OPERATIONS RESEARCH
Vol. 59, No. 2, March–April 2011, pp. 365–382
ISSN 0030-364X | EISSN 1526-5463 | 11 | 5902 | 0365

DOI 10.1287/opre.1110.0931

Discounted Robust Stochastic Games and an Application to Queueing Control

Erim Kardeş
Industrial and Systems Engineering Department, University of Southern California, Los Angeles, California 90089, kardes@usc.edu

Fernando Ordóñez
Industrial and Systems Engineering Department, University of Southern California, Los Angeles, California 90089; and Industrial Engineering Department, University of Chile, Santiago, Chile, fordon@usc.edu

Randolph W. Hall
Industrial and Systems Engineering Department, University of Southern California, Los Angeles, California 90089, rwhall@usc.edu

This paper presents a robust optimization model for n-person finite state/action stochastic games with incomplete information. We consider nonzero sum discounted stochastic games in which none of the players knows the true data of a game, and each player adopts a robust optimization approach to address the uncertainty. We call these games *discounted robust stochastic games*. Such games allow us to use simple uncertainty sets for the unknown data and eliminate the need to have an a-priori probability distribution over a set of games. We prove the existence of equilibrium points and propose an explicit mathematical programming formulation for an equilibrium calculation. We illustrate the use of discounted robust stochastic games in a single server queueing control problem.

Subject classifications: games: stochastic; dynamic programming/optimal control: Markov, finite state; queues: optimization.
Area of review: Optimization.
History: Received June 2007; revisions received May 2008, June 2009, February 2010, May 2010; accepted August 2010.

1. Introduction

Stochastic games are used to model situations where two or more decision makers, called players, take actions over time in a competitive manner to achieve their individual objectives. They consist of states and actions associated with each player. Players choose their respective actions in a state and receive payoffs. The play then moves into a next state, chosen according to a probability distribution, and continues thereon. Payoffs and transition probabilities are determined by the actions chosen and the state in which they are chosen. Therefore, stochastic games could be viewed as a generalization of Markov decision processes (MDPs) to more than one competing decision maker. They also generalize one-shot matrix games to a multistage, multistate setting.

In this paper, we consider n-person discounted stochastic games where no player knows the transition probabilities and/or payoffs of a game exactly. More specifically, we focus on incomplete information finite state/action stochastic games, where parameters belong to an uncertainty set over which one may not have probabilistic information. In our model, a player uses robust optimization to cope with the uncertainty, assuming that other players are robust optimizers as well. Players' robust approach is not with respect to the other players' strategies, but with respect to game data.

Stochastic games where uncertain data are known to belong to a given set can arise when the data are obtained through subjective judgments, there is a lack of data, there are random measurement errors on the data, or there are implementation imprecisions. Eliciting probability distributions through subjective judgments on unknown parameter values can be a difficult task, especially when there is lack of prior knowledge on the parameters. Decision makers may have difficulty in providing probabilistic information, and they may be more comfortable with providing simply ranges of values for the unknown data. As noted by Ben-Tal and Nemirovski (1998), identifying the underlying probability distributions can be problematic in a large-scale problem.

Hence, our motivation to consider a robust optimization approach is that it offers an alternative way to address uncertainty in the absence of distributional information. Furthermore, solutions to stochastic games may be sensitive to perturbations in payoff and transition probability data for which estimates from historical data could be inaccurate due to statistical errors. This stems from the fact that at equilibrium in a stochastic game, each player faces an MDP, and optimal solutions to MDPs can be sensitive with respect to data (Nilim and El Ghaoui 2005). Therefore, a robust optimization approach could be used to address parameter uncertainty in stochastic games.

A concrete example that could benefit from our model occurs in substitutable product inventory control problems, where different products are sold by different retailers, and customers can switch from one product to the other (see Avşar and Gürsoy 2002). A substitution rate, given by the probability that a customer switches from one product to the other, may be unknown for a new product, requiring subjective judgments. Another example can occur in a processor-sharing model, where customers are served simultaneously, and each arriving customer observes the current load and then chooses to join a shared system or uses an alternative service option. Each customer makes an individual decision, wishing to minimize his own service time (see Altman and Shimkin 1998). Due to service conditions, imprecisions in implementing a service rate, server malfunctions, and/or presence of other customers in the system, there may be uncertainty in service rates.

1.1. Summary of Contributions

Contributions of this paper are summarized as follows:

1. We provide a distribution-free robust optimization model for n-person finite state/action discounted stochastic games, extending the notion of robust one-shot games to stochastic games. Our work can also be seen as an extension of the infinite-horizon robust dynamic programming problem to a multiplayer setting.
2. We prove the existence of a robust Markov perfect equilibrium solution when players are restricted to play stationary strategies. Our proof extends the existence proof for robust one-shot games to a stochastic game setting. It extends the original result proved by Fink (1964) for the complete information case to an incomplete information setup.
3. We then show that a robust Markov perfect equilibrium point can be characterized as a solution to a multilinear system, when the uncertainty set of transition data is a polytope intersected with the probability simplex. Since players have different objective functions in general, their worst-case expectations from an uncertainty set will be different. Accordingly, our characterization provides equilibrium worst-case parameter values that a player expects from an uncertainty set.
4. We illustrate the use of discounted robust stochastic games in a queueing control problem.

1.2. Outline of the Paper

We conclude this introductory section with a literature review. In §2, we review basic ideas on discounted stochastic games and robust optimization, followed by the formulation of discounted robust stochastic games. In §3, we prove the existence of a robust Markov perfect equilibrium point when players use stationary strategies. In §4, we show that when the uncertainty set of transition data is a polytope intersected with the probability simplex, the robust equilibrium can be cast as a multilinear system formulation, the solution of which gives the set of equilibrium points of the discounted robust stochastic game. In §5, we illustrate the use of our approach on an example of a queueing control problem. Finally, §6 concludes the paper with remarks and future research directions. We include in Table A.1 in the appendix a summary of the notation used in the paper.

1.3. Literature Review

The first paper on finite state/action, two-person zero-sum stochastic games was written by Shapley (1953). Extensions to Shapley's basic model have received considerable attention in the literature. A comprehensive treatment of the subject is given by Filar and Vrieze (1997). Since stochastic games generalize MDPs and one-shot games, we next provide a literature review on addressing uncertainty in MDPs and game theory.

Robust optimization is used in the literature to cope with uncertainty in various contexts, including Markov decision processes and one-shot games. Nilim and El Ghaoui (2005), and Iyengar (2005) consider a robust optimization approach in MDPs with uncertainty in transition matrices to cope with the sensitivity of optimal policies to perturbations in data. They present independent proofs for the robust value iteration. Satia and Lave (1973), White and Eldeib (1994), and Givan et al. (1997) model an MDP, where the transition matrix lies in a given set, which is most typically a polytope. Bagnell et al. (2001) consider a similar problem and present the robust value iteration without proof. Different from a robust approach, a Bayesian approach is presented by Shapiro and Kleywegt (2002), where a prior distribution on the transition matrix should be known.

In the incomplete information game theory area, Harsanyi (1967, 1968a, b) model games where each player could use a prior probability distribution to obtain a conditional distribution on the unknown data of a game. Different from that method, a worst-case approach to payoff uncertainty is considered in one-shot games by Aghassi and Bertsimas (2006). In that paper, the authors present a distribution-free model of incomplete information one-shot games, in which the players use a robust optimization approach to contend with payoff uncertainty. The authors prove the existence of an equilibrium point when the payoff uncertainty set is bounded. They formulate the robust game by considering that the payoffs belong to a polytope, yielding a method to compute an equilibrium point. Our work extends this model and approach to stochastic games.

We note that there are precursors of a worst-case approach in game theory. Gilboa and Schmeidler (1989), Lo (1996), and Marinacci (2000) present an approach to uncertainty in the context of normal form games, based on a max-min criterion. Lo (1996) models each player as believing that his opponents' actions are realizations from an unknown probability distribution, which belongs to a set of known multiple priors. A player in that model wishes to maximize his minimum expected utility, where the minimization is with respect to the set of multiple priors. Lo (1996) is motivated by the possibility that the beliefs of a decision maker may not be representable by a probability measure and generalizes equilibrium concepts for normal form games to allow the beliefs of each player to be representable by a closed convex set of probability measures. Gilboa and Schmeidler (1989), and Marinacci (2000) adopt a similar approach, where nonadditive probability distributions are used instead of sets of multiple priors. Challenging expected utility models that could be used when probabilistic information is available, these authors argue that decision makers may lack distributional information. We note that these approaches are different from our model because they address complete information games and adopt a worst-case approach with respect to players' behaviors towards each other, rather than addressing data uncertainty.

The incomplete information case within the repeated games is first introduced by Aumann and Maschler (1968). Sorin's work (1984, 1985) considers a special class of stochastic games with incomplete information on one side that have a single nonabsorbing state. In Sorin (1984, 1985), the payoff matrix of the game is chosen according to a probability distribution. It is then shown that these games have a min-max and a max-min value. In a more recent effort, Rosenberg et al. (2004) consider two-person zero-sum games, where incomplete information is described by a finite set of stochastic games. A game is to be played out of this finite set, over which a probability distribution is specified. That paper focuses on stochastic games in which one player controls the transitions. We note that the approach adopted by Rosenberg et al. (2004) is based on the approach proposed by Aumann and Maschler (1968) and requires a probability distribution over a set of games.

2. Problem Setup

2.1. Stochastic Games

This section reviews the basics of stochastic game theory, as presented by Fink (1964) and Filar and Vrieze (1997). We denote the set of states by $\mathcal{S} = \{1, \ldots, S\}$, and the set of players by $\mathcal{I} = \{1, \ldots, I\}$. If the game is in state s, player i can choose the action a^i from a finite set of actions. We assume that each player has J actions in every state. The extension to the case where players have different number of actions in states is straightforward and involves no new insights—only more complex notation.

Suppose that each player makes a choice in a state, i.e., we have an action tuple $a = (a^1, \ldots, a^I) \in A$, where A is the set of all possible action tuples in any state. Then the game moves into state k with probability $P_{sak} \geqslant 0$, $\sum_{k=1}^{S} P_{sak} = 1$.

Let $\mathbf{x}_s^i$ be a probability distribution over player i's actions in state s. That is, the strategy of player i in state s is $\mathbf{x}_s^i = (x_{s1}^i, \ldots, x_{sJ}^i)$, which belongs to the J-dimensional probability simplex, $\Delta = \{\mathbf{x}_s^i \in \Re_+^J \mid \sum_{j=1}^{J} x_{sj}^i = 1\}$. Indeed, the set of mixed strategies for every player i in state s is Δ.

In this work, we consider stationary strategies, which prescribe a player the same probabilities for his choices each time the player visits a certain state, no matter what route he follows to reach that state. Stationary strategies have been prevalent in the study of stochastic games due to their mathematical tractability. It is known that the objective values the players achieve using stationary strategies in complete information discounted stochastic games are not worse than the values achieved using nonstationary strategies (Filar and Vrieze 1997).

Let us represent the stationary strategies of a player i by $\mathbf{x}^i = (\mathbf{x}_1^i, \ldots, \mathbf{x}_S^i)$ and denote the set of mixed strategies of all players in the state space of the game by $\mathbf{x} = (\mathbf{x}^1, \ldots, \mathbf{x}^I)$. We denote, for all states, the mixed strategies of all players except player i by $\mathbf{x}^{-i} = (\mathbf{x}^1, \ldots, \mathbf{x}^{i-1}, \mathbf{x}^{i+1}, \ldots, \mathbf{x}^I)$. The following notation is used to distinguish a mixed strategy of player i from those of others: $(\mathbf{x}^{-i}, \mathbf{u}^i) = (\mathbf{x}^1, \ldots, \mathbf{x}^{i-1}, \mathbf{u}^i, \mathbf{x}^{i+1}, \ldots, \mathbf{x}^I)$. We refer to the space of mixed strategies for all players and all states as X, e.g., $\mathbf{x} \in X$.

Suppose players choose their actions independently of each other at a given state. The probability that an action tuple a is chosen by the players in state s is denoted by

$$\pi_s^a(\mathbf{x}^{-i}, \mathbf{u}^i) = \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m u_{sa^i}^i.$$

Let C_{sa}^i be the immediate cost to player i induced by the action tuple a chosen by the players in state s. We note that the immediate costs to the players do not depend on the state to which the system transitions. Fink (1964) shows that, given all other players' strategies $\mathbf{x}^{-i}$, the unique value of the β-discounted stochastic game to player i starting the process in state s, denoted by v_s^i, is given by

$$v_s^i = \min_{u_s^i \in \Delta} \sum_{a \in A} \pi_s^a(\mathbf{x}^{-i}, \mathbf{u}^i) \left\{ C_{sa}^i + \beta \sum_{k=1}^{S} P_{sak} v_k^i \right\}, \quad \forall\, i \in \mathcal{I},\ s \in \mathcal{S}. \tag{1}$$

Equation (1) is a Bellman-type equation, and it is interpreted as follows: Given all other players' strategies fixed, if player i knew how to play optimally from the next stage on, then, at the current stage, he would select the strategy that minimizes the expected immediate cost at the current stage plus the total expected future costs. Hence, player i is not only concerned with the immediate outcome of his actions but also with the future consequences of his strategies in the current stage.

We now define equilibrium points in this setting, which are known as Markov perfect equilibria. We note that these types of equilibria form a small subset of the set of subgame perfect equilibria.

Definition 1. A point $\mathbf{x} \in X$ is a Markov perfect equilibrium point in a stochastic game if and only if, $\forall\, i \in \mathcal{I}$, $\forall\, s \in \mathcal{S}$, there exists a value v_s^i that satisfies Equation (1), such that,

$$\mathbf{x}_s^i \in \arg\min_{u_s^i \in \Delta} \left(\sum_{a \in A} \pi_s^a(\mathbf{x}^{-i}, \mathbf{u}^i) \left\{ C_{sa}^i + \beta \sum_{k=1}^{S} P_{sak} v_k^i \right\} \right). \qquad (2)$$

This definition states that $\mathbf{x}$ is an equilibrium point if and only if no player can improve his cost by unilaterally changing his strategy. That is, when Equation (1) is considered for every player and state, the minimization of the right-hand side yields back the same vector $\mathbf{x}$. Stated in another way, $\mathbf{x}$ is an equilibrium point if each player's strategy is a best response to the other players' strategies. Here, a player's best response to other players is obtained via Equation (1). We note that this equilibrium concept is the same as Nash. Fink (1964) proves that an equilibrium point exists in n-person nonzero sum discounted stochastic games, extending the original result by Nash (1950) to stochastic games.

