

# Learning based Cognitive Radio Access via Randomized Point-based Approximate POMDPs IEEE TCCN 2021: Addressing Reviewer Comments

Bharath Keshavamurthy and Nicolò Michelusi

April 2021

## 1 Addressing Reviewer-1 Comments

1. **Comment:** As claimed, PUs typically occupy a set of adjacent channels, which leads to frequency correlation. That is, the occupancy of frequency band  $k$  is related to the occupancy of the adjacent frequency bands  $k-1$  and  $k+1$ . However, the authors assume that the occupancy of frequency band  $k$  only depends on the occupancy of frequency band  $k-1$  in Eq. (6). Is that reasonable?

**Addressal:** Although we make a claim in our original manuscript that the Markov chain across frequency is reversible (“If the frequency correlation direction is changed, i.e., the occupancy of channel  $k+1$  influences the occupancy of channel  $k$  (bottom-up vs top-down correlation), our model and subsequent analyses still hold”), we fail to provide a mathematical basis for this claim. We address this comment in our revised manuscript by explicitly highlighting the *the reversibility of the Markov chain across frequency*:

$$\mathbb{P}(\vec{B}(i+1)|\vec{B}(i))=\mathbb{P}(B_1(i+1)|B_1(i))\prod_{k=2}^K\mathbb{P}(B_k(i+1)|B_{k-1}(i+1),B_k(i)); \quad (1)$$

Using Bayes’ rule,

$$\mathbb{P}(\vec{B}(i+1)|\vec{B}(i))=\mathbb{P}(B_K(i+1)|B_K(i))\prod_{k=2}^K\mathbb{P}(B_{k-1}(i+1)|B_k(i+1),B_{k-1}(i))\frac{\mathbb{P}(B_k(i+1),B_k(i),B_{k-1}(i))}{\mathbb{P}(B_{k-1}(i+1),B_k(i),B_{k-1}(i))}, \quad (2)$$

which can be re-written as

$$\mathbb{P}(\vec{B}(i+1)|\vec{B}(i)) = \mathbb{P}(B_K(i+1)|B_K(i)) \prod_{k=2}^K \mathbb{P}(B_{k-1}(i+1)|B_k(i+1), B_{k-1}(i)) \frac{\mathbb{P}(B_k(i+1)|B_k(i))\mathbb{P}(B_k(i)|B_{k-1}(i))}{\mathbb{P}(B_{k-1}(i+1)|B_{k-1}(i))\mathbb{P}(B_k(i)|B_{k-1}(i))}, \quad (3)$$

wherein after cancelling out the common terms, we get

$$\mathbb{P}(\vec{B}(i+1)|\vec{B}(i)) = \mathbb{P}(B_K(i+1)|B_K(i)) \prod_{k=2}^K \mathbb{P}(B_{k-1}(i+1)|B_k(i+1), B_{k-1}(i)). \quad (4)$$

Additionally, our inference is that the reviewer is also expecting a validation for a correlation model in which the incumbent occupancy in [channel  $k$  | time-slot  $i+1$ ] depends on the following (*simultaneous bi-directionality*):

- PU Occupancy in [channel  $k-1$  | time-slot  $i+1$ ],
- PU Occupancy in [channel  $k+1$  | time-slot  $i+1$ ], and
- PU Occupancy in [channel  $k$  | time-slot  $i$ ]; i.e.,

Mathematically,

$$\mathbb{P}(\vec{B}(i+1)|\vec{B}(i)) = \mathbb{P}(B_1(i+1)|B_2(i+1), B_1(i))\mathbb{P}(B_K(i+1)|B_{K-1}(i+1), B_K(i)) \prod_{k=2}^K \mathbb{P}(B_k(i+1)|B_{k-1}(i+1), B_{k+1}(i+1), B_k(i)). \quad (5)$$

In our revised manuscript, we have performed a BIC fit evaluation for this correlation model on the DARPA SC2 Active Incumbent PSD data.  
**Awaiting results...**

2. **Comment:** What is the definition of  $\Gamma$  in Eq. (8)? I cannot find the explanations for this variable.

