Constrained Markov Games: Nash Equilibria

Eitan ALTMAN
INRIA
2004 Route des Lucioles, B.P.93
06902 Sophia-Antipolis Cedex
France

Adam SHWARTZ
Electrical Engineering Department
Technion—Israel Institute of Technology
Haifa 32000
Israel

January 1998
To appear, Annals of Dynamic Games

Abstract

In this paper we develop the theory of constrained Markov games. We consider the expected average cost as well as discounted cost. We allow different players to have different types of costs. We present sufficient conditions for the existence of stationary Nash equilibrium. Our results are based on the theory of sensitivity analysis of mathematical programs developed by Dantzig, Folkman and Shapiro [9], which was applied to Markov Decision Processes in [3]. We further characterize all stationary Nash equilibria as fixed points of some coupled Linear Programs.

Keywords: Constrained Markov games, discounted cost, expected average cost, stationary Nash equilibrium.

1 Introduction

Constrained Markov decision processes arise in situations when a controller has more than one objective. A typical situation is when one wants to minimize one type of cost while keeping other costs lower than some given bounds. Such problems arise frequently in computer networks and data communications, see Lazar [20], Spieksma and Hordijk [16], Nain and Ross [23], Ross and Chen [26], Altman and Shwartz [1] and Feinberg and Reiman [12]. The theory of constrained MDPs goes back to Derman and Klein [10], and was developed by Hordijk and Kallenberg [15], Kallenberg [18], Beutler and Ross [6, 7], Ross and Varadarajan [27] Altman and Shwartz [2, 3], Altman [4, 5], Spieksma [32], Sennott [28, 29], Borkar [8], Feinberg [11], Feinberg and Shwartz [13, 14] and others.

In all these papers, a single controller was considered. A natural question is whether this theory extends to Markov games with several (say N) players, where player i wishes to minimize C_i^0, subject to some bounds $V_i^j, i = 1, ..., N$ on the costs $C_i^j,\ j = 1, \ldots, B_i$. Although a general theory does not exist, several applications to telecommunications have been analyzed, see [17, 19]. The problem studied in these references is related to dynamic decentralized flow control by several (selfish) users, each of which seeks to maximize its own throughput. Since voice and video traffic typically require the end-to-end delays to be bounded, this problem was posed as a constrained Markov game. For (static) games with constraints, see e.g. Rosen [24]. Some other theoretical results on zero-sum constrained Markov games were obtained by Shimkin [30]. A related theory of approachability for stochastic games was developed by Shimkin and Shwartz [31], and extended to semi-Markov games, with applications in telecommunications, by Levi and Shwartz [21, 22].

In this paper we present sufficient conditions for the existence of stationary Nash equilibria. Our results are based on the theory of sensitivity analysis of mathematical programs developed by Dantzig, Folkman and Shapiro [9]. This theory was applied to Markov Decision Processes in [3], and we rely on this application here. We further characterize all stationary Nash equilibria as fixed points of some coupled Linear Programs.

2 The model and main result

We consider a game with N players, labeled $1, \ldots, N$. Define the tuple $\{\mathbf{X}, \mathbf{A}, \mathcal{P}, c, V\}$ where

- $\mathbf{X}$ is a finite state space. Generic notation for states will be x, y.
- $\mathbf{A} = \{\mathbf{A}_i\}, i = 1, ..., N$ is a finite set of actions. We denote by $\mathbf{A}(x) = \{\mathbf{A}_i(x)\}_i$ the set of actions available at state x. A generic notation for a vector of actions will be $\mathbf{a} = (a_1, ..., a_N)$ where a_i stands for the action chosen by player i. Denote $\mathcal{K} = \{(x, \mathbf{a}) : x \in \mathbf{X},\ \mathbf{a} \in \mathbf{A}(x)\}$ and set $\mathcal{K}_i = \{(x, a_i) : x \in \mathbf{X},\ a_i \in \mathbf{A}_i(x)\}$.
- $\mathcal{P}$ are the transition probabilities; thus $\mathcal{P}_{x\mathbf{a}y}$ is the probability to move from state x to y if the vector $\mathbf{a}$ of actions is chosen by the players.

