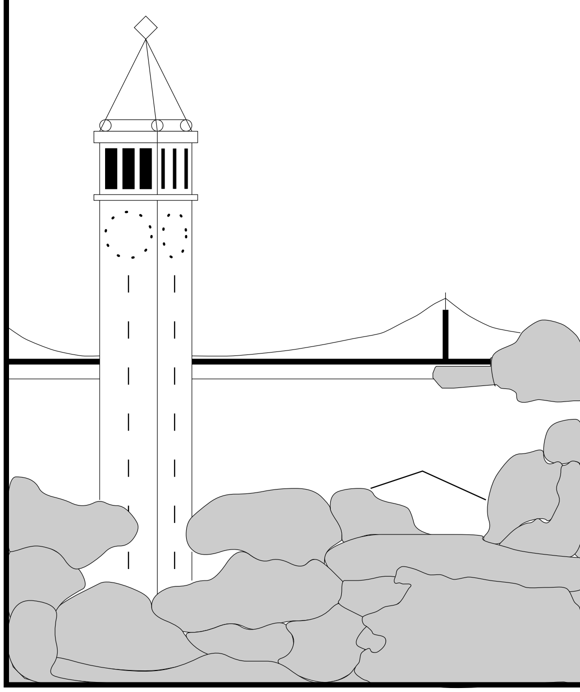
On Nash Equilibria in Stochastic Games

Krishnendu Chatterjee *Marcin Jurdziński* *Rupak Majumdar*



Report No. UCB/CSD-3-1281

October 2003

Computer Science Division (EECS)
University of California
Berkeley, California 94720

On Nash Equilibria in Stochastic Games*

Krishnendu Chatterjee, Marcin Jurdziński, and Rupak Majumdar

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, USA
{c_krish,mju,rupak}@eecs.berkeley.edu

Abstract. We study infinite stochastic games played by n-players on a finite graph with goals given by sets of infinite traces. The games are *stochastic* (each player simultaneously and independently chooses an action at each round, and the next state is determined by a probability distribution depending on the current state and the chosen actions), *infinite* (the game continues for an infinite number of rounds), *nonzero sum* (the players' goals are not necessarily conflicting), and undiscounted. We show that if each player has a reachability objective, that is, if the goal for each player i is to visit some subset R_i of the states, then there exists an ϵ-Nash equilibrium in memoryless strategies. We study the complexity of finding such Nash equilibria. Given an n-player reachability game, and a vector of values $(v_1, \ldots, v_n)$, we show it is NP-hard to determine if there exists a memoryless ϵ-Nash equilibrium where each player gets payoff at least v_i. On the other hand, for every fixed ϵ, the value can be ϵ-approximated in FNP.

We study two important special cases of the general problem. First, we study n-player *turn-based probabilistic* games, where at each state atmost one player has a nontrivial choice of moves. For turn-based probabilistic games, we show the existence of ϵ-Nash equilibria in pure strategies for all games where the goal of each player is a *Borel* set of infinite traces. We also derive the existence of pure exact Nash equilibria for n-player turn-based games where each player has an ω-regular objective.

Then we study the two player case and show that already for two-player games exact Nash equilibria may not exist. Our techniques for the general case also yield NP $\cap$ coNP ϵ-approximation algorithms for zero-sum reachability games, improving the previously known EXPTIME bound.

1 Introduction

The interaction of several agents is naturally modeled as non-cooperative games [25, 27]. The simplest, and most common interpretation of a non-cooperative game is that there is a single interaction among the players ("one-shot"), after which the payoffs are decided and the game ends. However, many, if not all, strategic endeavors occur over time, and in stateful manner. That is, the games

* This research was supported in part by the DARPA grant F33615-C-98-3614, the ONR grant N00014-02-1-0671, the NSF grants CCR-9988172 and CCR-0225610, the Polish KBN grant 7-T11C-027-20, and the EU RTN HPRN-CT-2002-00283.

progress over time, and the current game is decided based on the history of the interactions. Infinite *stochastic games* [30, 8] form a natural model for such interactions. A stochastic game is played over a *state space*, and is played in rounds. In each round, each player chooses an available action simultaneously with and independently from all other players, and the game moves to a new state under a possibly probabilistic transition relation based on the current state and the joint actions. For the verification and control of reactive systems, such games are infinite: play continues for an infinite number of rounds, giving rise to an infinite sequence of states, called the *outcome* of the game. The players receive a payoff based on a payoff function mapping infinite outcomes to a real in $[0, 1]$.

Payoffs are generally Borel measurable functions [23]. For example, the payoff set for each player is a Borel set B_i in the Cantor topology on S^ω (where S is the set of states), and player i gets payoff 1 if the outcome of the game is a member of B_i, and 0 otherwise. In verification, payoff functions are usually index sets of ω-*regular languages*. ω-regular sets occur in low levels of the Borel hierarchy (they are in $\Sigma_3 \cap \Pi_3$), but they form a robust and expressive language for determining payoffs for commonly used specifications [20]. The simplest ω-regular games correspond to safety (closed sets) or reachability (open sets) objectives.

Traditionally automata theory and verification has considered *zero sum* or strictly competitive versions of stochastic games. In these games there are two players with complementary objectives; so the payoff for one is one minus the payoff of the other. We argue that in many modeling instances this is too pessimistic an assumption. The environment of a component in a distributed system is not necessarily malicious. In fact, many natural interactions are modeled as a game between several components each with its own specification, and each component is interested solely in establishing its own specification without regard to the specification of other components. For example, consider a set of n processors each sending out data on a common network. At each round, each process can decide to send data or do nothing. If more than one process tries to send data simultaneously, then there is a conflict and the data is not sent; if there is a unique processor sending out data in a round, then its data is sent out. The game for two processors is schematically shown in Figure 1. It can be easily generalized for n processors. Each processor wishes to send an infinite number of data packets, that is, process i has the specification that the game visits the node i infinitely often.

Traditionally, the system will be modeled as a zero sum game between process i and an environment consisting of all other processes, and the requirement will be specified in a game logic such as alternating-time temporal logic [1] as $\langle\langle i \rangle\rangle \Box\Diamond i$, that is, we ask if player i has a strategy to visit node i infinitely often, against all strategies of the other players. This condition is too restrictive, and indeed, this cannot be proved for the network game (consider a strategy of the environment where all the other processors try to send at each round). We claim that the right way to model this system is as a *non-zero sum* game, where each processor i has the obligation $\Box\Diamond i$, and is solely interested in ensuring its specification

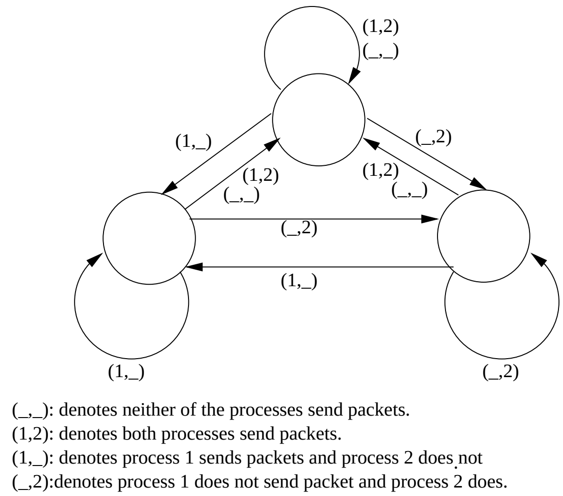


Fig. 1. The two processor game

without regard to the specifications of the other players. The solution concept in such games is a *Nash equilibrium* [13], that is, a strategy profile such that no processor can gain by deviating from the profile, assuming all other processors continue playing their strategies in the profile. However, the existence of Nash equilibria in infinite games is not clear.

