IEEE ComSoc ICC 2021 CRAIN Symposium

Learning-based Cognitive Radio Access via Randomized Point-Based Approximate POMDPs

Bharath Keshavamurthy*, IEEE Student Member
Nicolò Michelusi#, IEEE Senior Member

*PhD Student, School of Electrical and Computer Engineering, Purdue University
#Assistant Professor, School of Electrical, Computer and Energy Engineering, Arizona State University

Motivation (1/2)

THE WALL STREET JOURNAL.

TECH

Wireless Companies Share the Airwaves

New cellular service lets the military, companies and others use the same frequencies—provided they don't interfere with each other [1]

OPINION | LETTERS

U.S. Needs Better Spectrum Policy to Win in 5G World

To speed up 5G, we need to throw out spectrum auctions entirely. [2]

OPINION | BUSINESS WORLD

How Government Can Get Brave About Spectrum

Ignore the groups that gripe about a taxpayer rip-off. The public benefits when airwaves trade freely. [3]

---

[1] Drew Fitzgerald, The Wall Street Journal, Dec 28, 2019

[2] WSJ Opinion Letters, The Wall Street Journal, Feb 7, 2020

[3] Holman W. Jenkins Jr, The Wall Street Journal, June 14, 2019

Motivation (2/2)

4 

5 

4 5G Use Cases, Ericsson, [Online] Nov 26, 2015

5 Alison Gopnik, The Wall Street Journal, Oct 11, 2019

Problem Description[6]

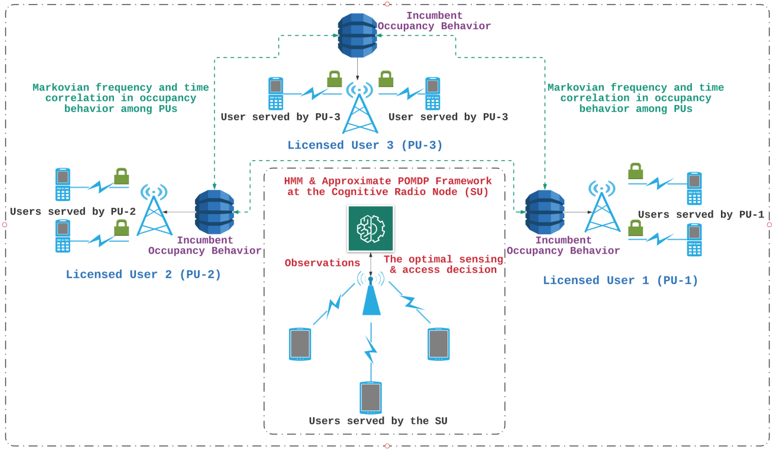

Figure 1: An example of a heterogeneous radio ecosystem in which cognitive radio technology is a game-changer

[6] Primary Users (PUs) are the licensed users/incumbents | Secondary Users (SUs) are the cognitive radio nodes

Challenges

- Noisy observations at the SU receiver hide the true occupancy states of the channels – impairing accurate white-space detection & access[7]
- Energy efficiency needs at the SU constrain the number of channels that it can sense in a time-slot[8]
- Occupancy behaviors of PUs are correlated across time AND frequency[9]
- This time-frequency correlation model is not known a priori
- The capability to regulate the trade-off between SU and PU(s) throughputs is crucial
- The curse of dimensionality

[7] F. Xu, et. al., "Architecture for Re-configurable Networks via Cognitive Radio", 2008 CROWNCOM

[8] S. Malecki, et. al., "Energy and throughput efficient strategies for spectrum sensing...", 2011 IEEE SPAWC

[9] Active Incumbent, DARPA SC2, [Online], Aug 28, 2019

Related Work | State-of-the-Art (SoA)

Custom heuristics[10] | Hidden Markov Models (HMMs)[11]
Multi-Armed Bandits (MABs)[12] | Reinforcement Learning (RL)[13]

Drawbacks in SoA	Our Answers
Noise-free [10] obs & Ignored [13] est errors	Noisy obs | HMMs
Failure to exploit [12] PU correlation	Successful exploitation
Apriori knowledge [13]	No apriori knowledge
Offline estimation [10]	Fully Online estimation
No support [10], [13] for tuning throughputs	Support via penalty tuning

Our framework is evaluated against [10], [11], and [13] (among others – including Deep-Q Networks (DQNs), the Viterbi Algorithm, and Neyman-Pearson Detection): We demonstrate superior performance.