2.2. Robust Optimization

This section briefly reviews the basics of robust optimization, as introduced by Ben-Tal and Nemirovski (1998). Consider the following optimization problem P_γ: $\min_{x \in \Re^n} f(\mathbf{x}, \boldsymbol{\gamma})$ s.t. $F(\mathbf{x}, \boldsymbol{\gamma}) \in K \subset \Re^m$, where $\boldsymbol{\gamma} \in \Re^M$ is the data vector, $\mathbf{x} \in \Re^n$ is the decision vector, and K is a convex cone. Suppose that the data of P_γ is uncertain and all that is known about the data is that it belongs to an uncertainty set $U \in \Re^M$. Now, consider the problem $P = \{P_\gamma\}_{\gamma \in U}$, where the constraints $F(\mathbf{x}, \boldsymbol{\gamma}) \in K$ must be satisfied no matter what the actual realization of $\boldsymbol{\gamma} \in U$ is. An optimal solution to the uncertain problem P is defined as a solution that gives the best possible guaranteed value under all possible realizations of constraints. Formally, it should be an optimal solution of the following program P_R: $\min_{x \in \Re^n} \{\sup_{\gamma \in U} f(\mathbf{x}, \boldsymbol{\gamma})$ s.t. $F(\mathbf{x}, \boldsymbol{\gamma}) \in K,\ \forall\, \boldsymbol{\gamma} \in U\}$. Problem P_R is called the robust counterpart of P, and its feasible and optimal solutions are called robust feasible and robust optimal solutions, respectively (Ben-Tal and Nemirovski 1998). Prior work (Bertsimas and Sim 2004, Ben-Tal and Nemirovski 1998) has shown that for many function types and uncertainty sets, the robust counterpart problem P_R can be solved as a single optimization problem of size comparable to a deterministic version of the problem.

2.3. Formulation of Robust Stochastic Games

In this section, we formalize our robust model for incomplete information stochastic games by considering that both payoffs and transition probabilities of a game belong to respective uncertainty sets. In discounted robust stochastic games, it is assumed that the players commonly know the uncertainty set of payoffs C_s at each state, the set of transition probabilities P_s out of each state, and that they all take a robust optimization approach. Different from, and as an alternative to the approach in Rosenberg et al. (2004), players do not have distributional information on the uncertainty, and they adopt a robust approach using stationary strategies.

We assume in our model that the uncertain parameters are realized anew every time a state is visited and that they are realized from the same, state dependent, uncertainty sets. Since the players cannot engage in learning this game beyond the uncertainty sets, stationary policies stand as a natural choice in our model.

We would like to note that assuming stationary strategies restricts our focus to an infinite horizon model. This is because optimal strategies in finite horizon models are nonstationary in general. Also, as demonstrated in the game "Big Match," for limiting average stochastic games, the value of a game does not need to exist within the class of stationary strategies (Blackwell and Ferguson 1968, Vrieze 2004). This result shows that, in general, nonstationary strategies are essential for limiting average stochastic games. We note that, compared to nonstationary strategies, whether using stationary strategies in discounted *robust* stochastic games results in a loss of optimality is an open research problem.

Now, in light of the results summarized in the previous section, we notice the following: If a player knew how to play in the robust stochastic game optimally from the next stage on, then, at the current stage, he would play with strategies that minimize the maximum sum of expected immediate costs at the current stage and expected costs possibly incurred in future stages. Hence, if optimal robust values for player i exist, given $\mathbf{x}^{-i}$, they must satisfy the following Bellman-type equation,

$$\omega_s^i = \min_{\mathbf{u}_s^i \in \Delta} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \sum_{a \in A} \pi_s^a(\mathbf{x}^{-i}, \mathbf{u}^i) \left\{ \tilde{C}_{sa}^i + \beta \sum_{k=1}^{S} \tilde{P}_{sak} \omega_k^i \right\}, \qquad (3)$$

where the inner maximization problem is with respect to the uncertain transition probabilities and immediate costs. Here, for each $s \in \mathcal{S}$, elements of the set C_s are vectors $\tilde{C}_s = [\tilde{C}_{sa}^i]_{i \in \mathcal{I},\, a \in A}$. That is, for each player i and action tuple a in state s, there is an uncertain immediate cost $\tilde{C}_{sa}^i$ that varies within the set C_s. The vectors $\tilde{P}_s = [\tilde{P}_{sak}]_{a \in A,\, k \in \mathcal{S}}$ are elements of the set P_s. We provide the conditions under which the robust values exist (and Equation (3) holds) in the next section when we study the existence of equilibrium (Theorem 2).

To ease the notation, let us define

$$\psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i) = \sum_{a \in A} \pi_s^a(\mathbf{x}^{-i}, \mathbf{u}^i) \left\{ \tilde{C}_{sa}^i + \beta \sum_{k=1}^{S} \tilde{P}_{sak} w_k^i \right\},$$

where $\mathbf{x}_s^{-i}$ denotes, for state s, the mixed strategies of all players except player i. We are now ready to state our definition of robust Markov perfect equilibrium in discounted robust stochastic games.

DEFINITION 2. A point $\mathbf{x}$ is a *robust Markov perfect equilibrium* point in a discounted robust stochastic game if and only if, $\exists \boldsymbol{\omega} = [\boldsymbol{\omega}^1, \ldots, \boldsymbol{\omega}^I]$ satisfying Equation (3), such that, $\forall\, i \in \mathcal{I}$, $\forall\, s \in \mathcal{S}$,

$$\mathbf{x}_s^i \in \arg\min_{\mathbf{u}_s^i \in \Delta} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s;\, \mathbf{x}_s^{-i}, \mathbf{u}_s^i;\, \boldsymbol{\omega}^i). \tag{4}$$

This definition states that $\mathbf{x}$ is a robust Markov perfect equilibrium if each player's strategy is a robust best response to the other players' strategies. While calculating his robust best response via Equation (3), player i assumes a worst-case perspective of the uncertain parameters. To interpret, given all other players' mixed strategies are fixed, player i assumes that for any mixed strategy considered, nature would choose parameters that constitute the worst-case for himself.

In general, since players have different objective functions, their worst-case expectations from an uncertainty set will be different. For example, if one player pays the other a fixed immediate cost for each action tuple (as in a zero-sum game) and there is uncertainty only in the transition data, players' resulting robust value expectations will not sum up to zero. We present such an example in §5. We note that this phenomenon is also observed in the context of robust one-shot games by Aghassi and Bertsimas (2006).

3. Existence of Equilibrium

Our proof of existence of equilibrium points in discounted robust stochastic games parallels Fink (1964). However, a different point-to-set mapping (correspondence) is defined that takes the robustness into account. This mapping uses a maximum expected total cost function with respect to mixed strategies. The proof is separated into two parts: The first part shows that for any strategy $\mathbf{x}$, there exists a unique robust value vector for a player. The second part uses Kakutani's fixed point theorem (Theorem 3) to show that the correspondence we consider has a fixed point that coincides with an equilibrium point.

Before we begin, we introduce some additional notation and present Kakutani's fixed point theorem (Kakutani 1941). In the following proofs, C_s and P_s are closed and bounded sets.

Let $W^i = \Re^{\mathcal{S}}$ be the space of robust game values for player i and let $W = \Pi_{i\in\mathcal{I}} W^i$. The infinity norm on W is given by $\|\boldsymbol{\omega}\|_\infty = \max_{i\in\mathcal{I},\, s\in\mathcal{S}} |\omega_s^i|$. Let $f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i)$ be the worst-case expected cost to player i in state s, i.e.,

$$f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i) = \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s;\, \mathbf{x}_s^{-i}, \mathbf{u}_s^i;\, \boldsymbol{\omega}^i).$$

We next define a best response function which, given the strategies of all other players $\mathbf{x}_s^{-i}$, takes any robust value vector $\boldsymbol{\omega}^i$ and minimizes the maximum expected total cost with respect to player i's mixed strategies. Let γ_s^i: $X_s^{-i} \times W^i \to \Re$ be defined by

$$\gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i) = \min_{\mathbf{u}_s^i \in \Delta} f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i),$$

where X_s^{-i} denotes the strategy space of all players in state s except player i.

Part I: The next theorem shows that the best response function for a player is a contraction mapping, and Theorem 2 below shows that a unique robust value vector exists for any given strategy of the players.

THEOREM 1. *For* $\mathbf{x} \in X$, *define* $\gamma_x(\boldsymbol{\omega})$: $W \to W$ *by* $(\gamma_x(\boldsymbol{\omega}))_{is} = \gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i)$. *The function* $\gamma_x(\boldsymbol{\omega})$ *is a contraction mapping.*

When all other players' strategies are fixed, player i faces a robust MDP. Hence, proof of Theorem 1 follows directly from Theorem 5 given by Iyengar (2005). For completeness, we also give an alternative proof of Theorem 1 in the appendix.

THEOREM 2 (APPLICATION OF BANACH'S THEOREM). *For any* $\mathbf{x} \in X$, *and* $\forall\, i \in \mathcal{I}$, $s \in \mathcal{S}$, *there exists a unique* ω_s^i *such that*

$$\omega_s^i = \min_{\mathbf{u}_s^i \in \Delta} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s;\, \mathbf{x}_s^{-i}, \mathbf{u}_s^i;\, \boldsymbol{\omega}^i) = \gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i).$$

PROOF. Note that $(W, \|\cdot\|_\infty)$ is a complete metric space and by Theorem 1, γ_x: $W \to W$ is a contraction mapping. Therefore, by Banach's Theorem, $\gamma_x(\boldsymbol{\omega})$ has a unique fixed point, $\boldsymbol{\omega}$. That is, there exists a unique vector, $\boldsymbol{\omega}$, such that $\gamma_x(\boldsymbol{\omega}) = \boldsymbol{\omega}$, which means

$$\begin{aligned}\omega_s^i &= \min_{\mathbf{u}_s^i \in \Delta} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s;\, \mathbf{x}_s^{-i}, \mathbf{u}_s^i;\, \boldsymbol{\omega}^i) \\ &= \gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i), \quad \forall\, i \in \mathcal{I},\ s \in \mathcal{S}. \quad \square\end{aligned}$$

We define the unique robust best response values for player i by

$$\boldsymbol{\tau}^i(\mathbf{x}^{-i}) = \left\{ \boldsymbol{\omega}^i = (\omega_1^i, \ldots, \omega_S^i): \right. \\ \left. \omega_s^i = \min_{\mathbf{u}_s^i \in \Delta} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s;\, \mathbf{x}_s^{-i}, \mathbf{u}_s^i;\, \boldsymbol{\omega}^i),\ s \in \mathcal{S} \right\},$$

and denote the sth element of $\boldsymbol{\tau}^i(\mathbf{x}^{-i})$ by $\tau_s^i(\mathbf{x}^{-i})$.

Part II: We now show the existence of an equilibrium point that satisfies conditions (4) by using Kakutani's fixed point theorem, which we present below.

DEFINITION 3. A correspondence ϕ: $X \to 2^X$ is upper semicontinuous if $\mathbf{y}^n \in \phi(\mathbf{x}^n)$, $\lim_{n\to\infty} \mathbf{x}^n = \mathbf{x}$, $\lim_{n\to\infty} \mathbf{y}^n = \mathbf{y}$ imply that $\mathbf{y} \in \phi(\mathbf{x})$.

Theorem 3 (Kakutani's Fixed Point Theorem). *If X is a closed, bounded, and convex set in a Euclidean space, and ϕ is an upper semicontinuous correspondence mapping X into the family of closed, convex subsets of X, then $\exists\, \mathbf{x} \in X$, s.t. $\mathbf{x} \in \phi(\mathbf{x})$.*

We will show that the fixed point of a suitably constructed correspondence is an equilibrium point. To this end, let

$$\phi(\mathbf{x}) = \left\{ \mathbf{y} \in X \mid \mathbf{y}_s^i \in \arg\min_{u_s^i \in \Delta} f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i), \right. \\ \left. \omega_s^i = \min_{u_s^i \in \Delta} f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i),\ \forall\, s \in \mathcal{S},\ i \in \mathcal{I} \right\}.$$

To show that this correspondence satisfies the assumptions of Kakutani's theorem, we first need several technical results. Let us denote, for all players, the space of mixed strategies in state s by X_s, i.e., $\mathbf{x}_s \in X_s$. Since $\mathbf{x}_s \in \Re^{IJ}$ and $\boldsymbol{\omega}^i \in \Re^S$, we consider the metrics induced by the infinity norm in each space:

$$d_{X_s}(\mathbf{x}_s, \mathbf{y}_s) = \max_{i \in \mathcal{I},\ j \in \{1,\ldots,J\}} |x_{sj}^i - y_{sj}^i|,$$

$$d_{W^i}(\boldsymbol{\omega}^i, \boldsymbol{\theta}^i) = \max_{s \in \mathcal{S}} |w_s^i - \theta_s^i|.$$

For the strategy vectors $\mathbf{x}_s$ and $\mathbf{u}_s$, and for the value vectors $\boldsymbol{\omega}^i$ and $\boldsymbol{\theta}^i$, let $\mathbf{p} = (\mathbf{x}_s, \boldsymbol{\omega}^i)$, $\mathbf{q} = (\mathbf{y}_s, \boldsymbol{\theta}^i)$. $d_1(\mathbf{p}, \mathbf{q}) = d_{X_s}(\mathbf{x}_s, \mathbf{y}_s) + d_{W^i}(\boldsymbol{\omega}^i, \boldsymbol{\theta}^i)$. We need the following lemma to show that f_s^i satisfies the properties needed to use Kakutani's theorem.