Notice that the network game has a Nash equilibrium where the processors are allocated time slots, and processor i only sends in time slot k where k mod $n = i$. Indeed this is a solution adopted in time triggered protocols in real time systems [16]. There is a (symmetric) equilibrium in the game as well: each processor rolls an n-sided dice, and sends data only if the dice shows 1. Then with probability 1, all processors can send data infinitely often. Interestingly, exponential backoff behavior implemented in real networks also have the above property, indeed, it is a Nash equilibrium where the strategy of a player is oblivious to the total number of processes participating in the game. The emergence of quite rich behavior in such a simple example shows the modeling power of stochastic games.

This work is motivated by the result by Secchi and Sudderth [29]. Secchi and Sudderth [29] proved that a Nash equilibrium exists for safety conditions where each player i has a subset of states S_i as their safe states and gets a payoff 1 if the play never leaves the set S_i and else get payoff 0. In the open (or reachability) game, each player i has a subset of states R_i as reachability targets. Player i gets payoff 1 if the outcome visits some state from R_i at some point, and 0 otherwise. Our main results on reachability games are summarised below.

1. We show that reachability games on finite state spaces always have ϵ-Nash equilibria in memoryless strategies. This is the best one can hope for: there

are two player non zero sum reachability games with no Nash equilibria [17]. In general equilibrium strategies require randomization.

2. We show that the problem of finding an ϵ-Nash equilibrium in memoryless strategies where each player gets at least some (specified) payoff is NP-hard. Related NP-hardness results appear in [5], but our results do not follow from theirs as our payoffs are restricted to be binary. Moreover, for any constant ϵ, we give an NP algorithm to approximate the value of some ϵ-Nash equilibrium in memoryless strategies. Already for two-person zero-sum games with reachability objective, values can be algebraic and there are simple examples when they are irrational [8]. Hence approximating the values is the best one can achieve.

Together with [29] this solves the existence question for the lowest level of the Borel hierarchy. We leave the existence of Nash equilibria in stochastic games where objectives are sets in higher levels of the Borel hierarchy as an interesting open question. We also study two important special cases of the general problem: *turn-based probabilistic games* [33, 28], where at each stage, at most one player has a nontrivial choice of actions, and the two-player case of general stochastic games.

For the special case of two person turn-based probabilistic zero sum games we prove a pure strategy determinacy theorem for all Borel payoff functions. The proof is a specialization of Martin's determinacy proof for stochastic games with Borel payoffs [23]. Using this, and a general construction of threat strategies [25], we show that ϵ-Nash equilibria exist for all turn based probabilistic games with arbitrary Borel set payoffs. Moreover, using further structural properties for turn-based probabilistic parity games, we show the existence of pure strategy Nash equilibria for parity payoffs. Since parity games are a canonical form for ω-regular properties [31], this proves that (exact) Nash equilibria exist for turn based probabilistic games with ω-regular payoffs. Using an NP $\cap$ co-NP strategy construction algorithm for parity games [3], we get an NP algorithm to find a Nash equilibrium in these games.

For the special case of two-player (concurrent) games, we show an improved NP $\cap$ co-NP upper bound to approximate the values for two-person zero sum reachability games within ϵ-tolerance for any constant ϵ, improving the previously best known EXPTIME upper bound [8]. This generalizes a result of Condon [4]. Notice that the solution of a zero-sum reachability game can be irrational, hence we can only hope to compute it to an ϵ-precision.

Related Work

Stochastic games were introduced by Shapley [30] and have been extensively studied in several research communities; the book of Filar and Vrieze [10] provides a unified treatment of the theories of stochastic games and Markov decision processes. Existence of Nash equilibria in (nonzero sum) discounted stochastic games was proved by Fink [11]. Since then, several results have appeared for special cases [32, 33]. One of the most important results in stochastic games in recent times is due to Vieille [34], [35] where he shows the existence of ϵ-Nash equilibria for two-person non-zero sum game with limit average criteria. The existence of Nash equilibria for n-person stochastic games with limit average criteria is still open. Our result shows in the special case of turn-based probabilistic n-person games ϵ-Nash equilibria exists as limit average criteria occurs in low levels of Borel hierarchy.

Infinite games with Borel winning conditions have been studied by descriptive set theorists [15]. Martin [23] proved the determinacy result for two-person stochastic zero sum games with Borel payoff, building on his earlier proof of Borel determinacy in perfect information games [22]. This result was extended by Maitra and Sudderth [19]. In the case of non-zero sum games the existence of Nash equilibria for Borel payoffs remain some of the most important questions in stochastic games. Secchi and Sudderth [29] showed the existence of Nash equlibria with safety conditions.

Computing the values of a Nash equlibria, when it exists, is another challenging problem [26, 36]. Recently [5] show hardness of several such questions. Condon [4] studies two-person turn-based probabilistic discounted games with reachability objective and showed that the values at a state can be computed in NP $\cap$ co-NP. Her result can be applied to show that the values of a two-person turn-based probabilistic zero-sum games with reachability objective can be computed in NP $\cap$ co-NP. We show that for the general case of two-person (concurrent) zero-sum games with reachability objective values can be approximated in NP $\cap$ co-NP. For zero-sum stochastic games with ω-regular objectives, [8] gives doubly exponential algorithms, and [3] gives more efficient algorithms for the turn-based case.

2 Definitions

An *n-person stochastic game* G consists of a finite, nonempty set of states S, n players $1, 2, \ldots, n$, a finite set of action sets $A_1, A_2, \ldots, A_n$ for the players, a conditional probability distribution p on $S \times (A_1 \times A_2 \times \cdots \times A_n)$ called the law of motion, and bounded, real valued payoff functions $\phi_1, \phi_2, \ldots, \phi_n$ defined on the history space $H = S \times A \times S \times A \cdots$, where $A = A_1 \times A_2 \times \cdots A_n$. The game is called a n-player deterministic game if for all states $s \in S$ and action choices $a = (a^1, a^2, \ldots, a^n)$ there is a unique state s' such that $p(s'|s, a) = 1$.

Play begins at an initial state $s_0 = s \in S$. Each player independently and concurrently selects a mixed action a_i^1 with a probability distribution $\sigma_i(s)$ belonging to $\mathcal{P}(A_i)$, the set of probability measures on A_i. Given s_0 and the chosen mixed actions $a^1 = (a_1^1, a_2^1, \ldots a_n^1) \in A$, the next state s_1 has the probability distribution $p(\cdot|s_0, a^1)$. Then again each player i independently selects a_i^2 with a distribution $\sigma_i((s_0, a^1, s_1))$ and given $a^2 = (a_1^2, a_2^2, \ldots, a_n^2)$, the next state s_2 has the probability distribution $p(\cdot|s_1, a^2)$. Play continues in this fashion thereby generating a random histroy $h = (s_0, a^1, s_1, a^2, \ldots) \in H$. Note that the game continues for an infinite number of steps [9], and the payoff is decided based on the infinite outcome. This is useful to model interactions between reactive systems [21].

A function π_i that specifies for each partial history $h' = (s_0, a^1, s_1, a^2, \ldots, s_k)$ the conditional distribution $\pi_i(h') \in \mathcal{P}(A_i)$ for player i's next action a_i^{k+1} is called a *strategy* for player i. A strategy profile $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ consists of a strategy π_i for each player i. A *selector* for a player i is a mapping $\sigma_i : S \to \mathcal{P}(A_i)$. A selector profile $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ consists of a selector σ_i for each player i. The *memoryless/stationary* strategy σ_i^∞ for player i is the strategy which choses mixed action $\sigma_i(s')$ each time the play visits s'. A strategy profile $\sigma^\infty = (\sigma_1^\infty, \sigma_2^\infty, \ldots, \sigma_n^\infty)$ is a memoryless startegy profile if all the strategies $\sigma_1^\infty, \sigma_2^\infty, \ldots, \sigma_n^\infty$ are memoryless. Given a memoryless strategy profile $\sigma^\infty = (\sigma_1^\infty, \sigma_2^\infty, \ldots, \sigma_n^\infty)$ we write $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ to denote the corresponding selector profile for the players. An initial state s and a strategy profile $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ together with the law of motion p determine a probability distribution $P_{s,\pi}$ on the history space. We write $E_{s,\pi}$ for the expectation operator associated with $P_{s,\pi}$.

