Baltzer Journals April 24, 1996

Non zero-sum stochastic games in admission, service and routing control in queueing systems

Eitan ALTMAN[1]

[1] *INRIA, B.P. 93*
2004 Route de Lucioles
06902 Sophia Antipolis Cedex
France
E-mail: altman@sophia.inria.fr

The purpose of this paper is to investigate situations of non-cooperative dynamic control of queueing systems by two agents, having different objectives. The main part of the paper is devoted to analyzing a problem of an admission and a service (vacation) control. The admission controller has to decide whether to allow arrivals to occur. Once the queue empties, the server goes on vacation, and controls the vacations duration (according to the state and past history of the queue). The immediate costs for each controller are increasing in the number of customers, but no convexity assumptions are made. The controllers are shown to have a stationary equilibrium policy pair, for which each controller uses a stationary threshold type policy with randomization in at most one state. We then investigate a problem of a non-zero sum stochastic game between a router into several queues, and a second controller that allocates some extra service capacity to one of the queues. We establish the equilibrium of a policy pair for which the router uses the intuitive "Join the shortest queue" policy.

Keywords: Non zero-sum stochastic games, control of queueing networks, admission control, service control, routing control.

1 Introduction

The control of queueing system has been an active area of research for more than thirty years. Typical problems that were investigated during these years were arrival control (see e.g. Naor [25], Stidham [31], Yechiali [37]), optimal scheduling of service [20] (for which the μc rule is obtained) and vacation control (Heyman [16], Yadin and Naor [35], Sobel [30]).

An important development in recent years in the field of dynamic control of queueing systems is the growing interest in situations of several controllers and several objectives, giving rise to non-cooperative dynamic game models.

One typical situation of game arises when considering control under some unknown perturbation, and choosing a worst case type criterion. It gives rise to the pessimistic (stochastic) zero-sum game approach, where an imaginary player, called nature, is playing against the controller, and tries to control the perturbed parameter in the worst possible way. Several queueing control problems of this type have been solved recently: flow control problems [1, 2, 4], service control [4, 5], customers' routing problems [3, 4, 6, 8] and server assignment models [4, 6]. Finally, an inventory control problem was solved in [22].

A second type of dynamic game situations in queueing problems arise in problems involving individual optimality. This typically occurs when the objective of an individual may be influenced by the (unknown) decisions that will be taken by future arrivals. Altman and Shimkin [7] analyze such a game arising in a problem of determining whether to use a fast computer which shares its computing capacity between all present users, versus a slow personal computer, that works on a single job at a time, which each user is assumed to possess. When a user wants to run a job, a decision on which facility to use should be taken according to the number of jobs already present in the fast computer. Since the run time of the job on the fast computer is influenced by the decision of future arrivals, a non-cooperative game situation arises. A model quite similar to that was studied by Haviv [15]. Other examples of this type of games are [8, 12, 14]. The solution concept is typically a Nash equilibrium, and one is interested in its existence, uniqueness, characterization, and computation. This type of game is dynamic, in the sense that decisions are state dependent, and the state itself is a stochastic process whose distribution is determined by the policies of all players. On the other hand, in such individual optimality problems, each decision maker makes only one decision, and then disappears. Related models appear also in fields other than queues, see e.g. [23] for an economic model, and [24] for a biological model.

The purpose of this paper is to study a third type of game, arising in cases where several agents control continuously a dynamic system, each of which having a different objective. Very little is known for such problems in queueing networks; the only work in this direction that we are aware of are the problem of flow control with several agents, see [19] and the paper on S-Modular games [36] (which considers static games). We thus formulate simple problems that concern routing, admission control and vacation control, and see how different types of structural results can be obtained in that framework. In our attempt to investigate monotone properties in this type of problems, we were not able to apply standard value iteration techniques, which were used in control of queues (or even in zero-sum stochastic games in queues, see [1, 2, 3]).

We present two queueing models that fall into the last category of non-cooperative stochastic non-zero sum games. We are interested again in the concept of Nash equilibrium, and especially in obtaining equilibria within policies that have a simple structure. Most of the paper is devoted to the study of the following problem with two agents: an admission and a service (vacation) controller, controlling a single queue. The admission controller has to decide whether to allow arrivals to occur. Once the queue empties, the server is forced to go on vacation, and it then controls the vacations duration (according to the state and past history of the queue). More precisely, it has the possibility at each instant to decide whether to continue its vacation, or to go back serving.

An appealing feature that makes the analysis of the above queueing model possible, is the fact that the state space can be represented by a one dimensional variable, rather than a vector.

We assume that the holding costs in the system are shared between the two controllers in the following fair way. The admission controller pays only for holding costs during the period it is active, i.e. during the busy periods of the server, and the server pays some (possibly different) holding cost only when it is active as controller, i.e. during its idle period. The holding costs for each controller are increasing in the number of customers, but no convexity assumptions are made. In addition, the admission controller receives a reward per unit of time that the server is busy. This corresponds to a reward proportional to the throughput. The server has an additional switch-on cost each time it restarts serving.

We consider the expected average cost criterion. Although in this model, the players are active at different epochs, their actions influence each others' objectives, and therefore the model is far from being a trivial extension of similar control problems. Note that even in the case of a single controller having two distinct objectives (e.g. minimizing holding costs and maximizing throughputs, for an admission controller), threshold policies turn out not to be optimal (see [18]).

Here are some intuitive considerations that the service controller might have in designing its control. (i) It may wish to avoid switching-on too frequently, since there is a switch-on cost; this aspect makes it advantageous to take long vacations. (ii) It has an incentive to end its vacation when there are many customers, since the holding cost becomes large. (iii) Since the average cost is considered and not just the cost until the end of a vacation, the service controller might, on the contrary, wish to have a large number of customers accumulate in the queue before ending its vacation, since it does not pay holding costs during its busy period. More precisely, consider a cycle containing a vacation period followed by a busy period. The expected average cost (for the case of stationary policies) equals the total expected cost per cycle divided by the expected cycle duration. The total expected cost per cycle equals the switch-on cost plus the total expected holding cost incurred during the vacation period. Thus, the longer the busy period is, the smaller the average cost will be. The expected cycle duration depends, however, not only on the policy of the server, but also on that of the admission control.

It seems intuitive that the interplay of the admission controller is between accepting more customers, which on one hand results in a higher throughput, which is desirable, and, on the other hand, results in a higher holding cost, which is undesirable. But, here again, the situation is more complicated, since the service controller has also an impact on both (i) the amount of customer found at a beginning of a busy period, and hence on the total expected holding cost that would be paid by the admission controller during a cycle, and (ii) the expected cycle duration.

Stochastic games in which different controllers are active at different time epochs are known as games with perfect information (see [28, 21] for the theory, and [3, 6, 22] for some applications, in the context of zero-sum games). As has been shown by Federgruen in [11] (p. 470), equilibria stationary policies for stochastic non-zero sum games with perfect information require, in general, randomization, and there need not exist stationary deterministic equilibria policies.