Lemma 1. *Given $\epsilon > 0$, $\exists \delta(\epsilon) > 0$ such that if for any $\mathbf{p}, \mathbf{q} \in X_s \times W^i$, $d_1(\mathbf{p}, \mathbf{q}) < \delta(\epsilon)$, then, $\forall\, \tilde{C}_s \in C_s$, $\forall\, \tilde{P}_s \in P_s$, $|\psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s, \boldsymbol{\omega}^i) - \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{y}_s, \boldsymbol{\theta}^i)| < \epsilon$.*

Proof. Since $\tilde{C}_s \in C_s$ and C_s is bounded $\forall\, s \in \mathcal{S}$, we have $|\tilde{C}_{sa}^i| \leqslant K$, where $K < \infty$. It is clear that robust values are bounded. Hence, we have $\forall\, i \in \mathcal{I}$, $s \in \mathcal{S}$, that $|\omega_s^i| \leqslant H$, where $H < \infty$. Note that

$$\begin{aligned}
&|\psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s, \boldsymbol{\omega}^i) - \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{y}_s, \boldsymbol{\theta}^i)| \\
&\quad = \left| \sum_{a \in A} \prod_{m=1}^{I} x_{sa^m}^m \tilde{C}_{sa}^i + \beta \sum_{a \in A} \left(\prod_{m=1}^{I} x_{sa^m}^m \right) \left(\sum_{k=1}^{S} \tilde{P}_{sak} \omega_k^i \right) \right. \\
&\qquad \left. - \sum_{a \in A} \prod_{m=1}^{I} y_{sa^m}^m \tilde{C}_{sa}^i - \beta \sum_{a \in A} \left(\prod_{m=1}^{I} y_{sa^m}^m \right) \left(\sum_{k=1}^{S} \tilde{P}_{sak} \theta_k^i \right) \right| \\
&\quad = \left| \sum_{a \in A} \tilde{C}_{sa}^i \left(\prod_{m=1}^{I} x_{sa^m}^m - \prod_{m=1}^{I} y_{sa^m}^m \right) \right. \\
&\qquad \left. + \beta \sum_{a \in A} \sum_{k=1}^{S} \tilde{P}_{sak} \left(\prod_{m=1}^{I} x_{sa^m}^m \omega_k^i - \prod_{m=1}^{I} y_{sa^m}^m \theta_k^i \right) \right| \\
&\quad \leqslant \left| \sum_{a \in A} \tilde{C}_{sa}^i \left(\prod_{m=1}^{I} x_{sa^m}^m - \prod_{m=1}^{I} y_{sa^m}^m \right) \right| \\
&\qquad + \left| \beta \sum_{a \in A} \sum_{k=1}^{S} \tilde{P}_{sak} \left(\prod_{m=1}^{I} x_{sa^m}^m \omega_k^i - \prod_{m=1}^{I} y_{sa^m}^m \theta_k^i \right) \right| \\
&\quad \leqslant K \sum_{a \in A} \left| \prod_{m=1}^{I} x_{sa^m}^m - \prod_{m=1}^{I} y_{sa^m}^m \right| \\
&\qquad + \beta \sum_{a \in A} \sum_{k=1}^{S} \left| \prod_{m=1}^{I} x_{sa^m}^m \omega_k^i - \prod_{m=1}^{I} y_{sa^m}^m \theta_k^i \right|.
\end{aligned}$$

The second to last inequality above follows from the triangle inequality. The last inequality follows because we have $|\tilde{C}_{sa}^i| \leqslant K$ and $\tilde{P}_{sak} \leqslant 1$, $\forall\, i \in \mathcal{I}$, $s \in \mathcal{S}$, $a \in A$, $k \in S$.

Let

$$\delta_1(\epsilon) = \frac{\min\{\epsilon, 1\}}{3K(2^I - 1)J^I}, \quad \delta_2(\epsilon) = \frac{\min\{\epsilon, 1\}}{3S\beta J^I},$$

$$\delta_3(\epsilon) = \frac{\min\{\epsilon, 1\}}{3HS\beta(2^I - 1)J^I},$$

and let $\delta(\epsilon) = \min\{\delta_1(\epsilon), \delta_2(\epsilon), \delta_3(\epsilon)\}$. Now, $d_1(\mathbf{p}, \mathbf{q}) < \delta(\epsilon)$ implies that, $\forall\, i \in \mathcal{I}$, $s \in \mathcal{S}$, and for all actions a^i, $x_{sa^m}^m = y_{sa^m}^m + \alpha_{sa^m}^m$ and $\omega_s^i = \theta_s^i + \gamma_s^i$, where $|\alpha_{sa^m}^m| < \delta(\epsilon)$, and $|\gamma_s^i| < \delta(\epsilon)$. We will make use of the following algebraic identity.

$$\left| \prod_{m=1}^{I} (y_{sa^m}^m + \alpha_{sa^m}^m) - \prod_{m=1}^{I} y_{sa^m}^m \right| = \left| \sum_{\substack{\Omega \subseteq \mathcal{I} \\ |\Omega| \geqslant 1}} \left(\prod_{m \in \Omega} \alpha_{sa^m}^m \right) \left(\prod_{m \in \Omega^C} y_{sa^m}^m \right) \right|,$$

where $\Omega^C = \mathcal{I} \backslash \Omega$. Note that $\prod_{m \in \Omega} |\alpha_{sa^m}^m| < (\delta_1(\epsilon))^{|\Omega|} \leqslant \delta_1(\epsilon)$, and that

$$\left| \sum_{\substack{\Omega \subseteq \mathcal{I} \\ |\Omega| \geqslant 1}} \left(\prod_{m \in \Omega} \alpha_{sa^m}^m \right) \left(\prod_{m \in \Omega^C} y_{sa^m}^m \right) \right| \leqslant \sum_{\substack{\Omega \subseteq \mathcal{I} \\ |\Omega| \geqslant 1}} \left| \prod_{m \in \Omega} \alpha_{sa^m}^m \right| \left| \prod_{m \in \Omega^C} y_{sa^m}^m \right|.$$

Hence, we have

$$\begin{aligned}
&K \sum_{a \in A} \left| \prod_{m=1}^{I} (y_{sa^m}^m + \alpha_{sa^m}^m) - \prod_{m=1}^{I} y_{sa^m}^m \right| \\
&\quad \leqslant K \sum_{a \in A} \sum_{\substack{\Omega \subseteq \mathcal{I} \\ |\Omega| \geqslant 1}} \left| \prod_{m \in \Omega} \alpha_{sa^m}^m \right| \left| \prod_{m \in \Omega^C} y_{sa^m}^m \right| \\
&\quad \leqslant K \sum_{a \in A} \sum_{\substack{\Omega \subseteq \mathcal{I} \\ |\Omega| \geqslant 1}} \left| \prod_{m \in \Omega} \alpha_{sa^m}^m \right| < K \sum_{a \in A} \sum_{\substack{\Omega \subseteq \mathcal{I} \\ |\Omega| \geqslant 1}} \gamma_1(\epsilon) = \frac{\epsilon}{3}.
\end{aligned}$$

We also have

$$\begin{aligned}
&\beta \sum_{a \in A} \sum_{k=1}^{S} \left| \prod_{m=1}^{I} x_{sa^m}^m \omega_k^i - \prod_{m=1}^{I} y_{sa^m}^m \theta_k^i \right| \\
&= \beta \sum_{a \in A} \sum_{k=1}^{S} \left| \prod_{m=1}^{I} (y_{sa^m}^m + \alpha_{sa^m}^m) \omega_k^i - \prod_{m=1}^{I} y_{sa^m}^m \theta_k^i \right| \\
&= \beta \sum_{a \in A} \sum_{k=1}^{S} \left| \prod_{m=1}^{I} y_{sa^m}^m (\omega_k^i - \theta_k^i) + \omega_k^i \sum_{\substack{\Omega \subseteq \mathcal{I} \\ |\Omega| \geqslant 1}} \prod_{m \in \Omega} \alpha_{sa^m}^m \prod_{m \in \Omega^C} y_{sa^m}^m \right| \\
&\leqslant \beta \sum_{a \in A} \sum_{k=1}^{S} \left| \prod_{m=1}^{I} y_{sa^m}^m \right| |(\omega_k^i - \theta_k^i)|
\end{aligned}$$

$$+\beta H\sum_{a\in A}\sum_{k=1}^{S}\sum_{\substack{\Omega\subseteq\mathcal{I}\\|\Omega|\geqslant 1}}\left|\prod_{m\in\Omega}\alpha_{sa^m}^m\right|\left|\prod_{m\in\Omega^C}y_{sa^m}^m\right| \tag{5}$$

$$\leqslant \beta\sum_{a\in A}\sum_{k=1}^{S}|\gamma_s^i|+\beta H\sum_{a\in A}\sum_{k=1}^{S}\sum_{\substack{\Omega\subseteq\mathcal{I}\\|\Omega|\geqslant 1}}\left|\prod_{m\in\Omega}\alpha_{sa^m}^m\right| \tag{6}$$

$$<\beta\sum_{a\in A}\sum_{k=1}^{S}\delta_2(\epsilon)+\beta H\sum_{a\in A}\sum_{k=1}^{S}\sum_{\substack{\Omega\subseteq\mathcal{I}\\|\Omega|\geqslant 1}}\delta_3(\epsilon) \tag{7}$$

$$=\frac{\epsilon}{3}+\frac{\epsilon}{3}=2\frac{\epsilon}{3}.$$

The inequality ($\leqslant$) in (5) above follows from the triangle inequality, from the fact that the robust values are bounded, and from the facts

$$\left|\prod_{m=1}^{I}y_{sa^m}^m(\omega_k^i-\theta_k^i)\right|\leqslant\left|\prod_{m=1}^{I}y_{sa^m}^m\right||(\omega_k^i-\theta_k^i)|$$

and

$$\left|\sum_{\substack{\Omega\subseteq\mathcal{I}\\|\Omega|\geqslant 1}}\prod_{m\in\Omega}\alpha_{sa^m}^m\prod_{m\in\Omega^C}y_{sa^m}^m\right|\leqslant\sum_{\substack{\Omega\subseteq\mathcal{I}\\|\Omega|\geqslant 1}}\left|\prod_{m\in\Omega}\alpha_{sa^m}^m\right|\left|\prod_{m\in\Omega^C}y_{sa^m}^m\right|.$$

The inequality ($\leqslant$) in (6) above follows because $|\prod_{m\in\Omega^C}y_{sa^m}^m|\leqslant 1$, and $\omega_s^i-\theta_s^i=\gamma_s^i$.

The inequality ($<$) in (7) follows because

$$|\gamma_s^i|<\delta(\epsilon)\leqslant\delta_2(\epsilon),\quad\text{and}$$

$$\left|\prod_{m\in\Omega}\alpha_{sa^m}^m\right|\leqslant|\alpha_{sa^m}^m|<\delta(\epsilon)\leqslant\delta_3(\epsilon).$$

Thus,

$$\begin{aligned}|\psi_s^i(\tilde{C}_s,\tilde{P}_s;\mathbf{x}_s,\boldsymbol{\omega}^i)-\psi_s^i(\tilde{C}_s,\tilde{P}_s;\mathbf{y}_s,\boldsymbol{\theta}^i)|&\\ \leqslant K\sum_{a\in A}\left|\prod_{m=1}^{I}x_{sa^m}^m-\prod_{m=1}^{I}y_{sa^m}^m\right|&\\ +\beta\sum_{a\in A}\sum_{k=1}^{S}\left|\prod_{m=1}^{I}x_{sa^m}^m\omega_k^i-\prod_{m=1}^{I}y_{sa^m}^m\theta_k^i\right|&\\ <\frac{\epsilon}{3}+2\frac{\epsilon}{3}=\epsilon.\quad\Box&\end{aligned}$$

Lemma 1 proves the equicontinuity of the set of functions $\{\psi_s^i(\tilde{C}_s,\tilde{P}_s;\mathbf{x}_s,\boldsymbol{\omega}^i),\tilde{C}_s\in C_s,\ \tilde{P}_s\in P_s\}$. This is a key result in our existence proof. It is first used to show the continuity of the function $f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$ in Lemma 2. The continuity of the function $f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$ is then used in the main existence theorem below (Theorem 4). Lemma 1 is also needed to establish the upper semicontinuity result used in Theorem 4.

Lemma 2 below follows from the basic real analysis result that states that the pointwise maximum of a family of equicontinuous functions is continuous. We also provide a detailed proof for our specific functions f_s^i and ψ_s^i in the appendix for completeness. Lemma 3 below follows from the definition of f_s^i.

LEMMA 2. *The function $f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$ is continuous in all of its variables $\forall\, i\in\mathcal{I}$, and $s\in\mathcal{S}$.*

LEMMA 3. *$f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$ is convex in $\mathbf{u}_s^i$ for fixed $\mathbf{x}_s^{-i}$ and $\boldsymbol{\omega}^i$.*

The following two technical results are the final ingredients needed to show the upper semicontinuity of the correspondence $\phi(\mathbf{x})$. Proof of Lemma 4 below follows directly from Fink (1964) and Lemma 1 above. Proof of Lemma 5 follows directly from Lemma 4 as shown in Fink (1964). These proofs are presented using our notation in the appendix. Lemma 4 is used to prove Lemma 5, and Lemma 5 is used to show the upper semicontinuity result required by Kakutani's fixed point theorem.

LEMMA 4. *$\gamma_s^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)$ is continuous in $\mathbf{x}_s^{-i}$. Furthermore, the set $\{\gamma_s^i(\cdot,\boldsymbol{\omega}^i)\mid\boldsymbol{\omega}^i$ is bounded$\}$ is equicontinuous.*

LEMMA 5. *If $\mathbf{x}^{-i,n}\to\mathbf{x}^{-i}$ and $\tau_s^i(\mathbf{x}^{-i,n})\to\omega_s^i$ as $n\to\infty$, then $\tau_s^i(\mathbf{x}^{-i})=\omega_s^i$.*

We are now ready to prove the main result of this section.

THEOREM 4 (EXISTENCE OF EQUILIBRIUM IN ROBUST STOCHASTIC GAMES). *Suppose that uncertain transition probabilities and payoffs in a discounted robust stochastic game belong to compact sets and that the set of actions and players, who use stationary strategies, are finite. Then, an equilibrium point of this game exists.*

PROOF. By Lemma 2, $f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$ is continuous in its variables. Since the minimum of this continuous function on a compact set Δ exists, $\arg\min_{\mathbf{u}_s^i\in\Delta}f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)\neq\varnothing$.

Also, by Theorem 2, the equality in the expression $\omega_s^i=\min_{\mathbf{u}_s^i\in\Delta}f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$ can be established. Therefore, $\phi(\mathbf{x})\neq\varnothing$. Note that by definition, $\phi(\mathbf{x})\subseteq X$, $\forall\,\mathbf{x}\in X$.