- $c = \{c_i^j\}, i = 1, ..., N,\ j = 0, 1, ..., B_i$ is a set of immediate costs, where $c_i^j : \mathcal{K} \to \mathbb{R}$. Thus player i has a set of $B_i + 1$ immediate costs; c_i^0 will correspond to the cost function that is to be minimized by that player, and $c_i^j,\ j > 0$ will correspond to cost functions on which some constraints are imposed.

- $V = \{V_i^j\}, i = 1, ..., N,\ j = 1, ..., B_i$ are bounds defining the constraints (see (3) below).

Let $M_1(G)$ denote the set of probability measures over a set G. Define a history at time t to be a sequence of previous states and actions, as well as the current state: $h_t = (x_1, \mathbf{a}_1, ..., x_{t-1}, \mathbf{a}_{t-1}, x_t)$. Let $\mathbf{H}_t$ be the set of all possible histories of length t. A policy u^i for player i is a sequence $u^i = (u_1^i, u_2^i, ...)$ where $u_t : \mathbf{H}_t \to M_1(\mathbf{A}_i)$ is a function that assigns to any history of length t, a probability measure over the set of actions of player i. At time t, the controllers choose independently of each other actions $\mathbf{a} = (a_1, ..., a_N)$, where action a_i is chosen by player i with probability $u_t(a_i|h_t)$ if the history h_t was observed. The class of all policies defined as above for player i is denoted by U^i. The collection $U = \times_{i=1}^N U^i$ is called the class of multi-policies ($\times$ stands for the product space).

A stationary policy for player i is a function $u^i : \mathbf{X} \to M_1(\mathbf{A}_i)$ so that $u^i(\cdot|x) \in M_1(\mathbf{A}_i(x))$. We denote the class of stationary policies of player i by U_S^i. The set $U_S = \times_{i=1}^N U_S^i$ is called the class of stationary multi-policies. Under any stationary multi-policy u (where the u^i are stationary for all the players), at time t, the controllers, independently of each other, choose actions $\mathbf{a} = (a_1, ..., a_N)$, where action a_i is chosen by player i with probability $u^i(a_i|x_t)$ if state x_t was observed at time t. Under a stationary multi-policy the state process becomes a Markov chain with transition probabilities $P_{xy}^w = \sum_{\mathbf{a}} \mathcal{P}_{x\mathbf{a}y} w(\mathbf{a}|x)$.

For $u \in U$ we use the standard notation u^{-i} to denote the vector of policies $u^k, k \neq i$; moreover, for $v^i \in U^i$, we define $[u^{-i}|v^i]$ to be the multi-policy where, for $k \neq i$, player k uses u^k, while player i uses v^i. Define $U^{-i} := \cup_{u \in U}\{u^{-i}\}$. For $\mathbf{a} \in \mathbf{A}(x)$ and $a \in \mathbf{A}_i(x)$, we use the obvious notation $\mathbf{a}^{-i}$, $[\mathbf{a}^{-i}|a]$, and the set $\mathbf{A}^{-i}(x)$, $i = 1, ..., N$, $x \in \mathbf{X}$.

A distribution β for the initial state (at time 1) and a multi-policy u together define a probability measure P_β^u which determines the distribution of the stochastic process $\{X_t,\ A_t\}$ of states and actions. The corresponding expectation is denoted as E_β^u.

Next, we define the cost criteria that will appear in the constrained control problem. For any policy u and initial distribution β, define the i,j-discounted cost by

$$C_\alpha^{i,j}(\beta, u) = (1-\alpha)\sum_{t=1}^{\infty} \alpha^{t-1} E_\beta^u c_i^j(X_t, A_t) \tag{1}$$

where $0 < \alpha < 1$ is fixed. The i,j-expected average cost is defined as

$$C_{ea}^{i,j}(\beta, u) = \overline{\lim_{T\to\infty}} \frac{1}{T}\sum_{t=1}^{T} E_\beta^u c_i^j(X_t, A_t). \tag{2}$$

We assume that all the costs $C^{i,j}$, $j = 1, ..., B_i$ corresponding to any given player i are either discounted costs with *the same discount factor* $\alpha = \alpha_i$ or they are all expected average costs. However, the discount factors α_i may vary between players, and some player may have discounted costs while others may have expected average costs.