Assume now that the payoff functions $\phi_i : H \to \mathbb{R}$ are bounded and measurable, where $\mathbb{R}$ is the set of reals. If the initial state of the game is s and each player i choses a strategy π_i, then the payoff to each player i is the expectation $E_{s,\pi}\phi_i$, where π is the strategy profile $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$.

For $\epsilon \geq 0$, an ϵ-equilibrium at the initial state s is a profile $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ such that, for all $i = 1, 2, \ldots, n$

$$E_{s,\pi}\phi_i \geq \sup_{\mu_i} E_{s,(\pi_1,\ldots,\pi_{i-1},\mu_i,\pi_{i+1},\ldots,\pi_n)}\phi_i - \epsilon$$

where μ_i ranges over the set of all strategies for player i. In other words, each π_i gurantees an expected payoff for player i which is within ϵ of the best possible expected payoff for player i when every other player $j \neq i$ playes π_j. A 0-equilibrium is called a *Nash equilibirum* and for every $\epsilon > 0$ an ϵ-equilibrium is called an *ϵ-Nash equilibrium* [13]. A strategy profile π for an ϵ-Nash equilibrium is referred as the ϵ-equilibrium profile. Similarly, a strategy profile π for a Nash equilibrium is referred as the Nash equilibrium profile.

Let $r_i : S \to \mathbb{R}$ be a daily reward function for player $i, i = 1, 2, \ldots, n$. It is known that Nash equilibria exist for some interesting payoff functions such as a discounted payoff

$$\phi_i(h) = \sum_{k=0}^{\infty} \beta^n r_i(s_k), 0 < \beta < 1$$

(cf. Mertens and Parthasarathy [24] and the references there), but need not exist for other payoff functions such as an average reward

$$\phi_i(h) = \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} r_i(s_k), 0 < \beta < 1$$

even for a two-person, zero-sum game with finite state space (cf. Gillete [12], Blackwell and Ferguson [2] for a famous counterexample and Vielle [34] and [35] for the existence of "equilibirum payoffs" in two person, non-zero sum games.) A game with a *total reward objective* is a game with payoff for player i (ϕ_i^T) defined as

$$\phi_i^T(h) = \sum_{k=0}^{\infty} r_i(s_k)$$

which assigns to a history a payoff that is the sum total of the reward of the states.

Secchi and Sudderth [29] proved that Nash equilibrium exist for safety conditions where each player i has a subset of states S_i as their safe states and gets a payoff 1 if the play never leaves the set S_i and else get payoff 0. That is, let $S_1^\infty, S_2^\infty, \ldots, S_n^\infty$ be the subsets of of H defined by

$$S_i^\infty = \{h = (s_0, a^1, s_1, a^2, \ldots) : s_k \in S_i \text{ for all } k = 0, 1, \ldots\}$$

and take the payoff function $\phi_i^{S_i}$ to be the indicator function of S_i^∞ for $i = 1, 2, \ldots, n$. The problem for Nash equilibrium for reachability objective was left open. In this work we show for every positive ϵ we have a ϵ-Nash equilibrium in n-player stochastic games with reachability objective. We now formally define the payoff functions. To define the payoff functions we study, let $R_1, R_2, \ldots, R_n$ be subsets of the state space S. The subet of states R_i is referred as the *target set* for player i. Then let $R_1^\infty, R_2^\infty, \ldots, R_n^\infty$ be the subsets of H defined by

$$R_i^\infty = \{h = (s_0, a^1, s_1, a^2, \ldots) : \exists k, s_k \in R_i\}$$

and take the payoff function $\phi_i^{R_i}$ to be the indicator function of R_i^∞ for $i = 1, 2, \ldots, n$. Thus each player recieves a payoff of 1 if the process of states $s_0, s_1, \ldots$ reaches a state in R_i and recieves payoff 0 otherwise. We call stochastic games with the payoff functions of this form *reach-a-set-games*.

3 Existence of ϵ-Nash Equilibria

We define a few more notations which we will use in our proofs below. Given a strategy profile $\tau = (\tau_1, \tau_2, \ldots, \tau_n)$ the strategy profile $\tau_{-i} = (\tau_1, \ldots, \tau_{i-1}, \tau_{i+1}, \ldots, \tau_n)$ is the strategy profile obtained by deleting the strategy τ_i from τ whereas for any strategy μ_i of player i, $\rho(\tau_{-i}, \mu_i) = (\tau_1, \ldots, \tau_{i-1}, \mu_i, \tau_{i+1}, \ldots, \tau_n)$ denotes the strategy profile where player i follows μ_i and the other players follows the strategy of τ_{-i}. Similar definitions hold for selector profiles as well. The main result of this section is the existence of ϵ-Nash equilibria.

Theorem 1 (ϵ-equilibrium). *A n-person reach-a-set-game G with a finite state space has an ϵ-Nash equilibrium at every initial state $s \in S$ for every positive ϵ. Moreover, there is a memoryless ϵ-equilibrium strategy profile.*

As Example 1 below shows, even for 2-player games with reachability objective Nash equilibrium need not exist. Hence ϵ-Nash equilibrium is the best one can achieve for n-person reach-a-set-games.

Example 1. [ϵ-equilibrium] Consider the following game, adapted from [9, 17]. The state space of the game is $S = \{s, t, u\}$. The action set for player 1 in state s is $\{a, b\}$ and for player 2 is $\{c, d\}$. The state t, u are absorbing states in the sense when the process of states reaches t, u it stays there for ever. The game has a deterministic law of motion p as follows:

$$p(s|s, a, c) = p(t|s, a, d) = p(t|s, b, c) = p(u|s, b, d) = 1.$$

The target set for player 1 is $\{t\}$ and for player 2 is $\{u\}$. For every $\epsilon > 0$ player 1 chooses move a, b with probability $1 - \epsilon$ and ϵ respectively to ensure reaching the state t with probability of $1 - \epsilon$ from s. However, player 1 has no strategy reach t with probability 1: if player 1 decides to play move b at the n-th round of the game, player 2 can play move d at the n-th round, so that the probability of reaching t is always less than 1. ∎

Definition 1 (β-discounted games). *Given a n-player game G we use G^β to denote a β-discounted version of the game G. The game G^β at each step halts with probability β (goes to a special sink state* halt *which has a reward 0 for every player) and continues as the game G with probability $1 - \beta$. β is called the discount-factor.* ∎

Definition 2 (Markov Decision Process (MDP) reach-a-set-game). *A Markov Decision Process (MDP) is a 1-player stochastic game. A MDP reach-a-set-game is a 1 player stochastic reach-a-set-game.* ∎

Definition 3 (Values of MDP). *Given a MDP reach-a-set-game G the value of the game at state s is denoted by*

$$v(s) = \sup_{\pi} E_{s,\pi} \phi_1^{R_1}$$

where π ranges over all strategy and $\phi_1^{R_1}$ is the reach-a-set-game payoff for the player in the game G. Similarly, we use

$$v^\beta(s) = \sup_{\pi} E_{s,\pi} \phi_1^{R}$$

to denote the value at state s in the game G^β, where G^β is the β-discounted version of the game G. In a similar way given a MDP G_T with a total reward objective we use the following notation

$$v_T(s) = \sup_{\pi} E_{s,\pi} \phi_1^{T}.$$

Also, $v_T^\beta(s)$ denote the value at state s for the game G_T^β which is the β-discounted version of the game G_T. ∎

Lemma 1. *Let G be a MDP reach-a-set-game and G^β be the β-discounted version of G. Then for all state $s \in S$ we have*

$$v(s) - v^\beta(s) \leq \beta.$$

Proof. Given a MDP reach-a-set-game G, let $R \subseteq S$ be the *target* set for the player. We construct a total reward game G_T as follows:

- *State Space:* $S_T = S \cup \{sink\}$
- *Reward Function:* $r(s) = 1$ if $s \in R$ else 0.
- *Law of Motion:* For all $s \in S \setminus R$, we have $p_T(s'|s) = p(s'|s)$ and for all $s \in R \cup \{sink\}$ we have $p_T(sink|s) = 1$.

