive to the system according to an arrival process with i.i.d. interarrival times, having distribution function  $G$ . Upon arrival each customer is routed to one of the  $M$  servers. If there is still a customer present at the server where the customer is routed to then the newly arriving customer is lost. All service times are exponentially distributed, with rate  $\mu_m$  for server  $m$ ,  $1 \leq m \leq M$ . The routing of new customers is static, i.e., it is not allowed to use current state information such as the availability of the servers. Our objective is to find the assignment policy that minimizes the average number of lost customers.

In Section 2 we formulate our problem as a stochastic control model with partial observations. The standard approach to this type of control problem is formulating an equivalent full observation model and solving that using one of the standard methods such as value iteration. (See Sections 6.4, 6.6 and 6.7 of Kumar & Varaiya [6] for general results.) In general this leads to problems which have state spaces that are too big to allow for numerical solution methods. For the current model we prove that a simpler representation as full observation problem exists, that be solved numerically for small  $M$ . This model has a  $M$ -dimensional state space, with as  $m$ th component the number of assignments ago that a customer was last sent to server  $m$ .

In Section 3 we analyze this full observation model analytically. Among other things we prove the periodicity of the optimal policies and the optimality of round robin in the case of symmetric servers. Finally in Section 4 we report on the numerical experiments. Because the full observation problem has a countable state space, we solve finite state upper and lower bound models. For state spaces sufficiently large the upper and lower bounds give the same solution for all models that

we consider, proving that these solutions are optimal for the original problems. We compare the optimal policies with other policies such as Bernoulli policies and myopic policies.

The contribution of this paper is twofold. Usually the numerical solution of partially observed control models is prohibited by the size of the state space of the equivalent full information model. The current model shows that this problem can be avoided for the current model by identifying, in the equivalent full observation model, the set of states that can actually occur. Besides this aspect, related to partial observation models, this paper adds to the literature on routing models. Several variants of our model have already been studied extensively. If the assignment is allowed to be dynamic, i.e., current state information can be used, then it is not difficult to show (using inductive arguments) that sending to the fastest available server is optimal (Koole [4], Section 1.3). The model with static assignment but with buffers at each of the servers has been studied in Combé & Boxma [1] and Hordijk et al. [3]. In [3] arguments based on partial observation are used as well, but an algorithm that does not guarantee optimality is used. A slotted version of the current model (i.e., with constant interarrival times) is the subject of Rosberg & Towsley [9]. They consider a heuristic assignment policy, based on the golden ratio, and prove asymptotic results for it.

## 2 The model and its equivalent full observation problem

In Koole [5] a new line of proof is developed for showing the equivalence between partial observation models and their full observation counterparts. Here we apply this method to the current assignment model. We start by formulating the system as a discrete time control problem. As we consider static routing in this paper, the control is not allowed to depend on the current state. For solving this *partial information* model directly there are no methods available. Therefore we formulate a second model, on a different state space, where the control is allowed to depend on the actual state. For the current routing problem this *full information* model has a special form, which allows it to be solved numerically. The main result of this section is the equivalence between the two models.

Let us now formulate the (partial information) model. As decision points we take the consecutive arrival times. The control problem is defined by  $(\mathcal{X}, \mathcal{A}, p, c)$ , with:

- $\mathcal{X} = \{0, 1\}^M$  the state space. For  $x = (x_1, \dots, x_M) \in \mathcal{X}$   $x_m = 1$  (0) indicates that there is a (no) customer at server  $m$  just before the next decision;
- $\mathcal{A} = \{1, \dots, M\}$  the set of possible assignments or actions, action  $a$  corresponding to sending the customer to server  $a$ ;
- $p(x, a, x')$  the transition probability of going from  $x$  to  $x'$  if action  $a$  is used. The actual value of  $p$  does not matter: as we shall see in the proof of Theorem 2.2 only transition probabilities for the state components are important (see also Remark 2.1 below). These marginal transition probabilities  $p_m$  are defined as follows. Let  $S$  be a r.v. with distribution  $G$ , and denote with  $E(\mu)$  an exponentially distributed r.v. with parameter  $\mu$ . Define  $q_m = \mathbb{P}(S < E(\mu_m))$ , i.e.,  $q_m$  is the probability that the customer at server  $m$  did not leave during an interarrival period. Now we have

$$p_m(x_m, a, x'_m) = \begin{cases} 1 & \text{if } x_m = 0, a \neq m \text{ and } x'_m = 0; \\ 0 & \text{if } x_m = 0, a \neq m \text{ and } x'_m = 1; \\ 1 - q_m & \text{if } x_m = 1 \text{ or } a = m \text{ and } x'_m = 0; \\ q_m & \text{if } x_m = 1 \text{ or } a = m \text{ and } x'_m = 1; \end{cases}$$

- $c(x, a)$  the direct costs, thus  $c(x, a) = 1$  (0) if  $x_a = 1$  (0).