We show in our example that both controllers have threshold type equilibrium policies with randomization in at most one state. We use the following approach to establish the existence of threshold type Nash. (i) We first show that each controller has an optimal threshold policy against any threshold policy of the other controller. To establish this result for the admission controller, we use dynamic programming techniques, following a non-standard method by Stidham and Weber [32], which uses some coupling ideas (and seem to be limited to a one dimensional state space). For the service controller, the optimality of threshold policies turns to be straight forward. (ii) We present some properties of the set of optimal threshold policies for each player, which are obtained as responses against some bounded class of threshold policies of the other player. We show that the responses can be chosen so as to remain in those bounded classes. Moreover, we establish continuity, convexity and compactness properties of the set of responses, with respect to the threshold policy of the other controller. This stage turns to be more involved than the first one, and requires, for the service controller, the explicit computation of several performance measures. (iii) A fixed point theorem is applied to conclude the existence of equilibria among threshold policies. (Note that even if for each player, an optimal stationary deterministic threshold policy exists as a response to the other player using a threshold policy, there need not exist pure stationary equilibrium threshold policies; it is essential for the fixed point theorem to hold, to have continuity and convexity properties of the cost with respect to the policies. These properties hold only among the larger class of randomized threshold policies).

We conclude the paper with a second problem of a non-zero sum stochastic game between a router into several queues, and a second controller that allocates some extra service capacity to one of the queues. We establish the equilibrium of a policy pair for which the router uses the intuitive "Join the shortest queue" policy.

2 Control of admission and of service vacations

We consider a discrete time, infinite buffer queue. There is an admission controller who has to decide whether to allow arrivals to occur. Once the queue empties, the second controller, the server, goes on vacation, and controls the vacations duration (according to the state and past history of the queue). We assume that there are two phases of control. The admission controller is active only during the "on" time, i.e. when the server is active. The server may take a control decision only when it is on vacation, on whether to continue the vacation or to pass to the "on" phase (a busy period).

Time is slotted, and is considered at discrete intervals of length Δ, called time slots. At the beginning of time slot t (i.e. at time $t \cdot \Delta$), a customer may arrive with probability λ. During "on" periods, we assume that
(1) the admission controller has the possibility to allow an arrival to occur or not, at the beginning of each time-slot;
(2) if the queue is nonempty at the beginning of the time slot then a service succeeds with probability ν, and then a customer leaves the queue just before the end of the time slot. If service does not succeed, then the customer remains in the queue. Denote

$$\rho = \frac{\lambda}{\nu}, \qquad \tilde{\rho} = \frac{\lambda(1-\nu)}{\nu(1-\lambda)}. \tag{1}$$

During the off phases, we assume that the flow is not controlled, and all arrivals are accepted. The admission controller gets a reward per unit of time that the server is busy. This corresponds to a reward proportional to the throughput. In addition, it has to pay a holding cost, which is increasing in the number of customers in the system. In particular, if a linear holding cost is chosen, it is known to be related to the expected waiting time per customer (see [18]).

Once the system empties, the server controller leaves for a vacation. It may resume serving at the beginning of any time slot, resulting in a switch-on cost. In that case, the admission controller may immediately take an admission decision, at the beginning of the same time slot, and allow or reject arriving customers. The server controller has to pay in addition a holding cost which is an increasing function of the number of customers.

We shall now use the stochastic games formalism in order to describe the control problem, with more details. We consider the tuple $\{\mathcal{X}, \mathbf{A}, \mathbf{B}, \mathcal{P}, c, d\}$, where

- $\mathcal{X} = \{..., -2, -1, 0, 1, 2, ...\}$ denotes the state space, where for $i > 0$, the state i means that there are i customers in the queue and that the server is on, and $i \leq 0$ means that there are $|i|$ customers in the queue and that the server is off. We shall denote $\mathbf{X} = \{1, 2, ...\}$, and $\underline{\mathbf{X}} = \{0, -1, -2, ...\}$.
- $\mathbf{A} = \{0, 1\}$ are the actions of the admission controller, where 0 means rejecting an arriving customer, and 1 means that if an arrival occurs (w.p. λ) then it would be accepted. Both 0 and 1 are available at any state in $\mathcal{X}$. However, if the server is off, then the actions of player 1 have no influence.
- $\mathbf{B} = \{0, 1\}$, where 0 means vacation, and 1 means turning (or remaining) on. At states $\mathbf{X}$, only 1 is available, and at state 0, only 0 is available. The server is thus forced to leave on vacation when the system empties. We assume that both 0 and 1 are available at states $-1 \geq x \geq -L$ where $L \leq \infty$ is some nonnegative integer, and if there are more than L customers then the server must resume service, i.e. only 1 is available for $x < -L$.
- $\mathcal{P}$ are the transition probabilities. We have

$$\mathcal{P}_{xaby} = \begin{cases} \lambda\overline{\nu} & \text{if } y = |x| + 1, a = 1, b = 1 \\ \overline{\lambda}\overline{\nu} + \lambda\nu & \text{if } y = |x|, a = 1, b = 1 \\ \overline{\lambda}\nu & \text{if } y = |x| - 1, a = 1, b = 1 \\ \overline{\nu} & \text{if } y = |x|, a = 0, b = 1 \\ \nu & \text{if } y = |x| - 1, a = 0, b = 1 \\ \lambda & \text{if } x \in \underline{\mathbf{X}}, y = x - 1, a = 1, b = 0 \\ \overline{\lambda} & \text{if } x \in \underline{\mathbf{X}}, y = x, a = 1, b = 0 \end{cases}$$

  Above, for any variable a, $\overline{a}$ stands for $1 - a$.
- The immediate cost for the admission controller is a separable function $c : \mathcal{X} \times \mathbf{A} \to \mathbb{R}$, where $c(x, a) = c_1(x) + c_2(a)$; c_1 is the holding cost, that is zero for $x \leq 0$, and is strictly increasing (not necessarily convex); there exists some $\delta > 0$ such that

$$c_1(x + 1) - c_1(x) \geq \delta. \tag{2}$$

  $c_2(a) = -r1\{a = 1\}$, where r is some positive reward related to admitting customers, and thus to a reward related to the throughput.
- The immediate cost for the service controller is $d(x, b) = d_1(x) + \gamma\{b = 1, x < 0\}$. γ is a switch-on cost. and $d_1 : \mathcal{X} \to \mathbb{R}$ is a holding cost; it is zero at states $x > 0$, and it is strictly decreasing at $x \leq 0$ (i.e., $d(-1) \leq d(-2) \leq d(-3)...$, so that d increases as there are more customers in the system). We require that

$$\overline{\lim_{x \to -\infty}} d_1(x) > \lambda\gamma. \tag{3}$$

We denote by U the class of policies for the admission controller, and by V the class of policies for the server controller. A policy v for the service controller is a sequence $v = (v_0, v_1, v_2, ...)$ where v is a probability measure over $\mathbf{B}$ conditioned on the history of previous actions of both players, previous states, as well as the current state. A policy u for the admission controller is defined similarly.

Any policy pair (u, v), together with an initial state x define a probability measure $P_x^{u,v}$ with respect to which the state process $\{X_t\}_{t\in\mathbb{N}}$ and the action processes $\{A_t\}_{t\in\mathbb{N}}$ and $\{B_t\}_{t\in\mathbb{N}}$ are defined. (Thus, X_t is the state at time t, and A_t and B_t are the actions at time t.) Let $E_x^{u,v}$ denote the expectation operator with respect to that measure.

Let $U(S)$ and $V(S)$ denote the stationary (randomized) strategies for the players. For a stationary policy, say u, we shall sometimes identify u with the probability of choosing action 1. When using stationary policies, we shall sometimes use the notation

$$\mathcal{P}_{xuvy} = \sum_{a\in\mathbf{A}} \sum_{b\in\mathbf{B}} u(a)v(b)\mathcal{P}_{xaby}, \qquad c(x,u) = \sum_{a\in\mathbf{A}} u(a)c(x,a),$$

and define similarly $d(x, v)$.

