Utility Maximization in Cognitive Radio Networks using POMDP Approximate Value Iteration methods

Bharath Keshavamurthy, Nicolò Michelusi

[NM: title is a bit too long..]

[NM: Please use this template! And work directly on this file. I noticed that the revision was missing the commands "nm", "sst" and "add" which I created..]

***Abstract*—**

***Index Terms*—Hidden Markov Model, POMDP, and the PERSEUS Algorithm**

I. Introduction

II. System Model

[NM: add figure with system model]

A. The Observation Model

We consider a network consisting of one licensed user termed the Primary User (PU)**[NM: why one PU? I think it is more plausible to assume mulitple PUs that access the spectrum and one SU.. this is more consistent with our model of spectrum occupancy]** and one cognitive radio node termed the Secondary User (SU), ~~which is~~ equipped with a spectrum sensor. The goal of the SU is to opportunistically access portion of the spectrum left unused by the PU, with the goal to maximize its own throughput. To this end, the SU should learn to intelligently access spectrum holes (white-spaces) in order to maximize its throughput while maintaining strict non-interference compliance with incumbent transmissions. The wideband signal observed at the SU receiver is denoted as $y(n)$ and is given by,

$$y(n) \;=\; \sum_{p=1}^{P}\sum_{m=0}^{M_p-1} h_p(m)x_p(n-m)+v(n), \tag{1}$$

where~~Here,~~ $y(n)$ is expressed as a convolution of the ~~PU~~ signal $x_p(n)$ of the pth PU with the channel impulse response $h_p(n)$, and $v(n)$ denotes additive white Gaussian noise (AWGN) with variances σ_v^2.~~added with a noise term v(n).~~ ~~Equation~~Eq. (1) can be written in the frequency domain by taking a K-point DFT which decomposes the observed wideband signal into ~~K~~ K discrete narrow-band components as shown below,

$$Y_k(i) \;=\; \sum_{p=1}^{P} H_{p,k}X_{p,k}(i)+V_k(i) \tag{2}$$

where $i \in \{1,2,3,...,T\}$ represents the index of the observation; $k \in \{1,2,3,...,K\}$ represents the index of the ~~channel~~ components in the frequency domain; $V_k(i) \sim \mathcal{CN}(0,\sigma_V^2)$ represents the circularly symmetric additive complex Gaussian noise sample, i.i.d across channel indices and across time indices; $X_{p,k}(i)$ is the signal of the pth PU in the frequency domain, and $H_{p,k}$ is its frequency domain channel.**[NM: is the chennl constant across i?]** The~~se~~ noise samples are assumed to be independent of the occupancy state of the channels. We further assume that the P PUs employ an orthogonal access to the spectrum (e.g., OFDMA) so that $X_{p,k}(i)X_{q,k}(i)=0,\ \forall p \neq q$. Thus, letting p_k be the index of the PU that contributes to the signal in the kth spectrum band (possibly, $p_k=0$ if no PU is transmitting in the kth spectrum band), and letting $H_k=H_{p_k,k}$ and $X_k(i)=X_{p_k,k}(i)$, we can rewrite (2) as

$$Y_k(i) \;=\; H_kX_k(i)+V_k(i). \tag{3}$$

Thus, H_k~~$H_k \sim \mathcal{CN}(0,\sigma_H^2)$~~ represents the ~~$k^{th}$~~ kth DFT coefficient of the impulse response $h_{p_k,k}(n)$ of the channel in between the PU operating on the kth spectrum band and the SU; we model it as a zero-mean circular symmetric complex Gaussian random variable with variance σ_H^2, independent across frequency bands, $H_k \sim \mathcal{CN}(0,\sigma_H^2)$, ~~These impulse response samples are assumed to be~~ and independent of the occupancy state of the channels.