A multi-policy u is called i-feasible if it satisfies:

$$C^{i,j}(\beta, u) \le V_i^j, \qquad \text{for all } j = 1, ..., B_i. \tag{3}$$

It is called feasible if it is i-feasible for all the players $i = 1, ..., N$. Let U_V be the set of feasible policies. A policy $u \in U_V$ is called constrained Nash equilibrium if for each player $i = 1, ..., N$ and for any v^i such that $[u^{-i}|v^i]$ is i-feasible,

$$C^{i,0}(\beta, u) \le C^{i,0}(\beta, [u^{-i}|v^i]). \tag{4}$$

Thus, any deviation of any player i will either violate the constraints of the ith player, or if it does not, it will result in a cost $C^{i,0}$ for that player that is not lower than the one achieved by the feasible multi-policy u.

For any multi-policy u, u^i is called an optimal response for player i against u^{-i} if u is i-feasible, and if for any v^i such that $[u^{-i}|v^i]$ is i-feasible, (4) holds. A multi-policy v is called an optimal response against u if for every $i = 1, ..., N$, v^i is an optimal response for player i against u^{-i}.

We introduce the following assumptions

- <u>(Π_1) Ergodicity:</u> If there is at least one player that uses the expected average cost criteria, then the unichain ergodic structure holds, i.e. under any stationary multi-policy u, the state process is an irreducible Markov chain with one ergodic class (and possibly some transient states).

- (Π_2) Strong Slater condition: For any stationary multi-policy u, and for any player i, there exists some v^i such that

$$C^{i,j}(\beta,[u^{-i}|v^i]) < V_i^j, \qquad \text{for all } j=1,...,B_i. \tag{5}$$

We are now ready to introduce the main result.

Theorem 2.1 *Assume that Π_1 and Π_2 hold. Then there exists a stationary multi-policy u which is constrained-Nash equilibrium.*

3 Proof of main result

We begin by describing the way an optimal stationary response for player i is computed for a given stationary multi-policy u. Fix a stationary multi-policy u. Denote the transition probabilities induced by players other than i, when player i uses action a_i, by $\mathcal{P}^{u,i} = \{\mathcal{P}^{u,i}_{xa_iy}\}$ where

$$\mathcal{P}^{u,i}_{xa_iy} := \sum_{\mathbf{a}^{-i}\in\mathbf{A}^{-i}} \prod_{l\neq i} u^l_x(a_l|x)\mathcal{P}_{x\mathbf{a}y}, \qquad \mathbf{a}=[\mathbf{a}^{-i}|a_i].$$

That is, we consider new transition probabilities to go from x to y as a function of the action a_i of player i, for fixed stationary policies of players l, $l\neq i$. Similarly, define

$$c^{j,u}_i(x,a_i) := \sum_{\mathbf{a}^{-i}\in\mathbf{A}^{-i}} \prod_{l\neq i} u^l_x(a_l|x)c^j_i(x,\mathbf{a}) \qquad \mathbf{a}=[\mathbf{a}^{-i}|a_i].$$

Next we present a Linear Program (LP) for computing the set of all optimal responses for player i against a stationary policy u^{-i}. The following LP will be related to both the discounted and average case; if the ith player uses discount costs, then α_i below will stand for its discount factor. If it uses the expected average costs then α_i below is set to 1. Recall that β is the initial distribution.