Next, we introduce the expected average cost functions. Denote

$$C(x,u,v) = \overline{\lim_{t\to\infty}} \frac{1}{t+1} \sum_{s=0}^{t} E_x^{u,v} c(X_s, A_s), \qquad D(x,u,v) = \overline{\lim_{t\to\infty}} \frac{1}{t+1} \sum_{s=0}^{t} E_x^{u,v} d(X_s, B_s),$$
$$C(x,v) = \inf_u C(x,u,v), \qquad D(x,u) = \inf_v C(x,u,v). \tag{4}$$

We assume that both controllers (players) want to minimize their cost. A policy pair (u^*, v^*) is said to be a Nash equilibrium if no player can benefit by deviating unilaterally, i.e.

$$C(x,u^*,v^*) \le C(x,u^*,v), \ \forall v \in V, \tag{5}$$
$$D(x,u^*,v^*) \le C(x,u,v^*), \ \forall u \in U. \tag{6}$$

Let $M(R)$ denote the class of probability measures over a set R, endowed with the weak convergence topology which we denote $G(R)$.

A policy $u \in U$ is called threshold with parameter (K_u, q_u), where $K_u \ge$ is an integer and $0 \le q_u \le 1$, if it is stationary, and it satisfies for $x \in \mathbf{X}$:

$$u(x) = \begin{cases} 0 & \text{if } x > K_u \\ q_u & \text{if } x = K_u \\ 1 & \text{if } x < K_u \end{cases}$$

A policy $v \in V$ is called threshold with parameter (K_v, q_v), where $K_v > 0$ is an integer and $0 \le q_v \le 1$, if it is stationary, and it satisfies

$$v(x) = \begin{cases} 1 & \text{if } L \le x < -K_v \\ q_v & \text{if } x = -K_v \\ 0 & \text{if } 0 > x > -K_v \end{cases}$$

We shall always identify $(K_u, 1) = (K_u + 1, 0)$, and $(K_v, 1) = (K_v + 1, 0)$. We shall often denote a threshold policy u (or v) by its parameters, i.e. (K_u, q_u) will be identified as a policy with the corresponding threshold.

Our main result is

Theorem 1
For arbitrary L (possibly infinite), there exists a stationary Nash (u,v) for both players, such that both u and v are threshold policies.

3 The admission controller

We show in this section that for any stationary policy of the server, the admission controller has an optimal threshold policy. The method of proof follows the ideas of Stidham and Weber in [32], and is related to the approach used by Shimkin [27]. Before proving the Theorem, we need some definitions and Lemmas.

Denote $\tau(y) = \inf\{t : X_t = y\}$, $y = 0, 1, 2, \ldots$. Following Stidham and Weber [32], we define the total g-revised cost as

$$Y^g(x,u) = E_x^u \sum_{t=0}^{\tau(0)-1} [c(X_s, A_s) - g].$$

Further define

$$Y^g(x,y;u) = E_x^u \sum_{t=0}^{\tau(y)-1} [c(X_s, A_s) - g], \tag{7}$$

where $x \in \mathbf{X}$. Define

$$Y^g(x) = \inf_{u \in U(S)} Y^g(x,u), \ x \in \mathbf{X}, \tag{8}$$

$$Y^g(x,y) = \inf_{u \in U(S)} Y^g(x,y;u), \ x \in \mathbf{X}. \tag{9}$$

A stationary policy that achieves the inf in (8) for all $x \in \mathbf{X}$ is called an optimal policy (for the admission controller) for the g-revised problem. A stationary policy that achieves the inf in (9) for all given x and y is called an (x,y) optimal policy (for the admission controller) for the g-revised problem.

It turns out that $Y^g(x,u)$ and $Y^g(x,y;u)$ are well defined since $\lim_{y\to\infty} c_1(y) = \infty$, see Lemma 6 in [32].

We see from the definition of the g-revised cost that they are not functions of the strategy of the other player. This is not the case for the expected average cost, of course. This property makes it appealing to analyze the g-revised problem, and only later to pass to the average cost. We shall first obtain the structure of optimal policies for the g-revised problem. This structure will be shown to hold for any value of g. This fact will enable us to obtain the structure of optimal policies for the expected average cost criterion.

Note that in order to move from state x to state $y < x - 1$, one has to go through state $x - 1$. It then follows from (7) that

$$Y^g(x,u) = \sum_{y=1}^{x} Y^g(y, y-1; u), \qquad Y^g(x,y;u) = Y^g(x,u) - Y^g(y,u). \tag{10}$$

The following Lemma summerizes some properties of the g-revised cost. The proof is the same as in [32] Section 1.2.

Lemma 2
(i) For any $g < \infty$, $h^g(x) = Y^g(x)$ is a solution for the dynamic programming equation:

$$\begin{aligned} h^g(0) &= 0, \\ h^g(x) &= \min_a \left\{ c(x,a) - g + \sum_{y \in \mathbf{X}} P_{xa1y} h^g(y) \right\} \\ &= \min_{u \in M(\mathbf{A})} \left\{ c(x,u) - g + \sum_{y \in \mathbf{X}} P_{xu1y} h^g(y) \right\}, \ x \in \mathbf{X}. \end{aligned} \tag{11}$$

(ii) A stationary policy u is optimal for the g-revised problem ($g < \infty$), i.e. $Y^g(x) = Y^g(x,u)$, if and only if it is conservative, i.e. the support of $u(x)$ are actions minimizing the right hand side of (11) at x (when we substitute $Y^g(x)$ for $h^g(x)$).
(iii) A stationary policy u is optimal for the g-revised problem (with $g < \infty$) if it minimizes $Y^g(y, y-1; u)$ for all $y \in \mathbf{X}$.

Lemma 3
Fix any $g < \infty$. Then
(i) $Y^g(x+1,x)$ is strictly increasing in x: $Y^g(x+1,x) - Y^g(x,x-1) \geq \delta/\nu$ (where δ is a bound defined in (2)).
(ii) There exists an optimal stationary policy u for the admission controller in the g-revised problem, that chooses action 0 at $x \geq \mathcal{K}_1^*(g)$, where

$$\mathcal{K}_1^*(g) = \inf \left\{ x : \nu c_2(1) + \lambda \left(c_1(x) + c_2(1) - g \right) > 0 \right\}. \tag{12}$$

(iii) Any stationary policy u for the admission controller, such that the support of $u(x)$ is among the actions that minimize the right hand side of (11) when substituting $h^g(x) = Y^g(x)$, is a threshold policy and is optimal for the g-revised problem. Moreover, $K_u < \mathcal{K}_1^*(g)$.