That is, in the total reward game G_T, defined on the same state space S with a special sink state, for every state in R the game goes to the sink state and stays there for ever. The reward is 1 for every state in R and 0 elsewhere. Let G_T^β be the β-discounted version of the game G_T. It is easy to notice that for all state s we have $v(s) = v_T(s)$ and $v^\beta(s) = v_T^\beta(s)$. It follows from the continuity of the values of MDP's with total reward objective with $\beta \to 0$ for positive total reward (Theorem 4.4.1, pg-197 Filar-Vrieze [10]) and the Lipschitz continuity of the values of MDP's with total reward objective (Theorem 4.3.7, pg-185 Filar-Vrieze [10]) that for all $s \in S$ we have $v_T(s) - v_T^\beta(s) \leq \beta$. The required result follows. ∎

Definition 4 (Stopping time of history in β-discounted games). *Consider the stopping time T defined on histories $h = (s_0, a^1, s_1, a^2, \dots)$ by*

$$T(h) = \inf\{k \geq 0 : x_k = halt\}$$

where as usual the infimum of the empty set is $+\infty$. ∎

Lemma 2. *Let G^β be a n-player β-discounted stochastic game. Then, for all initial states s and all strategy profiles π we have*

$$P_{s,\pi}[T > m] \leq (1-\beta)^m$$

Proof. At each step of the game G^β the game reaches the halt state with probability β. Hence the probability of not reaching the halt state in m steps is $\leq (1-\beta)^m$. ∎

The proof of the next Lemma is similar to the proof of Lemma 2.2 of Stay-in-a-set games of Secchi and Sudderth [29].

Lemma 3. *There exist selectors $\sigma_i : S \to \mathcal{P}(A_i), i = 1, 2, \dots, n$, such that the memoryless profile $\sigma^\infty = (\sigma_1^\infty, \sigma_2^\infty, \dots, \sigma_n^\infty)$ is a Nash equilibrium profile in G^β for every $s \in S$.*

Proof. Regard each n-tuple $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ of selectors as a vector in a compact, convex subset K of the appropriate Eucledian space. Then define a correspondence λ that maps each element σ of K to the set $\lambda(\sigma)$ of all elements $g = (g_1, g_2, \dots, g_n)$ of K such that, for $i = 1, 2, \dots, n$ and all $s \in S$, $g_i^\infty(s)$ is an optimal response for player i in G^β against $\sigma_{-i}^\infty = (\sigma_1^\infty, \dots, \sigma_{i-1}^\infty, \sigma_{i+1}^\infty, \dots, \sigma_n^\infty)$.

Clearly, it suffices to show that there is a $\sigma \in K$ such that $\lambda(\sigma) = \sigma$. To show this, we will verify the Kakutani's Fixed Point Theorem [14]:

1. For every $\sigma \in K$, $\lambda(\sigma)$ is closed, convex and nonempty;
2. If, for $k = 1, 2, \ldots, g^(k) \in \lambda(\sigma^{(k)}), \lim_{k\to\infty} g^{(k)} = g$ and $\lim_{k\to\infty} \sigma^{(k)} = \sigma$ then $g \in \lambda(\sigma)$.

To verify condition 1., fix $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n) \in K$ and $i \in \{1, 2, \ldots, n\}$. For each $s \in S$, let $v(s)$ be the maximal payoff that player i can achieve in G^β against σ_{-i}^∞. Since fixing the strategy for all the other player the game becomes a MDP we know that g_i^∞ is an optimal response to σ_{-i}^∞ if and only if, for each $s \in S$, $g_i(s)$ puts positive probability only on actions $a_i \in A_i$ that maximize the expectation of $v(s)$, namely,

$$\sum_{s'} v(s)p(s'|s, (a_i, \sigma_{-i}(s)))$$

Hence condition 1. follows easily.

Condition 2. is an easy consequence of the continuity mapping

$$\sigma \mapsto E_{s,\sigma^\infty(s)}\phi_i^{R_i}$$

from K to the real line. It follows from Lemma 2 that the mapping is continuous. ∎

Definition 5 (Memoryless strategy profile, MDP and Markov Chains). *Given a n-player stochastic game G let $\sigma^\infty = (\sigma_1^\infty, \sigma_2^\infty, \ldots, \sigma_n^\infty)$ be a memoryless strategy profile and $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ be the corresponding selector profile. Then the game G_σ is a Markov chain where the law of motion p_σ is defined by the functions in selector profile σ and the law of motion p of the game G. Similarly, $G_{\sigma_{-i}}$ is a Markov Decision process where the mixed action of each player $j \neq i$ at a state s is fixed according to the selector function $\sigma_j(s)$. The law of the motion $p_{\sigma_{-i}}$ of the MDP is determined by the selectors in σ_{-i} and law of motion p of G.* ∎

Lemma 4. *Given a n-player stochastic game G and a positive ϵ there is a memoryless profile $\sigma^\infty = (\sigma_1^\infty, \sigma_2^\infty, \ldots, \sigma_n^\infty)$ such that σ^∞ is an ϵ-Nash equilibrium profile in G.*

Proof. Given the game G we construct a game G^ϵ which is a discounted version of G with discount-factor ϵ. It follows from Lemma 3 that there is a memoryless strategy profile σ^∞ in G^ϵ such that σ^∞ is a Nash equilibrium profile in the game G^ϵ. We show that the profile σ^∞ is an ϵ-equilibrium profile for G. Let $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ be the selector functions corresponding the strategy profile $\sigma^\infty = (\sigma_1^\infty, \sigma_2^\infty, \ldots, \sigma_n^\infty)$. Consider any player i and the strategy profile σ_{-i}^∞. The game $G_{\sigma_{-i}}$ is a MDP where the mixed actions of all the other players are fixed according to the σ_{-i}. Also, $G_{\sigma_{-i}}^\epsilon$ is the MDP which is the ϵ-discounted version of the game $G_{\sigma_{-i}}$. It follows from Lemma 1 that σ^∞ is an ϵ-equilibrium profile in the game G. ∎

Lemma 4 yields Theorem 1.

4 Complexity of computing equilibrium values

Let π be an ϵ-equilibrium profile. Then, the value at a state s for a player i for the equilibrium profile π, denoted $v_i^\pi(s)$, is $E_{s,\pi}\phi_i^{R_i}$. The value of an ϵ-equilibrium profile π at a state s is the value vector $\boldsymbol{v}^\pi(s) = (v_1^\pi(s), v_2^\pi(s), \ldots, v_n^\pi(s))$. Our main results about the computational complexity of computing the value of any ϵ-equilibrium profile within a tolerance of ϵ are summarized below.