The system is to be controlled from 1 to  $\infty$ . In general a policy or control law  $\phi$  is a set of decision rules  $\phi_1, \phi_2, \dots$ , with  $\phi_t : \mathcal{H}_t \rightarrow \mathcal{A}$ , with  $\mathcal{H}_t = (\mathcal{X} \times \mathcal{A})^{t-1} \times \mathcal{X}$ . Here we restrict to

policies such that  $\phi_t : \mathcal{A}^{t-1} \rightarrow \mathcal{A}$ , meaning that no state information is used. Denote the set of these assignment policies by  $\Phi$ . Note that this choice of  $\Phi$  makes the control problem a partial observation problem.

Define, for a fixed policy  $\phi$ , the r.v.'s  $X_t$  and  $A_t$  as follows (sometimes we write  $X_t(\phi)$ , etc., to denote the dependence on the policy). Assume that the system is initially in the empty state  $\{0, \dots, 0\}$ , which defines the initial distribution  $X_1$ .

For each realization  $h_t = (x_1, a_1, \dots, x_{t-1}, a_{t-1}, x_t) \in \mathcal{H}_t$  of  $(X_1, A_1, \dots, X_t) = H_t$ ,  $A_t$  is given by  $a_t = \phi_t(h_t)$ , then, given  $H_t = h_t$ ,  $X_{t+1}$  has the value  $x$  with probability  $p(x_t, a_t, x)$ . For obvious reasons  $h_t$  is called the history of the system at  $t$ . We write  $h$  for  $h_\infty$ .

Now define  $C(\phi) = \limsup_{T \rightarrow \infty} \frac{1}{T} E \sum_{t=1}^T c(X_t, A_t)$ , the average expected costs under  $\phi$ . The problem is to find a policy  $\phi^*$  (if one exists) such that  $C(\phi^*) = \min\{C(\phi) \mid \phi \in \Phi\}$ . Note that the average costs under a policy  $\phi$  are equal to the average probability that a customer is blocked. If we are interested in the average number of blocked customers, then we should multiply  $C(\phi)$  by the average number of arrivals per time unit, equal to  $(\mathbb{E}S)^{-1}$ .

**Remark 2.1** It is tempting to state that  $p(x, a, x') = \prod_{m=1}^M p_m(x_m, a, x'_m)$ . This holds only true for  $S$  a.s. constant, because  $\{E(\mu_1) \leq S\}$  and  $\{E(\mu_2) \leq S\}$  are dependent events in all other cases. As we shall see in the proof of Theorem 2.2 however it is only  $p_m$  that matters, and thus we might as well assume that  $p$  has the above form.

Now we define the full observation problem  $(\tilde{\mathcal{X}}, \tilde{\mathcal{A}}, \tilde{p}, \tilde{c})$  as follows (all variables related to it are indicated by  $\tilde{\phantom{x}}$ ):

- $\tilde{\mathcal{X}} = \mathbb{N}^M$  is the state space. For  $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_M) \in \tilde{\mathcal{X}}$   $\tilde{x}_m = k$  denotes that  $k$  arrivals ago a customer was assigned to server  $m$ ;
- $\tilde{\mathcal{A}} = \{1, \dots, M\}$  is the set of possible assignments or actions in each state, action  $\tilde{a}$  corresponding to sending the customer to server  $\tilde{a}$ ;
- $\tilde{p}(\tilde{x}, \tilde{a}, \tilde{x}')$  is the transition probability of going from  $\tilde{x}$  to  $\tilde{x}'$  if action  $\tilde{a}$  is used.

$$\tilde{p}(\tilde{x}, \tilde{a}, \tilde{x}') = \begin{cases} 1 & \text{if } \tilde{x}'_{\tilde{a}} = 1 \text{ and } \tilde{x}'_m = \tilde{x}_m + 1 \text{ for all } m \neq \tilde{a}; \\ 0 & \text{otherwise;} \end{cases}$$

- $\tilde{c}$ , the direct costs, are equal to  $\tilde{c}(\tilde{x}, \tilde{a}) = q_{\tilde{a}}^{\tilde{x}_{\tilde{a}}}$ , the probability that the currently arriving customer is blocked.

The class of allowable policies  $\tilde{\Phi}$  consists of all policies  $\tilde{\phi}$  with decision rules  $\tilde{\phi}_t : \tilde{\mathcal{H}}_t \rightarrow \tilde{\mathcal{A}}$ , where  $\tilde{\mathcal{H}}_t = (\tilde{\mathcal{X}} \times \tilde{\mathcal{A}})^{t-1} \times \tilde{\mathcal{X}}$ . Thus the control is allowed to depend on current and past states and actions.