Proof
We begin by proving (i). For any stationary policy u, define the shifted policy $\theta(u)$ as the one that uses $u(x+1)$ at state x. Let u^* be an optimal $(x+1,x)$ stationary policy. Then,

$$\begin{aligned} & Y^g(x+1,x) - Y^g(x,x-1) \geq Y^g(x+1,x) - Y^g(x,x-1;\theta(u^*)) \\ &= E_{x+1}^{u^*} \sum_{t=0}^{\tau(x)-1} [c(X_s,A_s) - g] - E_x^{\theta(u^*)} \sum_{t=0}^{\tau(x-1)-1} [c(X_s,A_s) - g] \\ &= E_{x+1}^{u^*} \sum_{t=0}^{\tau(x)-1} [c(X_s,A_s) - c(X_s - 1, A_s)] \\ &\geq \delta E_{x+1}^{u^*} \tau(x) \geq \delta/\nu. \end{aligned}$$

Next, we prove (ii). Let x be such that $x \geq \mathcal{K}_1^*(g)$ and assume that u^* is conservative for (11) (and thus optimal). The following follows by (10):

$$
\begin{aligned}
Y^g(x,x-1) &= \min_{u \in U(S)} Y^g(x,x-1;u) \\
&= -g + c_1(x) + \min_a \{\overline{\nu} Y^g(x,x-1), \\
&\quad c_2(1) + \lambda\overline{\nu}(Y^g(x+1,x) + Y^g(x,x-1)) + (\overline{\lambda}\overline{\nu} + \lambda\nu)Y^g(x,x-1)\} \\
&= -g + c_1(x) + \overline{\nu} Y^g(x,x-1) \\
&\quad + \min\{0, c_2(1) + \lambda(\overline{\nu} Y^g(x+1,x) + \nu Y^g(x,x-1)\} \qquad (13) \\
&= -g + c_1(x) + \overline{\nu} Y^g(x,x-1) + \qquad (14) \\
&\quad \min\left\{0, c_2(1) + \lambda\left(\overline{\nu} E_{x+1}^{u^*} \sum_{t=0}^{\tau(x)-1} [c(X_s,A_s) - g] + \nu E_x^{u^*} \sum_{t=0}^{\tau(x-1)-1} [c(X_s,A_s) - g]\right)\right\}
\end{aligned}
$$

Denote by $Z(x)$ the second term in the minimization. Note that (12) implies that $c_2(1) - g + c_1(x) > 0$, and recall that $c_2(1) < 0$. Then

$$
\begin{aligned}
Z(x) &\geq c_2(1) \\
&\quad + \lambda\left(\overline{\nu} E_{x+1}^{u^*} \sum_{t=0}^{\tau(x)-1} [c_1(x) + c_2(1) - g] + \nu E_x^{u^*} \sum_{t=0}^{\tau(x-1)-1} [c_1(x) + c_2(1) - g]\right) \\
&= c_2(1) + \lambda\left(c_1(x) + c_2(1) - g\right)\left(\overline{\nu} E_{x+1}^{u^*} \tau(x) + \nu E_x^{u^*} \tau(x-1)\right) \\
&\geq c_2(1) + \lambda\left(c_1(x) + c_2(1) - g\right)/\nu > 0. \qquad (15)
\end{aligned}
$$

It follows from Lemma 2 (i),(iii) that an action a is optimal at state x if and only if a corresponds to the minimum in (13) (i.e., $a = 0$ is optimal if and only if the first term in the minimization in (13) is smaller than or equal to the second term; $a = 1$ is optimal if and only if the second term in the minimization in (13) is smaller than or equal to the first one). This, together with (15) establishes (ii).

Next we prove (iii). Assume that u is conservative but not threshold. Then there exists some state x for which (a) u uses action 0 (rejects) with strictly positive probability, and (b) u uses action 1 with strictly positive probability at state $x+1$. From (a) and (13) it follows that $0 \leq Z(x)$. From (b) and (13) it follows that $0 \geq Z(x+1)$, so we conclude that $Z(x) \geq Z(x+1)$. On the other hand, it follows from (i) and from (13) that $Z(x) < Z(x+1)$. This contradiction thus concludes (iii). $\square$

Theorem 4
Assume that v is a threshold policy for the server with a finite threshold. Then
(i) $C(x,v) \leq c_1(K_v + 1) < \infty$.
(ii) Any stationary policy u for the admission controller which is conservative for (11) (i.e. for all $x \in \mathbf{X}$, the support of $u(x)$ contains only actions minimizing the right hand side of (11) at x), where $g = C(x,v)$, and $h^g(x) = Y^g(x)$, is expected average optimal against v, and is a threshold policy. Moreover, the parameter K_u of any such policy satisfies

$$
K_u < \mathcal{K}_1^*(C(x,v)) \leq \inf\left\{x : \nu c_2(1) + \lambda\left(c_1(x) + c_2(1) - c_1(K_v+1)\right) > 0\right\}. \qquad (16)
$$

Proof
(i) The bound is obtained since $C(x,v) \le C(x,u_0,v)$, where u_0 is the policy that always rejects customers; since all states greater than K_v+1 are transient under (u_0,v), we have for all $t > T$, where T is some random variable with finite expectation, $X_t \le K_v+1$, $P_x^{u_0,v}$ a.s. so that $c(X_t,A_t) \le c_1(K_v+1)$, $P_x^{u_0,v}$ a.s. This implies (i).

Let u be a stationary policy as in (ii). It follows from Lemma 3 (iii) that it is threshold and, together with (i), that (16) holds. It remains to show that it is average optimal. Consider the dynamic programming equation:

$$\begin{aligned} h(0)=0, \quad h(x) &= \min_a \left\{ c(x,a) - g + \sum_{y\in\mathcal{X}} \mathcal{P}_{xavy} h(y) \right\} \\ &= \min_{u(x)\in M(\mathbf{A})} \left\{ c(x,u(x)) - g + \sum_{y\in\mathcal{X}} \mathcal{P}_{x,u(x),v,y} h(y) \right\}, \quad x \in \mathcal{X} \end{aligned} \quad (17)$$

where $h:\mathcal{X}\to\mathbb{R}, \mathrm{g}\in\mathbb{R}$. We shall show that it has a unique solution (h^*,g^*) which satisfies

$$g^* = C(x,v), \quad h(x) = Y^g(x) \text{ for } x\in\mathbf{X}. \quad (18)$$

Let $h(x)=Y^g(x)$. (We do not assume yet that $g=C(x,v)$.) Assume that the threshold policy u^* is conservative for (11) when $h^g(x)=Y^g(x)$. Lemma 2 (ii) implies that $h^g(x) = Y^g(x,u^*)$, so that $h(x)=Y^g(x,u^*)$. For simplicity of presentation, we shall assume that the threshold of policy v satisfies $q_v=0$. The same type of arguments hold for general q_v. Denote $T^u(x) := E_x^u\tau(0)$ the expected time until the system reaches state 0 starting at state $x\in\mathbf{X}$, when the admission controller uses policy u. Then for $0\ge x > -K_v$ we have by (17)

$$h(x) = -g + \overline{\lambda} h(x) + \lambda h(x-1)$$

from which we conclude that

$$h(x) = -\frac{xg}{\lambda}, \qquad 0 \ge x \ge -K_v. \quad (19)$$

It follows again from (17) that

$$\begin{aligned} h(-K_v) &= -g+\gamma+\lambda\overline{\nu} h(K_v+1) + (\lambda\nu+\overline{\lambda}\overline{\nu})h(K_v) + \overline{\lambda}\nu h(K_v-1) \\ &= -g+\gamma+\lambda\overline{\nu} Y^g(K_v+1,u^*) + (\lambda\nu+\overline{\lambda}\overline{\nu})Y^g(K_v,u^*) + \overline{\lambda}\nu Y^g(K_v-1,u^*) \\ &= -g+\gamma+\lambda\overline{\nu} Y^0(K_v+1,u^*) + (\lambda\nu+\overline{\lambda}\overline{\nu})Y^0(K_v,u^*) + \overline{\lambda}\nu Y^0(K_v-1,u^*) \\ &\quad -g\left(\lambda\overline{\nu} T^{u^*}(K_v+1) + (\lambda\nu+\overline{\lambda}\overline{\nu})T^{u^*}(K_v) + \overline{\lambda}\nu T^{u^*}(K_v-1)\right). \end{aligned}$$