Theorem 2 (Computing values of a memoryless equilibrium profile). *For n-player deterministic reach-a-set-game G, a initial state s and a value vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ it is NP-hard to determine whether there is a Nash equilibrium profile π such that the value for every player i from the state s for the profile π ,i.e. $v_i^\pi(s)$ is greater than equal to v_i. Given a fixed ϵ there is a NP algorithm to compute if there is an ϵ-equilibrium profile π in memoryless strategies such that for all player i we have $v_i^\pi(s) \geq v_i - \epsilon$.*

4.1 Reduction of 3-SAT to computing equilibrium values

We first prove it is NP-hard to compute a memoryless Nash equilibrium profile of n-player deterministic reach-a-set games by reduction from 3-SAT. Given a 3-SAT formula ψ with n-clauses and m-variables we will construct a n-player deterministic reach-a-set-game G_ψ. Let the variables in the formula ψ be $x_1, x_2, \ldots, x_m$ and the clauses be $C_1, C_2, \ldots, C_n$. In the game G_ψ each clause is a player. The state space S, the law of motion and the target states are defined as follows:

- *State Space:*

$$S = \{1, 2, \ldots, m, m+1, (1,0), (1,1), (2,0), (2,1), \ldots, \\ (i,0), (i,1), \ldots, (m,0), (m,1), sink\}.$$

- *Law of Motion:* For any state $(i,0), (i,1)$ the game always moves to the state $i+1$. Let $C_i = \{C_{i_1}, C_{i_2}, \ldots, C_{i_k}\}$ be the set of clauses in which variable x_i occurs. Then, in state i players $i_1, i_2, \ldots, i_k$ have a choice of moves between $\{0,1\}$. If all the players chose move 0 the game proceeds to state $(i,0)$, if all the players chose move 1 the game proceeds to state $(i,1)$, else the game goes to the sink state. Once the game reaches the sink state or the state $m+1$ it remains there for ever.
- *Target States:* The target set for the players is defined as follows: let $C_i^0 = \{C_{k_1}^0, C_{k_2}^0, \ldots, C_{k_l}^0\}$ be the set of clauses that are satisfied assigning $x_i = 0$, then the state $(i,0)$ is a target state for players $k_1, k_2, \ldots, k_l$. Similarly, let $C_i^1 = \{C_{k_1'}^1, C_{k_2'}^1, \ldots, C_{k_j'}^1\}$ be the set of clauses that are satisfied by assigning the variable $x_i = 1$ then the state $(i,1)$ is a target state for players $k_1', k_2', \ldots, k_j'$. States $1, 2, \ldots, m+1$ and the *sink* state is not a target state for any player.

The game is illustrated in Figure 2.

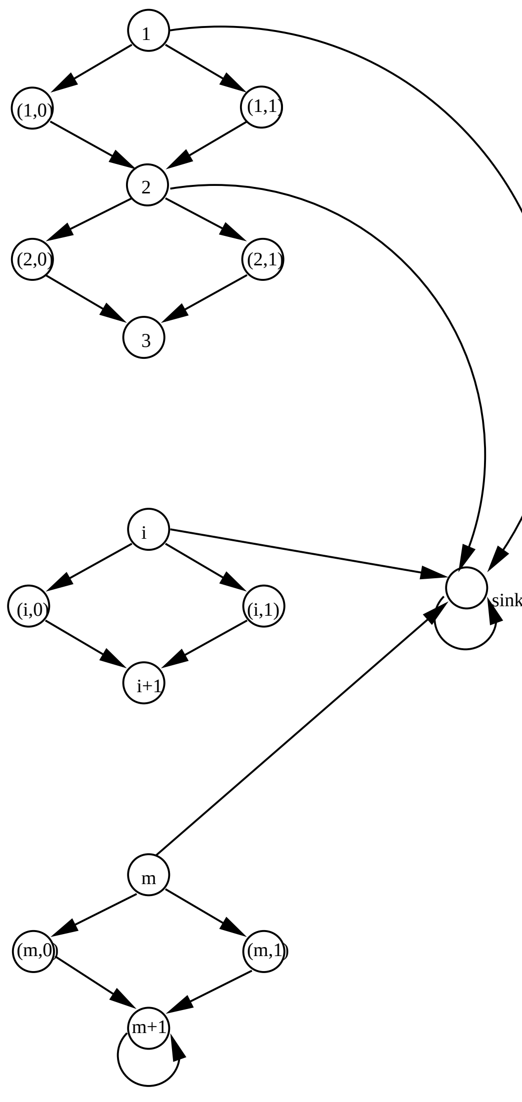


Fig. 2. The game G^{ψ}

Lemma 5 (NP-hardness). *Consider a n-player reach-a-set-game G, an initial state s and a value vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$. It is NP-hard to determine whether there is a Nash equilibrium profile such that the value of each player i at state $s \geq v_i$.*

Proof. We reduce the 3-SAT problem to the problem of determining whether there is an equilibrium such that each player has a value $\geq v_i$ at state s. Given a 3-SAT formula ψ we construct the game G_ψ as described above. Each player gets a value 1 at state 1 iff the formula ψ is satisfiable. If the formula is satisfiable then consider a satisfying assignment to the variables. Then at each state i all the players chose the move as specified by the satisfying assignment and hence every player get a payoff 1. If all the players get an payoff 1 in the game G_ψ it follows from the construction of G_ψ that there is an assignment such that every clause is satisfied and hence the 3-SAT formula ψ is satisfiable. ∎

The Nash equilibrium condition in memoryless strategies can be written as a sentence in the first order theory of reals with addition and multiplication $((\mathbb{R}, +, \cdot))$. The length of the sentence is polynomial in the size of the game and the depth of the quantifiers is constant. This gives an EXPTIME procedure for the following decision problem: given a game G and a value vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ is there an ϵ-Nash equilibrium in memoryless strategy profile such that each player i gets payoff $\geq v_i$. Notice that the reduction to the theory of reals with addition and multiplication allows us to solve other problems in a similar way. For example, given a game G whether there is an ϵ-Nash equilibrium in memoryless strategy profile such that player i gets a payoff at least v_i can be solved in time exponential in the game and polynomial in $\log(\frac{1}{\epsilon})$ using binary search in the interval $[0, 1]$.

Since the number of Nash equilibria where each player gets a payoff 1 is exactly the number of satisfying assignments, the following corollary is immediate.

Corollary 1. *Counting the number of Nash equilibria in reachability games where each player gets at least a given payoff is #P-hard.*

4.2 Approximating equilibrium value in NP

We will show that the memoryless ϵ-equilibrium profile can be approximated by a k-uniform memoryless strategy profile. We will use a result by Lipton et.al. [18]. Given a n-player stochastic reach-a-set-game G we use $|S|$ to denote the size of the state space and l to denote the maximum number of moves available to any player at any state of G.

Definition 6 (Pure selector). *A selector function σ_i for player i is pure if for all states $s \in S$ we have that there is an action $a_i \in A_i$ such that $\sigma_i(a_i|s) = 1$.* ∎

Definition 7 (k-uniform selector and k-uniform memoryless strategy). *A selector function σ_i^k for player i is a k-uniform selector if for all states $s \in S$ we have σ_i^k is the uniform distribution on a multiset M of pure selectors with*

$|M| = k$. *A selector profile* $\sigma^k = (\sigma_1^k, \sigma_2^k, \ldots, \sigma_n^k)$ *is* k*-uniform if all the selectors* σ_i^k *is a* k*-uniform selector for all* $i \in \{1, 2, \ldots, n\}$. *A memoryless strategy profile* $\sigma^{k,\infty} = (\sigma_1^{k,\infty}, \sigma_2^{k,\infty}, \ldots, \sigma_n^{k,\infty})$ *is* k*-uniform is the selector profile* $\sigma^k = (\sigma_1^k, \sigma_2^k, \ldots, \sigma_n^k)$ *corresponding to the strategy profile* $\sigma^{k,\infty}$ *is* k*-uniform.* ∎

Lemma 6 ([18]). *Let* J *be a matrix-game (1 Step game) with* n*-players and each player has atmost* l *moves. Let* π *be a Nash-equilibrium strategy/selector. Then for every* $\epsilon > 0$ *there exists, for every* $k \geq \frac{3n^2 \ln n^2 l}{\epsilon^2}$, *a set of* k*-uniform strategy/selector such that the deviation from any pure strategy/selector with positive support in* π *is less than* ϵ.