Next, we show that $\phi(\mathbf{x})$ is a convex set. Suppose that $(\mathbf{z}^1,\ldots,\mathbf{z}^I),(\mathbf{v}^1,\ldots,\mathbf{v}^I)\in\phi(\mathbf{x}^1,\ldots,\mathbf{x}^I)$. Then, $\forall\,\mathbf{u}_s^i$, and $s\in\mathcal{S}$, $i\in\mathcal{I}$, $\omega_s^i=f_s^i(\mathbf{x}_s^{-i},\mathbf{z}_s^i;\boldsymbol{\omega}^i)=f_s^i(\mathbf{x}_s^{-i},\mathbf{v}_s^i;\boldsymbol{\omega}^i)\leqslant f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$. Hence, for any $\lambda\in[0,1]$ and $\forall\, i\in\mathcal{I}$, $s\in\mathcal{S}$,

$$\begin{aligned}\omega_s^i&=\lambda f_s^i(\mathbf{x}_s^{-i},\mathbf{z}_s^i;\boldsymbol{\omega}^i)+(1-\lambda)f_s^i(\mathbf{x}_s^{-i},\mathbf{v}_s^i;\boldsymbol{\omega}^i)\\&\leqslant f(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i).\end{aligned}$$

By the convexity of $f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)$, we obtain

$$\begin{aligned}f_s^i(\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i)&\geqslant\omega_s^i\\&=\lambda f_s^i(\mathbf{x}_s^{-i},\mathbf{z}_s^i;\boldsymbol{\omega}^i)+(1-\lambda)f_s^i(\mathbf{x}_s^{-i},\mathbf{v}_s^i;\boldsymbol{\omega}^i)\\&\geqslant f_s^i(\mathbf{x}_s^{-i},((\lambda)\mathbf{z}_s^i+(1-\lambda)\mathbf{v}_s^i);\boldsymbol{\omega}^i)\geqslant\omega_s^i,\end{aligned}$$

and hence, $(\lambda)(\mathbf{z}^1,\ldots,\mathbf{z}^I)+(1-\lambda)(\mathbf{v}^1,\ldots,\mathbf{v}^I)\in\phi(\mathbf{x}^1,\ldots,\mathbf{x}^I)$.

Finally, we must show that $\phi(\mathbf{x})$ is an upper semicontinuous correspondence. Suppose $\mathbf{x}_n\to\mathbf{x}$, $\mathbf{y}_n\to\mathbf{y}$, and $\mathbf{y}_n\in\phi(\mathbf{x}_n)$. Taking a subsequence, we can consider

$\tau_s^i(\mathbf{x}^{-i,n}) \to \omega_s^i$. Using the triangle inequality, we have $\forall\, i \in \mathcal{I}$, $s \in \mathcal{S}$ that

$$\begin{aligned}
&|f_s^i(\mathbf{x}_s^{-i}, \mathbf{y}_s^i; \boldsymbol{\omega}^i) - \omega_s^i| \\
&\quad \leqslant |f_s^i(\mathbf{x}_s^{-i}, \mathbf{y}_s^i; \boldsymbol{\omega}^i) - f_s^i(\mathbf{x}_s^{-i,n}, \mathbf{y}_s^{i,n}; \boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))| \\
&\qquad + |f_s^i(\mathbf{x}_s^{-i,n}, \mathbf{y}_s^{i,n}; \boldsymbol{\tau}^i(\mathbf{x}^{-i,n})) - \omega_s^i| \\
&\quad = |f_s^i(\mathbf{x}_s^{-i}, \mathbf{y}_s^i; \boldsymbol{\omega}^i) - f_s^i(\mathbf{x}_s^{-i,n}, \mathbf{y}_s^{i,n}; \boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))| \\
&\qquad + |\tau_s^i(\mathbf{x}^{-i,n}) - \omega_s^i| \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Therefore, $\omega_s^i = f_s^i(\mathbf{x}_s^{-i}, \mathbf{y}_s^i; \boldsymbol{\omega}^i)$. By Lemma 5, we also have $\tau_s^i(\mathbf{x}^{-i}) = w_s^i$. Thus, we obtain that

$$\omega_s^i = f_s^i(\mathbf{x}_s^{-i}, \mathbf{y}_s^i; \boldsymbol{\omega}^i) = \tau_s^i(\mathbf{x}^{-i}) = \min_{\mathbf{u}_s^i \in \Delta} f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i).$$

Therefore, $\mathbf{y} \in \phi(\mathbf{x})$, completing the proof that ϕ is an upper semicontinuous correspondence. The fact that $\phi(\mathbf{x})$ is a closed set for any $\mathbf{x}$ follows from the fact that it is an upper-semicontinuous correspondence. Therefore, ϕ satisfies the assumptions of Kakutani's fixed point theorem. □

4. A Multilinear System Formulation for Robust Markov Perfect Equilibria

Now that we have proved the existence of an equilibrium point in a discounted robust stochastic game, our next step is to calculate such a point. We will show that when the uncertainty in the probability transition data of a game belongs to a polytope intersected with the probability simplex, the problem of finding an equilibrium point could be cast as a multilinear system formulation. The characterization result we present here generalizes a previous result for normal form one-shot games by Aghassi and Bertsimas (2006) to stochastic games. For simplicity, we only consider uncertainty in the transition data in constructing a formulation that characterizes equilibrium points. An analogous approach can be used to consider uncertainty both in payoffs and transition data of a game.

Recall the definition of a robust Markov perfect equilibrium and conditions (4). These conditions are equivalent to the requirement that $\forall\, i \in \mathcal{I}, s \in \mathcal{S}, \exists \omega_s^i \in \Re$ such that $(\mathbf{x}_s^i, \omega_s^i)$ is an optimizer of the following robust mathematical program P_R:

$$(P_R) \quad \left\{ \min_{u_s^i, q_s^i} q_s^i \colon q_s^i \geqslant \max_{\tilde{P}_s \in P_s} \psi_s^i(C_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i),\ \mathbf{u}_s^i \in \Delta \right\}.$$

Here, $(\mathbf{x}^{-i}, \boldsymbol{\omega}^i)$ is treated as data. Define the uncertain probability transition matrix induced by a strategy $(\mathbf{x}^{-i}, \mathbf{u}^i)$:

$$\tilde{\mathbf{P}}(\mathbf{x}^{-i}, \mathbf{u}^i) = \left[\sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m u_{sa^i}^i \tilde{P}_{sak} \right]_{s=1,\, k=1}^{S,\, S}.$$

Denote the sth row of $\tilde{\mathbf{P}}(\mathbf{x}^{-i}, \mathbf{u}^i)$ by the vector $[\tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)]'$. Let $\tilde{\mathbf{p}}_s$ denote the uncertain transition probability vector associated with the starting state s, that is, $\tilde{\mathbf{p}}_s = [\tilde{P}_{sak}]_{a \in A;\, k \in S}$. Let $\mathbf{1}$ be a column vector of ones of appropriate dimension. Let $\mathbf{E}_s^i(\mathbf{x}_s^{-i}, C^i) \in \Re^{(J^{I-1}) \times J}$ denote the matrix whose rows are given by the vectors

$$\left[\prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m C_{s(a^{-i}, a^i)}^i \right]_{a^i \in \{1, \ldots, J\}}. \tag{8}$$

Note that we represent an action tuple a by (a^{-i}, a^i). Here, a^{-i} denotes an action tuple comprising all players' actions except player i. That is, $a^{-i} = (a^1, \ldots, a^{i-1}, a^{i+1}, \ldots, a^I)$. The number of possible action tuples for all players except player i is J^{I-1}, and hence the matrix $\mathbf{E}_s^i(\mathbf{x}_s^{-i}, C^i)$ has J^{I-1} rows. We have the following requirement in P_R:

$$\begin{aligned}
q_s^i &\geqslant \max_{\tilde{P}_s \in P_s} \psi_s^i(C_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i) \\
&= \max_{\tilde{P}_s} \beta [\tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)]' \boldsymbol{\omega}^i + \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, C^i) \mathbf{u}_s^i.
\end{aligned} \tag{9}$$

Note that, on the right-hand side of the equality in expression (9), the first term yields the discounted total expected future cost and the second term yields the total expected immediate cost in state s, induced by the strategy $(\mathbf{x}^{-i}, \mathbf{u}^i)$.

We assume that the uncertain transition probabilities out of a state s belong to a polytope intersected with the probability simplex. That is, transition probabilities belong to the following uncertainty set:

$$P = \{\tilde{\mathbf{p}}_s, s \in \mathcal{S} \colon \mathbf{A}_s \tilde{\mathbf{p}}_s \geqslant \mathbf{b}_s, \mathbf{Q}_s \tilde{\mathbf{p}}_s = \mathbf{1},\ \tilde{\mathbf{p}}_s \geqslant 0\}, \tag{10}$$

where $\mathbf{A}_s \in \Re^{l \times SJ^I}$. The number of rows l of $\mathbf{A}_s$ can be chosen appropriately to obtain a desired polytope. In this uncertainty set, $\mathbf{Q}_s \in \Re^{J^I \times SJ^I}$ is a matrix of 0s and 1s (such that, when multiplied by $\tilde{\mathbf{p}}_s$, each of its rows corresponding to a pure strategy tuple $\mathbf{a}$ yields $\sum_{k \in S} \tilde{P}_{sak} = 1$).

Consider the maximization problem in P_R, where $(\mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i)$ is regarded as data. Given that the uncertainty set is as stated, for fixed $(\mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i)$, this maximization problem is equivalent to the following LP:

$$\left\{ \max_{\tilde{\mathbf{p}}_s} \beta [\tilde{\mathbf{p}}_s(\mathbf{x}^{-i}, \mathbf{u}^i)]' \boldsymbol{\omega}^i \colon \mathbf{A}_s \tilde{\mathbf{p}}_s \geqslant \mathbf{b}_s, \mathbf{Q}_s \tilde{\mathbf{p}}_s = \mathbf{1}, \tilde{\mathbf{p}}_s \geqslant 0 \right\}. \tag{11}$$

Let us denote the set of all action tuples for all players except player i by A^{-i}, i.e., $a^{-i} \in A^{-i}$. Define the column vector $\mathbf{z}_s^i \in \Re^{J^I S}$ as

$$\mathbf{z}_s^i = \left[\prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m u_{sa^i}^i \omega_k^i \right]_{a^{-i} \in A^{-i};\, a^i \in A^i;\, k \in S} \tag{12}$$

such that the indices of $\mathbf{z}_s^i$ match the indices of $\tilde{\mathbf{p}}_s$. Let $\mathbf{Y}_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i) \in \Re^{J^I S \times J}$ be the matrix such that

$$\mathbf{Y}_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i) \mathbf{u}_s^i = \mathbf{z}_s^i. \tag{13}$$

Let $\mathbf{m}_s^i$ and $\mathbf{n}_s^i$ be the dual variable vectors of problem (11). The dual of problem (11) is

$$\left\{\min_{m_s,\,n_s}[[\mathbf{b}_s]';[\mathbf{1}]']\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}:\begin{bmatrix}\mathbf{A}_s\\ \mathbf{Q}_s\end{bmatrix}'\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}\geqslant\beta\mathbf{Y}_s^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)\mathbf{u}_s^i,\ \mathbf{m}_s^i\leqslant 0\right\}. \tag{14}$$

By the definition of our uncertainty set, problem (11) is feasible, and it is clear that it is bounded. By strong duality, problem (14) is bounded, feasible, and its optimal objective value is equal to that of problem (11). Therefore, if $(\mathbf{x}^{-i},\mathbf{u}^i,\boldsymbol{\omega}^i)$ satisfies condition (9), then (9) is equivalent to the condition that $\exists\mathbf{m}_s^i\in\Re^l$ and $\exists\mathbf{n}_s^i\in\Re^{J^l}$ such that

$$q_s^i-\mathbf{1}'\mathbf{E}_s^i(\mathbf{x}_s^{-i},C^i)\mathbf{u}_s^i\geqslant[[\mathbf{b}_s]';[\mathbf{1}]']\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix} \tag{15}$$

$$\begin{bmatrix}\mathbf{A}_s\\ \mathbf{1}_s\end{bmatrix}'\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}\geqslant\beta\mathbf{Y}_s^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)\mathbf{u}_s^i$$

$$\mathbf{m}_s^i\leqslant 0.$$

Conversely, if condition (15) is satisfied, then problem (14) is feasible. Then by weak duality, any feasible solution $[[\mathbf{b}_s]';[\mathbf{1}]'][\begin{smallmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{smallmatrix}]$ of problem (14) is greater than or equal to any solution $\beta[\tilde{\mathbf{p}}_s(\mathbf{x}^{-i},\mathbf{u}^i)]'\boldsymbol{\omega}^i$ of problem (11), so,

$$q_s^i-\mathbf{1}'\mathbf{E}_s^i(\mathbf{x}_s^{-i},C^i)\mathbf{u}_s^i\geqslant[[\mathbf{b}_s]';[\mathbf{1}]']\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}\geqslant\max_{\tilde{\mathbf{p}}_s}\beta[\tilde{\mathbf{p}}_s(\mathbf{x}^{-i},\mathbf{u}^i)]'\boldsymbol{\omega}^i.$$

Therefore conditions (9) and (15) are equivalent. This proves the following lemma:

Lemma 6. *Condition* (9) *is equivalent to condition* (15).

Let

$$\mathbf{T}^i(\mathbf{x})=\left[\sum_{a\in A}\prod_{m=1}^{I}x_{sa^m}^m t_{sak}^i\right]_{s=1,\,k=1}^{S,\,S}, \tag{16}$$

and denote the sth row of $\mathbf{T}^i(\mathbf{x})$ by the vector $[\mathbf{t}_s^i(\mathbf{x})]'$. Let $\mathbf{t}_s^i$ denote the variables representing the transition probabilities adopted by player i according to i's worst-case perspective, associated with the starting state s. That is, $\mathbf{t}_s^i=[t_{sak}^i]_{a\in A;\,k\in S}$.