Combining this with (19), we obtain

$$g = \frac{\gamma+\lambda\overline{\nu} Y^0(K_v+1,u^*) + (\lambda\nu+\overline{\lambda}\overline{\nu})Y^0(K_v,u^*) + \overline{\lambda}\nu Y^0(K_v-1,u^*)}{K_v+1+\left(\lambda\overline{\nu} T^{u^*}(K_v+1) + (\lambda\nu+\overline{\lambda}\overline{\nu})T^{u^*}(K_v) + \overline{\lambda}\nu T^{u^*}(K_v-1)\right)}.$$

Finally, for $x < -K_v$, we have

$$h(x) = -g+\gamma+\lambda\overline{\nu} h(-x+1) + (\lambda\nu+\overline{\lambda}\overline{\nu})h(-x) + \overline{\lambda}\nu h(-x-1).$$

We thus obtained explicitly, the uniques solution of (17) that satisfies $h(x) = Y^g(x)$. Next, we show that $g = C(x,v)$. Iterating (17) we get

$$\begin{aligned} h(x) &= c(x,u^*) - g + E_x^{u^*,v} h(X_1) \\ &= c(x,u^*) - g + E_x^{u^*,v} c(X_1,u^*) - g + E_x^{u^*,v} h(X_2) \\ &= \sum_{j=0}^{n-1} E_x^{u^*,v} c(X_j,u^*) - ng + E_x^{u^*,v} h(X_n). \end{aligned}$$

Since $X_n \le \max(x, K_u + 1), P^{u^*,v} a.s.$, it follows that $E_x^{u^*,v} h(X_n) = o(n)$. Hence, when dividing by n and taking the limit as $n \to \infty$, we get $C(x,u^*,v) = g$. We shall now use a similar procedure to show that u^* is optimal. It turns out that our model satisfies the Weber-Stidham conditions [34], which implies the existence of an optimal stationary policy u with finite expected average cost, for which the Markov chain has one positive recurrent class. This implies in particular that the expected time between two consecutive visits to state zero, which we denote by $T_{u,v}^{00}$, is finite. Let $R_{u,v}^{00}$ be the total expected cost between two consecutive visits of state zero. Then it follows from the reward renewal theorem [26] that

$$C(x,u,v) = \frac{R_{u,v}^{00}}{T_{u,v}^{00}}.$$

Since $C(x,u,v)$ and $T_{u,v}^{00}$ are finite, so is $R_{u,v}^{00}$. It follows by iterating (17) that

$$\begin{aligned} h(x) &\le c(x,u) - g + E_x^{u,v} h(X_1) \\ &\le c(x,u) - g + E_x^{u,v} c(X_1,u) - g + E_x^{u,v} h(X_2) \\ &\le \sum_{j=0}^{n-1} E_x^{u,v} c(X_j,u) - ng + E_x^{u,v} h(X_n) \\ &= \sum_{j=0}^{n-1} E_x^{u,v} c(X_j,u) - ng + E_x^{u,v} Y^g(X_n,u^*) \\ &\le \sum_{j=0}^{n-1} E_x^{u,v} c(X_j,u) - ng + E_x^{u,v} Y^g(X_n,u). \end{aligned}$$

The last inequality follows since, by definition of u^*, it minimizes $Y^g(x,u)$ for all u. Since $R_{u,v}^{00} < \infty$, it follows from [26] p. 79 that $E_x^u Y^g(X_n,u) = o(n)$. Thus dividing again by n and taking the liminf as $n \to \infty$, we obtain $g \le C(x,u,v)$. Since we have already shown that $g = C(x,u^*,v)$, this implies that u^* is expected average optimal against v. $\square$

Remark 1
The last part of the proof could serve as an alternative way of deriving sufficient conditions for existence of stationary policies for countable Markov chains with unbounded cost; the standard way, being through the discounted cost. Our conditions are very similar to those of Cavazos-Cadena [9], the difference being that in [9], it is required that a stationary policy should exist with finite expected average cost, for which the whole Markov chain is irreducible and ergodic; we did not have to check that condition above, and used in fact a

stationary policy, candidate for being optimal, which has finite expected average cost and for which all but a finite number of states are transient.

4 The controller of service

We establish in this section results analogous to those of the previous section, this time for the service controller.

Lemma 5
For any threshold policy u with finite threshold of the admission controller, the service controller has an optimal threshold policy.

Proof
The proof is straight forward. Against any threshold policy u with finite threshold, the service controller is faced with a Markov Decision Process which is easily seen to satisfy the Weber-Stidham conditions [34]. We conclude that it has an optimal stationary deterministic policy v. Let v be such an optimal deterministic stationary policy, and assume it is not a threshold policy. Let $x \in \underline{\mathbf{X}}$ be given by:

$$x = \max\{y \in \underline{\mathbf{X}} : v(y) = 1\}$$

(where above we understand $\max\{\emptyset\} := 0$). This means that if the number of customers is smaller than $|x|$, then v takes a vacation, and at state x it serves. Then all states for which the number of customers (during a vacation) is larger than $|x|$ (i.e. all $y \in \underline{\mathbf{X}}$ such that $y < x$) become transient. Hence, the policy that coincides with v for $y \geq x$ and uses $v(y) = 0$ for $y < x$ has the same expected average cost as the policy v, and is thus optimal. This implies the Lemma. $\square$

The above result is not sufficient to conclude that a threshold Nash exist; we still need some continuity, compactness and convexity properties in order to apply some fixed point theorem. We thus further investigate some properties of optimal threshold policies. In addition to convexity and continuity properties, we establish that the value of optimal thresholds K_v of the service controller, when the admission controller restricts to threshold policies, are upper bounded by some constant $\mathcal{K}_2^*$, uniformly in all threshold of the admission controller.

Denote $T^u(x) := E_x^u \tau(0)$ the expected time until the system reaches state 0 starting at state $x \in \mathbf{X}$, when the admission controller uses policy u.

Lemma 6
Assume that the admission controller uses the threshold policy u with parameter (K_u, q_u). Then $T^u(x)$ is continuous in (K_u, q_u), and concave increasing in x.

Proof
We have $T^u(x) =$

$$\begin{cases} 0 & x = 0 \\ 1 + \lambda\overline{\nu}T^u(x+1) + (\lambda\nu + \overline{\lambda}\overline{\nu})T^u(x) + \overline{\lambda}\nu T^u(x-1) & 0 < x < K_u \\ 1 + q_u\lambda\overline{\nu}T^u(x+1) + (q_u\lambda\nu + \overline{(q_u\lambda)}\overline{\nu})T^u(x) + \overline{(q_u\lambda)}\nu T^u(x-1) & x = K_u \\ 1 + \overline{\nu}T^u(x) + \nu T^u(x-1) & x > K_u \end{cases}$$

From the fourth relation above we obtain

$$T^u(x) = T^u(K_u) + \frac{x - K_u}{\nu}, \qquad X > K_u$$

and from the third equality we get

$$T^u(K_u) = 1 + q_u\lambda\overline{\nu}[T^u(K_u) + \nu^{-1}] + (q_u\lambda\nu + \overline{(q_u\lambda)}\overline{\nu})T^u(K_u) + \overline{(q_u\lambda)}\nu T^u(K_u - 1)$$

so that

$$T^u(K_u) = T^u(K_u - 1) + \frac{1 + q_u\rho\overline{\nu}}{\overline{(q_u\lambda)}\nu}. \tag{20}$$

The second equality is a homogeneous difference equation, and when taking into account the boundary condition $T^u(0) = 0$, its solution for $x \le K_u$ is

$$T^u(x) = \begin{cases} c_u\left(\tilde{\rho}^{-x} - 1\right) + \dfrac{x}{\nu\overline{\lambda} - \lambda\overline{\nu}} & \rho \neq 1 \\ -\dfrac{x^2}{2\lambda\overline{\nu}} + b_u x & \rho = 1 \end{cases} \tag{21}$$