Furthermore assume that  $\tilde{X}_1 = (\infty, \dots, \infty)$  a.s. It is now time to look at the relation between the two models.

**Theorem 2.2** *The partial observation problem  $(\mathcal{X}, \mathcal{A}, p, c)$  with allowable policies  $\Phi$  and the full observation problem  $(\tilde{\mathcal{X}}, \tilde{\mathcal{A}}, \tilde{p}, \tilde{c})$  are equivalent in the sense that:*

- i) *Optimal policies exist for both models;*
- ii) *The expected average costs of the optimal policies are equal;*
- iii) *The optimal policy for the partial observation model consists of the consecutive assignments done by the optimal policy of the full observation model.*

**Proof** The structure of the proof is as follows. After some preliminary work we define a function  $\Gamma : \Phi \rightarrow \tilde{\Phi}$ , and we show that  $C(\phi) = \tilde{C}(\Gamma(\phi))$  for all  $\phi \in \Phi$ . Then we define an equivalence relation on  $\tilde{\Phi}$ , for which we show that policies in the same equivalence class have the same costs,

and that each class has a “representative” which is the image of a policy in  $\Phi$ . This shows that if an optimal policy exists for the full observation problem, then there is an optimal policy for the original problem with the same value. How this policy looks like follows directly from the proof. So far we follow the method of [5], Section 2. Finally we show the existence of an optimal policy for the full observation problem using a condition from Ross [10].

Thus let us first consider the structure of an arbitrary policy  $\phi$ . Define the string of actions  $a_1, a_2, \dots$  as follows:  $a_1 = \phi_1$ ,  $a_2 = \phi_2(a_1)$ ,  $a_3 = \phi_3(a_1, a_2)$ , etc. Note that these are the actions that are actually chosen, thus the history  $h = (a_1, a_2, \dots)$  occurs a.s. Now we define the function  $\Gamma : \Phi \rightarrow \tilde{\Phi}$  as follows:  $\tilde{\phi} = \Gamma(\phi)$  is given by  $\tilde{\phi}_t(\tilde{h}_t) = a_t$  for all  $\tilde{h}_t$ . Thus  $\tilde{\phi}$  uses at  $t$  for each history the action that is taken a.s. by  $\phi$ . Note that also under an arbitrary  $\tilde{\phi}$  a single history occurs a.s., as the transition probabilities are either 1 or 0. Consider for some  $\phi$  and  $\tilde{\phi} = \Gamma(\phi)$  the histories  $h$  and  $\tilde{h}$  that occur a.s. It is readily verified that  $(\tilde{x}_t)_m = t - \max_s \{a_s = m\}$  if such an  $s$  exists,  $\infty$  otherwise. Therefore from the definition of  $p_m$  we see that  $\mathbb{P}(x_m = 1) = q_m^{(\tilde{x}_t)_m}$ , and thus  $\mathbb{E}c(X_t, a) = q_a^{(\tilde{x}_t)_a} = \mathbb{E}\tilde{c}(\tilde{X}_t, a)$ , and finally  $C(\phi) = \tilde{C}(\Gamma(\phi))$ .

Define an equivalence relation  $\sim$  on  $\tilde{\Phi}$  as follows:  $\tilde{\phi} \sim \tilde{\phi}'$  if  $\tilde{\phi}_t(\tilde{h}_t) = \tilde{\phi}'_t(\tilde{h}_t)$  for  $\tilde{h}_t$  the history that occurs a.s., and for all  $t$ . It is easily verified that this is indeed an equivalence relation. It follows directly that  $\tilde{C}(\tilde{\phi}) = \tilde{C}(\tilde{\phi}')$ . The policy  $\tilde{\phi}$  such that  $\tilde{\phi}_t$  is constant for each  $t$  is the representative of its class.

Finally we prove the existence of an optimal policy for the full observation problem. According to Theorem 2.2 of Ch. 5 of [10], it suffices to show that  $|\tilde{V}^\alpha(\tilde{x}) - \tilde{V}^\alpha(z)|$  is uniformly bounded in  $\alpha$  and  $\tilde{x}$ , for some  $z$ , where  $\tilde{V}^\alpha(\tilde{x})$  are the minimal  $\alpha$ -discounted costs. Recall that a policy (for a given initial state) is equivalent to an assignment sequence. Note that when the same assignment sequence is used for two different initial states, then the  $\alpha$ -discounted difference cannot be bigger than  $M$ , as  $0 \leq \alpha^t \leq 1$  for all  $t$  and  $0 \leq \tilde{c}(\tilde{x}, \tilde{a}) \leq 1$  for all  $\tilde{x}$  and  $\tilde{a}$ , and because differences in costs for an action can only occur the first time that that action occurs in the assignment sequence.