Theorem 5. *A stationary strategy* $\mathbf{x}$ *is a robust Markov perfect equilibrium point with the robust value vector* $\boldsymbol{\omega}^i$, *iff* $\forall\, i\in\mathcal{I}, s\in\mathcal{S}, \exists\mathbf{m}_s^i\in\Re^l, \mathbf{n}_s^i\in\Re^{J^l}, \mathbf{t}_s^i\in\Re^{SJ^l}$ *such that for* $j=1,\ldots,J$, $(\boldsymbol{\omega}^i,\mathbf{x}_s,\mathbf{m}_s^i,\mathbf{n}_s^i,\mathbf{t}_s^i)$ *satisfies*

$$\omega_s^i=\mathbf{1}'\mathbf{E}_s^i(\mathbf{x}_s^{-i},C^i)\mathbf{x}_s^i+\beta[\mathbf{t}_s^i(\mathbf{x})]'\boldsymbol{\omega}^i$$

$$\mathbf{e}_j'\mathbf{E}_s'^i(\mathbf{x}_s^{-i},C^i)\mathbf{1}+\beta\mathbf{e}_j'\mathbf{Y}_s'^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)\mathbf{t}_s^i\geqslant\omega_s^i\geqslant[[\mathbf{b}_s]';[\mathbf{1}]']\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}+\mathbf{1}'\mathbf{E}_s^i(\mathbf{x}_s^{-i},C^i)\mathbf{x}_s^i$$

$$\begin{bmatrix}\mathbf{A}_s\\ \mathbf{Q}_s\end{bmatrix}'\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}-\beta\mathbf{Y}_s^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)\mathbf{x}_s^i\geqslant 0$$

$$\mathbf{1}'\mathbf{x}_s^i=1,\quad \mathbf{m}_s^i\leqslant 0,\quad \mathbf{x}_s^i\geqslant 0,\quad \mathbf{A}_s\mathbf{t}_s^i\geqslant\mathbf{b}_s,\quad \mathbf{Q}_s\mathbf{t}_s^i=\mathbf{1},$$

where $\mathbf{e}_j$ *is the jth unit column vector of dimension J*, $\mathbf{1}$ *is a column vector of all ones of appropriate dimension*, $\mathbf{E}_s^i(\mathbf{x}_s^{-i},C^i)$ *is obtained from expression* (8), $\mathbf{A}_s$, $\mathbf{Q}_s$, *and* $\mathbf{b}_s$ *are as given in the uncertainty set* (10), $\mathbf{Y}_s^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)$ *is the matrix given by Equation* (13), *and* $\mathbf{t}_s^i(\mathbf{x})$ *is given by Equation* (16).

Proof. Recall problem P_R. By conditions (4) and Lemma 6, if $\mathbf{x}$ is a robust Markov perfect equilibrium point, given $(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)$, $\forall\, i\in\mathcal{I}, s\in\mathcal{S}, \exists\mathbf{m}_s^{i*}\in\Re^l, \mathbf{n}_s^{i*}\in\Re^{J^l}$ such that $(\mathbf{x}_s^i,\omega_s^i,\mathbf{m}_s^{i*},\mathbf{n}_s^{i*})$ is an optimizer of

$$\min_{u_s^i,\,q_s^i,\,m_s^i,\,n_s^i} q_s^i \tag{17}$$

$$q_s^i-\mathbf{1}'\mathbf{E}_s^i(\mathbf{x}_s^{-i},C^i)\mathbf{u}_s^i\geqslant[[\mathbf{b}_s]';[\mathbf{1}]']\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{A}_s\\ \mathbf{Q}_s\end{bmatrix}'\begin{bmatrix}\mathbf{m}_s^i\\ \mathbf{n}_s^i\end{bmatrix}\geqslant\beta\mathbf{Y}_s^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)\mathbf{u}_s^i,$$

$$\mathbf{m}_s^i\leqslant 0,\ \mathbf{1}'\mathbf{u}_s^i=1,\ \mathbf{u}_s^i\geqslant 0.$$

The dual of the above is

$$\max_{\nu_s^i,\,t_s^i}\nu_s^i\colon\ \mathbf{A}_s\mathbf{t}_s^i\geqslant\mathbf{b}_s,\quad \mathbf{Q}_s\mathbf{t}_s^i=\mathbf{1},$$

$$\nu_s^i\leqslant\mathbf{e}_j'\mathbf{E}_s'^i(\mathbf{x}_s^{-i},C^i)\mathbf{1}+\beta\mathbf{e}_j'\mathbf{Y}_s'^i(\mathbf{x}_s^{-i},\boldsymbol{\omega}^i)\mathbf{t}_s^i,\quad j=1,\ldots,J. \tag{18}$$

The statement in the theorem follows from strong duality and Theorem 2. For the other direction, suppose that $\forall\, i\in\mathcal{I}, s\in\mathcal{S}$, $(\boldsymbol{\omega}^{i*},\mathbf{x}_s^*,\mathbf{m}_s^{i*},\mathbf{n}_s^{i*},\mathbf{t}_s^{i*})$ satisfies the system in the statement of the theorem. Let

$$\nu_s^i=\min_{j\in\{1,\ldots,J\}}\mathbf{e}_j'\mathbf{E}_s'^i(\mathbf{x}_s^{-i*},C^i)\mathbf{1}+\beta\mathbf{e}_j'\mathbf{Y}_s'^i(\mathbf{x}_s^{-i*},\boldsymbol{\omega}^{i*})\mathbf{t}_s^{i*},$$

$$q_s^i=[[\mathbf{b}_s]';[\mathbf{1}]']\begin{bmatrix}\mathbf{m}_s^{i*}\\ \mathbf{n}_s^{i*}\end{bmatrix}+\mathbf{1}'\mathbf{E}_s^i(\mathbf{x}_s^{-i*},C^i)\mathbf{x}_s^{i*}.$$

Then, given $(\mathbf{x}_s^{-i*},\boldsymbol{\omega}^{i*})$, $\forall\, i\in\mathcal{I}, s\in\mathcal{S}$, $(\mathbf{x}_s^{i*},q_s^i,\mathbf{m}_s^{i*},\mathbf{n}_s^{i*})$ is feasible for problem (17), and $(\nu_s^i,\mathbf{t}_s^{i*})$ is feasible for problem (18) with $\nu_s^i\geqslant q_s^i$. By weak duality $\nu_s^i\leqslant q_s^i$, so $\nu_s^i=q_s^i$. Hence, $(\mathbf{x}_s^{i*},q_s^i,\mathbf{m}_s^{i*},\mathbf{n}_s^{i*})$ is optimal for problem (17). Equivalently, $(\mathbf{x}_s^{i*},q_s^i)$ is optimal in P_R. Therefore, $\mathbf{x}^*$ is an equilibrium point of the discounted robust stochastic game. □

Theorem 5 states that when the uncertainty set for the transition probability data of a game is a polytope intersected with the probability simplex, the set of robust Markov perfect equilibria can be characterized as a multilinear system formulation. In a discounted zero-sum stochastic game, there can be multiple equilibrium points with the same equilibrium value. Consequently, equilibria in zero-sum games are simpler to compute (Filar and Vrieze 1997). However, there may be multiple equilibrium points and nonunique equilibrium values in robust stochastic games. Therefore, the robust version of even the zero-sum case becomes more difficult to solve than its complete information version.

The formulation in Theorem 5 can be used to solve large-scale problems to the extent that we can solve large scale-multilinear systems. In general, solving these nonconvex problems is a difficult task and developing efficient algorithms is an active area of research. Recent algorithms using nonlinear methods to solve multilinear systems that arise from one-shot games are able to find solutions to problems with up to four players and four actions per player in less than five minutes (Datta 2003, Aghassi and Bertsimas 2006). Similarly, homotopy methods have been used to solve multilinear systems that arise from complete information stochastic games for two players, two actions per player, and five states in less than a minute (Herings and Peeters 2004).

In the next section, we use an existing solver LOQO (Vanderbei 2006) to solve the problem defined in Theorem 5 for an application in queuing control. This problem is a two-person stochastic game with 30 states and two actions per player in each state. LOQO is a generic solver for nonlinear, nonconvex problems and is able to solve the instances of our problem within four minutes on a Dell workstation with a CPU at 3.20 GHz, 2 GB RAM, using the Red Hat Linux 3 operating system.

5. A Queueing Control Application

Game theoretical analysis has been widely applied to queueing control problems (Altman and Shimkin 1993, 1998; Heyman 1968; Sobel 1969, Stidham and Weber 1989; Yechiali 1971; Altman 1994a, b; Altman and Hordijk 1995). In this section, we present an application of our robust model for incomplete information stochastic games to a single-server exponential queuing system.

Consider a queueing control problem encountered in packet switched networks. The most well-known packet switched networks are the Internet and local area networks (Peterson 2007). In these networks, packets (blocks of data) are routed among nodes over data links shared with other traffic. Here, a node is a server connected to the network. For example, a server can be a computer or a personal digital assistant. The service rate of a server can be set to different levels and is controlled by a service provider (player 1)—for instance, by increasing or decreasing the processing capacity of a server at a node. The service rate may change in time in an unpredictable way (see Altman 1994a, b). The reasons for this can be the imprecision during the implementation of an intended service rate due to operating conditions, service conditions, presence of other packets in the system, and/or server malfunctions.

In packet switched networks, packets are routed by a programmable physical device, called a router (player 2). A router dynamically controls the flow of arriving packets into a finite buffer at a server. The rates that the service provider and the router choose depend on the number of packets in the system. This allows players to choose rates in order to address congestion and throughput (Altman 1994b). In fact, it is to the benefit of a service provider to increase the amount of packets processed in the system. However, such an increase may result in an increase in packets' waiting times in the buffer (called latency), and routers are used to reduce packets' waiting times. The game theoretical nature of the problem arises because the service provider and the router have conflicting objectives. We model this problem as a zero-sum stochastic game, where a player is incurred payoffs that are modeled as being paid to the other player. In our model, having uncertainty in the service rates results in uncertainty in the transition probabilities. This affects both players, who use stationary strategies and deal with the uncertainty using a robust optimization approach. Next, we present the details of our model.

The state space is $\mathcal{S} = \{0, 1, \ldots, S\}$, where $S < \infty$ is the maximum number of packets that can be present in the system. Only one packet can be in service at any time, with the remaining packets waiting for service in the buffer. The router admits one packet into the system at a time. Every time a state is visited, the service provider and the router simultaneously choose a service rate μ and an admission rate λ from their respective finite set of rates. We assume that any selected service rate μ can deviate unpredictably within a given interval, i.e., $\mu \in (\mu_{\min}, \mu_{\max})$. For instance, if the service provider intends to have a service rate of 1 packet per 20 seconds, in practice he may observe that the actual service rates vary between 1 packet per 19 seconds to 1 packet per 21 seconds. Suppose that there are s packets in the system and the players choose the action tuple (μ, λ). Then the router is incurred a holding cost $h(s)$, and a cost $\theta(\mu, \lambda)$ associated with having packets served at rate μ when it admits packets at rate λ. If there are no packets in the system, this cost represents the setup cost of the server. These payoffs are modeled as being paid to the service provider, since the players' objectives are in conflict. The service provider, in turn, pays the router $\rho(\mu, \lambda)$, which represents the reward to the router for choosing the rate λ. It can also be interpreted as the setup cost of the router.

We assume that the time until the admission of a new packet and the next service completion are both exponentially distributed with means $1/\lambda$ and $1/\mu$, respectively. We can, therefore, represent the number of packets in the system with a birth and death process, which has the following state transition probabilities:

$$\tilde{p}(k \mid s, \mu, \lambda) = \begin{cases} \mu/(\lambda+\mu), & 1 \leqslant s \leqslant S, k = s-1 \\ \lambda/(\lambda+\mu), & 0 \leqslant s \leqslant S-1, k = s+1 \\ 1, & s = 0, k = 1 \\ 1, & s = S, k = S-1. \end{cases} \tag{19}$$

Although there is uncertainty only in the service rates, both players face uncertainty in the transition data among the states. This is so because the transition probabilities depend not only on the admission rates but also on the service rates. For instance, if the players choose μ and λ in a state, with $\mu \in (\mu^{\min}, \mu^{\max})$, then the transition probability to the next state would be minimum $\lambda/(\lambda+\mu^{\max})$ and maximum $\lambda/(\lambda+\mu^{\min})$.

For every fixed strategy of the players in a state, the uncertainty in service rates results in interval uncertainty on the transition probabilities. Note that since this uncertainty is resolved after the players select their strategies, the entire transition probability interval is valid. Therefore, the uncertainty set on the transition probabilities out of each state is represented by a polyhedron formed by the intersection of these intervals and the probability simplex.

5.1. Problem Description

We set up an example of the above problem with the router and the service provider having two (pure) actions in each state. For simplicity, we keep the two actions for each player the same for every state. However, since the players can use mixed strategies, the rates they choose in effect may differ over the states. The router's first action, denoted by $\bar{\lambda}$, is to admit a packet into the system every 10 seconds. Its second action, $\underline{\lambda}$, is to admit a packet every 25 seconds. The service provider's first action is to serve a packet every 11 seconds. We call this rate the intended service rate. For instance, due to the uncertainty, if the service rate is set to 11 seconds, it may vary between 10 and 12 seconds. We denote his first action by $\bar{\mu}$; hence we have $\bar{\mu} \in [1/12, 1/10]$. The service provider's second action is to serve a packet every 20 seconds, which may vary between 19 and 21 seconds due to uncertainty. Therefore, we have $\underline{\mu} \in [1/21, 1/19]$. Hence, the interval length within which the service rate varies is 2 seconds per packet for each action.

In this example, we use an exponential holding cost $h(s) = ab^{\alpha s}$, $s \geqslant 1$ with $a = 1.2$, $b = 1.9$, and $\alpha = 0.2$. The holding cost when there are no packets in the system is 0. For each state, we let $\theta(\bar{\mu}, \bar{\lambda}) = \theta(\bar{\mu}, \underline{\lambda}) = 110$, $\theta(\underline{\mu}, \bar{\lambda}) = \theta(\underline{\mu}, \underline{\lambda}) = 90$, $\rho(\bar{\mu}, \bar{\lambda}) = 60$, $\rho(\bar{\mu}, \underline{\lambda}) = 30$, $\rho(\underline{\mu}, \bar{\lambda}) = 20$, $\rho(\underline{\mu}, \underline{\lambda}) = 70$.

This payoff scenario indicates that the router pays the service provider more when the service rate is higher. It also indicates that the reward that the router receives is higher when both players choose their first actions (higher rates), or second actions (lower rates). In other words, the router receives a smaller reward when the admission and service rates are inconsistent; that is, when the service rate is relatively higher compared to the admission rate, or vice versa.

We use $S = 30$. Actions for the router at the boundary state 30 represent the setup options for the router so that it can control admissions at specified rates when the system is not full. When the system is full, packets are dropped, and ρ represents the setup cost of the router. Similarly, actions of the server at state 0 represent the setup options. If there are no packets in the system, θ represents the setup cost of the server so that, given an admission rate, it can be controlled to operate at a required rate at the next state. When the system is full or empty, players still choose a strategy based on the setup costs. However, their strategies at the boundary states 0 and 30 do not affect the transition probabilities out of these states, which are 1 according to formula (19). Therefore, there can be at most 30 packets in the system.