(where $\tilde{\rho}$ is given in (1)). Thus $T^u(x)$ is concave increasing in x for $\rho = 1$. We show that it is also concave increasing for $\rho \neq 1$. Combining (20) and (21), we get,

$$\begin{aligned} c_u &= \left(\frac{1 + q_u\overline{\nu}\rho}{\overline{(q_u\lambda)}\nu} - \frac{1}{\nu\overline{\lambda} - \lambda\overline{\nu}}\right)\frac{\tilde{\rho}^{K_u}}{1 - \tilde{\rho}}, \\ b_u &= \frac{1 + q_u\rho\overline{\nu}}{\overline{(q_u\lambda)}\nu} + 1 - \frac{1}{\lambda\overline{\nu}}. \end{aligned}$$

c_u is negative for any $\rho \neq 1$. Indeed, for $\rho < 1$ we have

$$c_u \le \left(\frac{1}{\overline{\lambda}\nu} - \frac{1}{\nu\overline{\lambda} - \lambda\overline{\nu}}\right)\frac{\tilde{\rho}^{K_u}}{1 - \tilde{\rho}} < 0.$$

For $\rho > 0$ it is straight forward to see that $c_u < 0$. Hence $T_x(u)$ is concave increasing for any ρ. The continuity follows immediately fro (21). $\square$

Let $\mathcal{T}_{K_v}(x)$ and $\mathcal{D}_{K_v}(x)$ denote the total expected time and total expected cost (for the server controller), from the moment that the system is at state $x \in \underline{\mathbf{X}}$, till the server switches on (not including that moment), given that a threshold policy v is used with parameter $(K_v, 1)$. Then,

Lemma 7
$\mathcal{T}_{K_v}(0)$ is linear in K_v, and $\mathcal{D}_{K_v}(0)$ is strictly convex in K_v.

Proof

$$\mathcal{T}_{K_v}(x) = \begin{cases} 1 + \lambda \mathcal{T}_{K_v}(x-1) + \overline{\lambda} \mathcal{T}_{K_v}(x) & x < K_u \\ 0 & x \geq K_u \end{cases}$$

$$\mathcal{D}_{K_v}(x) = \begin{cases} d_1(x) + \lambda \mathcal{D}_{K_v}(x-1) + \overline{\lambda} \mathcal{D}_{K_v}(x) & x < K_u \\ 0 & x \geq K_u \end{cases}$$

Hence,

$$\mathcal{T}_{K_v}(0) = \frac{K_v}{\lambda}, \qquad \mathcal{D}_{K_v}(0) = \frac{\sum_{i=0}^{K_v-1} d_1(i)}{\lambda}.$$

We conclude that $\mathcal{T}_{K_v}(0)$ is indeed linear in K_v, and

$$\mathcal{D}_{K_v+1}(0) - \mathcal{D}_{K_v}(0) = \frac{d_1(K_v)}{\lambda}$$

which is strictly increasing, since d_1 is strictly increasing. This establishes the convexity of $\mathcal{D}_{K_v}(0)$. ☐

Lemma 8
(i) Assume that u is threshold with parameter (K_u, q_u) and that v is threshold with parameter $(K_v, 1)$, where K_v is finite. Then $D(x, (K_u, q_u), (K_v, 1))$ is continuous in (K_u, q_u) and does not depend on x. Moreover, $D(x, (K_u, q_u), (K_v, 1))$ is strictly integer convex in K_v.
(ii) Define

$$\mathcal{K}_2^* := \inf \left\{ k : \frac{\sum_{y=0}^{k-1} d_1(-y)}{\lambda} > k\gamma \right\} = \sup \left\{ k : \frac{\mathcal{D}_k(0)}{\lambda} > k\gamma \right\}. \tag{22}$$

($\mathcal{K}_2^*$ is finite due to (3)). Then, against any threshold policy u of the admission controller, any optimal threshold policy of the server has $K_v \leq \mathcal{K}_2^*$.

Proof
(i) We have by the renewal reward theorem [26]

$$D(x, u, (K_v, 1)) = \frac{\mathcal{D}_{K_v}(0) + \gamma}{\mathcal{T}_{K_v}(0) + T^u(K_v)} \tag{23}$$

which is independent of x. It then follows from Lemma 6 and 7 that $D(x, u, (K_v, 1))$ is integer convex in K_v. The only expression in (23) that depends on u is $T^u(K_v)$ which is continuous in (K_u, q_u). This implies the continuity of $D(x, u, v)$.
(ii) Because of the concavity of the denominator of (23), it follows that for $k \geq \mathcal{K}_2^*$,

$$D(x, u, (k, 1)) = \frac{\mathcal{D}_k(0) + \gamma}{\mathcal{T}_k(0) + T^u(k)} > \frac{(k+1)\gamma}{(k+1)[\mathcal{T}_1(0) + T^u(1)]} = D(x, u, (1, 1)).$$

Hence, the cost for the server when using the threshold policy $(k, 1)$ is strictly higher than the policy $(1, 1)$, for all $k \geq \mathcal{K}_2^*$. This establishes (ii). $\square$

5 Proof of the main result

The main result will follow from the following Lemma, which provides an extension of Kakutani fixed point theorem which was obtained by Glicksberg [13] and Fan [10]. We say that a point to set mapping $\Phi : \mathcal{U} \to 2^{\mathcal{U}}$ is upper semi continuous, if for any sequence $\zeta(n) \in \mathcal{U}$ converging to some $\zeta \in \mathcal{U}$, and any $\xi(n) \in \Phi(\zeta(n))$ such that $\xi(n)$ converges to some ξ, we have $\xi \in \Phi(\zeta)$.

Lemma 9
Given an upper semi-continuous point to convex set mapping $\Phi : \mathcal{U} \to 2^{\mathcal{U}}$, defined on a convex compact subset U of a linear Hausdorff locally convex topological space, there exists a point $x \in \Phi(x)$.

Define
$$\mathcal{K}_1^* = \inf \left\{ x : \nu c_2(1) + \lambda \left(c_1(x) + c_2(1) - c_1(\mathcal{K}_2^* + 1) \right) > 0 \right\}.$$
It follows from Theorem 4, that if the server controller uses a threshold policy $v = (K_v, q_v)$ with $K_v \leq \min\{\mathcal{K}_2^*, L\}$, then the admission controller has a nonempty set of threshold policies that are optimal against that v and are conservative in (11), which we denote by $\phi(K_v, q_v)$, all of which have thresholds no larger than $\mathcal{K}_1^*$.

On the other hand, it follows from Lemma 5 and 8, that against any threshold policy $u = (K_u, q_u)$ with $K_u \leq \mathcal{K}_1^*$ (in fact against any threshold policy) of the admission controller, the service controller has a nonempty set of threshold policies, which we denote by $\psi(K_u, q_u)$, all of which have thresholds no larger than $\min\{\mathcal{K}_2^*, L\}$.

Denote $\mathcal{A} = \{0, ..., \mathcal{K}_1^*\} \times [0, 1]$, $\mathcal{B} = \{1, ..., \min(\mathcal{K}_2^*, L)\} \times [0, 1]$. We identify $\mathcal{A}$ with threshold policies of the admission controller, and $\mathcal{B}$ with threshold policies of the service controller.