**Addressal:** Unfortunately, this was a typographical error in our original manuscript:  $\Gamma$  (the number of model parameters) was incorrectly identified as  $\gamma$  in the description immediately following Eq. (8). In our revised manuscript, this has been corrected to accurately identify  $\Gamma$  as the number of model parameters in our Bayesian Information Criterion (BIC) model fit validation.

3. **Comment:** In Sec. IV, the SINRs of SU and PU under each scenario are assumed to be a constant regardless of the PU index, channel index, and time-slot index. This assumption may not be practical and should be addressed.

**Addressal:** **Awaiting results...**

- In our revised manuscript, we have addressed this problem by performing rate adaptation at both the SUs as well as the PUs: we model the LoS and NLoS components for the transmissions from these users (to their respective sinks) and their corresponding probabilities ( $P_{\text{LoS}}$ ,  $P_{\text{NLoS}}$ ); in addition to the distance-based path-loss modeling for each of these links encapsulating large-scale fading, we model their small-scale fading as follows: the LoS small-scale fading component is modeled as a Rician, while that of the NLoS link is modeled as a Rayleigh random variable.
  - With this channel model, we perform rate adaptation at each user (Tx) via the bisection method. Finally, the user throughput is evaluated using the LoS & NLoS probabilities ( $P_{\text{LoS}}$ ,  $P_{\text{NLoS}}$ ), and the corresponding optimal rates ( $R^*(\beta_{\text{LoS}}(d), K)$  [Rice],  $R^*(\beta_{\text{NLoS}}(d), 0)$  [Rayleigh]).
4. **Comment:** In Sec. V, the update process for the aggregated-ranked list keeps repeating until a consensus is reached. The time consumed by this repetitive process should also be addressed.  
**Addressal: Awaiting results...**
- In our revised manuscript, we have incorporated a Big-O algorithmic computational complexity analyses of our neighbor discovery and channel access order allocation schemes in multi-agent deployments.
  - In addition to these complexity analyses, based on mobility data from the DARPA SC2 “Payline” (*disaster response*: 10 mobiles nodes) and “Alleys of Austin” (*military settings*: 10 mobile nodes [9 soldiers and 1 UAV]) scenarios, as a part of these revisions, we have provided practical results about the computational complexity of these distributed algorithms in highly mobile multi-agent distributed deployments.
5. **Comment:** One relevant issue for future bench-marking would be some information on the computation complexity of the proposed strategy.  
**Addressal: Awaiting results...**
- In our revised manuscript, we have included Big-O computational complexity analysis for our parameter estimation algorithm (*Baum-Welch*), approximate POMDP value iteration algorithm (*Fragmented PERSEUS with Hamming distance state filters*), and the concurrent combination of the two.
  - Additionally, in our revised manuscript, we have also bench-marked (“computation time” v “number of channels in the discretized spectrum of interest”) our complete solution (*Baum-Welch concurrent with Fragmented PERSEUS with Hamming distance state filters*) against relevant works in the state-of-the-art (*Adaptive DQN*, *TD-SARSA with LFA*).

## 2 Addressing Reviewer-2 Comments

1. **Comment:** The authors claimed that the proposed scheme can achieve an optimal sensing performance, which is a very strong statement, requiring a detailed analysis and proof.

**Addressal:**

- Unfortunately, this was an oversight on our part: since PERSEUS is an approximate, randomized, point-based POMDP value-iteration algorithm, the optimal sensing & access policy obtained through it *cannot* be claimed to be optimal.
- However, we have incorporated the “normalized sub-optimality” gap in Fig. 4 of our original manuscript (vis-à-vis an “Oracle” with perfect knowledge of incumbent occupancies throughout the simulation period) AND demonstrated that the loss in performance due to the lack of apriori correlation model (MDP transition model) knowledge does not dent our agent’s performance significantly (Fig. 5 of our original manuscript).

