Collaborative Research: Enhanced Radio Spectrum via Information Acquisition and Learning

Bayesian Rate Control Over Markovian Channel States

Parisa Mansourifard*, Bhaskar Krishnamachari*, Tara Javidi+

[*] University of Southern California [+]University of California, San Diego

parisama@usc.edu, bkrishna@usc.edu, tjavidi@ucsd.edu

Introduction

- The channel bandwidth: stochastic sequence, $B(t)$
- The states $i = 1,\dots,M$
- Transition matrix, P
- Partially Observable Markov Decision Process
- Finite horizon T, time steps $t = 1,\dots,T$
- Our belief vector at time t, $b_t = [b_t(1),\dots,b_t(M)]$

$$b_t(i) = \Pr(B(t) = i), i = 1,\dots,M$$

- Choose an action (rate) a; if higher than B(t), we will find out the exact value of B(t), but incur a penalty.
- Evolution of belief vector :

$$b_{t+1} = \begin{cases} T_a b_t P, & if\ a \le B(t) \\ I_{B(t)} P, & if\ a > B(t) \end{cases} \qquad T_a b(i) = \begin{cases} 0, & if\ i < a \\ \dfrac{b(i)}{\sum_{j=a}^{M} b(j)}, & if\ i \ge a \end{cases}$$

- Immediate reward :

$$R(B(t); a) = \begin{cases} a, & if\ a \le B(t) \\ B(t) - C(a - B(t)), & if\ a > B(t) \end{cases}$$

Problem Formulation

- Policy vector: $\pi = [\pi(1),\dots,\pi(T)]$
- Selecting an action $\pi(t) = a_t \in \{1,\dots,M\}$
- Maximizing total discounted expected reward:

$$\max_{\pi} E^{\pi}[\sum_{t=1}^{T} \beta^{t-1} R(b_t; a_t) \mid b_1]$$

- defining value function, i.e. maximum expected remaining reward starting from time t: $V_t(b)$

Dynamic Programming

$$V_t(b_t) = \max_{a=1,\dots,M} V_t(b_t; a), \quad \forall t = 1,\dots,T$$

$$V_T(b_T; a) = \overline{R}(b_T; a),$$

$$V_t(b_t; a) = \overline{R}(b_t; a) + \beta V_t^{f}(b_t; a)$$

$$V_t^{f}(b_t; a) = E\{V_{t+1}(b_{t+1}) \mid a\}$$

$$= \sum_{i=a}^{M} b_t(i) V_{t+1}(T_a b_t P) + \sum_{i=1}^{a-1} b_t(i) V_{t+1}(I_i P)$$

Optimal Policy

Our goal is to find (at least structural properties of) the optimal policy. $a_t^{Optimal} = \arg\max_{a=1,\dots,M} V_t(b_t; a)$

Lower Bound Action

ignoring the impact of the current action on the future reward, myopic policy is given by

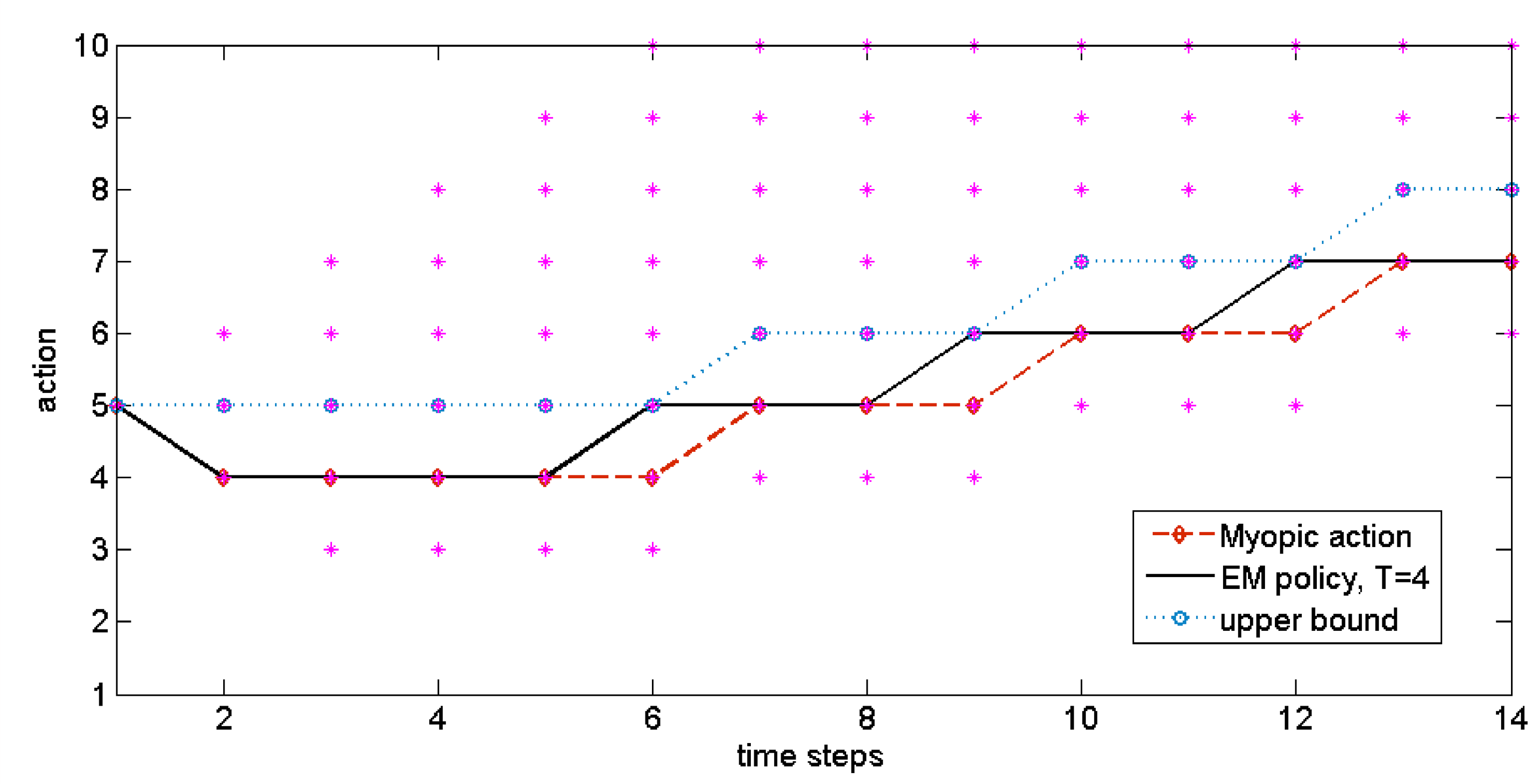
$$a_t^{Myopic}(b_t) = \arg\max_{a=1,\dots,M} \overline{R}(b_t; a) = \min\{a \mid \sum_{i=1}^{a} b_t(i) \ge \frac{1}{1+C}\}$$