Remark 2
If we restrict to $u \in \mathcal{A}$, $v \in \mathcal{B}$, then all states which are not in
$$\overline{\mathbf{X}} = \{-1 - \min(\mathcal{K}_2^*, L), ..., 0, 1, ..., \mathcal{K}_1^* + 1\}$$
are transient. Moreover, it takes finite expected time and finite expected cost to reach $\overline{\mathbf{X}}$ from any state. Hence the stochastic game behaves as if it is reduced to a finite state space.

Lemma 10
$\phi : \mathcal{B} \to 2^{\mathcal{A}}$ and $\psi : \mathcal{A} \to 2^{\mathcal{B}}$ are convex and compact valued and upper semi continuous.

Proof
We restrict to $u \in \mathcal{A}$, $v \in \mathcal{B}$. We first show that the solution (g^*, h^*) of (11) are continuous in v. It follows from Remark 2 that $C(x, u, v)$ does not depend on x, that it can

be computed by considering an equivalent model with a finite state $\overline{\mathbf{X}}$. The following continuity properties hold.

- $C(x,u,v)$ is seen to be continuous in v, since the steady state probabilities are continuous in v (see e.g. Lemma 10.2 in Hordijk [17]). Since $\mathcal{B}$ is compact, it follows that the continuity is uniform in v. One can then deduce that $C(x,v)$ are uniformly continuous in v (see e.g. [33] Theorem 2.1 and Remark 2.2).

- $Y^g(x,u)$ is uniformly continuous in v (see e.g. Proposition 5.1 (viii) in Spieksma [29]). This implies (see e.g. [33] Theorem 2.1 and Remark 2.2) that $Y^g(x)$ are uniformly continuous in v.

- $c(x,u,v)$ is uniformly continuous in v.

- (18) and the above imply that

$$\left\{ c(x,u(x)) - g_v^* + \sum_{y\in\mathcal{X}} \mathcal{P}_{x,u(x),v,y} h_v^*(y) \right\} \tag{24}$$

  is uniformly continuous in v.

Applying again [33] Theorem 2.1, we deduce that the minimizer of (24), and hence ϕ, are upper semi-continuous. The compactness and convexity of ϕ follows, since ϕ represent conservative policies for (24), so that for each state x, a policy is in $\phi(v)$ if and only if it is a minimizer of the linear function (24) over a compact interval.

Next we prove for ψ. Since for threshold policies v with parameter $(K_v,1)$, $D(x,u,v)$ is strictly integer convex in K_v (by Lemma 8), then it is minimized either by
(a) a single threshold policy, say with $(K_v,1)$ and in that case, for any $q_v<1$, the policies (K_v,q_v) and (K_v+1,q_v) are not optimal; or by
(b) two different thresholds of the form $(K_v,1)$ and $(K_v+1,1)$. In that case, all threshold (K_v+1,q_v), $q_v\in[0,1]$ are optimal. Both (a) and (b) follow e.g. from Key Lemma in [29], p. 168. Hence $\psi(K_u,q_u)$ is either a singleton, or of the form $\{(K_v,q_v),v\in[0,1]\}$, and hence compact and convex. The upper semi-continuity is a straight forward consequence of the continuity of $D(x,u,v)$ in (K_u,q_u) (Lemma 8), and hence the uniform continuity (since u is restricted to a compact set). $\square$

Proof of Theorem 1:
It follows from the discussion in the beginning of the section, that it suffices to search for an equilibrium policy within $u\in\mathcal{A}$, $v\in\mathcal{B}$. If we find one, it will be equilibrium among all policies. The result now follows by combining Lemmas 9 and 10, and by identifying $\mathcal{U}$ in Lemma 9 with $\mathcal{A}\times\mathcal{B}$. $\blacksquare$

6 Stochastic game between a router and an extra service capacity

In the previous model that we analyzed, both controllers had a cost that grew as the number of customers in the system increased. This cost structure induced some type of cooperation, in the sense that both controllers benefited from the fact that the server did not allow too many customers to accumulate before restarting service. This type of situations is in contrast with zero-sum stochastic games, where typically, nature would benefit mostly if it could force the controlled Markov chain to be null recurrent or transient. The following model that we shall analyze is an example of a conflict between two players, even though the cost is not zero-sum; the immediate cost of one player increases (in some sense) in the state, and the immediate cost of the other decreases. Such situations may occur when, for example, one player represents a customer, who pays some cost to a second player. In practice, the profit of the second player is less than the amount received from the first player, due to overhead, taxes etc.

The model

There are N parallel queues, each queue $1 \le i \le N$ having a capacity $L_i \le \infty$. Denote $L = (L_1, \dots, L_N)$. Customers arrive to the system according to a Poisson process with rate λ.

Player one is assumed to control service allocation in the following way. Each queue has an exponential dedicated server with rate μ. An extra service rate of Δ is distributed by the controller between the queues. Thus, a rate of a_i is allocated to queue i such that $\sum_{i=1}^N a_i = \Delta$.

The decisions of the router, player two, is taken immediately after an arrival to the system occurs, and its decisions are to which queue should the arriving customer be routed an action $b \in \{1, ..., N\}$ means routing an arriving customer to queue b.

Let the state of the system $\mathcal{X}$ be the number of customers in the different queues (including in service) just prior to a transition. The immediate costs are assumed to be a function of the state only; i.e. $c(x,a,b) = c_1(x)$, and $d(x,a,b) = d_1(x)$.

Using uniformization, the system is observed at time instants when either arrival or a (possibly fictitious, if the queue is empty) service completion occurs, which form a Poisson process with rate $\delta = \lambda + N\mu + \Delta$. The probability that a transition corresponds to an arrival, an uncontrolled end of service in queue j, or an end of a controlled service in queue j are λ/δ, μ/δ and b_j/δ respectively. Let $\mathcal{A}_i y$ and $\mathcal{D}_i y$ be the state obtained by an arrival to queue i and by a departure (possibly dummy) from queue i, when the state was previously y. The transition probabilities are thus

$$P_{xaby} = \begin{cases} \lambda/\delta & \text{if } y = \mathcal{A}_a x \le L \text{ or } x = y = L\,, \\ (\mu + b_i)/\delta & \text{if } y = \mathcal{D}_i x\,. \end{cases}$$

We define the class of policies U and V as in Section 2, and we consider the probability $P_x^{u,v}$ induced by each policy pair (u,v) and an initial state x. We denote again by $\{X_t, A_t, B_t\}$ ths stochastic processes representing the states and the actions at time t, $t = 0, 1, ...$

We introduce the following properties of a function $f : \mathcal{X} \to \mathbb{R}$.

C1 (monotonicity) $f(z) \le f(y)$ for $z \le y$ (where we mean by $z \le y$ a componentwize ordering: $z_i \le y_i, 1 \le i \le N$).

C2 (symmetry) $f(y) = f(\pi y)$ for any permutation πy of y such that $\pi y \in \mathcal{X}$.

C3 $f(\mathcal{A}_k y) \le f(\mathcal{A}_l y)$ for y such that $y_k \le y_l$, $\mathcal{A}_k \mathcal{A}_l y \le L$.

We shall consider the following cost criteria:

(i) the finite horizon costs

$$C^t(x,u,v) = \sum_{s=0}^{t} E_x^{u,v} c_1(X_s), \qquad D^t(x,u,v) = \sum_{s=0}^{t} E_x^{u,v} d_1(X_s);$$

(ii) the discounted costs with discount factors α^i, $0 < \alpha_i < 1$, $i = 1, 2$,

$$C^\alpha(x,u,v) = \sum_{s=0}^{\infty} \alpha^s E_x^{u,v} c_1(X_s), \qquad D^\alpha(x,u,v) = \sum_{s=0}^{\infty} \alpha^s E_x^{u,v} d_1(X_s);$$

(iii) the expected average cost (defined in (4)).
We define equilibria in the sense of Nash, i.e. as in (5) and (6).