Definition 8 (Difference of two MDP's). *Let* G_1 *and* G_2 *be two MDP's defined on the same state space* S. *The difference of the two MDP's, denoted* $err(G_1, G_2)$, *is defined as*

$$err(G_1, G_2) = \sum_{s,s' \in S^2} |p_1(s|s') - p_2(s|s')|$$

That is, $err(G_1, G_2)$ *is the sum of the difference of the probabilities of all the edges of the MDP's.* ∎

Lemma 7. *Let* G^β *be a discounted* n*-player stochastic reach-a-set-game and* σ^∞ *be a memoryless Nash equilibrium profile with selector profile* $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$. *Then for every* $\epsilon > 0$, *there exists, for every* $k \geq \frac{3n^2 \ln n^2 l}{(\frac{\epsilon}{n \cdot |S|^2})^2}$, *a set of* k*-uniform memoryless strategy profile* $\sigma^{k,\infty}$ *(with selector profile* σ^k*) such that the following holds:*

– *for any player* i, *the MDP's* $G_{\sigma_{-i}}$ *and* $G_{\sigma_{-i}^k}$ *satisfy*

$$err(G_{\sigma_{-i}}, G_{\sigma_{-i}^k}) \leq \epsilon$$

Proof. It follows from Lemma 6 that there is a selector profile σ^k such that for any player i the deviation (or error) of σ_i^k from any pure strategy with positive support of σ_i at any state $s \in S$ is atmost $\frac{\epsilon}{n \cdot |S|^2}$. Since there are n players for any edge the difference in probabilities in $G_{\sigma_{-i}}$ and $G_{\sigma_{-i}^k}$ is atmost $\frac{\epsilon}{|S|^2}$. Since there can be atmost $|S|^2$ edges the result follows. ∎

Lemma 8. *Given a* n*-player discounted stochastic reach-a-set-game* G^β *then for every* ϵ *there exists, for every* $k \geq \frac{3n^4 |S|^4 \ln n^2 l}{\epsilon^2}$ *there is a* k*-uniform memoryless strategy profile* $\sigma^{k,\infty} = (\sigma_1^{k,\infty}, \sigma_2^{k,\infty}, \ldots, \sigma_n^{k,\infty})$ *such that* $\sigma^{k,\infty}$ *is an* ϵ*-Nash equilibrium profile in the game* G^β.

Proof. The result follows from Lemma 7 and Lipschitz continuity of values of MDP's with respect to *err* (Theorem 4.3.7,pg-185 Filar-Vrieze [10]). ∎

Lemma 9. *Given a* n*-player stochastic reach-a-set-game* G *for every* $\epsilon > 0$, *there exists, for every* $k \geq \frac{12n^4 |S|^4 \ln n^2 l}{\epsilon^2}$ *a* k*-uniform memoryless strategy profile* $\sigma^{k,\infty}$ *such that* $\sigma^{k,\infty}$ *is a* ϵ*-equilibrium profile for the game* G.

Proof. Let $G^{\frac{\epsilon}{2}}$ be a discounted version of the G with a discount factor $\frac{\epsilon}{2}$. Let $\sigma^{k,\infty}$ be a k-uniform memoryless strategy profile such that $\sigma^{k,\infty}$ is a $\frac{\epsilon}{2}$-Nash equilibrium profile in $G^{\frac{\epsilon}{2}}$. Existence of such a $\sigma^{k,\infty}$ follows from Lemma 8. Then $\sigma^{k,\infty}$ is a strategy profile which is an ϵ-equilibrium profile in the game G. ∎

Lemma 10. *Given a constant ϵ the value of an ϵ-equilibrium with a memoryless strategy profile of a n-player stochastic reach-a-set-game can be approximated within ϵ tolerance by an NP algorithm.*

Proof. The NP algorithm guesses a k-uniform selector σ^k for a k-uniform memoryless ϵ-equilibrium strategy profile $\sigma^{k,\infty}$. It then verifies that the value for the MDP's $G_{\sigma^k_{-i}}$ for every state $s \in S$ and each player i is within ϵ-tolerance as compared to the value of the Markov Chain define by G_{σ^k}. Since the computation of values of a MDP can be achieved in polynomial time (using Linear program solution) it follows that the approximation within ϵ tolerance can be achieved by a NP-algorithm. ∎

Lemmas 5 and 10 yield Theorem 2.

5 Games with Turns

An n-person stochastic game is *turn-based* if at each state, there is exactly one player who determines the next state. Formally, we extend the action sets A_i for $i = 1, \ldots, n$ to be state dependent, that is, for each state $s \in S$, there are action sets A_{is} for $i = 1, \ldots, n$, and we restrict the action sets so that for any $s \in S$, there is at most one $i \in \{1, \ldots, n\}$ such that $|A_{is}| > 1$. A strategy π_i for player i is *pure* if for every history $h = (s_0, a^1, s_1, \ldots, a^k, s_k)$ there is a action $a_k \in A_{is_k}$ such that $\pi_i(a) = 1$. In other words, a strategy is pure if for every history the strategy chooses one action rather than a probability distribution over the action set. A strategy profile is pure if all the strategies of the profile are pure.

We consider payoff functions that are index sets of *Borel sets* (see e.g., [15] for definitions), that is, given a Borel set B, we consider a payoff function χ_B that assigns a payoff 1 to a play that is in the set B, and 0 to a play that is not in the set B. With abuse of notation, we identify the set B with the payoff function χ_B. We consider turn based games in which each player is given a Borel payoff B_i. If $n = 2$, we call the game two-player. A two-player Borel game is *zero sum* if the payoff set B of one player is the complement $S^\omega \setminus B$ of the other player, that is, the players have strictly opposing objectives. Borel sets are studied in descriptive set theory for their rich structural properties. A deep result by Martin shows that two player zero sum infinite stochastic games with Borel payoffs have a value [23]. The proof constructs, for each real $v \in (0, 1]$ a zero sum turn-based deterministic infinite-state game with Borel payoff such that a (pure) winning strategy for player 1 in this game can be used to construct a (mixed) winning strategy in the original game that assures player 1 a payoff of at least v. From the determinacy of turn-based deterministic games with Borel payoffs [22], the existence of value in zero sum stochastic games with Borel payoffs follows. Moreover, the proof constructs ϵ-optimal mixed winning strategies. A careful inspection of Martin's proof in the special case of turn-based probabilistic games shows that the ϵ-optimal strategies of player 1 are pure. The mixed strategies are derived from solving certain one-shot concurrent games at each round. In our special case these one-shot games have pure winning strategies since only one player has a choice of moves.

Lemma 11. *Pure memory determinacy For each $\epsilon > 0$ there is a pure strategy π_1 of player 1 such that for all strategies π_2 of player 2 $E_s^{\pi_1,\pi_2}\{\, f \,\} \geq v - \epsilon$.*

Theorem 3. *For each $\epsilon > 0$ there exists an ϵ-Nash equilibrium in every n-player turn based probabilistic games with Borel payoffs.*

Proof. Our construction is based on a general construction from repeated games. The basic idea is that player i plays optimal strategies in the zero sum game against all other players, and any deviation by player i is punished indefinitely by the other players by playing ϵ-optimal spoiling strategies in the zero sum game against player i (see, e.g., [25,34]). Let player i have the payoff set B_i, for $i = 1, \ldots, n$. Consider the n zero sum games played between i and the team $[n]\setminus\{i\}$, with the winning objective B_i for i. By lemma 11 here is a pure ϵ-optimal strategy π_i^i for player i in this game, and a pure ϵ-optimal spoiling strategy for players $j \neq i$. This spoiling strategy induces a strategy π_i^j for each player $j \neq i$. Now consider the strategy τ^i for player i as follows. Player i plays the strategy π_i^i as long as all the other players j play π_j^j and switch to π_j^i as soon as some player j deviates. Since the strategies are pure, any deviation is immediately noted. The strategies τ^i for $i = 1, \ldots, n$ form an ϵ-Nash equilibrium. ∎

Notice that the construction above for probabilistic Borel games guarantees only ϵ-optimality. As a special case, using the determinacy result of [22], we get that turn based deterministic games (perfect information games) with payoffs corresponding to Borel sets have Nash equilibria.