$\mathbf{LP(i,u)}$:

Find $z^* := \{z^*(y,a)\}_{y,a}$ that minimizes $\mathcal{C}^{i,0}(z) := \sum_{y\in\mathbf{X}}\sum_{a\in\mathbf{A}_i(y)} c^{0,u}_i(y,a)z(y,a)$ subject to:

$$\sum_{y\in\mathbf{X}}\sum_{a\in\mathbf{A}_i(y)} z(y,a)\left[\delta_r(y)-\alpha_i\mathcal{P}^{u,i}_{yar}\right] = [1-\alpha_i]\beta(r) \qquad r\in\mathbf{X} \tag{6}$$

$$\mathcal{C}^{i,j}(z) := \sum_{y\in\mathbf{X}} \sum_{a\in\mathbf{A}_i(y)} c_i^{j,u}(y,a)z(y,a) \leq V_i^j \qquad 1 \leq j \leq B_i \tag{7}$$

$$z(y,a) \geq 0, \qquad \sum_{y\in\mathbf{X}} \sum_{a\in\mathbf{A}_i(y)} z(y,a) = 1 \tag{8}$$

Define $\Gamma(i,u)$ to be the set of optimal solutions of $\mathbf{LP}(i,u)$.

Given a set of nonnegative real numbers $z = \{z(y,a), y \in \mathbf{X}, a \in \mathbf{A}_i(y)\}$, define the point to set mapping $\gamma(z)$ as follows: If $\sum_a z(y,a) \neq 0$ then $\gamma_y^a(z) := \{z(y,a)[\sum_a z(y,a)]^{-1}\}$ is a singleton: for each y, we have that $\gamma_y(z) = \{\gamma_y^a(z) : a \in \mathbf{A}_i(y)\}$ is a point in $M_1(\mathbf{A}_i(y))$. Otherwise, $\gamma_y(z) := M_1(\mathbf{A}_i(y))$, i.e. the (convex and compact) set of all probability measures over $\mathbf{A}_i(y)$. Define $g^i(z)$ to be the set of stationary policies for player i that choose, at state y, action a with probability in $\gamma_y^a(z)$.

For any stationary multi-policy v define the occupation measures

$$f(\beta,v) := \{f^i(\beta,v;y,a) : y \in \mathbf{X},\, a \in \mathbf{A}_i(y),\, i = 1,...,N\}$$

as follows. Let $P(v)$ be the transition probabilities of the Markov chain representing the state process when the players use the stationary multi-policy v. If player i uses the discounted cost $\alpha_i < 1$, then

$$f^i(\beta,v;y,a) := (1-\alpha_i) \sum_{x\in\mathbf{X}} \beta(x) \left(\sum_{s=1}^{\infty} \alpha_i^{s-1} [P(v)^s]_{xy} \right) v^i(a|y). \tag{9}$$

If player i uses the expected average cost then

$$f^i(\beta,v;y,a) := \pi^v(y) v^i(a|y),$$

where π^v is the steady state (invariant) probability of the Markov chain describing the state process, when policy v is used (which exists and is unique by Assumption Π_1).

Proposition 3.1 *Assume Π_1-Π_2. Fix any stationary multi-policy u.*
(i) If z^* is an optimal solution for $\mathbf{LP}(i,u)$ then any element w in $g^i(z^*)$ is an optimal stationary response of player i against the stationary policy u^{-i}. Moreover, the multi-policy $v = [u^{-i}|w]$ satisfies $f^i(\beta,v) = z^*$.
(ii) Assume that w is an optimal stationary response of player i against the stationary policy u^{-i}, and let $v := [u^{-i}|w]$. Then $f^i(\beta,v)$ is optimal for $\mathbf{LP}(i,u)$.

(iii) The optimal sets $\Gamma(i,u)$, $i = 1,...,N$ are convex, compact, and upper semi-continuous in u^{-i}, where u is identified with points in $\times_{i=1}^{N} \times_{x\in\mathbf{X}} M_1(\mathbf{A}_i(x))$.
(iv) For each i, $g^i(z)$ is upper semi-continuous in z over the set of points which are feasible for $\mathbf{LP}(i,u)$ (i.e. the points that satisfy constraints (6)-(8)).