2. **Comment:** It seems that the authors proposed their methods by borrowing some ideas or methods from other references. Did the authors make some innovations when applying other methods into their work? If the authors just put the existing methods into their scheme, the contribution of the proposed method is limited. If the authors made some changes, a more detailed explanation is required.

**Addressal:**

- Firstly, in addition to a novel POMDP formulation facilitated by a noisy observation model and limited observations, the PERSEUS algorithm employed to solve for the spectrum sensing & access policy in this formulation has been modified to incorporate *Fragmentation* and *Hamming distance state filters* in order to alleviate the computational intractability involved in the inherent/constituent belief update processes.
- Secondly, leveraging distributed multi-processing and multi-threading capabilities of software frameworks, we have significantly reduced the convergence time involved in determining the policy in non-stationary settings wherein the cognitive radio nodes do not possess apriori knowledge about the underlying MDP’s transition model: a fully-online parameter estimator executes concurrently with simplified PERSEUS algorithm.
- Finally, the novel reward formulation facilitates a critical capability: regulating the trade-off between SU and PU network throughputs (Fig. 5 of our original manuscript). This reward formulation simplifies the channel access decision at the SU to a rudimentary threshold-based decision heuristic involving the posterior belief (Eq. (13) of original manuscript).

- We have highlighted these contributions in more detail in our revised manuscript in order to convey them better to the reader.

3. **Comment:** The authors claimed that their proposed work can achieve better performance than Q-learning which is a very basic tool for optimization. How about the comparison with deep Q learning or other deep reinforcement learning networks?

**Addressal:**

- The original manuscript does compare our solution against an *Adaptive DQN* strategy: Experiential Replay (Memory Size  $C=10^6$ ), 2048 input neurons, 4096 neurons with ReLU activation functions in each of the 2 hidden layers, a Mean-Squared Error cost function with an Adam Optimizer, a Fixed Exploration Factor  $\epsilon=0.1$ , a Learning Rate of  $\alpha=10^{-4}$ , a Batch Size of  $W=32$ , and a sensing restriction of 6 – our solution offers a 9% improvement over this strategy.
- We have stressed this comparison (in addition to that against *TD-SARSA with LFA*) further in our revised manuscript – specifically, in the Abstract and Introduction.

4. **Comment:** There are only five simulation results in the paper. More simulation figures are required to comprehensively demonstrate the performance of the proposed scheme. Moreover, it is better for the authors to use the commonly accepted sensing performance metrics, e.g.,  $P_A$  or  $P_D$ .

**Addressal:** In our revised manuscript, in the course of addressing the comments made by Reviewer-1, we have added the following additional figures (**Awaiting results...**):

- “Computation Time” v “Number of channels in the discretized spectrum of interest” plot to bench-mark the convergence performance of our framework against other relevant works in the state-of-the-art – namely, *Adaptive DQN* and *TD-SARSA with LFA*; and
- “Number of neighbor discovery & access order messages” v “Simulation/Emulation Time” plot to evaluate the computational load associated with the neighbor discovery & channel access rank allocation schemes involved in our distributed (*custom*) and centralized (*DARPA SC2 Payline* and *Alleys of Austin*) multi-agent deployments;

Additionally, our revised manuscript also contains the following plots (**Awaiting results...**):

- “Normalized PU(s) Throughput” v “SU Network Throughput” under varying noise power levels (at the SU receiver) for our framework ( $\lambda=1$ ) – in addition to those corresponding to other works in the

state-of-the-art (*MEM with GC-CCE and MPE*, *MEM with MEI-CCE and MPE*, *Viterbi*, *Neyman-Pearson Detection*, *TD-SARSA with LFA*, and *Adaptive DQN*); and

- Receiver Operating Characteristics (ROC) of our framework (“Missed Detection Probability” v “False Alarm Probability”) – in addition to those corresponding to other works in the state-of-the-art (*MEM with GC-CCE and MPE*, *MEM with MEI-CCE and MPE*, *Viterbi*, *Neyman-Pearson Detection*, *TD-SARSA with LFA*, and *Adaptive DQN*).