B. ~~*The Correlation Model*~~*PU Spectrum occupancy model*

We now introduce the model of PU occupancy over time and across the frequency domain. We model each $X_k(i)$ as

$$X_k(i)=\sqrt{P_{tx}}B_k(i)S_k(i),$$

where P_{tx} is the transmission power of the PUs, $S_k(i)$ is the transmitted symbol, modeled as a constant amplitude signal, $|S_k(i)| = 1$, i.i.d. over time and across frequency bands;[1] $B_k(i) \in \{0,1\}$ is the binary spectrum occupancy variable, with $B_k(i)=1$ if the kth spectrum band is occupied by a PU at time i, and $B_k(i)=0$ otherwise.~~The PU occupancy behavior in each sub-band is modelled as $X_k \in \{0,1\}$ taking two possible values 0 ($Idle$) and 1 ($Occupied$).~~ Therefore, the PU occupancy behavior in the entire wideband spectrum of interest discretized into narrow-band frequency components at time index i can be modeled as the vector~~a vector as shown below.~~

$$\vec{B}(i) \;=\; [B_1(i),B_2(i),B_3(i),\cdots,B_K(i)]^T \in \{0,1\}^K. \tag{4}$$

This research has been funded by —–.

The authors are with the School of Electrical and Computer Engineering, Purdue University. email: {bkeshava,michelus}@purdue.edu.

[1] In the case where $S_k(i)$ does not have constant amplitude, we may approximate $H_kS_k(i)$ as complex Gaussian, without any modification to the following analysis.

[NM: use B instead of X for the occupancy] ~~We assume that the PU employs an OFDMA access strategy and therefore, this spectrum occupancy vector has a sparse support.~~ PUs join and leave the spectrum at random times. To capture this temporal correlation in the spectrum occupancy dynamics of PUs, we model the spectrum occupancy dynamics as a Markov process: ~~with the following transition model.~~ given $\vec{X}(i)$, the spectrum occupancy state at time index i, $\vec{X}(i+1)$ is independent of the past, $\vec{X}(j),\ j<i;\ j,i \in \{1,2,3,...,T\}$, i.e.

$$\mathbb{P}(\vec{X}(i+1)|\vec{X}(j),\ \forall j \leq i) \ = \ \mathbb{P}(\vec{X}(i+1)|\vec{X}(i)). \tag{5}$$

Additionally, when joining the spectrum pool, PUs occupy a number of adjacent spectrum bands, and may vary their spectrum needs depending on traffic demands, channel conditions, etc. To capture this behavior, we model $\vec{X}(i)$ as having~~Additionally, the spectrum occupancy vector $\vec{X}(i)$ exhibits~~ Markovian correlation across ~~the~~ sub-bands as,

$$\mathbb{P}(\vec{X}(i+1)|\vec{X}(i)) \ = \ \prod_{k=1}^{K} \mathbb{P}(B_{k+1}(i+1)|B_{k+1}(i), B_k(i)). \tag{6}$$

[NM: I dont understand, didnt we say that $B_{k+1}(i+1)$ should depend on $B_{k+1}(i)$ and $B_k(i+1)$? (instead of $B_k(i)$)] That is, $B_{k+1}(i+1)$ depends on the occupancy state of the adjacent spectrum band $B_k(i+1)$ at the same time, and that of the same spectrum band k in the previous block i. The true states encapsulate the actual occupancy behavior of the PU and the measurements at the SU are noisy observations of these true states which are modeled to be the observed states of a Hidden Markov Model (HMM). Owing to physical design limitations at the SU's spectrum sensor**[NM: citation?]**, not all sub-bands in the discretized spectrum can be sensed. Let κ with $1 \leq \kappa \leq K$ be the number of spectrum bands that can be sensed by the SU at any time, capturing the spectrum sensing constraint. Let $\mathcal{K}_i \subseteq \{1,2,\dots,K\}$ with $|\mathcal{K}_i| \leq \kappa$ be the set of indices corresponding to the spectrum bands sensed by the SU, which is part of our design. Then,~~Given this constraint,~~ we model the emission process of the HMM as~~shown below. The observation vector at time index i is given by,~~

