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

Bharath Keshavamurthy and Nicolò Michelusi

May 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 highlighting the reversibility of the Markov chain across frequency [1]. In other words, the frequency correlation inherent in the occupancy behavior of incumbents in the network can be captured either by a forward chain, i.e.,

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

or by a backward chain, i.e.,

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

To prove this reversibility in our Markov chain across frequency, we start with the forward chain already detailed in our original manuscript,

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

and re-write the temporal chain dependence to retain the conditioning on the entire occupancy vector in the previous time-step,

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

Next, apply Bayes' rule to re-write the above equation as

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

Taking the product operation inside, we can write the above equation as

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

which upon simplification yields,

$$\mathbb{P}(\vec{B}(i{+}1)|\vec{B}(i)){=}\mathbb{P}(B_1(i{+}1)|B_1(i))\prod_{k=1}^{K-1}\left\{\mathbb{P}(B_k(i{+}1)|B_{k+1}(i{+}1),\vec{B}(i))\right\} \\ \frac{\mathbb{P}(B_K(i{+}1),\vec{B}(i))}{\mathbb{P}(B_1(i{+}1),\vec{B}(i))}.$$

Finally, filtering out temporally independent conditions, we get

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

which upon further simplification gives us our final result concerning the reversibility of the Markov chain across frequency, i.e.,

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

In our revised manuscript, we have performed a Bayesian Information Criterion (BIC) fit evaluation for this reversed dependency across frequency on the DARPA Spectrum Collaboration Challenge (SC2) [12] Active Incumbent [2] PSD data.

Manuscript Revision Locations: Highlighted descriptions (in blue) on pages X and Y

2. **Comment**: What is the definition of Γ in Eq. (8)? I cannot find the explanations for this variable.
   Addressal: Unfortunately, this was a typographical error in our original manuscript: Γ (the number of model parameters) was incorrectly identified as γ in the description immediately following Eq. 8. In our revised manuscript, this has been corrected to accurately identify Γ as the number of model parameters in our BIC model fit validation.
   Manuscript Revision Location: Typo fixed (in blue) on page X

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:

   - In our revised manuscript, we have addressed this problem by performing rate adaptation at both the SUs and the PUs. Additional developments to our channel model in order to achieve this capability are detailed in the following items.
   - For any given communication link between a user (Tx: PU/SU) and its corresponding sink (Rx), we model the channel coefficient as $\hbar{=}\sqrt{\psi}\omega$, where ψ and ω encapsulate the large and small scale channel variations, respectively – with $\mathbb{E}[|\omega|^2]{=}1$.
   - Denoting ψ_0 as the reference path-loss at a distance of 1 m from the Tx, d as the Euclidean distance between the Tx and the Rx, $\chi{\in}(0, \frac{\pi}{2}]$ radians as the angle of elevation between the Tx and the Rx, $\iota{\in}(0, 1]$ as the additional NLoS attenuation, $2{\leq}\mu{\leq}\tilde{\mu}$ as the path-loss exponents, and $z_1, z_2, f_1,$ & f_2 as parameters specific to the propagation environment, we explicitly address the model developments corresponding to LoS and NLoS link components as follows [16].
   - For LoS components:
     - ψ is modeled as $\psi_{\text{LoS}}(d){=}\psi_0 d^{-\mu}$,
     - ω is modeled as a Rician with K-factor $\mathbb{K}(\chi){=}f_1 e^{f_2\chi}$, and
     - The LoS probability is modeled as $P_{\text{LoS}}(\chi){=}\frac{1}{1+z_1 e^{-z_2(\chi-z_1)}}$.
   - For NLoS components:
     - ψ is modeled as $\psi_{\text{NLoS}}(d){=}\iota\psi_0 d^{-\tilde{\mu}}$,
     - ω is modeled as a Rayleigh random variable, i.e., Rician with K-factor $\mathbb{K}{=}0$, and
     - The NLoS probability is modeled as $P_{\text{NLoS}}(\chi){=}\frac{z_1 e^{-z_2(\chi-z_1)}}{1+z_1 e^{-z_2(\chi-z_1)}}$.
   - Now that we have modeled the channel coefficient $\hbar$ corresponding to a certain communication link, we can define the link capacity as $C(\hbar){=}W\log_2\left(1{+}\frac{\hbar P_T}{\sigma_V^2}\right)$, where P_T is the transmission power, W is the channel bandwidth, and σ_V^2 is the noise power as described in Sec. II-A of our original manuscript.