Fix  $\alpha$  and  $\tilde{x}$ , and let  $\tilde{\phi}^z$  be the  $\alpha$ -optimal assignment sequence for initial state  $z$ . Write  $\tilde{V}^\alpha(\tilde{\phi}^z, \tilde{x})$  for the value of the same assignments done for initial state  $\tilde{x}$ . Then

$$\tilde{V}^\alpha(\tilde{x}) - \tilde{V}^\alpha(z) \leq \tilde{V}^\alpha(\tilde{\phi}^z, \tilde{x}) - \tilde{V}^\alpha(z) = \tilde{V}^\alpha(\tilde{\phi}^z, \tilde{x}) - \tilde{V}^\alpha(\tilde{\phi}^z, z) \leq M.$$

The same argument applies to  $\tilde{V}^\alpha(z) - \tilde{V}^\alpha(\tilde{x})$ . □

Now that we have proven the equivalence we use both models without making a difference. Therefore we drop the  $\tilde{\cdot}$  from the notation.

**Remark 2.3** For each well-formulated partial observation model there exists an equivalent full observation problem (e.g., see [6], Ch. 6). However, in general its state space consists of all probability distributions on the original state space  $\mathcal{X}$ . Compared to that we were able in Theorem 2.2 to reduce the complexity of the state space considerably, from a subset of  $[0, 1]^{2^M - 1}$  to  $\mathbb{N}^M$ . In Section 4 we will see that Theorem 2.2 even allows us to compute optimal policies. The equivalence also shows us that optimal policies for partial observation models do not randomize. Therefore we only considered deterministic policies.

### 3 Structure of the optimal policy

The equivalence in the previous section is mainly developed for numerical purposes: in the next section we will compute optimal policies with it. However, it can also be used to prove structural results, such as the optimality of the round-robin policy if  $\mu_1 = \dots = \mu_M$ . This type of result is the subject of this section. We start by showing that every server is used infinitely often.

**Theorem 3.1** *Each optimal assignment sequence  $\phi = (a_1, a_2, \dots)$ ,  $a_t \in \{1, \dots, M\}$ , has the following property:  $\sup_t \{a_t = a\} = \infty$  for all  $a \in \{1, \dots, M\}$ .*

**Proof** Suppose that for some  $a$  there is a  $K$  such that  $\max_t \{a_t = a\} < K$ . Then  $\phi' = (a'_1, a'_2, \dots) = (a_K, a_{K+1}, \dots)$  is also optimal; note that from  $K$  on the total difference in costs between  $\phi'$  and  $\phi$  is limited by  $M$  (as in the proof of Theorem 2.2). Thus there exists an optimal policy  $\phi'$  (with average costs  $C(\phi') > 0$ ) that does not assign to server  $a$ . Now take  $s$  such that  $q_a^s < C(\phi')$ , and construct  $\phi'' = (b_1, b_2, \dots)$  as follows:

$$b_t = \begin{cases} a & \text{if } t \text{ is a multiple of } s; \\ a'_{t+\lfloor t/s \rfloor} & \text{otherwise.} \end{cases}$$

It is readily seen that  $c(X_{t+\lfloor t/s \rfloor}(\phi''), A_{t+\lfloor t/s \rfloor}(\phi'')) \leq c(X_t(\phi'), A_t(\phi'))$  for all  $t$  not a multiple of  $s$ : the assignments are equal, but occur later under  $\phi''$ . Thus  $C(\phi'') \leq \frac{1}{s}q_a^s + \frac{s-1}{s}C(\phi') < C(\phi')$ , and thus is  $\phi'$  not optimal. Then  $\phi$  is neither optimal, giving the contradiction.  $\square$

This result can be used to completely characterize the structure for  $M = 2$ .

**Theorem 3.2** *For  $M = 2$  and  $\mu_2 \geq \mu_1$  the optimal assignment sequence repeats  $12 \cdots 2$ , where the length of this string is equal to  $k$  for  $k$  the integer minimizing  $\frac{1}{k}q_1^k + \frac{k-2}{k}q_2 \frac{1}{k}q_2^2$ .*