Proof: When all players other than i use u^{-i}, then player i is faced with a constrained Markov decision process (with a single controller). The proof of (i) and (ii) then follows from [3] Theorems 2.6. The first part of (iii) follows from standard properties of Linear Programs, whereas the second part follows from an application of the theory of sensitivity analysis of Linear Programs by Dantzig, Folkman and Shapiro [9] in [3] Theorem 3.6 to $\mathbf{LP}(i,u)$. Finally, (iv) follows from the definition of $g^i(z)$. ■

Define the point to set map

$$\Psi : \times_{i=1}^{N} M_1(\mathcal{K}_i) \to 2^{\left\{\times_{i=1}^{N} M_1(\mathcal{K}_i)\right\}}$$

by

$$\Psi(\mathbf{z}) = \times_{i=1}^{N} \Gamma(i, g(z))$$

where $\mathbf{z} = (z_1, \ldots, z_N)$, each z_i is interpreted as a point in $M_1(\mathcal{K}_i)$ and $g(z) = (g^1(z_1), \ldots, g^N(z_N))$.

Proposition 3.2 *Assume Π_1-Π_2. In the case that no player uses the expected average cost criterion, assume further that, for some $\eta > 0$, the initial distribution satisfies $\beta(x) > \eta$ for all x. Then there exists a fixed point $\mathbf{z}^* \in \Psi(\mathbf{z}^*)$.*

Proof: By proposition 3.1(i) and (9), it follows that under the stated conditions, all solutions z^* of $\mathbf{LP}(i,u)$ satisfy $\sum_a z^*(x,a) > \eta'$ for some $\eta' > 0$. But under this restriction the set $\gamma_y(z)$ is a singleton, and hence the range of the function $g^i(z)$ is also single points. The proof now follows directly from Kakutani's fixed point theorem applied to Ψ. Indeed, by Proposition 3.1 (iii) and (iv), $\Gamma(i, g(z))$ is a composition of two upper semi-continuous functions: $g(\cdot)$, and Γ, which have convex compact ranges. Hence Ψ is upper semi-continuous in $\mathbf{z}$, and has a compact range. Since $g(\cdot)$ can be considered as a regular (that is, not set valued) function, the composition also has a convex range. Therefore the conditions of Kakutani's Theorem hold. ■

Proof of Theorem 2.1: Under the conditions of Proposition 3.2, the proof is obtained by combining Proposition 3.1 (i) with Proposition 3.2. Indeed, Proposition 3.1 (i) implies that for any fixed point $\mathbf{z}$ of Ψ, the stationary multi-policy $g = \{g^i(z^i); i = 1, ..., N\}$ is constrained Nash equilibrium. It therefore remains to treat the case where all players use a discounted criterion and, moreover, the initial distribution β satisfies $\beta(x) = 0$ for some x.

Given such β, let β_n be a sequence of initial distributions satisfying $\beta_n(x) > \eta_n > 0$ and $\beta_n(x) \to \beta(x)$ for each x (so that $\eta_n \to 0$). Let u_n be a constrained Nash equilibrium multi-policy for the problem with initial distribution β_n: this was just shown to exist. If we identify the multi-policies u_n with points in $\times_{i=1}^{N} \times_{x \in \mathbf{X}} M_1(\mathbf{A}_i(x))$, then they lie in a compact set. Let u be any limit point. We claim that u is a constrained Nash equilibrium multi-policy for the problem with initial distribution β. To show this we need to establish that

(i) u satisfies (3) for each i, and

(ii) if $[u^{-i}|v^i]$ is i-feasible, then (4) holds.

From (1) it follows that all costs are linear (hence continuous) functions of the frequencies $f^i(\beta, v)$ which are, in turn, continuous functions of (β, v) (see (9)). In fact, the costs are linear in β. Therefore the costs $C^{i,j}(\beta, u)$ are continuous in (β, u). Since (3) holds for (β_n, u_n), the continuity implies that it holds for (β, u) and (i) is established.