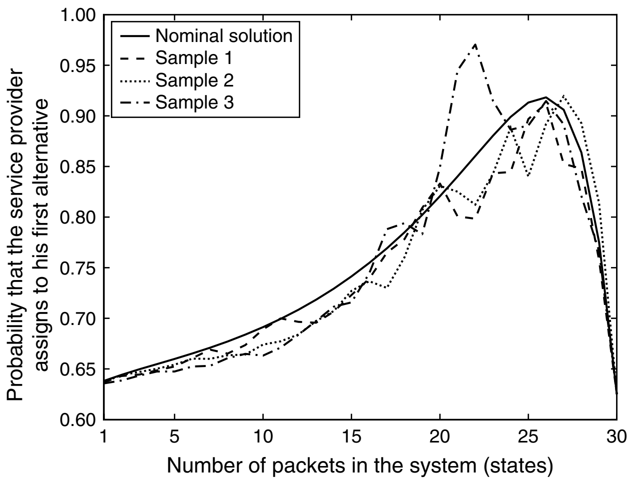
We generate four more instances of this problem by increasing the interval length for service rates to 6, 10, 14, and 18 seconds per packet, keeping the intended service rates the same. Intervals of service rates for each instance are depicted in Table 1. For instance, when the interval length is 10 seconds per packet (for the instance number 3), we have $\bar{\mu} \in [1/16, 1/6]$, and $\underline{\mu} \in [1/25, 1/15]$ for states 1 through 30. Therefore, given the players choose an action tuple in a state, the difference between the minimum and maximum transition probabilities to another state becomes larger for larger interval lengths. For example, when the interval length is 10, if the players choose their second actions in a state, the transition probability to the next state would be minimum 0.3753, and maximum 0.5.

The purpose of Figure 1 below is to demonstrate that the nominal solution of this example is sensitive to perturbations in service rates. To this end, for simplicity, we use an a-priori sample for each state and service rate action. To generate Figure 1, we solve the nominal problem using the nonlinear programming formulation for zero-sum stochastic games given by Filar and Vrieze (1997). We disregard the uncertainty in the nominal problem and assume that the possible service rates are fixed at $\bar{\mu} = 1/11$ and $\underline{\mu} = 1/20$, the midpoints of their respective intervals. The probability that the service provider assigns to his first action (i.e., the nominal solution strategy) is represented by the solid line in Figure 1. Then, for each state and first service rate action, we take a service rate sample from the uniform distribution defined on the first interval of instance 3, i.e., on $[1/16, 1/6]$. We do the same for the service provider's second action, using the second interval of instance 3, i.e., using $[1/25, 1/15]$. Therefore, we obtain a sample for each

Table 1. Intervals of rates for different instances.

Instance number	1	2	3	4	5
$\bar{\mu}$	[1/12, 1/10]	[1/14, 1/8]	[1/16, 1/6]	[1/18, 1/4]	[1/20, 1/2]
$\underline{\mu}$	[1/21, 1/19]	[1/23, 1/17]	[1/25, 1/15]	[1/27, 1/13]	[1/29, 1/11]

Figure 1. Sensitivity of the nominal solution.

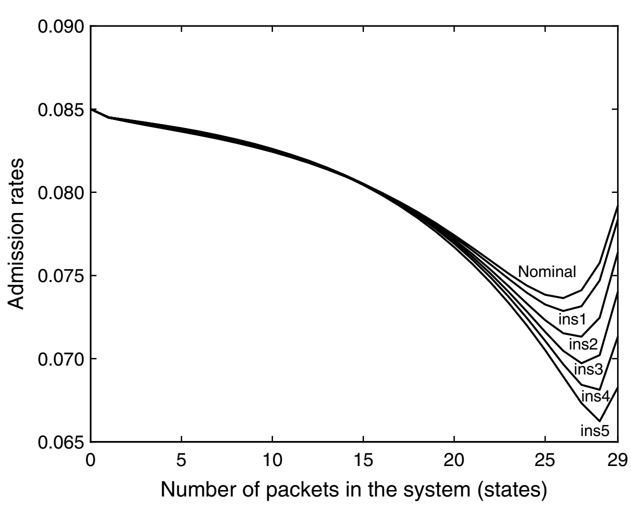


state and service rate action, solve the resulting nominal game using the probabilities given by the sampled service rates, and plot the solution represented by a dashed line (sample 1) in Figure 1. We repeat the process for two more sets of samples represented by dashed lines. We observe that the server equilibrium strategy is sensitive to the sampled values of the service rates, with markedly different solutions depending on the samples.

5.2. Equilibrium Solution

Now we investigate how the robust and nominal solutions differ for different instances of our problem. In particular, we are interested in calculating the admission and service rates, average number of packets, average amount of time a packet spends in the system, and the average value of the game for the players given by the nominal and robust solutions. To this end, for each instance, we calculate the intervals that the transition probabilities belong to. We then

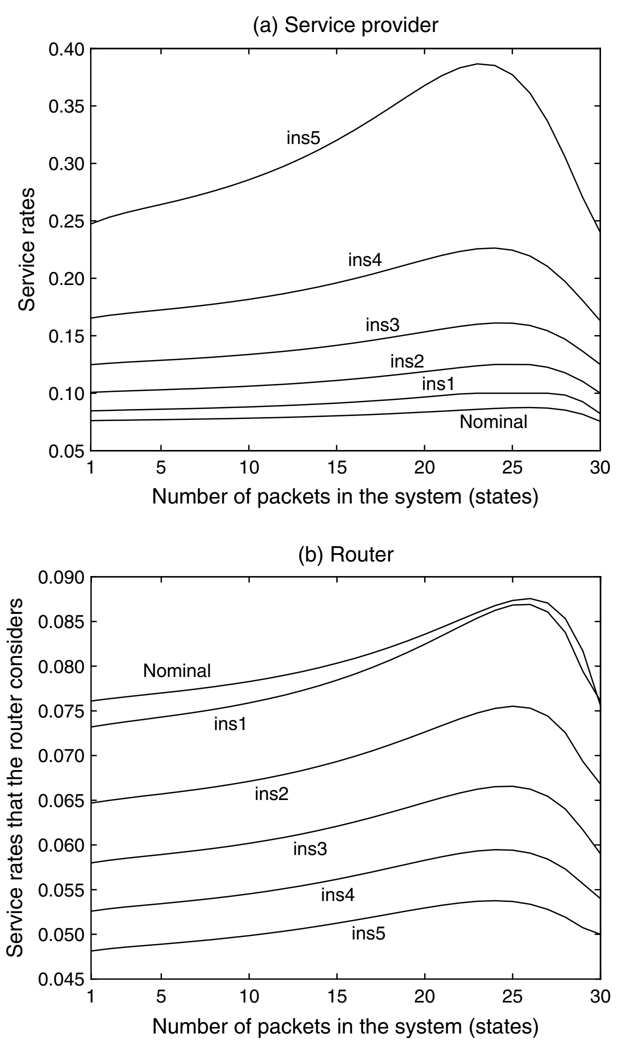
Figure 2. Admission rates.



solve each instance using Theorem 5 with the calculated lower and upper bounds on the transition data, the payoffs given above, and with the discount factor $\beta = 0.95$. A feature of Theorem 5 is that it yields the equilibrium strategies $\mathbf{x}_s^i$ and the probabilities $\mathbf{t}_s^i$ that each player expects from the uncertainty set.

Figure 2 depicts the equilibrium admission rates for the nominal and robust solutions. This admission rate is obtained by combining the two possible admission rate actions with the optimal router strategy. The service rate cannot be computed directly from the server strategy because it also depends on players' worst-case transition data expectations. That is, each player expects a different service rate at the equilibrium. We present in Figure 3 the equilibrium service rates expected by the server and the router, respectively. These rates are computed by using the formula of transition probabilities for a birth-death process (19) along with the solution weighted transition probabilities of each player given by $\mathbf{T}^i(\mathbf{x})$ in Lemma 6.

Figure 3. Service rates expected by the service provider and the router.



We note from these figures that the router expects that the service provider decreases the service rates as the uncertainty gets larger, and therefore tends to decrease the admission rates to protect himself against congestion. On the other hand, service rates the service provider expects increase as the uncertainty set becomes larger, which is a pessimistic approach, because this might result in a decrease in the queue size and, consequently, in a decrease in his overall profit.

Next, we calculate steady state probabilities, the average number of packets in the system (L), the average amount of time a packet spends in the system (W), and the average value (AV) of the game from the point of view of each player. The AV for a player is calculated by weighting the value to that player in a state by the respective steady state probabilities from that player's perspective and taking the summation over states.

The results are depicted in Figures 4 and 5. Figure 4(a) indicates that as far as the service provider is concerned, the average number of customers in the system is less than that of the nominal solution when there is uncertainty in the system. From the router's perspective, L increases as the intervals become larger. Note that these are pessimistic points of views for both players, because an increase in the number of packets in the system would be an advantage for the service provider, whereas it would be a disadvantage for the router. Figure 4(b) indicates that the average waiting time for a packet decreases from the service provider's perspective and increases from the router's perspective. The reason for this is that in equilibrium, the service provider assumes the pessimistic perspective of having fewer customers in the system, whereas the router assumes the opposite. Accordingly, as the uncertainty sets get larger, AV of the game to the service provider (i.e., the overall profit that he makes) decreases, whereas AV to the router (i.e., the overall cost of the game to the router) increases. Note that, although this stochastic game is zero-sum, the resulting values differ for each player when they play robustly. This example illustrates that although the problem is a zero-sum game, formulations for zero-sum games cannot be used to

Figure 4. Average number of packets in the system (L) and average time a packet spends in the system (W).

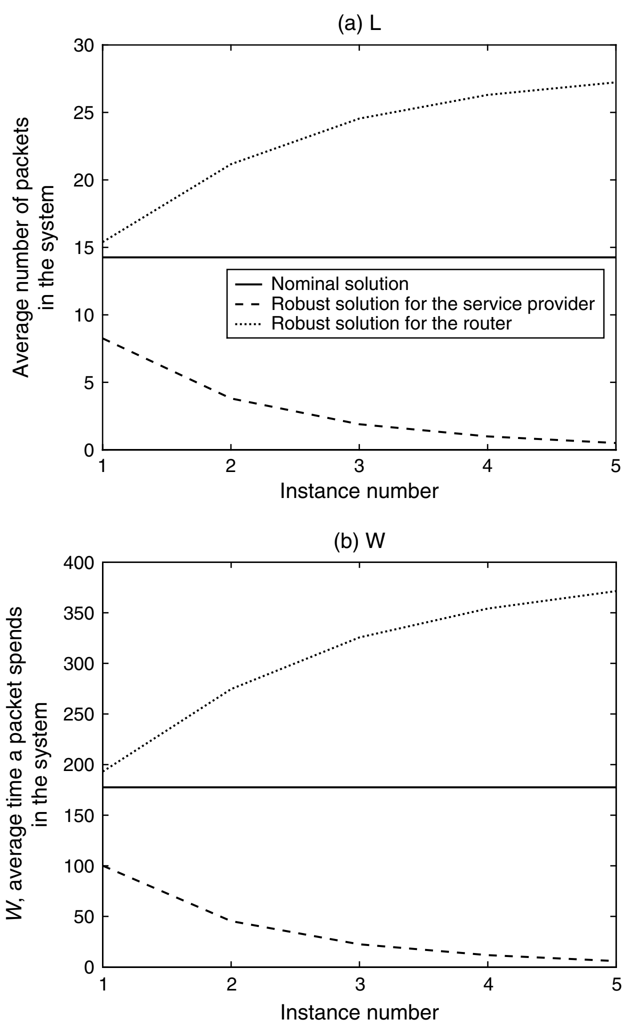


Figure 5. Average values for the players.

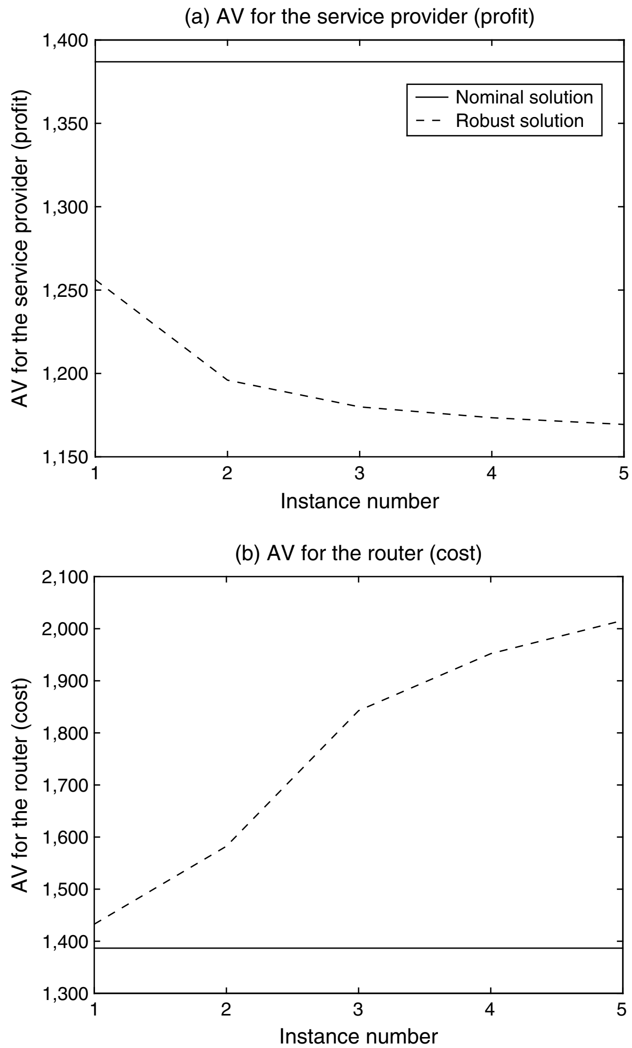


Table 2. Means and standard deviations of the values obtained in simulation for different equilibrium strategies.

Instance number	Nominal Mean	Nominal Stdv	Robust Mean	Robust Stdv	% stdv reduction
1	1,194	16.1	1,192	15.8	1.9
2	1,196	21.0	1,196	20.5	2.6
3	1,194	16.8	1,193	15.0	10.6
4	1,194	17.9	1,193	15.5	13.7
5	1,192	17.8	1,192	16.8	5.8

solve discounted robust stochastic games, despite the fact that one player pays the other player a fixed amount.