**Proof** Due to Theorem 3.1 the assignment 12 occurs infinitely often. At the first decision epoch after the occurrence of 12 the full observation model state is equal to  $(2, 1)$ . As the state will be equal after the next 12, the assignments between two consecutive 12's will be equal too. To avoid 12 from occurring in this string it must be of the form  $2 \cdots 21 \cdots 1$ , possibly with 0 1's or 2's. By considering two consecutive strings we see that the assignment sequence repeats indefinitely the string  $1 \cdots 12 \cdots 2$ , with  $k_1 \geq 1$  1's and  $k_2 \geq 1$  2's. The average costs of this policy are equal to  $(q_1^{k_2+1} + (k_1-1)q_1 + q_2^{k_1+1} + (k_2-1)q_2)/(k_1+k_2)$ . If  $k_1 \geq 2$  and  $k_2 \geq 2$  then interchanging the first and last assignment gives  $21 \cdots 12 \cdots 21$ , with costs  $(q_2^2 + q_1^2 + (k_1-2)q_1 + q_2^{k_1} + (k_2-2)q_2 + q_1^{k_2})/(k_1+k_2)$ , which is smaller than the expression above, because  $q_1^n$  and  $q_2^n$  are convex in  $n$ . But the string  $21 \cdots 12 \cdots 21$  can be rewritten as  $1 \cdots 12 \cdots 212$ , and thus the assignments between two consecutive 12's are not equal. Therefore the optimal policy repeats either  $12 \cdots 2$  or  $1 \cdots 12$ . Given that the total length is  $k$ , the average costs are  $\frac{1}{k}q_1^k + \frac{k-2}{k}q_2 + \frac{1}{k}q_2^2$  and  $\frac{1}{k}q_2^k + \frac{k-2}{k}q_1 + \frac{1}{k}q_1^2$ , respectively. From  $q_2 < q_1$  it follows that the first is smaller (because  $q^k - (k-2+q)q$  is decreasing in  $q$  if  $0 < q < 1$ ).  $\square$

Let  $\phi = (a_1, a_2, \dots)$  be an optimal assignment sequence of Theorem 3.2 with length  $k$ . Then  $\phi$  can also be defined by the function  $a_n = \lfloor (n+1)/k \rfloor - \lfloor n/k \rfloor + 1$ . This type of assignment sequence (with  $k$  not necessary integer) is the subject of Hajek [2], where a routing model to two queues with infinite waiting room is considered. For  $k$  non-integer sequences that repeat for example 12122 can also occur (here we took  $k = \frac{5}{2}$ ); in our model they are only optimal if both 12 and 122 are optimal, thus if in state  $(2, 1)$  action 1 and 2 are both optimal. (The additional +1 in our formula comes from the fact that we have a 12-valued sequence, while [2] considers 01-sequences.)

Another interesting case is when two or more of the service rates are equal. We show that it is optimal to assign to these servers in a cyclic manner. Thus assume that there is a set of indices  $\mathcal{I} \subset \{1, \dots, M\}$  ( $|\mathcal{I}| = I > 1$ ) and  $\mu > 0$  such that  $\mu_i = \mu$  for all  $i \in \mathcal{I}$ . We say that an assignment sequence is cyclic on  $\mathcal{I}$  if there is some permutation  $\Pi$  of  $\mathcal{I}$  such that the subsequence consisting of all assignments to servers in  $\mathcal{I}$  repeats  $\Pi(1) \cdots \Pi(I)$  indefinitely.

**Theorem 3.3** Assume that  $\mu_i$  is constant on some set of indices  $\mathcal{I}$ . Then there exists an optimal policy  $\phi$  such that  $\phi$  is cyclic on  $\mathcal{I}$ .

**Proof** Let  $\phi_{\mathcal{I}} = (a_1, a_2, \dots)$  be the subsequence of  $\phi$  containing all assignments to servers in  $\mathcal{I}$ , and let  $n$  be the first non-cyclical assignment (with  $a_n = i$ ). Thus either  $n \leq I$  and  $a_k = a_n$  for some  $k < n$ , or  $n > I$  and  $a_n \neq a_{n-I}$ . We construct a new policy  $\phi'_{\mathcal{I}}$  with lower total costs by interchanging  $i$  with another action  $j$  from  $n$  on. If  $n \leq I$  then choose  $j$  arbitrary from  $\mathcal{I} \setminus \{a_1, \dots, a_{n-1}\}$ , if  $n > I$  then we take  $j = a_{n-I}$ .

Define  $k_i = \max\{k < n \mid a_k = i\}$  and  $k_j = \min\{k > n \mid a_k = j\}$  (we take  $k_j = \infty$  if  $j$  does not occur after  $n$ ). Let  $g_1$  ( $g_2$ ) be the number of assignments in  $\phi$  other than to  $\mathcal{I}$  between assignment  $k_1$  and  $n$  ( $n$  and  $k_2$ ) in  $\phi_{\mathcal{I}}$ . Consider first the case  $n \leq I$ . Note that for the first assignment to a server the costs are 0. The costs at  $n$  and  $k_j$  go from  $q_{\mathcal{I}}^{n-k_i+g_1}$  to  $q_{\mathcal{I}}^{k_j-k_i+g_1+g_2}$  (with  $q_{\mathcal{I}} = q_i$  for  $i \in \mathcal{I}$ ), the costs at other epochs remain unchanged. From  $k_i < n < k_j$  it follows that the costs decrease. For the case that  $n > I$  the costs go from  $q_{\mathcal{I}}^{n-k_i+g_1} + q_{\mathcal{I}}^{k_j-n+I+g_0+g_1+g_2}$  to  $q_{\mathcal{I}}^{I+g_0+g_1} + q_{\mathcal{I}}^{k_j-k_i+g_1+g_2}$ , which decreases the costs because  $n - k_i < I, k_j - k_i < k_j - n + I$  and because  $q_{\mathcal{I}}^k$  is a convex function of  $k$ . Here  $g_0$  is defined as  $g_1$  but for assignments between  $n - I$  and  $k_1$ . Note also that in both cases the costs at  $n$  decrease. Therefore the average costs from 1 to  $t$  under  $\phi'_{\mathcal{I}}$  are lower for any  $t$ , and thus the same holds for its limsup, the long-run average costs. Repeating this argument gives the optimality of cyclic routing.  $\square$