$$\vec{Y}(i) \ = \ [Y_k(i)]_{k \in \mathcal{K}_i}. \tag{7}$$

~~where, $y_k(i) = \phi$ indicates that the SU did not sense sub-band k at time index i. Therefore, the observation probability termed as the emission probability of the HMM is given by,~~ Given the spectrum occupancy vector $\vec{B}(i)$ and the set of sensed spectrum bands $\mathcal{K}_i$, the probability density function of $\vec{Y}(i)$ is expressed as

$$f(\vec{Y}(i)|\vec{B}(i), \mathcal{K}_i) = \prod_{k \in \mathcal{K}_i} f(Y_k(i)|B_k(i)), \tag{8}$$

owing to the channels, noise and transmitted symbols across frequency bands. ~~where,~~Moreover,

$$Y_k(i)|B_k(i) \sim \mathcal{CN}(0, \sigma_H^2 P_{tx} B_k(i) + \sigma_V^2). \tag{9}$$

[NM: no need to define m_r] Now, we model the spectrum access scheme of the SU as a Partially Observable Markov Decision Process (POMDP) wherein the goal of the POMDP agent is to devise an optimal sensing and access policy in order to maximize its throughput while maintaining strict non-interference compliance with incumbent transmissions.

C. *The POMDP Agent Model*

The agent's limited observational capabilities coupled with its noisy observations result in an increased level of uncertainty at the agent's end about the occupancy state of the spectrum under consideration and the exact effect of executing an action on the radio environment. The transition model of the underlying MDP as described in the ~~Correlation Model~~Sec. XXXX of this paper, is denoted by A and is learnt by the agent by interacting with the radio environment. The emission model B is given by (8), with $f(Y_k(i)|B_k(i))$ given by (8). We model the POMDP as a tuple $(\mathcal{X}, \mathcal{A}, \mathcal{Y}, \mathcal{B}, A, B)$ where~~,~~ $\mathcal{X} \equiv \{0,1\}^K$ represents the state space of the underlying MDP with states $\vec{x}$ ~~which are~~ given by all possible realizations of the spectrum occupancy vector as given by (4)~~ and $|\mathcal{X}| = 2^K$~~, $\mathcal{A}$ represents the action space of the agent, given by all $\begin{pmatrix} K \\ \kappa \end{pmatrix}$ possible combinations in which the κ spectrum bands are chosen to be sensed out of K at any given time block;~~ considering the imposed sensing limitations,~~ $\mathcal{Y}$ represents the observation space of the agent based on the Observation Model outlined earlier in the paper, and $\mathcal{B}$**[NM: Do you need to define the belief space? It is a function of the underlying model A, B]** represents the belief space of the agent.

The run-time or interaction time of the agent is quantized into discrete time-steps termed as episodes**[NM: what is one episode, is it one block? If that is the case, just go with blocks]**. At the beginning of each episode **[NM: time i? Be more precise]**, the agent selects κ spectrum bands out of K, thus defining the sensing set $\mathcal{K}_i$, performs spectrum sensing on these spectrum bands, ~~executes an action $a \in \mathcal{A}$,~~ observes $\vec{y} \in \mathcal{Y}$, and updates its belief $\forall \vec{x}'$ as follows.

$$b_a^{\vec{y}}(\vec{x}') \ = \ \mathbb{P}(\vec{x}'|\vec{y}, a, \vec{b}) \ = \ \frac{\mathbb{P}(\vec{y}|\vec{x}', a)}{\mathbb{P}(\vec{y}|a, \vec{b})} \sum_{\vec{x} \in \mathcal{X}} \mathbb{P}(\vec{x}'|\vec{x}, a) b(\vec{x}) \tag{10}$$

where, $\mathbb{P}(\vec{y}|a, \vec{b})$ is the normalization constant given by,**[NM: no need to provide this formula it is obvious]**

$$\mathbb{P}(\vec{y}|a, \vec{b}) \ = \ \sum_{\vec{x}' \in \mathcal{X}} \mathbb{P}(\vec{y}|\vec{x}', a) \sum_{\vec{x} \in \mathcal{X}} \mathbb{P}(\vec{x}'|\vec{x}, a) b(\vec{x}) \tag{11}$$