- Assuming that ψ and $\mathbb{K}$ are known throughout the simulation period, and denoting the data rate at the Tx as R, we can define the outage probability as

$$P_{\text{out}}(R,\psi,\mathbb{K}){=}\mathbb{P}(C(\hbar){<}R|\psi,\mathbb{K}){=}\mathbb{P}\left(|\omega|^2{<}\frac{\sigma_V^2(2^{\frac{R}{W}}{-}1)}{\psi P_T}\right)$$
$$={1-}Q_1\left(\sqrt{2\mathbb{K}},\sqrt{2(\mathbb{K}{+}1)\frac{\sigma_V^2(2^{\frac{R}{W}}{-}1)}{\psi P_T}}\right),$$

where Q_1 denotes the standard Marcum Q-function.

- Next, the expected throughput is given by

$$C(R,\psi,\mathbb{K}){=}R(1{-}P_{\text{out}}(R,\psi,\mathbb{K})){=}RQ_1\left(\sqrt{2\mathbb{K}},\sqrt{2(\mathbb{K}{+}1)\frac{\sigma_V^2(2^{\frac{R}{W}}{-}1)}{\psi P_T}}\right).$$

- The rate adaptation optimization problem can now be described as,

$$R^*(\psi,\mathbb{K}){=}\operatorname*{arg\,max}_{R\geq 0} C(R,\psi,\mathbb{K}).$$

- Defining $U(R,\psi){\triangleq}\frac{\sigma_V^2(2^{\frac{R}{W}}{-}1)}{\psi P_T}$ and re-formulating this optimization problem in $Z{\triangleq}\sqrt{\frac{2\psi P_T U(R,\psi)}{\sigma_V^2}}$, we can write $R{=}f(Z){=}W\log_2\left(1{+}\frac{Z^2}{2}\right)$ and

$$Z^*{=}\operatorname*{arg\,max}_{Z\geq 0}\log_e\left[f(Z)Q_1\left(\sqrt{2\mathbb{K}},\sqrt{2(\mathbb{K}{+}1)U(f(Z),\psi)}\right)\right]$$
$$=\operatorname*{arg\,min}_{Z\geq 0}-\log_e f(Z)-\log_e Q_1\left(\sqrt{2\mathbb{K}},\sqrt{\frac{(\mathbb{K}{+}1)\sigma_V^2}{\psi P_T}Z}\right),$$

with optimal expected throughput $C^*(R^*(\psi,\mathbb{K}),\psi,\mathbb{K})$ attained at a rate $R^*(\psi,\mathbb{K}){=}f(Z^*)$ – obtained efficiently via a straight-forward bisection method [15].

- Finally, the average throughput across a communication link – with this rate adaptation scheme in place – considering both LoS and NLoS components, is given by

$$\bar{C}(d,\chi){=}P_{\text{LoS}}(\chi)C^*(\psi_{\text{LoS}}(d),\mathbb{K}(\chi)){+}P_{\text{NLoS}}(\chi)C^*(\psi_{\text{NLoS}}(d),0).$$

- With these developments in place, we have included an additional "Normalized PU Network Throughput" v "SU Network Throughput" curve in Fig. 6 of our revised manuscript.
- Additional simulation setup developments for numerical evaluations:

$\xi{=}0,\mu{=}2.0,\tilde{\mu}{=}2.8,\iota{=}0.2,W{=}160\text{kHz},f_1{=}1.0,f_2{=}0.0512,z_1{=}9.12,z_2{=}0.16.$

Manuscript Revision Locations: Highlighted descriptions (in blue) on pages X and Y; Additional curve in Fig. 6 (L)

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:

   - 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 mobile nodes per team) and Alleys of Austin (military settings: 10 mobile nodes per team – 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 settings.