Generalizations to non-exponential service times are possible, but fall outside the scope of this paper. The following corollary is immediate.

**Corollary 3.4** If  $\mu_1 = \dots = \mu_M$ , then the cyclic assignment policy (also known as the round robin policy) minimizes the average costs.

This policy is well-known to be optimal in the situation with waiting room at the queues (see e.g. Liu & Towsley [7]).

For more than two queues and all service rates different we have no results on the structure of the optimal policy. However, we can show that the optimal policy is *periodic*. A policy  $\phi = (a_1, a_2, \dots)$  is called periodic if there are  $T$  and  $k$  such that  $a_t = a_{t+k}$  for all  $t \geq T$ . We call  $k$  the length of the period.

**Theorem 3.5** For each model there exists an optimal periodic assignment sequence.

**Proof** Consider a policy  $\phi$  and a certain server  $a$ . Let  $d_1, d_2, \dots$  be the lengths of the consecutive intervals between two assignments to server  $a$ . Assume that  $\sup_i d_i = \infty$ . We construct an assignment policy  $\phi'$  with similarly defined distances  $d'_1, d'_2, \dots$ , for which  $\sup_i d'_i < \infty$  and  $C(\phi') \leq C(\phi)$ .

Consider for  $\phi$   $M + 1$  consecutive assignments. At least one server occurs twice: the costs for the second assignment (which is one of the last  $M$ ) is at least  $(\min_m q_m)^M$ . Take  $n$  such that  $2q_a^n < (\min_m q_m)^M$ . Consider for the assignment sequence  $\phi$  a  $d_i$  for which  $d_i \geq 2n + M$ . Choose from the middle  $M$  assignments an assignment (say to  $a'$ ) for which the costs are at least  $(\min_m q_m)^M$ . Due to the argument above at least one such assignment exists. Replace this assignment by  $a$ . The decrease in costs are larger than  $(\min_m q_m)^M - 2q_a^n > 0$ , as the costs for the next assignment to  $a'$  decrease as well. Repeating this gives a policy  $\phi'$  for which  $\sup_i d'_i < 2n + M$ , and for which the total costs are smaller, and thus  $C(\phi') \leq C(\phi)$ . If  $\phi$  is optimal, then so is  $\phi'$ , with  $\sup_i d'_i < \infty$ .

For the full information model this result means that there exists an optimal policy that only uses states  $x$  in the full information model for which  $x_a \leq 2n + M$ . Repeating this for all servers shows that there is an optimal policy that only visits a finite number of states. Thus there are one or more states that are visited infinitely often. Because the optimal policy can be chosen to be stationary and because the state transitions are deterministic the policy repeats the same sequence infinitely often.  $\square$

## 4 Numerical experiments

In this section we report on the numerical experiments. Theorem 3.5 shows us that the state space to be considered is finite. However, the bounds are so big that numerical solution is still impossible; therefore we consider another method for restricting the size of the state space. This concerns upper and lower bound models, which already give equal optimal policies for reasonable state space sizes.

**Upper and lower bounds** We define, for some integer  $B$ , the upper and lower bound models  $(\mathcal{X}^u, \mathcal{A}^u, p^u, c^u)$  and  $(\mathcal{X}^l, \mathcal{A}^l, p^l, c^l)$  as follows: the state and action spaces and the transition are equal to those of the full observation problem  $(\tilde{\mathcal{X}}, \tilde{\mathcal{A}}, \tilde{p}, \tilde{c})$ , but the costs are given by:

$$c^u(x, a) = q_a^{\min\{x_a, B\}}; \quad c^l(x, a) = \begin{cases} q_a^{x_a} & \text{if } x_a < B; \\ 0 & \text{otherwise.} \end{cases}$$

As the same policies are allowable under all three models and  $c^l(x, a) \leq \tilde{c}(x, a) \leq c^u(x, a)$  for all  $(x, a)$ , it is clear that it concerns indeed upper and lower bound models, and that if some policy is optimal for the upper and the lower bound, both having the same costs, then this policy is optimal as well for the original problem.