$\vec{b} \in \mathcal{B}$ represents the belief vector of the agent, i.e. a probability distribution over all states, in the previous time-step, $b(\vec{x}) \in \vec{b}$ is termed the belief and it represents the degree of certainty assigned to world state $\vec{x} \in \mathcal{X}$ by the belief vector $\vec{b}$. The belief, by definition being a probability measure, has to satisfy the Kolmogorov's axioms, i.e.**[NM: this is also obvious, no need to state..]**

$$\sum_{\vec{x} \in \mathcal{X}} b(\vec{x}) \ = \ 1 \tag{12}$$

$$0 \leq b(\vec{x}) \leq 1$$

Considering sub-band k, the set of available actions to the agent in an episode i is given by,

$$a_k(i) \;=\; \begin{cases} 1, & \text{sense and access sub-band } k, \\ 0, & \text{do nothing with respect to sub-band } k \end{cases} \tag{13}$$

We define $\kappa < K$ to be the number of the channels the SU can sense simultaneously. Based on this sensing constraint, the size of the action space is given by, $\mathcal{A} \;=\; K^{\kappa}$**[NM: No, it is $K!/\kappa!/(K-\kappa)!$].** **[NM: You need to define the spectrum access model first, before defining the reward. How does the SU access the spectrum given its belief?]** The reward to the agent is modelled as follows based on the number of truly idle sub-bands found which accounts for the throughput maximization aspect of our end-goal and a penalty for missed detections which accounts for the incumbent non-interference constraint.

$$R(\vec{x}(i), a(i)) \;=\; (1 - P_{FA}(i)) + \lambda P_{MD}(i) \tag{14}$$

where, $P_{FA}(i)$ represents the False Alarm Probability across all channels in episode i, $P_{MD}(i)$ represents the Missed Detection Probability across all channels in episode i, and $\lambda < 0$ represent the cost term penalizing the agent for missed detections, i.e. interference with the incumbent. The action policy of the agent $\pi : \mathcal{B} \rightarrow \mathcal{A}$ maps the belief vectors $\vec{b} \in \mathcal{B}$ to actions $a \in \mathcal{A}$ and is characterized by a Value Function,

$$V^{\pi}(\vec{b}) \;=\; \mathbb{E}_{\pi}\Big[\sum_{i=0}^{\infty} \gamma^{i} R(\vec{b}_i,\ \pi(\vec{b}_i))|\vec{b}_0 = \vec{b}\Big] \tag{15}$$

where, $0 < \gamma < 1$ is the discount factor, $\pi(\vec{b}_i)$ is the action taken by the agent in episode i under policy π, and $\vec{b}_0$ is the initial belief vector. The optimal policy π^* specifies the optimal action to take in the current episode assuming that the agent behaves optimally in future episodes as well. It is evident from equation (15) that we have an infinite-horizon discounted reward problem formulation and in order to solve for the optimal policy we need to solve the modified Bellman equation given as follows. $\forall \vec{b} \in \mathcal{B}$,

$$V^*(\vec{b}) \;=\; \max_{a \in \mathcal{A}} \Big[\sum_{\vec{x} \in \mathcal{X}} R(\vec{x}, a) b(\vec{x}) + \gamma \sum_{\vec{y} \in \mathcal{Y}} \mathbb{P}(\vec{y}|a, \vec{b}) V^*(\vec{b}_a^{\vec{y}})\Big] \tag{16}$$

Given the high dimensionality of the spectrum sensing and access problem, i.e. the number of states of the underlying MDP scales exponentially with the number of sub-bands, solving equation (16) using Exact Value Iteration and Policy Iteration algorithms is computationally infeasible. Additionally, solving for the optimal policy from equation (16) requires prior knowledge about the underlying MDP's transition model. Therefore, in this paper we present a framework to estimate the transition model of the underlying MDP and then utilize this learned model to solve for the optimal policy by employing Randomized Point-Based Value Iteration techniques, namely, the PERSEUS algorithm.

[NM: All the citations should go into the file ref.bib with the specific bibtex format (you can typically get the entry from IEEE Xplore)]