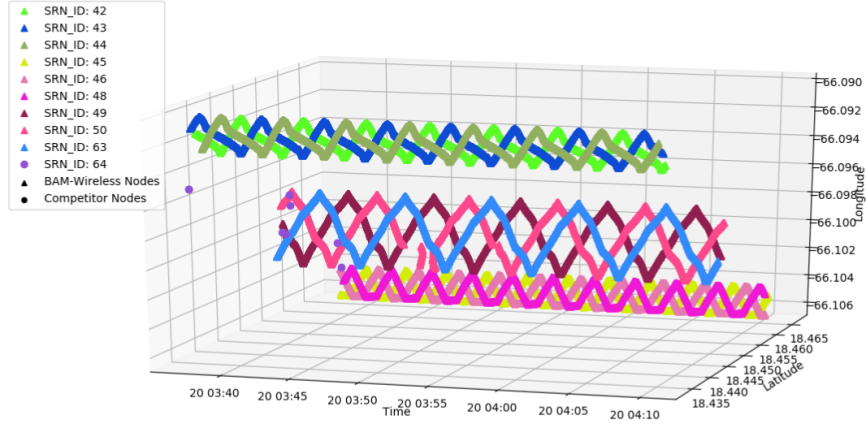


Figure 1: A 3-D visualization of the node movements during the DARPA SC2 Payline scenario

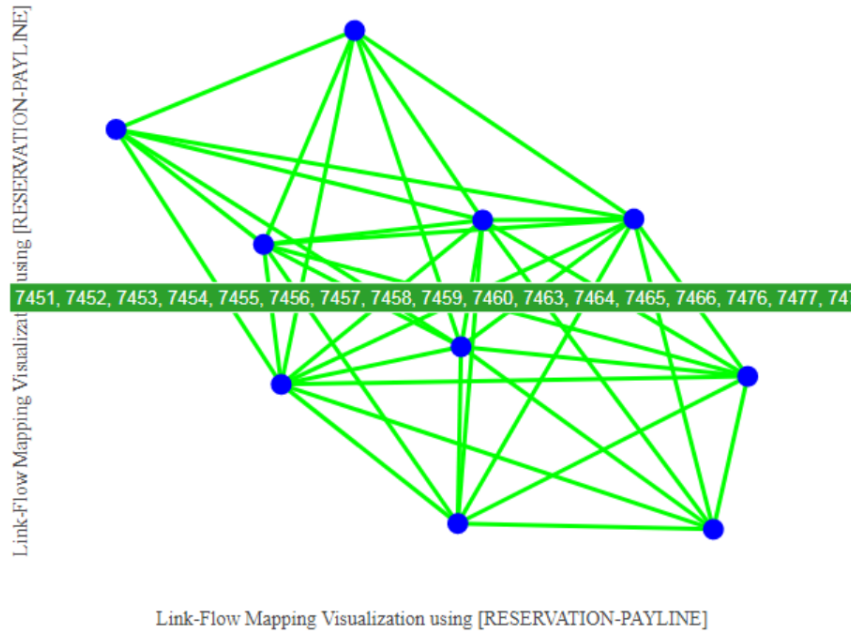


Figure 2: A graph of network links within our Purdue BAM! Wireless radio network, during the DARPA SC2 Payline scenario