The main result

Theorem 11
Assume that the holding costs c_1, and $(-d_1)$, both satisfy **C1,C2,C3**. *Consider a routing strategy u^o that always joins a (non-full) shortest queue. Consider a server's strategy r^o that always serves a shortest queue at the maximum rate $\mu + \Delta$. Then (r^o, u^o) are equilibrium for*
(i) the finite cost criterion criterion,
(ii) the infinite horizon discounted cost and expected average cost criteria, provided that the immediate costs c_1 and d_1 are polynomially bounded.

Proof
Altman and Shimkin prove in [8] the Theorem for the case of finite queues and the finite horizon criterion, where $c_1 = -d_1$ (i.e. the zero-sum stochastic game). Their proof is generalized to the other cost criteria in [4]. Although in our case $c_1 \neq -d_1$, it follows that the proof in [8] still applies. Indeed, the proof in [8] is based on solving two independent MDPs. First, it is shown that if the server restricts to r^o, then u^o is an optimal response of the router, i.e. it minimizes the router's cost. Conversely, it is shown that, if the router restricts to u^o, then r^o is an optimal response of the server, i.e. it maximizes the server's cost. For each MDP, the proof follows just by assuming that the properties **C1,C2,C3** hold for the appropriate immediate cost. This proof therefore establishes that (u^o, r^o) are equilibrium policies for the non-zero sum game as well. The extensions to the other cost criteria can now be done following [4] (considering, again, two separate MDPs). □

References

[1] E. Altman, Flow control using the theory of zero-sum Markov games, *IEEE Trans. Automatic Control*, **39**, No. 4 (1994) 814–818.
[2] Monotonicity of optimal policies in a zero sum game: a flow control model, *Advances of dynamic games and applications*, **1** (1994) 269–286.
[3] E. Altman, A Markov game approach for optimal routing into a queueing network, INRIA report No. 2178, submitted (1994).
[4] E. Altman and A. Hordijk, Zero-sum Markov games and worst-case optimal control of

queueing systems, invited paper, *QUESTA* **21**, special issue on optimization of queueing systems, Ed. S. Stidham (1995) 415–447.

[5] E. Altman, A. Hordijk and F. M. Spieksma, Contraction conditions for average and α-discount optimality in countable state Markov games with unbounded rewards, submitted (1994) Research report No. TW-93-16, University of Leiden.

[6] E. Altman and G. Koole, Stochastic scheduling games with Markov decision arrival processes, *Journal Computers and Mathematics with Appl.*, 3rd special issue on "Differential Games" (1993) 141–148.

[7] E. Altman and N. Shimkin Individually Optimal Dynamic Routing in a Processor Sharing System: Stochastic Game Analysis, EE Pub No. 849, August (1992), submitted to *Operations Research.*

[8] E. Altman and N. Shimkin, Worst-case and Nash routing policies in parallel queues with uncertain service allocations, IMA Preprint Series No. 1120, Institute for Mathematics and Applications, University of Minnesota, Minneapolis, USA (1993) submitted to *Operations Research.*

[9] R. Cavazos-Cadena, Weak conditions for the existence of optimal stationary policies in average decision chains with unbounded costs, *Kybernetika*, **25**, No. 3 (1989) 145–156.

[10] K. Fan, Fixed-point and minimax theorems in locally convex topological linear spaces, *Proc. Natn, Acad. Sci.* **38** (1952) 121–126.

[11] A. Federgruen, On N-person stochastic Games with denumerable state space, *Adv. Appl Prob*, **10** (1978) 452–471.

[12] A. Glazer and R. Hassin, Stable priority purchasing in queues, *Operations Research Letters*, **6** (1986) 285–288.

[13] I. Glicksberg, A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points, *Proc. Amer. Math. Soc.* **3** (1952) 170–174.

[14] R. Hassin and M. Haviv, Equilibrium strategies and the value of information in a two line queueing system with threshold jockeying, *Commun. Statist. - Stochastic Models*, **10** No. 2 (194) 415–435.

[15] M. Haviv, Stable strategies for processor sharing systems, *European J. of Operations Research* **52** (1991) 103–106.

[16] D. P. Heyman, Optimal operating policies for M/G/1 queueing systems, *Operations Research* **16** (1968) 362–382.

[17] A. Hordijk, *Dynamic Programming and Markov Potential Theory*, Second Edition, Mathematical Centre Tracts 51, Mathematisch Centrum, Amsterdam (1977).

[18] A. Hordijk and F. Spieksma, Constrained admission control to a queueing system, *Adv. Appl. Prob.* **21**, (1989) 409–431.

[19] M. T. Hsiao and A. A. Lazar, A game theoretic approach to decentralized flow control of Markovian queueing Networks, *Performance '87*, Courtois & Latouche (eds.) (1988) 55–73.

[20] M. M. Jaiswal, *Priority Queues*, Academic Press (1968).

[21] H.-U. Küenle, *Stochastiche Spiele und Entscheidungesmodelle*, Tebuner-Texte, Band 89 (1986).

[22] H.-U. Küenle, On the optimality of (s,S)-strategies in a minimax inventory model with average cost criterion, *Optimization* **22** No. 1 (1991) 123–138.

[23] S. A. Lippman and J. W. Mamer, Preemptive innovation, *J. of Economic Theory*, **61**, No. 1 (1993) 104–119.

[24] J. M. McNamara, S. Merad and E. J. Collins, The Hawk-Dove game as an average-cost problem, *Adv. Appl. Prob.* **23** (1991) 667–682.

[25] P. Naor, On the regulation of queueing size by levying tolls, *Econometrica*, **37** (1969) 15–24.

[26] S. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, 1970.

[27] N. Shimkin, A sample path approach to admission and flow control, private communication.

[28] T. E. S. Raghavan and J. A. Filar, Algorithms for stochastic games - a survey, *ZOR*, **35**

(1991) 437–472.

[29] F. M. Spieksma, *Geometrically Ergodic Markov Chains and the Optimal Control of Queues*, Ph.D. thesis, University of Leiden (1990).

[30] M. J. Sobel, Optimal Average-Cost Policy for a Queue with Start-Up and Shut Down Costs, *Operations Research*, **17** (1969) 145–162.

[31] S. Stidham, Socially and individually optimal control of arrivals to a GI/M/1 queue, *Management Science*, **24**, (1970) 1598–1610.

[32] S. Stidham and R. R. Weber, Monotonic and insensitive optimal policies for control of queues with undiscounted cost, *Operations Research* **87**, No. 4 (1989) 611–625.

[33] M. Tidball and E. Altman, Approximations in dynamic zero-sum games, I, *SIAM J. Control and Optimization*, **34**, No. 1, (1996) 311–328.

[34] R. R. Weber and S. Stidham, Optimal control of service rates in networks of queues, *Advances in Applied Probability*, **19** (1987) 202–218.

[35] M. Yadin and P. Naor, Queueing systems with removable service station, *Operations Research Quaterly*, **14** (1963) 393–405.

[36] D. D. Yao, S-modular games, with queueing applications, invited paper, *QUESTA* **21**, special issue on optimization of queueing systems, Ed. S. Stidham (1995) 449–475.

[37] U. Yechiali, On optimal balking rules and toll charges in a GI/M/1 queueing process, *Operations Research*, **19** (1971) 349–370.