We change the upper and lower bound model as follows. Using for example dynamic programming we can show that the value in states such as  $x$  and  $x + e_m$  (with  $e_m$  the  $m$ th unit vector) is equal if  $x_m \geq B$ . Therefore we can restrict  $\mathcal{X}^u$  and  $\mathcal{X}^l$  to  $\{1, \dots, B\}^M$  without changing optimal policies and values, by changing  $p$  such that a state component that arrives at  $B$  rests at  $B$  unless a customer is assigned to the corresponding server. The rest of this section is devoted to the numerical results.

**Value iteration** The results presented below are computed using value iteration. The current problem (i.e., the upper and lower bounds) has an interesting structure: it can be shown to be *weakly communicating* (Puterman [8], p. 348). The set of states that is accessible from any other state consists of all states such that one of the components is equal to 1, and no two components are equal (unless they are equal to  $B$ ). It is easy to construct an assignment sequence that leads us into such a state in a finite number of steps. All other states are transient. Then, as stated in [8], p. 483, value iteration converges, given that the problem is aperiodic. This is not the case, for example a policy that repeats 12 with states (1, 2) and (2, 1) has period 2. Therefore we apply the standard aperiodicity transformation (described in [8], p. 371) to make the system aperiodic.

The numerical results presented here are mainly for  $M = 3$ . We start with interarrival times that are exponential (with parameter  $\lambda$ ). The state space bound  $B$  is chosen such that the upper and lower bounds are equal, giving the optimal policy for the original countable state model. For most models it was sufficient to take  $B = 5$  or 10, but sometimes, when the optimal sequence was found to be long, we had to take  $B \approx 20$ . Note that for exponential interarrival times  $q_m = \lambda/(\lambda + \mu_m)$ . We compare the optimal policies with myopic and Bernoulli policies. We introduce these first.

**Myopic policies** A policy is called *myopic* if it takes in every state the action that minimizes the direct costs. Thus for the current model the myopic policy is individually optimal, it minimizes for each arrival individually its blocking probability. The socially optimal policy, which is by definition the average optimal policy, has also to take into account the effect of an assignment on customers that arrive later. This results in sending more customers to the faster queues. We see this effect in Table 1 below.

**Bernoulli policies** A Bernoulli policy is a policy that assigns each customer to the servers using the same probability distribution. The class of Bernoulli policies is not a subclass of  $\tilde{\Phi}$ , as we did not allow for randomization. However, in Remark 2.3 we argued that there exists an optimal deterministic policy, and therefore a Bernoulli policy cannot perform better than the optimal assignment policy for the class  $\tilde{\Phi}$ .

Let us try to find the optimal Bernoulli policy for  $M = 2$  and exponential interarrival times. A single server with arrival rate  $\lambda$  and service rate  $\mu$  has a stationary blocking probability equal to  $\lambda/(\lambda + \mu)$ . If  $f$  is the fraction of arrivals that is assigned to server 1, then it follows that the average costs are equal to

$$\frac{\lambda f^2}{\mu_1 + \lambda f} + \frac{\lambda(1-f)^2}{\mu_2 + \lambda(1-f)}.$$

Standard calculations (using Maple) showed that the minimum is attained at  $f = \mu_1/(\mu_1 + \mu_2)$ , with value  $\lambda/(\lambda + \mu_1 + \mu_2)$ . We see that for  $M = 2$  the optimal Bernoulli policy performs as a single server with parameters  $\lambda$  and  $\mu_1 + \mu_2$ . Repetition of this argument shows that for  $M$  arbitrary the optimal Bernoulli policy assigns a fraction  $\mu_m/\sum_i \mu_i$  to server  $m$ . This policy has value  $\lambda/(\lambda + \sum_i \mu_i)$ , the blocking probability of a single server with service rate  $\sum_i \mu_i$ . This leads to the surprising result that the optimal assignment policy from the set  $\Phi$ , for arbitrary  $M$ , performs better than a system with one server having the joint service capacity.

**Two servers** For two servers we naturally find the structure proven in the previous section. For example, for  $\lambda = 1$ ,  $\mu_1 = 1$  and  $\mu_2 = 5$  the optimal policy repeats 1222 with costs 0.105903, the myopic policy repeats 122 with costs 0.106481, and the optimal Bernoulli policy, which sends 1 out of 6 customers to server 1, has value 0.142857. This and other experiments show that the myopic policy performs close to optimal, while the optimal Bernoulli policy performs much worse.