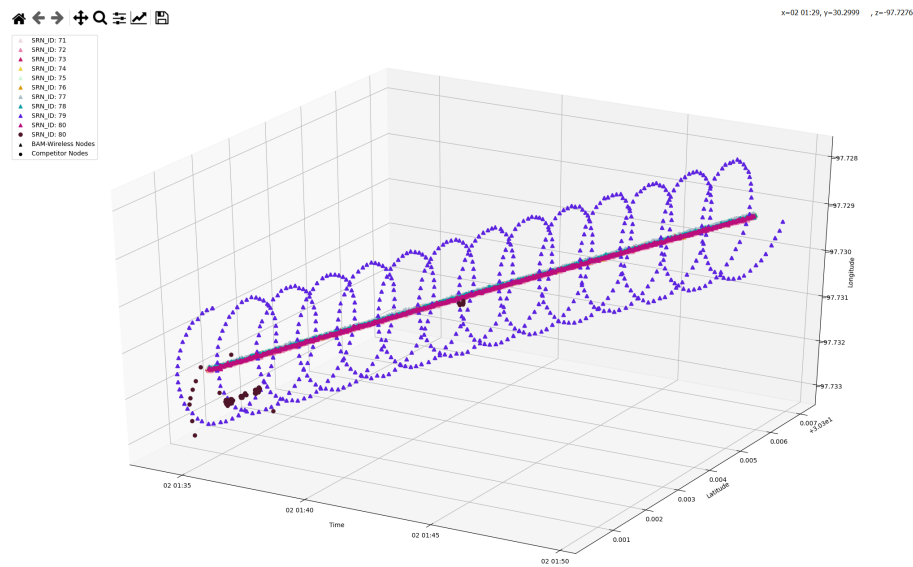


Figure 3: A 3-D visualization of the node movements during the DARPA SC2 Alleys of Austin scenario



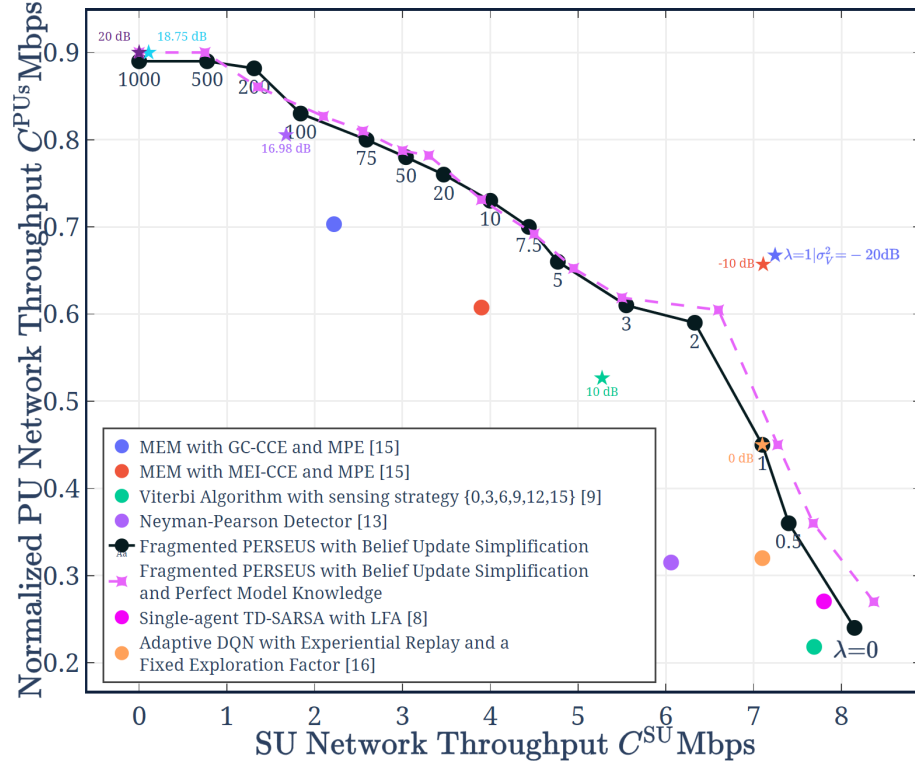


Figure 4: “Normalized PU(s) Throughput” v “SU Network Throughput” under varying noise power levels (at the SU receiver) for our framework ( $\lambda=1$ ) – in addition to penalty-based tuning and state-of-the-art comparisons