Answer to a question on p. 47 (the chapter by L. Kallenberg "Finite State and Action MDPs")

December 3, 2001

The open problem on p. 47 consists of two questions regarding constrained average reward MDPs with finite state and action sets: (i) how to find the best randomized stationary and (ii) how to find the best pure stationary policy. The latter problem is NP-hard. Therefore, in view of the current state of the knowledge in the area of $P = NP$?, there is little hope to find a good algorithm to compute the best pure stationary policy.

This NP-hardness result follows from the paper by Filar and Krass, *Math Oper. Res.* **19**, 223-237, 1994, where it was shown that the Hamiltonian Cycle Problem for a graph with N nodes can be reduced to a constrained average reward unichain MDP with N states and no more than N actions in each state.

In Feinberg, *Math Oper. Res.* **25**, 130-140, 2000, it was shown that finding the best pure stationary policy in a discounted MDP is an NP-hard problem. This also implies the NP-hardness result for unichain average reward MDPs. Indeed, for a discounted MDP, one may consider an average reward MDP with the same state and action sets, with the same reward functions, and with the transition probabilities $p^{\beta}(i|j,a) = \beta p(i|j,a) + (1-\beta)\delta_{i,j}$, where i is the initial state. Then the average rewards per unit time for this MDP are equal to the total discounted rewards for the original MDP multiplied by $(1-\beta)$. This construction reduces a constrained discounted MDP to a constrained average reward unichain MDP.

Communicated by Eugene Feinberg