   Manuscript Revision Locations: Highlighted descriptions and Big-O complexity analyses (in blue) on pages X and Y; A new figure labeled Fig. 9 – with four sub-plots

5. **Comment**: One relevant issue for future bench-marking would be some information on the computation complexity of the proposed strategy.
   Addressal:

   - In our revised manuscript, we have included Big-O computational complexity analyses for our parameter estimation algorithm (Baum-Welch [11]), our approximate POMDP value iteration algorithm (Fragmented PERSEUS [13] 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 (including against Adaptive DQN [17] and TD-SARSA with LFA [7]).

   Manuscript Revision Locations: Highlighted descriptions and Big-O complexity analyses (in blue) on pages X and Y; A new sub-plot labeled Fig. 4 (R) – appended to Fig. 4

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 sensing & access policy obtained through it *cannot* be claimed to be optimal.
   - However, in Fig. 4 of our original manuscript, we have established that our *approximate* scheme achieves a low sub-optimality gap. In other words, vis-à-vis an "Oracle" with perfect knowledge of incumbent occupancies throughout the simulation period, we have shown that the normalized sub-optimality gap $\mathbb{E}\left[\frac{|R_P(\vec{B}(i),\hat{\beta}_i)-R_O(\vec{B}(i))|}{R_O(\vec{B}(i))}\right]$=5%.
   - Also, in Fig. 6 of our original manuscript, we have demonstrated that the loss in performance due to the lack of apriori correlation model knowledge (MDP transition model) does not dent our agent's performance significantly.

   Manuscript Revision Locations: Corrected the claim and highlighted it (in blue) – along with additional comments – in the Abstract (page X), Introduction (page Y), and Numerical Evaluations (page Z) sections

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 – driven by an HMM construction [11] – emanating from settings constituting noisy observations and sensing restrictions, the PERSEUS algorithm [13] 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 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 node does not possess apriori knowledge about the underlying MDP's transition model: a fully-online parameter estimator executes concurrently with the simplified

PERSEUS algorithm. This is in contrast to works in the state-of-the-art that assume apriori model knowledge [10], or assume model ignorance but learn this model via an offline estimation involving pre-loaded databases [5] that do not scale well to non-stationary settings.

- Thirdly, we scale this single-agent solution involving HMM EM and PERSEUS to centralized and distributed multi-agent deployments, with neighbor discovery and channel access order allocation schemes – including comparative analyses of our framework against competitors [3, 6, 14] in the DARPA SC2 Active Incumbent scenario [2]. Also, we demonstrate the implementation feasibility of this multi-agent POMDP model on a distributed ad-hoc testbed of ESP32 radios [4].
- Finally, unlike every other approach in the state-of-the-art [5, 8, 9, 10], our novel reward formulation facilitates a critical capability: regulating the trade-off between SU and PU network throughputs (Fig. 5 of our original manuscript). Also, 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.

Manuscript Revision Locations: Highlighted additional comments (in blue) about our contributions on pages X and Y

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 with 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 ϵ=0.1, a learning rate of α=10^{-4}, a batch size of W=32, and a sensing restriction of 6 – our solution offers a 9% improvement over this strategy [17].
- We have stressed this comparison (in addition to that against TD-SARSA with LFA [7]) further in our revised manuscript – specifically, in the Abstract and the Introduction.

Manuscript Revision Locations: Highlighted our comparisons (in blue) against adaptive DQN and TD-SARSA with LFA on pages X and Y

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:

- "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 [17] and TD-SARSA with LFA [7] (Fig. 4 (R));
- Receiver Operating Characteristics (ROC) of our framework ("Missed Detection Probability" v "False Alarm Probability") – in addition to those corresponding to works in the state-of-the-art (MEM with GC-CCE and MPE [5], MEM with MEI-CCE and MPE [5], Viterbi [10], Neyman-Pearson Detection [9], TD-SARSA with LFA [7], and Adaptive DQN [17]) (Fig. 6 (R)); 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 a customized evaluation of highly-mobile DARPA SC2 Payline and Alleys of Austin multi-agent scenarios (Fig. 9)).