5.3. Simulation

Since solutions to stochastic games can be sensitive to changes in data, one motivation for using a robust approach is to reduce the sensitivity of a solution to data perturbations. Our first purpose in this section is to compare the means and standard deviations of the values given by nominal and robust strategies, when the service rates vary within their respective intervals. The second purpose of this section is to study the effect of the uncertainty sets on the robust value estimates and the actual expected values players achieve when the service rates vary.

Accordingly, for each instance, we fix the robust and nominal solution strategies for both players, sample the service rates, and calculate the corresponding value of the game via a discrete event simulation, starting from the empty (initial) state. Every time a state is visited in a simulation, service rates are sampled from uniform distributions given by the corresponding interval for each service rate action. We ran 300 simulations for each instance, terminating each simulation run after the value of the game starting at the empty state converges to the 10th decimal point. Means and standard deviations of the values starting the process from the empty initial state for each instance are depicted in Table 2.

Table 2 depicts that means obtained using robust strategies are about the same as those obtained using nominal strategies. It presents that the standard deviations obtained using robust strategies are lower than those obtained using nominal strategies. This is to the benefit of both players as lower standard deviations imply less sensitivity to perturbations in the data or more stability in the system. In this example, we observe that the standard deviation reduction seems to increase with the size of the uncertainty set; however, as instance 5 shows, this is not always the case as the robust equilibrium could have players adopting overly conservative strategies.

The next set of results show the differences between robust value estimates and actual expected values achieved in the simulation for this example. The second and third columns in Table 3 give robust value estimates for the service provider's profit (RobS) and the router's cost (RobR), respectively, for the initial state 0. The fourth column is the difference between these estimates. The fifth and sixth columns present the percent differences between robust value estimates and actual expected values achieved. They are calculated using the robust value estimates and the means given in Table 2 by fixing the robust strategies in the simulation. The seventh column presents for each instance the frequency that a simulated value is greater than the robust value estimate for the service provider, when robust strategies are used in the simulation. Similarly, the last column depicts the frequency that a simulated value is less than the robust value estimate for the router. The last two columns indicate that, when both players are using robust strategies, the frequencies with which their simulation output outperforms their robust value estimates increase with the size of the uncertainty.

In this zero-sum model, router minimizes (over the mixed strategies) the maximum expected costs, where the maximum is taken with respect to the transition uncertainty. On the other hand, the service provider maximizes over his mixed strategies the minimum expected costs. At equilibrium, player 1 considers one point in the uncertainty set and player 2 considers a different point, which provide their worst-case expected values. Therefore, as the uncertainty set gets larger, the worst-case outcomes of the uncertainty become farther apart, and consequently the difference between the robust value estimates and sampled expected values players achieve increases.

6. Concluding Remarks and Future Research

In this paper, we consider n-person, nonzero-sum discounted stochastic games in which none of the players

Table 3. Differences between robust value estimates and actual expected values achieved in simulation.

Inst no	RobS	RobR	Diff	$\frac{\text{(mean} - \text{RobS)}}{\text{RobS}}$	$\frac{\text{(RobR} - \text{mean)}}{\text{RobR}}$	Freq val > RobS	Freq val < RobR
1	1,185.3	1,192.2	6.9	0.6	0.02	0.65	0.57
2	1,179.8	1,200.6	20.8	1.4	0.4	0.83	0.76
3	1,175.4	1,210.5	35.0	1.5	1.5	0.94	0.85
4	1,171.9	1,221.4	49.5	1.8	2.4	0.99	0.95
5	1,168.8	1,232.9	64.1	2.0	3.3	0.99	0.96

knows the true transition probabilities and/or payoffs of a game and each player adopts a robust optimization approach to data uncertainty. We offer an alternative equilibrium concept for stochastic games with incomplete information. We propose a distribution-free model that lends itself to computational results via a multilinear system formulation that characterizes equilibrium points. We finally illustrate the use of discounted robust stochastic games in a queueing control example.

We observe the following points in this research. First, an equilibrium exists even if there exist players who do not adopt a robust optimization approach. This stems from the fact that when there are no uncertainty sets for the data of a stochastic game, best response functions are already continuous, as shown in Fink (1964). Hence, we can construct a correspondence that satisfies Kakutani's theorem and takes into account the players who disregard uncertainty. Second, if the uncertainty in a discounted stochastic game is a common set for the players, the zero-sum property of the equilibrium values do not necessarily hold. This is so since if there is uncertainty in any data of a game, the players' approach to this uncertainty may differ. This is the case even if a game is zero-sum and there is uncertainty only in transition data. Therefore, formulations for zero-sum stochastic games could not be used for analyses, and properties that pertain to zero-sum games may not hold in the presence of uncertainty.

In the example of §5, the robust solution's strategies yield lower standard deviations compared to those of the nominal solution. This may not always be the case because a robust solution can cause the players to adopt overly conservative strategies or there may be instances that are not sensitive to perturbations in the data. When players are overly conservative and the service rates are sampled from large intervals, there may be substantial discrepancies between the sampled rates and the rates that the players expect. Therefore, high standard deviations can be observed as a result of overly conservative strategies. Future research should investigate in more detail the problem conditions and/or a-priori measures to identify when a robust formulation of a stochastic game is preferable. Another research direction would be taking the players' risk attitudes into account and investigating the player behavior for which a robust approach is most beneficial. We note that whether using stationary strategies in robust stochastic games results in a loss of optimality is not known. A future research direction would be to tackle this problem and to extend the model proposed in this paper to finite horizon and limiting average stochastic games.

Appendix

To clarify the notation used throughout the paper, we provide a notation summary in Table A.1. In this paper, matrices are denoted by boldfaced capital, and vectors by boldfaced lowercase letters. For a matrix $\mathbf{A}$, $\mathbf{A}'$ denotes its transpose. The remainder of this appendix presents the statement of Banach's contraction mapping theorem, as well as the detailed proofs for Lemmas 2, 4, and 5.

Theorem (Banach's Contraction Mapping Theorem). *Let (W, ρ) be a complete metric space and let γ: $W \to W$ be a contraction mapping. Then there exists a unique fixed point of the function γ.*

Lemma 2. *The function $f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i)$ is continuous in all of its variables $\forall\, i \in \mathcal{I}$, and $s \in \mathcal{S}$.*

Proof. Recall that $f_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i) = \max_{\tilde{C}_s \in C_s,\, \tilde{P}_s \in P_s} \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i)$.

Let $\max_{\tilde{C}_s \in C_s,\, \tilde{P}_s \in P_s} \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i) = a$.

Let $\mathbf{p} = (\mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i)$, $\mathbf{q} = (\mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i)$. Lemma 1 states that given $\epsilon > 0$, $\exists \delta(\epsilon) > 0$ such that for any $\mathbf{p}, \mathbf{q} \in X_s \times W^i$, with $d_1(\mathbf{p}, \mathbf{q}) < \delta(\epsilon)$, then, $\forall\, \tilde{C}_s \in C_s$, $\forall\, \tilde{P}_s \in P_s$, $|\psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i) - \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i)| < \epsilon$.

Therefore, $\forall\, \tilde{C}_s \in C_s$, $\forall\, \tilde{P}_s \in P_s$, $\psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i) \leqslant \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i) + \epsilon \leqslant a + \epsilon$.

Hence, $\max_{\tilde{C}_s \in C_s,\, \tilde{P}_s \in P_s} \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i) \leqslant a + \epsilon$.

For the other direction, take $\tilde{C}_s^0 \in C_s$ and $\tilde{P}_s^0 \in P_s$ such that $\psi_s^i(\tilde{C}_s^0, \tilde{P}_s^0; \mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i) \geqslant a - \epsilon/2$. From Lemma 1, we have that since $d_1(\mathbf{p}, \mathbf{q}) < \delta(\epsilon)$, $|\psi_s^i(\tilde{C}_s^0, \tilde{P}_s^0; \mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i) - \psi_s^i(\tilde{C}_s^0, \tilde{P}_s^0; \mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i)| < \epsilon$. Then,

$$\begin{aligned}\max_{\substack{\tilde{C}_s \in C_s\\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i) &\geqslant \psi_s^i(\tilde{C}_s^0, \tilde{P}_s^0; \mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i)\\ &\geqslant \psi_s^i(\tilde{C}_s^0, \tilde{P}_s^0; \mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i) - \epsilon\\ &\geqslant a - 3\epsilon/2.\end{aligned}$$

In conclusion, we have that $\forall\, \epsilon > 0, \exists \delta(\epsilon)$ such that if $d_1(\mathbf{p}, \mathbf{q}) < \delta(\epsilon)$ then

$$\begin{aligned}a - 2\epsilon \leqslant a - 3\epsilon/2 &\leqslant \max_{\substack{\tilde{C}_s \in C_s\\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{y}_s^{-i}, \mathbf{z}_s^i, \boldsymbol{\theta}^i)\\ &\leqslant a + \epsilon \leqslant a + 2\epsilon,\end{aligned}$$

which completes the proof. □

Lemma 4 (Fink 1964). *$\gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i)$ is continuous in $\mathbf{x}_s^{-i}$. Furthermore, the set $\{\gamma_s^i(\cdot, \boldsymbol{\omega}^i) \mid \boldsymbol{\omega}^i$ is bounded$\}$ is equicontinuous.*

Proof. Let

$$\gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i) = \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}; \boldsymbol{\omega}^i),$$

$$\gamma_s^i(\mathbf{y}_s^{-i}, \boldsymbol{\omega}^i) = \psi_s^i(C_s^i(\mathbf{y}_s^{-i}, \mathbf{z}_s^{*i}), P_s^i(\mathbf{y}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{y}_s^{-i}, \mathbf{z}_s^{*i}; \boldsymbol{\omega}^i).$$

Furthermore,

$$\begin{aligned}&\gamma_s^i(\mathbf{y}_s^{-i}, \boldsymbol{\omega}^i) - \gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i)\\ &\quad\leqslant \psi_s^i(C_s^i(\mathbf{y}_s^{-i}, \mathbf{u}_s^{*i}), P_s^i(\mathbf{y}_s^{-i}, \mathbf{u}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{y}_s^{-i}, \mathbf{u}_s^{*i}; \boldsymbol{\omega}^i)\\ &\qquad - \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}; \boldsymbol{\omega}^i),\\ &\gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i) - \gamma_s^i(\mathbf{y}_s^{-i}, \boldsymbol{\omega}^i)\\ &\quad\leqslant \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}; \boldsymbol{\omega}^i)\\ &\qquad - \psi_s^i(C_s^i(\mathbf{y}_s^{-i}, \mathbf{z}_s^{*i}), P_s^i(\mathbf{y}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{y}_s^{-i}, \mathbf{z}_s^{*i}; \boldsymbol{\omega}^i).\end{aligned}$$

If $\boldsymbol{\omega}^i$ is restrained to be in a bounded region, then the right-hand sides can be made uniformly small because of the uniform continuity of ψ_s^i on compact sets. □

Table A.1. Notation summary used in the paper.

Notation	Definition
Sets	
$\mathscr{I}$ and $\mathscr{S}$	Set of players: $\mathscr{I}=\{1,\ldots,I\}$. Set of states: $\mathscr{S}=\{1,\ldots,S\}$.
A	Set of all action tuples in a state for all players.
A^{-i}	Set of all action tuples in a state for all players except i.
Δ	J-dimensional probability simplex, i.e., strategy set for player i in s.
X	Strategy set for all players and states.
X_s	Strategy set for all players and only state s.
C_s and P_s	Uncertainty sets for immediate costs and transition probability, resp.
Actions	
J	No. of actions for any $i\in\mathscr{I}$, $s\in\mathscr{S}$.
a^i and a	a^i: Action chosen by i. a: Action tuple, $a=(a^1,\ldots,a^I)$.
a^{-i}	Action tuple of all players except i, $a^{-i}=(a^1,\ldots,a^{i-1},a^{i+1},\ldots,a^I)$.
Strategies	
x^i_{sj}	Probability that i assigns to action j in state s.
$\mathbf{x}^i_s$	Strategy of player i in state s, $\mathbf{x}^i_s=(x^i_{s1},\ldots,x^i_{sJ})\in\Delta$.
$\mathbf{x}^i$, $\mathbf{x}_s$, and $\mathbf{x}$	$\mathbf{x}^i=(\mathbf{x}^i_1,\ldots,\mathbf{x}^i_S)$, $\mathbf{x}_s=(\mathbf{x}^1_s,\ldots,\mathbf{x}^I_s)\in X_s$, and $\mathbf{x}=(\mathbf{x}^1,\ldots,\mathbf{x}^I)\in X$.
$\mathbf{x}^{-i}$	Strategy of all players except player i, $\mathbf{x}^{-i}=(\mathbf{x}^1,\ldots,\mathbf{x}^{i-1},\mathbf{x}^{i+1},\ldots,\mathbf{x}^I)$.
$(\mathbf{x}^{-i},\mathbf{u}^i)$	Distinguishes i's strategy, $(\mathbf{x}^{-i},\mathbf{u}^i)=(\mathbf{x}^1,\ldots,\mathbf{x}^{i-1},\mathbf{u}^i,\mathbf{x}^{i+1},\ldots,\mathbf{x}^I)$.
Parameters	
β	Discount factor, $0\leqslant\beta<1$.
C^i_{sa} and P_{sak}	Immediate cost and transition probability, resp.
$\tilde{C}^i_{sa}$ and $\tilde{P}_{sak}$	Uncertain immediate cost and transition probability resp.
$\tilde{C}_s\in C_s$ and $\tilde{P}_s\in P_s$	Vectors of uncertain costs and transition probabilities, resp.
Values	
v^i_s and ω^i_s	Value and robust value to i in s, resp.
$\boldsymbol{\omega}^i$	Robust value vector for i, $\boldsymbol{\omega}^i=[\omega^i_s]_{s\in\mathscr{S}}$.
Other quantities	
$\pi^a_s(\mathbf{x})$	Probability that a is chosen in s under $\mathbf{x}$.
$\psi^i_s(\tilde{C}_s,\tilde{P}_s;\mathbf{x}^{-i}_s,\mathbf{u}^i_s;\boldsymbol{\omega}^i)$	Exp cost function for i in s, i.e., the objective function.
$f^i_s(\mathbf{x}^{-i}_s,\mathbf{u}^i_s;\boldsymbol{\omega}^i)$	Worst-case exp cost function for i in s.
Matrices	
P	Uncertainty set for transition probabilities. $P=\{\tilde{\mathbf{p}}_s, s\in\mathscr{S}\colon \mathbf{A}_s\tilde{\mathbf{p}}_s\geqslant\mathbf{b}_s, \mathbf{Q}_s\tilde{\mathbf{p}}_s=\mathbf{1}, \tilde{\mathbf{p}}_s\geqslant 0\}$.
$\tilde{\mathbf{p}}_s$	$\tilde{\mathbf{p}}_s=[\tilde{P}_{sak}]_{a\in A;\,k\in S}$.
$\mathbf{e}_j$	jth unit column vector.
$\mathbf{1}$	Column vector of ones of appropriate dimension.
$\mathbf{E}^i_s(\mathbf{x}^{-i}_s,C^i)$	$(J^{I-1})\times J$ matrix associated with immediate costs (see expression (8)).
$\mathbf{z}^i_s$	$\mathbf{z}^i_s\in\Re^{J^I S}$, vector associated with future costs (see Equation (12)).
$\mathbf{Y}^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)$	$\mathbf{Y}^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)\in\Re^{J^I S\times J}$, $\quad\mathbf{Y}^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)\mathbf{u}^i_s=\mathbf{z}^i_s$.
Other notation	
μ, λ	Service and arrival rates, resp.
$h(\cdot)$	Holding cost.
$\theta(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$	Cost of service and admission reward to the router, resp.