Now fix some i and suppose that v^i is such that $[u^{-i}|v^i]$ is i-feasible. Note that it is possible that, for some i, $[u_n^{-i}|v^i]$ is not i-feasible for any n large. However, by assumption Π_2, we can find some $\tilde{v}^i$ so that (5) holds. Fix an arbitrary ϵ and note that (5) holds also if we replace $\tilde{v}^i$ by $v^i_\epsilon = \epsilon \tilde{v}^i + (1-\epsilon) v^i$. By the continuity and the linearity we established in proving (i) above, this implies that $[u_n^{-i}|v^i_\epsilon]$ is i-feasible for all n large enough. Therefore,

$$
\begin{aligned}
C^{i,0}(\beta, u) &= \lim_{n\to\infty} C^{i,0}(\beta_n, u_n) \\
&\leq \lim_{n\to\infty} C^{i,0}(\beta_n, [u_n^{-i}|v^i_\epsilon]) \\
&= C^{i,0}(\beta, [u^{-i}|v^i_\epsilon]).
\end{aligned}
$$

Using the continuity again, we have

$$
C^{i,0}(\beta, [u^{-i}|v^i]) = \lim_{\epsilon\to 0} C^{i,0}(\beta, [u^{-i}|v^i_\epsilon])
$$

and (ii) follows. ∎

Remark 3.1 (i) The Linear Program formulation $\mathbf{LP}(i,u)$ is not only a tool for proving the existence of a constrained Nash equilibrium; in fact, due to Proposition 3.1 (ii), it can be shown that any stationary constrained Nash equilibrium w has the form $w = \{g^i(z^i); i = 1,...,N\}$ for some $\mathbf{z}$ which is a fixed point of Ψ. Indeed, if w is a constrained Nash equilibrium then it follows from Proposition 3.2 that $f(\beta, w)$ is a fixed point of Ψ.
(ii) It follows from [3] Theorems 2.4 and 2.5 that if $\mathbf{z} = (z^1,...,z^N)$ is a fixed point of Ψ, then any stationary multi-policy g in $\times_{i=1}^N g^i(z^i)$ satisfies $C^{i,j}(\beta, g) = \mathcal{C}^{i,j}(z), i = 1,...,N, j = 0,...,B_i$. Conversely, if w is a constrained Nash equilibrium then

$$C^{i,j}(\beta, w) = \sum_{y\in\mathbf{X}} \sum_{a\in\mathbf{A}_i(y)} f^i(\beta, w; y, a) c_i^{j,w}(y, a)$$

(and $f(\beta, w)$ is a fixed point of Ψ).

Acknowledgment. Research of the second author was supported in part by the Israel Science Foundation, administered by the Israel Academy of Sciences and Humanities.

References

[1] E. Altman and A. Shwartz, "Optimal priority assignment: a time sharing approach", *IEEE Transactions on Automatic Control* Vol. AC-34 No. 10, pp. 1089-1102, 1989.

[2] E. Altman and A. Shwartz, "Markov decision problems and state-action frequencies," *SIAM J. Control and Optimization.* **29**, No. 4, pp. 786-809, 1991

[3] E. Altman and A. Shwartz, "Sensitivity of constrained Markov Decision Problems", *Annals of Operations Research*, **32**, pp. 1-22, 1991.

[4] E. Altman, "Denumerable constrained Markov Decision Processes and finite approximations", *Math. of Operations Research*, **19**, No. 1, pp. 169-191, 1994.

[5] E. Altman, "Asymptotic Properties of Constrained Markov Decision Processes", *ZOR - Methods and Models in Operations Research*, **37**, Issue 2, pp. 151-170, 1993.

[6] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint", *J. Mathematical Analysis and Applications* **112**, 236-252, 1985.

[7] F. J. Beutler and K. W. Ross, "Time-Average Optimal Constrained Semi-Markov Decision Processes", *Advances of Applied Probability* **18**, No. 2, pp. 341-359, 1986.

[8] V. S. Borkar, "Ergodic control of Markov Chains with constraints – the general case", *SIAM J. Control and Optimization.* **32**, No. 1, pp. 176-186, 1994.

[9] Dantzig G. B., J. Folkman and N. Shapiro, "On the continuity of the minimum set of a continuous function", *J. Math. Anal. and Applications,* Vol. 17, pp. 519-548, 1967.