**Three servers** In Table 1 there are some results for  $M = 3$ . Note that for many of the policies there are equivalent sequences having the same costs but disjoint states: for example 123 is equivalent to 132, but the sets of the recurrent states are  $\{(1, 3, 2), (2, 1, 3), (3, 2, 1)\}$  and  $\{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}$ , respectively. We see for several sequences the properties proven in Theorem 3.3.

parameters				optimal policy		myopic policy		Bernoulli policy
$\lambda$	$\mu_1$	$\mu_2$	$\mu_3$	sequence	value	sequence	value	value
1	1	1	1	132	0.125000	132	0.125000	0.250000
1	1	1	2	1323	0.086806	1323	0.086806	0.200000
1	1	1	10	13323	0.033988	1323	0.035382	0.076923
1	1	4	4	123232132323	0.025271	13232	0.025450	0.100000
1	1	4	7	132323	0.017350	12323	0.019366	0.076923
10	1	1	10	1323333333	0.427109	13233333	0.429127	0.454545
10	1	4	4	132323232	0.468243	1232323213232323	0.468299	0.526316
10	1	4	7	1323232323	0.390657	13232323	0.391413	0.454545

Table 1. Results for  $M = 3$  and exponential interarrival times.

**Constant interarrival times** Finally we compute the optimal policy for constant interarrival times. It is easily seen that in this case  $q_m = \exp(-\mu_m \mathbb{E}S)$ , which is the probability of 0 events of a Poisson process with rate  $\mu_m S$ , where  $S$  is the length of the constant interarrival time. The optimal sequences and policies can be found in Table 2. The impact of the change of interarrival time distribution is remarkable, we see that in all cases the value is lower for constant interarrival times. This is not surprising: it is easy to show that  $\mathbb{P}(s < E(\mu)) < \mathbb{P}(E(1/s) < E(\mu))$ , and thus for the same policy the costs are lower at each epoch under constant interarrival times, while the expected interarrival times are equal.

parameters				$S$ exponential		$S$ constant	
$\mathbb{E}S$	$\mu_1$	$\mu_2$	$\mu_3$	sequence	value	sequence	value
1	1	1		12	0.250000	12	0.135335
1	1	2		12	0.180555	122	0.067813
1	1	3		122	0.145833	1222	0.030092
1	1	5		1222	0.105903	122222	0.004913
1	1	1	1	132	0.125000	132	0.049787
1	1	1	2	1323	0.086806	1323	0.018315
1	1	1	10	13323	0.033988	133333333323	0.000031
1	1	4	4	123232132323	0.025271	132323232	0.000239
1	1	4	7	132323	0.017350	1323232323	0.000105
0.1	1	1	1	132	0.751315	132	0.740818
0.1	1	1	10	1323333333	0.427109	13333333333323	0.317333

Table 2. Results for constant and exponential interarrival times.

**More than three servers** We did not do computations for more than three servers. Theoretically this is well possible, but run times grow fast. Indeed, per iteration a new value has to be computed for each state. Because the number of states is equal to  $B^M$ , the computation time per iteration is exponential in the number of servers. As the number of iterations tend to increase with the size of the state space, we conclude the same for the total run time. For moderate  $B$  a value of  $M = 5$  should be possible. Our experience is that convergence is slow for long cycle lengths.

**Acknowledgement** I would like to thank both referees and Sandjai Bhulai for their helpful comments.

## References

- [1] M.B. Combé and O.J. Boxma. Optimization of static traffic allocation policies. *Theoretical Computer Science*, 125:17–43, 1994.
- [2] B. Hajek. Extremal splitting of point processes. *Mathematics of Operations Research*, 10:543–556, 1985.
- [3] A. Hordijk, G.M. Koole, and J.A. Loeve. Analysis of a customer assignment model with no state information. *Probability in the Engineering and Informational Sciences*, 8:419–429, 1994.

- [4] G.M. Koole. *Stochastic Scheduling and Dynamic Programming*. CWI, Amsterdam, 1995. CWI Tract 113.
- [5] G.M. Koole. A transformation method for stochastic control problems with partial observations. Technical Report BS-R9601, CWI, Amsterdam, 1996. Electronically available as [www.cs.vu.nl/~koole/papers/BSR9601.ps](http://www.cs.vu.nl/~koole/papers/BSR9601.ps).
- [6] P.R. Kumar and P. Varaiya. *Stochastic Systems*. Prentice-Hall, 1986.
- [7] Z. Liu and D. Towsley. Optimality of the round-robin routing policy. *Journal of Applied Probability*, 31:466–475, 1994.
- [8] M.L. Puterman. *Markov Decision Processes*. Wiley, 1994.
- [9] Z. Rosberg and D. Towsley. Customer routing to parallel servers with different rates. *IEEE Transactions on Automatic Control*, 30:1140–1143, 1985.
- [10] S.M. Ross. *Introduction to Stochastic Dynamic Programming*. Academic Press, 1983.