Corollary 2. *Every turn-based deterministic game with payoffs corresponding to Borel sets has a Nash equilibrium with pure strategy profile.*

A particularly interesting case of turn based probabilistic games is when each payoff function B is an ω-regular set [21]. Games with ω-regular winning conditions are used in the verification and control of (probabilistic) systems [1, 31, 6]. In the special case of turn-based probabilistic games with parity winning conditions, pure and memoryless optimal winning strategies exist for two player zero-sum case [3]. Moreover, the pure memoryless optimal strategies can be computed in NP $\cap$ coNP. Therefore we have the following.

Proposition 1. *There exists a Nash equilibrium with pure strategy profile in every turn-based probabilistic game with parity payoff conditions. The value profile of some Nash equilibrium can be computed in FNP.*

6 Games with Two Players

In this section we consider the special case of two-player reach-a-set-games, namely two-player constant-sum games. For this special cases we prove a NP $\cap$ coNP bound to approximate the value of a ϵ-equilibrium profile, given a fixed ϵ.

6.1 Two-player constant-sum reach-a-set-games

We now define a two-player constant-sum reach-a-set-game.

Definition 9 (Two-player constant-sum reach-a-set-games). *For a two-player reach-a-set-game G let Π denote the set of all ϵ-equilibrium strategy profile for $\epsilon \geq 0$. We use the following notation*

$$v_1(s) = \sup_{\pi \in \Pi} E_{s,\pi} \phi_1^{R_1} \text{ and}$$

$$v_2(s) = \sup_{\pi \in \Pi} E_{s,\pi} \phi_2^{R_2}.$$

The game is constant-sum if for all state $s \in S$ we have the following conditions:

- $v_1(s) + v_2(s) = 1$.
- *for all $\pi \in \Pi$ we have $E_{s,\pi}\phi_1^{R_1} + E_{s,\pi}\phi_1^{R_2} = 1$.* ∎

We now prove computing the values $v_1(s)$ and $v_2(s)$ within a ϵ-tolerance, given a fixed ϵ can be achieved in NP $\cap$ coNP.

Lemma 12. *Let G be a two-player constant-sum reach-a-set-game, s an initial state and v^1 and v^2 be two values. For a fixed ϵ it can be determined in NP $\cap$ coNP whether*

$$v_1(s) \geq v^1 - \epsilon.$$

Proof. It follows from Lemma 9 that there is a k-uniform memoryless ϵ-equilibrium profile $\sigma^{k,\infty} = (\sigma_1^{k,\infty}, \sigma_2^{k,\infty})$ with selector profile $\sigma^k = (\sigma_1^k, \sigma_2^k)$. Since two-player constant-sum reach-a-set game is a special case of n-player stochastic reach-a-set-game it follows from Lemma 10 that the two-player constant-sum reach-a-set-game unique equilibrium value can be approximated by an algorithm in NP.

To prove that there is a coNP algorithm consider the case when $v_1(s) < v^1 - \epsilon$. The coNP algorithm guesses the k-uniform selector σ_2^k for player 2 and verifies that the value of player 1 in the state s in the MDP $G_{\sigma_{-1}^k}$ is less than $v^1 - \epsilon$. Since the value of a MDP at any state can be computed in polynomial time (using a Linear program solution) the required result follows. ∎

Theorem 4 (Two-player constant-sum reach-a-set-games). *Given a fixed ϵ the value of an ϵ-equilibrium profile of two-player stochastic constant-sum reach-a-set-games can be computed in NP $\cap$ coNP.*

6.2 Concurrent Reachability Games

We now show that the values of two player concurrent reachability games (zero-sum reach-a-set-games) can be approximated within ϵ tolerance in NP $\cap$ coNP. The previous best known algorithm was exponential [8].

A two-player concurrent reachability game [7] G is a two-player stochastic game with $R_1 \subseteq S$ as a target set of states for player 1. Given a random history $h = (s_0, a^1, s_1, a^2, \dots)$ player 1 gets a payoff 1 if the history contains a state in R_1, else the player 2 gets an payoff 1. In other words, player 1 plays a reachability game with target set R_1 and player 2 plays a safety game with its safe set of state S_2, where $S_2 = S \setminus R_1$. Let Π_1 and Π_2 be the set of all strategies of player 1 and player 2 respectively. Then for any state $s \in S$ we use the following notation

$$v_1(s) = \sup_{\pi_1 \in \Pi_1} \inf_{\pi_2 \in \Pi_2} E_{s,\pi_1,\pi_2} \phi_1^{R_1}$$

$$v_2(s) = \sup_{\pi_2 \in \Pi_2} \inf_{\pi_1 \in \Pi_1} E_{s,\pi_1,\pi_2} \phi_2^{S_2}$$

It follows from determinacy of Blackwell games [23] that for all states $s \in S$, we have $v_1(s) + v_2(s) = 1$. Let $W_2 = \{s | v_2(s) = 1\}$ and $W_1 = \{s | v_1(s) = 1\}$. We prove that the concurrent reachability game can be reduced to a two-player reach-a-set game G_R with W_1 and W_2 as the target set of states for player 1 and player 2, respectively and also all the states in W_1 and W_2 are absorbing states or sink states. In the proof below we use the following notation

$$v_1^{\pi_1,\pi_2}(s) = E_{s,\pi_1,\pi_2} \phi_1^{R_1}$$

$$reach^{\pi_1,\pi_2}(W_2)(s) = E_{s,\pi_1,\pi_2} \phi_2^{R_2 = W_2}$$

$$reach^{\pi_1,\pi_2}(W_1)(s) = E_{s,\pi_1,\pi_2} \phi_1^{R_1 = W_1}$$

$$reach(W_2)(s) = \sup_{\pi_2 \in \Pi_2} \inf_{\pi_1 \in \Pi_1} reach^{\pi_1,\pi_2}(W_2)(s)$$

Lemma 13. *Let G be a concurrent reachability game and G_R be the two-player reach-a-set-game with the target set for player 1 and player 2 being $R_1 = W_1, R_2 = W_2$, respectively. Also every state in $W_1 \cup W_2$ is an absorbing state and once the process of states reaches a state in W_1 or W_2 it remains there forever. Then, for all states $s \in S$ we have $v_2(s) = reach(W_2)(s)$.*

Proof. From every state $s \in W_2$ there is a strategy π_2' such that player 2 can stay in its safety set $S \setminus R_1$ with probability 1. Hence combining a strategy to reach the set W_2 with the strategy π_2' we get that $v_2(s) \geq reach(W_2)(s)$.