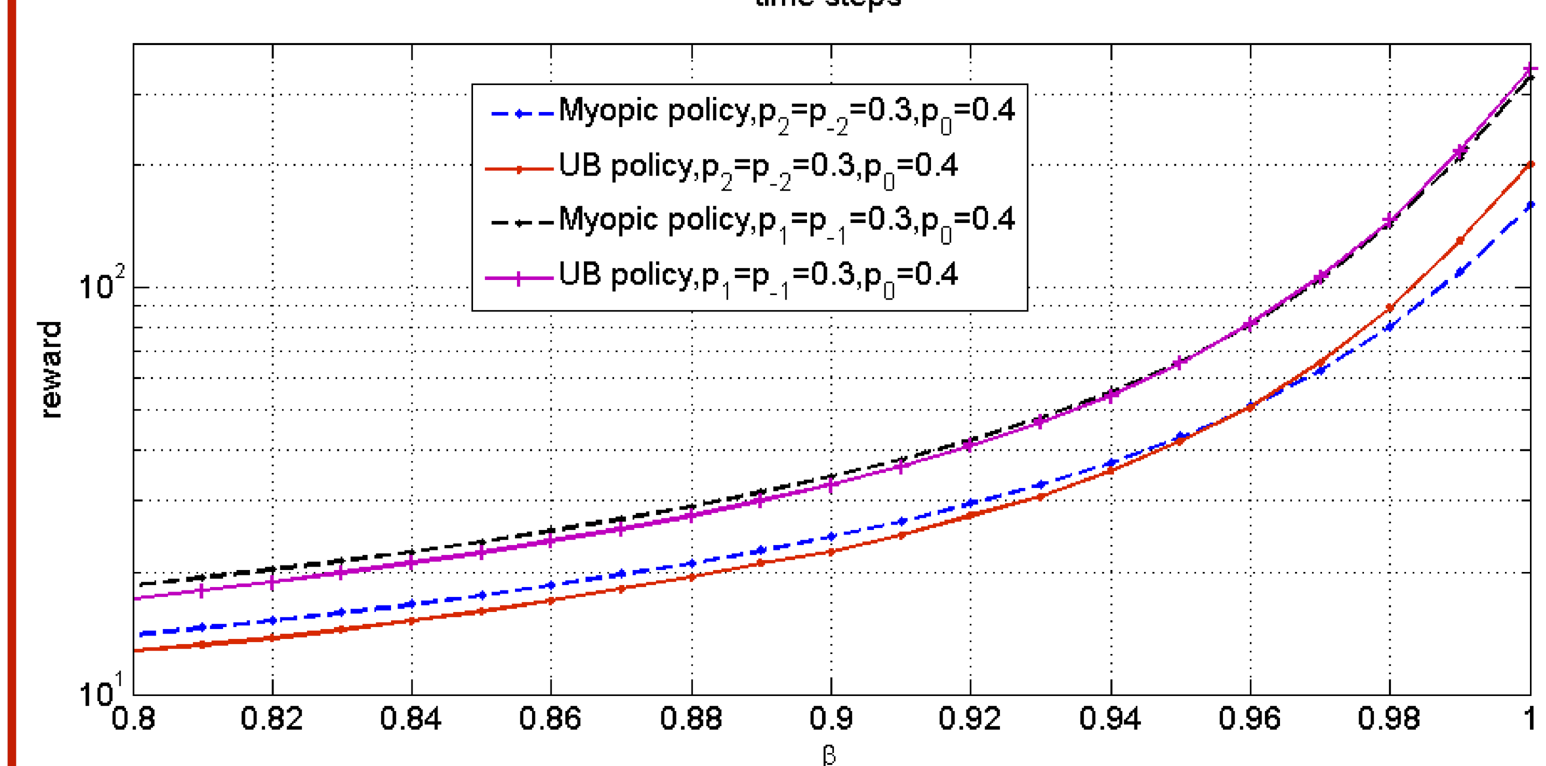
Upper Bound Action

$$a_t^{UB}(b_t) = \min\{a : \sum_{i=a+1}^{M} [\frac{\beta}{1-\beta}(i-a) + 1 + C] b_t(i) \le C\}$$

- Assumption: The P matrix satisfies the State-Independent State Change (SISC) property.

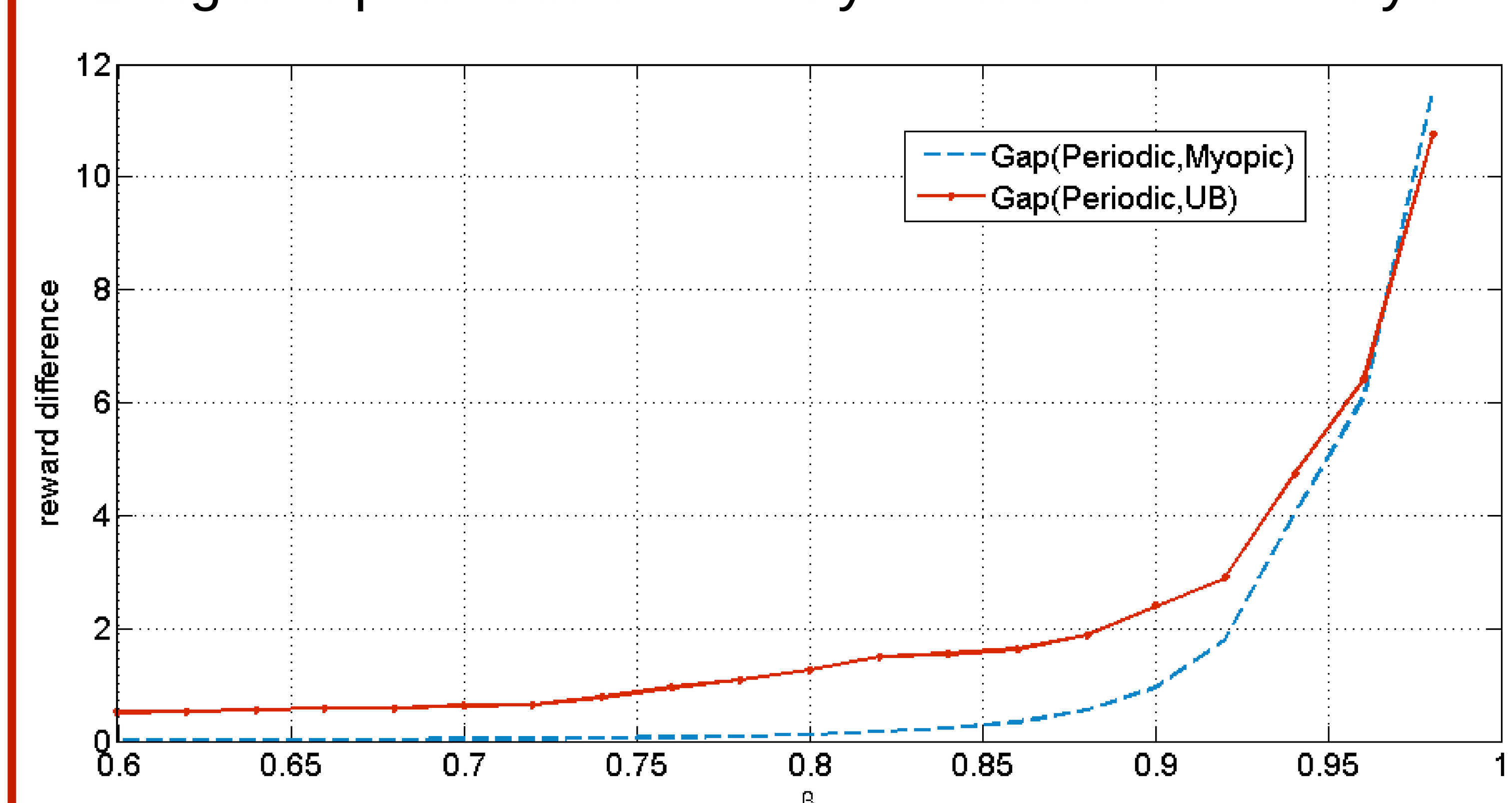
Simulation Results





Heuristic: Periodic Myopic-UB policy (some steps myopic and one step UB)

Length of period increases by C and decreases by beta



This work is funded by the U.S. National Science Foundation through grants ECCS-EARS Grants **1247995** and **1248017**