Lemma 5 (Fink 1964). *If $\mathbf{x}^{-i,n}\to\mathbf{x}^{-i}$ and $\tau^i_s(\mathbf{x}^{-i,n})\to\omega^i_s$ as $n\to\infty$, then $\tau^i_s(\mathbf{x}^{-i})=\omega^i_s$.*

Proof. We have

$$
\begin{aligned}
&|\omega^i_s-\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)|\\
&\quad\leqslant|\omega^i_s-\tau^i_s(\mathbf{x}^{-i,n})|+|\tau^i_s(\mathbf{x}^{-i,n})-\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))|\\
&\qquad+|\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))-\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)|.
\end{aligned}
$$

Now, by assumption, as $n\to\infty|\omega^i_s-\tau^i_s(\mathbf{x}^{-i,n})|\to 0$ and $|\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))-\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)|\to 0$. Note that $|\tau^i_s(\mathbf{x}^{-i,n})-\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))|=|\gamma^i_s(\mathbf{x}^{-i,n}_s,\boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))-\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\tau}^i(\mathbf{x}^{-i,n}))|\to 0$ as $n\to\infty$ by Lemma 4. Hence, $|\omega^i_s-\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)|\to 0$ as $n\to\infty$. $\square$

An alternative proof of Theorem 1 follows.

Theorem 1. *For $\mathbf{x}\in X$, define $\gamma_x(\boldsymbol{\omega})\colon W\to W$ by $(\gamma_x(\boldsymbol{\omega}))_{is}=\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)$. The function $\gamma_x(\boldsymbol{\omega})$ is a contraction mapping.*

Proof. Let $\boldsymbol{\omega},\boldsymbol{\theta}\in W$. For $\mathbf{x}^{-i}_s$ fixed, $\forall\, i\in\mathscr{I}$, $s\in\mathscr{S}$,

$$
\begin{aligned}
\gamma^i_s(\mathbf{x}^{-i}_s,\boldsymbol{\omega}^i)&=\min_{\mathbf{u}^i_s\in\Delta}\max_{\substack{\tilde{C}_s\in C_s\\ \tilde{P}_s\in P_s}}\psi^i_s(\tilde{C}_s,\tilde{P}_s;\mathbf{x}^{-i}_s,\mathbf{u}^i_s;\boldsymbol{\omega}^i)\\
&=\psi^i_s(C^i_s(\mathbf{x}^{-i}_s,\mathbf{u}^{*i}_s),P^i_s(\mathbf{x}^{-i}_s,\mathbf{u}^{*i}_s,\boldsymbol{\omega}^i);\mathbf{x}^{-i}_s,\mathbf{u}^{*i}_s;\boldsymbol{\omega}^i),
\end{aligned}
$$

where $\mathbf{u}_s^{*i}$ is the minimizer, and $C_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}) \in C_s$ and $P_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}, \boldsymbol{\omega}^i) \in P_s$ are the optimizers that now depend on $(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i})$. Similarly, with $\mathbf{z}_s^{*i}$ and $C_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}) \in C_s$, $P_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\theta}^i) \in P_s$, we have

$$\begin{aligned}\gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\theta}^i) &= \min_{\mathbf{z}_s^i \in \Delta} \max_{\substack{\tilde{C}_s \in C_s \\ \tilde{P}_s \in P_s}} \psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{z}_s^i; \boldsymbol{\theta}^i) \\ &= \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\theta}^i); \mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}; \boldsymbol{\theta}^i).\end{aligned}$$

Now,

$$\begin{aligned}&\gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i) - \gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\theta}^i) \\ &= \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{x}_s^{-i}, \mathbf{u}_s^{*i}; \boldsymbol{\omega}^i) \\ &\quad - \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\theta}^i); \mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}; \boldsymbol{\theta}^i) \\ &\leqslant \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\omega}^i); \mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}; \boldsymbol{\omega}^i) \\ &\quad - \psi_s^i(C_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}), P_s^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \boldsymbol{\theta}^i); \mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}; \boldsymbol{\theta}^i) \\ &= \sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{s,a^m}^m z_{s,a^i}^{*i} \left\{ C_{sa}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}) + \beta \sum_{k=1}^{S} P_{sak}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \omega_k^i) \omega_k^i \right\} \\ &\quad - \sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{s,a^m}^m z_{s,a^i}^{*i} \left\{ C_{sa}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}) + \beta \sum_{k=1}^{S} P_{sak}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \theta_k^i) \theta_k^i \right\} \\ &\leqslant \sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m z_{sa^i}^{*i} \beta \left\{ \sum_{k=1}^{S} P_{sak}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \omega_k^i)(\omega_k^i - \theta_k^i) \right\} \\ &\leqslant \sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m z_{sa^i}^{*i} \beta \left(\sum_{k=1}^{S} P_{sak}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \omega_k^i) \right) \|\boldsymbol{\omega} - \theta\|_\infty \\ &= \beta \|\boldsymbol{\omega} - \theta\|_\infty.\end{aligned}$$

The second to the last inequality above follows from the fact that

$$\begin{aligned}&\sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m z_{sa^i}^{*i} \sum_{k=1}^{S} P_{sak}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \omega_k^i) \theta_k^i \\ &\qquad \leqslant \sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m z_{sa^i}^{*i} \sum_{k=1}^{S} P_{sak}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \theta_k^i) \theta_k^i,\end{aligned}$$

because for a given $(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \theta_k^i)$, $[P_{sak}^i(\mathbf{x}_s^{-i}, \mathbf{z}_s^{*i}, \theta_k^i)]_{k=1,\dots,S}$ is the maximizer of $\psi_s^i(\tilde{C}_s, \tilde{P}_s; \mathbf{x}_s^{-i}, \mathbf{z}_s^i; \boldsymbol{\theta}^i)$ over $\tilde{P}_s \in P_s$.

Similar to the above arguments, we have for $\mathbf{x}_s^{-i}$ fixed that, $\forall\, i \in \mathcal{I}$, $s \in \mathcal{S}$,

$$\gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\theta}^i) - \gamma_s^i(\mathbf{x}_s^{-i}, \boldsymbol{\omega}^i) \leqslant \beta \|\boldsymbol{\omega} - \theta\|_\infty.$$

Thus, $\|\gamma_x(\boldsymbol{\omega}) - \gamma_x(\boldsymbol{\theta})\|_\infty \leqslant \beta \|\boldsymbol{\omega} - \boldsymbol{\theta}\|_\infty$. $\square$

Acknowledgments

The authors thank two anonymous referees and an anonymous associate editor for their comments, which substantially improved this paper. This research was supported by the United States Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE) under grant 2007-ST-061-000001. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of the United States Department of Homeland Security. The authors thank Michelle Aghassi for the insights she provided on the existence proof for robust one-shot n-person normal form games, and Jerzy Filar for his valuable comments.

I, the first author, dedicate this paper to my grandparents Nazire Atak and Ragıp Atak, whose support I am always going to feel throughout my life.

References

Aghassi, M., D. Bertsimas. 2006. Robust game theory. *Math. Programming, Ser. B* **107**(1) 231–273.

Altman, E. 1994a. Flow control using the theory of zero-sum Markov games. *IEEE Trans. Automatic Control* **39**(4) 814–818.

Altman, E. 1994b. Monotonicity of optimal policies in a zero sum game: A flow control model. *Advances in Dynamic Games and Applications*, Vol. 1. Birkhäuser, Boston.

Altman, E., A. Hordijk. 1995. Zero-sum Markov games and worst-case optimal control of queueing systems. *QUESTA* **21**(3) 415–447.

Altman, E., N. Shimkin. 1993. Worst-case and Nash routing policies in parallel queues with uncertain service allocations. Technical report, IMA Preprint No. 1120, Institute for Mathematics and Applications, University of Minnesota, Minneapolis.

Altman, E., N. Shimkin. 1998. Individually optimal dynamic routing in a processor sharing system: Stochastic game analysis. *Oper. Res.* **46**(6) 776–784.

Aumann, R. J., M. B. Maschler. 1968. Repeated games of incomplete information: The zero-sum extensive case. Technical report, U.S. Arms Control and Disarmament Agency, ST-143, Washington, DC, chapter III, 37116.

Avşar, Z. M., M. B. Gürsoy. 2002. Inventory control under substitutable demand: A stochastic game application. *Naval Res. Logist.* **49**(4) 359–375.

Bagnell, J., A. Ng, J. Schneider. 2001. Solving uncertain Markov decision problems. Technical Report CMU-RI-TR-01-25, Robotics Institute, Carnegie Mellon University. Pittsburgh.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Math. Oper. Res.* **23**(4) 769–805.

Bertsimas, D., M. Sim. 2004. The price of robustness. *Oper. Res.* **52**(1) 35–53.

Blackwell, D., T. Ferguson. 1968. The big match. *Ann. Math. Statist.* **39**(1) 159–163.

Datta, R. 2003. Using computer algebra to find Nash equilibria. *Intl. Symp. Symb. Alg. Comp.* ACM Press, New York, 74–79.

Filar, J., K. Vrieze. 1997. *Competitive Markov Decision Processes*. Springer-Verlag, New York.

Fink, A. M. 1964. Equilibrium in a stochastic n-person game. *J. Sci. Hiroshima Univ., Ser. A–I* **28**(1) 89–93.

Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with a nonunique prior. *J. Math. Econom.* **18**(2) 141–153.

Givan, R., S. Leach, T. Dean. 1997. Bounded parameter Markov decision processes. *Fourth Eur. Conf. Planning*. Springer-Verlag, London, 234–246.

Harsanyi, J. C. 1967. Games with incomplete information played by "Bayesian" players, part I. The basic model. *Management Sci.* **14**(3) 159–182.

Harsanyi, J. C. 1968a. Games with incomplete information played by "Bayesian" players, part II. Bayesian equilibrium points. *Management Sci.* **14**(5) 320–324.

Harsanyi, J. C. 1968b. Games with incomplete information played by "Bayesian" players, part III. The basic probability distribution of the game. *Management Sci.* **14**(7) 486–502.

Herings, P. J., R. J. A. P. Peeters. 2004. Stationary equilibria in stochastic games: Structure, selection, and computation. *J. Econom. Theory* **118**(1) 32–60.

Heyman, D. P. 1968. Optimal operating policies for $M/G/1$ queuing systems. *Oper. Res.* **16**(2) 362–382.

Iyengar, G. 2005. Robust dynamic programming. *Math. Oper. Res.* **30**(2) 1–21.

Kakutani, S. 1941. A generalization of Brouwer's fixed point theorem. *Duke Math. J.* **8**(3) 457–459.

Lo, K. C. 1996. Equilibrium in beliefs under uncertainty. *J. Econom. Theory* **71**(2) 443–484.

Marinacci, M. 2000. Ambiguous games. *Games Econom. Behav.* **31**(2) 191–219.

Nash, J. 1950. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **36**(1) 48–49.

Nilim, A., L. El Ghaoui. 2005. Robust control of Markov decision processes with uncertain transition matrices. *Oper. Res.* **53**(5) 780–798.

Peterson, L. L. 2007. *Computer Networks: A Systems Approach*. Morgan Kaufmann Publishers.

Rosenberg, D., E. Solan, N. Vieille. 2004. Stochastic games with a single controller and incomplete information. *SIAM J. Control Optim.* **43**(1) 86–110.

Satia, J. K., R. L. Lave. 1973. Markov decision processes with uncertain transition probabilities. *Oper. Res.* **21**(3) 728–740.

Shapiro, A., A. J. Kleywegt. 2002. Minimax analysis of stochastic problems. *Optim. Methods Software* **17**(1) 523–592.

Shapley, L. S. 1953. Stochastic games. *Proc. Natl. Acad. Sci. USA* **39**(10) 1095–1100.

Sobel, M. J. 1969. Optimal average-cost policy for a queue with start-up and shut-down costs. *Oper. Res.* **17**(1) 145–162.

Sorin, S. 1984. Big match with lack of information on one side I. *Internat. J. Game Theory* **13**(4) 201–255.

Sorin, S. 1985. Big match with lack of information on one side II. *Internat. J. Game Theory* **14**(3) 173–204.

Stidham, S., Jr., R. R. Weber. 1989. Monotonic and insensitive optimal policies for control of queues with undiscounted costs. *Oper. Res.* **37**(4) 611–625.

Vanderbei, R. J. 2006. LOQO user's manual—Version 4.05. Technical report, Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ.

Vrieze, O. J. 2004. Stochastic games and stationary strategies. Stochastic games and applications. A. Neyman, S. Sorin, eds. *Proc. NATO Adv. Study Inst., NATO Sci. Ser. C*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

White, C. C., H. K. Eldeib. 1994. Markov decision processes with imprecise transition probabilities. *Oper. Res.* **42**(4) 739–749.

Yechiali, U. 1971. On optimal balking rules and toll charges in the $GI/M/1$ queuing process. *Oper. Res.* **19**(2) 349–370.