Suppose $v_2(s) > reach(W_2)(s)$. It follows from [8] that player 2 has an optimal memoryless strategy in the concurrent reachability game G. Let π_2 be an optimal memoryless strategy for player 2 in the concurrent reachability game G. Fixing the memoryless optimal strategy π_2 for player 2 in the game G_R we get an MDP G_{R,π_2} where at each state player 2 plays according to the strategy π_2. Let an optimal memoryless strategy of player 1 against the strategy π_2 in the game G be π_1. The game G_{π_1,π_2} is a Markov chain. Let C be any terminal strongly connected component of the Markov chain G_{π_1,π_2}. If $C \cap R_1 \neq \emptyset$ then from every state $s \in C$ player 1 wins with probability 1, and if $C \cap R_1 = \emptyset$ then from every state $s \in C$ player 2 wins with probability 1. Since π_2 is an optimal strategy and π_1 is an optimal strategy against π_2 we have that every terminal strongly connected component is a subset of W_1 or W_2. Hence in the Markov chain G_{R,π_1,π_2} we have

$$reach^{\pi_1,\pi_2}(W_1)(s) + reach^{\pi_1,\pi_2}(W_2)(s) = 1$$

Now if $v_2(s) > reach^{\pi_1,\pi_2}(W_1)(s)$ then we have $reach^{\pi_1,\pi_2}(W_1)(s) < 1 - v_2(s) = v_1(s)$. Since π_2 is an optimal strategy for player 2 and π_1 is an optimal strategy against it we must have $reach^{\pi_1,\pi_2}(W_1)(s) = v_1(s)$. Hence this is a contradiction. Therefore, we have $v_2(s) \leq reach(W_2)(s)$. Hence proved that $v_2(s) = reach(W_2)(s)$. ∎

Lemma 14. *Given a fixed ϵ, the values $v_1(s)$ and $v_2(s)$ of a concurrent reachability game can be approximated within ϵ tolerance in NP $\cap$ coNP.*

Proof. It follows from Lemma 13 that a concurrent reachability game can be reduced to a two-player stochastic reach-a-set game with target set for player 1 and player 2 being W_1 and W_2 respectively. It follows from the result of deAlfaro and Henzinger [6] that the sets W_1 and W_2 can be computed in polynomial time. It follows from the result of Martin on determinacy of Blackwell games [23] that this game is a constant-sum two-player stochastic reach-a-set-game. The result then follows from Lemma 12. ∎

Corollary 3 (Two-player concurrent reachability games). *The value of a two-player concurrent reachability game can be approximated within ϵ-tolerance in NP $\cap$ coNP, given a fixed ϵ.*

The natural question at this point is whether there is a polynomial time algorithm for concurrent zero sum reachability games. Since simple stochastic games [4] can be easily reduced to concurrent reachability games, a polynomial time algorithm for this problem will imply a polynomial time algorithm for simple stochastic games and mean payoff games [37]. These have been long standing open problems in the area.

Acknowledgments.

We thank Antar Bandyopadhyay for many insightful discussions.

References

1. R. Alur, T.A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.

2. D. Blackwell and T.S. Ferguson. The big match. *Annals of Mathematical Statistics*, 39:159–163, 1968.
3. K. Chatterjee, M. Jurdziński, and T.A. Henzinger. Quantitative stochastic parity games. In *SODA 04: ACM-SIAM Symposium on Discrete Algorithms*, 2004. (To appear): Technical Report: UCB/CSD-3-1280 (October 2003).
4. A. Condon. The complexity of stochastic games. *Information and Computation*, 96(2):203–224, 1992.
5. V. Conitzer and T. Sandholm. Complexity results about Nash equilibria. In *IJCAI 03: International Joint Conference on Artificial Intelligence*, pages 765–771, 2003.
6. L. de Alfaro and T.A. Henzinger. Concurrent omega-regular games. In *Proceedings of the 15th Annual Symposium on Logic in Computer Science*, pages 141–154. IEEE Computer Society Press, 2000.
7. L. de Alfaro, T.A. Henzinger, and O. Kupferman. Concurrent reachability games. In *FOCS 98: Foundations of Computer Science*, pages 564–575. IEEE Computer Society Press, 1998.
8. L. de Alfaro and R. Majumdar. Quantitative solution of omega-regular games. In *STOC 01: Symposium on Theory of Computing*, pages 675–683. ACM Press, 2001.
9. H. Everett. Recursive games. In *Contributions to the Theory of Games III*, volume 39 of *Annals of Mathematical Studies*, pages 47–78, 1957.
10. J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, 1997.
11. A.M. Fink. Equilibrium in a stochastic n-person game. *Journal of Science of Hiroshima University*, 28:89–93, 1964.
12. D. Gillete. Stochastic games with zero stop probabilitites. In *Contributions to the Theory of Games III*, pages 179–188. Princeton University Press, 1957.
13. J.F. Nash Jr. Equilibrium points in n-person games. *Proceedings of the National Academny of Sciences USA*, 36:48–49, 1950.
14. S. Kakutani. A generalization of Brouwer's fixed point theorem. *Duke Journal of Mathematics*, 8:457–459, 1941.
15. A. Kechris. *Classical Descriptive Set Theory*. Springer, 1995.
16. H. Kopetz. *Real-Time Systems: Design Principles for Distributed Embedded Applications*. Kluwer Academic Publishers, 1997.
17. P.R. Kumar and T.H. Shiau. Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games. *SIAM J. Control and Optimization*, 19(5):617–634, 1981.
18. R.J. Lipton, E. Markakis, and A. Mehta. Playing large games using simple strategies. In *EC 03: Electronic Commerce*, pages 36–41. ACM Press, 2003.
19. A. Maitra and W. Sudderth. Finitely additive stochastic games with borel measurable payoffs. *International Journal of Game Theory*, 3:257–267, 1998.
20. Z. Manna and A. Pnueli. *The Temporal Logic of Reactive and Concurrent Systems: Specification*. Springer-Verlag, 1992.
21. Z. Manna and A. Pnueli. *Temporal Verification of Reactive Systems: Safety*. Springer-Verlag, 1995.
22. D.A. Martin. Borel determinacy. *Annals of Mathematics*, 102(2):363–371, 1975.
23. D.A. Martin. The determinacy of Blackwell games. *The Journal of Symbolic Logic*, 63(4):1565–1581, 1998.
24. J-F. Mertens and T.P. Parthasarathy. Non zero-sum stochastic games. In T.E.S. Raghavan et al, editor, *Stochastic Games and related topics*, pages 145–148. Kluwer Academic Publishers, 1991.
25. G. Owen. *Game Theory*. Academic Press, 1995.

26. C.H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *Journal of Computer and Systems Sciences*, 48(3):498–532, 1994.
27. C.H. Papadimitriou. Algorithms, games, and the internet. In *STOC 01: Symposium on Theory of Computing*, pages 749–753. ACM Press, 2001.
28. T.E.S. Raghavan and J.A. Filar. Algorithms for stochastic games — a survey. *ZOR — Methods and Models of Operations Research*, 35:437–472, 1991.
29. P. Secchi and W.D. Sudderth. Stay-in-a-set games. *International Journal of Game Theory*, 30:479–490, 2001.
30. L.S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci. USA*, 39:1095–1100, 1953.
31. W. Thomas. On the synthesis of strategies in infinite games. In *STACS 95: Theoretical Aspects of Computer Science*, volume 900 of *Lecture Notes in Computer Science*, pages 1–13. Springer-Verlag, 1995.
32. F. Thuijsman. *Optimality and Equilibria in Stochastic Games.* CWI-Tract 82, CWI, Amsterdam, 1992.
33. F. Thuijsman and T.E.S. Raghavan. Perfect information stochastic games and related classes. *International Journal of Game Theory*, 26:403–408, 1997.
34. N. Vieille. Two player stochastic games I: a reduction. *Israel Journal of Mathematics*, 119:55–91, 2000.
35. N. Vieille. Two player stochastic games II: the case of recursive games. *Israel Journal of Mathematics*, 119:93–126, 2000.
36. B. von Stengel. Computing equilibria for two-person games. *Chapter 45, Handbook of Game Theory*, 3:1723–1759, 2002. (editors R.J. Aumann and S. Hart).
37. U. Zwick and M.S. Paterson. The complexity of mean payoff games on graphs. *Theoretical Computer Science*, 158:343–359, 1996.