[10] C. Derman and M. Klein, "Some remarks on finite horizon Markovian decision models," *Oper. Res.* Vol. 13, pp. 272–278, 1965.

[11] E. A. Feinberg, "Constrained Semi-Markov Decision Processes With Average Rewards", *ZOR* **39**, pp. 257–288, 1993.

[12] E. A. Feinberg and M. I. Reiman, "Optimality of randomized trunk reservation", *Probability in the Engineering and Informational Sciences.* **8**, pp. 463-489, 1994.

[13] E. A. Feinberg and A. Shwartz "Constrained Markov decision models with weighted discounted rewards," *Math. of Operations Research* **20** pp. 302–320, 1995.

[14] E. A. Feinberg and A. Shwartz "Constrained discounted dynamic programming," *Math. of Operations Research* **21**, pp. 922–945, 1996.

[15] A. Hordijk and L. C. M. Kallenberg, "Constrained undiscounted stochastic dynamic programming", *Mathematics of Operations Research,* **9**, No. 2, pp. 276-289, 1984.

[16] A. Hordijk and F. Spieksma, "Constrained admission control to a queuing system" *Advances of Applied Probability* Vol. 21, pp. 409-431, 1989.

[17] M. T. Hsiao and A. A. Lazar, "Optimal Decentralized Flow Control of Markovian queueing Networks with Multiple Controllers", *Performance evaluation*, **13**, 181-204, 1991.

[18] L. C. M. Kallenberg, *Linear Programming and Finite Markovian Control Problems*, Mathematical Centre Tracts 148, Amsterdam, 1983.

[19] Y. A. Korilis and A. Lazar, "On the Existence of Equilibria in Noncooperative Optimal Flow Control", *J. of the Association for Computing Machinery*, **42**, No. 3, pp. 584-613, 1995.

[20] A. Lazar, "Optimal flow control of a class of queuing networks in equilibrium", *IEEE Transactions on Automatic Control*, Vol. 28 no. 11, pp. 1001-1007, 1983.

[21] R. Levi and A. Shwartz, "A theory of approachability and throughput-cost tradeoff in a queue with impatient customers," EE Pub. 936, Technion, 1994.

[22] R. Levi and A. Shwartz, "Throughput-Delay tradeoff with impatient arrivals," Proceedings of the 23rd Allerton Conference on Communications, Control and Computing, Allerton, IL 1994.

[23] P. Nain and K. W. Ross, "Optimal Priority Assignment with hard Constraint," *Transactions on Automatic Control,* Vol. 31 No. 10, pp. 883-888, October 1986.

[24] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, **33**, pp. 520-534, 1965.

[25] K. W. Ross, "Randomized and past-dependent policies for Markov decision processes with multiple constraints", *Operations Research* **37**, No. 3, pp. 474-477, 1989.

[26] K. W. Ross and B. Chen, "Optimal scheduling of interactive and non interactive traffic in telecommunication systems", *IEEE Trans. on Auto. Control,* Vol. 33 No. 3 pp. 261-267, 1988.

[27] K. W. Ross and R. Varadarajan, "Markov Decision Processes with Sample path constraints: the communicating case", *Operations Research*, **37**, No. 5, pp. 780-790, 1989.

[28] L. I. Sennott, "Constrained discounted Markov decision chains", *Probability in the Engineering and Informational Sciences*, **5**, pp. 463-475, 1991.

[29] L. I. Sennott, "Constrained average cost Markov decision chains", to appear in *Probability in the Engineering and Informational Sciences.*

[30] N. Shimkin, "Stochastic games with average cost constraints", *Annals of the International Society of Dynamic Games, Vol. 1: Advances in Dynamic Games and Applications*, Eds. T. Basar and A. Haurie, Birkhauser, 1994.

[31] N. Shimkin and A. Shwartz, "Guaranteed performance regions for Markovian systems with competing decision makers," *IEEE Trans. Auto. Control* Vol. 38, pp. 84–95, 1993.

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