Manuscript Revision Locations: New sub-plots labeled Fig. 4 (R) and Fig. 6 (R), along with a new figure labeled Fig. 9, illustrating additional evaluations of our framework and its constituent algorithms; Highlighted descriptions (in blue) for these additional evaluations on pages X and Y;

References

[1] Brill, P. H.; Cheung, Chi ho; Hlynka, Myron; and Jiang, Q. "Reversibility Checking for Markov Chains". In: Communications on Stochastic Analysis 12.2 (2018). DOI: 10.31390/cosa.12.2.02.

[2] DARPA. "Active Incumbent Scenario Specifications". In: DARPA SC2 (2019). URL: https://sc2colosseum.freshdesk.com/support/solutions/articles/22000239489-active-incumbent-.

[3] DARPA. "Purdue University BAM! Wireless Radio". In: DARPA SC2 (2019). URL: https://archive.darpa.mil/sc2/news/spectrum-collaboration-challenge-awards-four-teams-with-half-prizes.

[4] Espressif Systems (Shanghai) Co. Ltd. "Espressif ESP32: A Different IoT Power and Performance". In: Espressif (2019). URL: https://www.espressif.com/en/products/hardware/esp32/overview.

[5] M. Gao et al. "Fast Spectrum Sensing: A Combination of Channel Correlation and Markov Model". In: 2014 IEEE MilCom. Oct. 2014, pp. 405–410. DOI: 10.1109/MILCOM.2014.73.

[6] S. Giannoulis et al. "Dynamic and Collaborative Spectrum Sharing: The SCATTER Approach". In: 2019 IEEE DySPAN. 2019, pp. 1–6.

[7] J. Lundén et al. "Multiagent Reinforcement Learning Based Spectrum Sensing Policies for Cognitive Radio Networks". In: IEEE JSTSP 7.5 (Oct. 2013), pp. 858–868. DOI: 10.1109/JSTSP.2013.2259797.

[8] N. Michelusi et al. "Multi-Scale Spectrum Sensing in Dense Multi-Cell Cognitive Networks". In: IEEE Transactions on Communications 67.4 (Apr. 2019), pp. 2673–2688. DOI: 10.1109/TCOMM.2018.2886020.

[9] S. Mosleh, A. A. Tadaion, and M. Derakhtian. "Performance analysis of the Neyman-Pearson fusion center for spectrum sensing in a Cognitive Radio network". In: IEEE EUROCON 2009. May 2009, pp. 1420–1425. DOI: 10.1109/EURCON.2009.5167826.

[10] C. Park et al. "HMM Based Channel Status Predictor for Cognitive Radio". In: 2007 Asia-Pacific Microwave Conference. Dec. 2007, pp. 1–4. DOI: 10.1109/APMC.2007.4554696.

[11] Lawrence R. Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". In: Proceedings of the IEEE 77, no. 2 (Feb. 1989), pp. 257–286.

[12] Mark Rosker. "Spectrum Collaboration Challenge (SC2)". In: DARPA SC2 (2018). URL: https://www.darpa.mil/program/spectrum-collaboration-challenge.

[13] Matthijs T. J. Spaan and Nikos A. Vlassis. "Perseus: Randomized Point-based Value Iteration for POMDPs". In: CoRR abs/1109.2145 (2011). arXiv: 1109.2145. URL: http://arxiv.org/abs/1109.2145.

[14] D. Stojadinovic et al. "SC2 CIL: Evaluating the Spectrum Voxel Announcement Benefits". In: 2019 IEEE DySPAN. 2019, pp. 1–6.

[15] Yin Sun, Árpád Baricz, and Shidong Zhou. "On the Monotonicity, Log-Concavity, and Tight Bounds of the Generalized Marcum and Nuttall Q-Functions". In: IEEE Transactions on Information Theory 56.3 (2010), pp. 1166–1186. DOI: 10.1109/TIT.2009.2039048.

[16] David Tse and Pramod Viswanath. Fundamentals of Wireless Communication. 2005.

[17] S. Wang et al. "Deep Reinforcement Learning for Dynamic Multichannel Access in Wireless Networks". In: IEEE TCCN 4.2 (2018), pp. 257–265.