[10] M. Gao, et. al., "Fast Spectrum Sensing...", 2014 IEEE MilCom

[11] C. Park, et. al., "HMM Based Channel Status Predictor for Cognitive Radio", 2007 APAC MW Conference

[12] K. Cohen, et. al., "Restless Multi-Armed Bandits...", 2014 Asilomar

[13] J. Lundén, et. al., "Multiagent Reinforcement Learning...", IEEE Journal of Selected Topics in SigProc, 2013

System Model

OFDMA; AWGN observation model; Rayleigh fading channel; Sensing limits

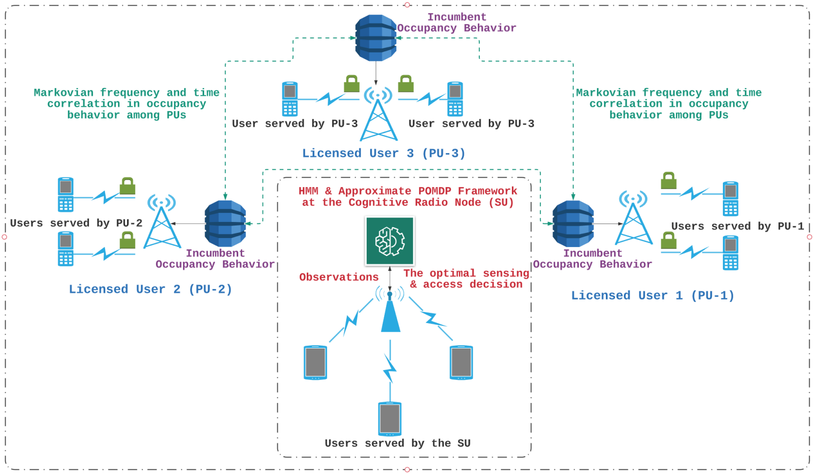

Figure 2: An exemplification of the radio ecosystem under analysis: 18 channels, 3 PUs, and 1 SU with sensing limit - 3 channels per time-slot

PU Occupancy Model[14]

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

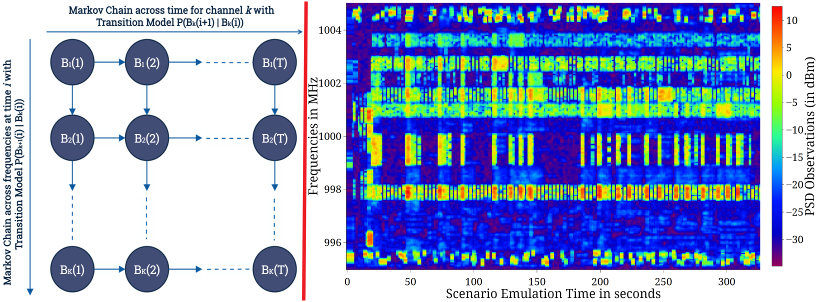

Figure 3: Markovian time-frequency correlation of PU occupancies | Model Validation through DARPA SC2 Active Incumbent

14 Active Incumbent, DARPA SC2, [Online], Aug 28, 2019

SU Spectrum Sensing Model

- SU can sense at most κ out of K spectrum bands at any given time, with $1 \leq \kappa \leq K$
- HMM framework:

$$\vec{Y}(i) = [Y_k(i)]_{k \in \mathcal{K}_i} \text{ is the observation vector}$$

$$f(\vec{Y}(i) | \vec{B}(i), \mathcal{K}_i) = \prod_{k \in \mathcal{K}_i} f(Y_k(i) | B_k(i)) \text{ is its PDF}$$

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

POMDP Agent Model (1/2)

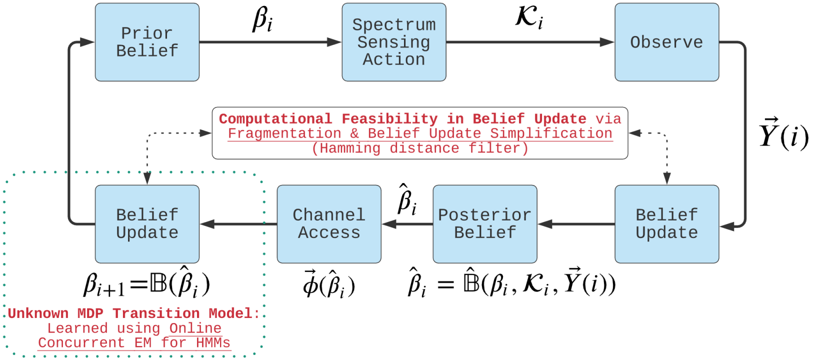

Figure 4: The POMDP Formulation

POMDP Agent Model (2/2)

- Goal: Find optimal spectrum sensing policy that maximizes the infinite-horizon discounted reward:

$$\pi^{*} = \arg\max_{\pi} V^{\pi}(\beta) \triangleq \mathbb{E}_{\pi}\left[\sum_{i=1}^{\infty} \gamma^{i} R(\vec{B}(i), \hat{\beta}_i) | \beta_0 = \beta\right]$$

- Access decision:

$$\vec{\phi}(\hat{\beta}_i) \triangleq \arg\max_{\vec{B} \in \mathcal{B}} \hat{\beta}_i(\vec{B})$$

- Reward formulation:

$$R(\vec{B}(i), \hat{\beta}_i) = \sum_{k=1}^{K} (1 - B_k(i))(1 - \phi_k(\hat{\beta}_i)) - \lambda B_k(i)(1 - \phi_k(\hat{\beta}_i))$$

- π^{*} is the solution to $V^{*} = \mathcal{H}[V^{*}]$; Bellman operator $V_{t+1} = \mathcal{H}[V_t]$ is $\forall \beta$:

$$V_{t+1}(\beta) = \max_{\mathcal{K} \in \mathcal{A}} \sum_{\vec{B} \in \mathcal{B}} \beta(\vec{B}) \mathbb{E}_{\vec{Y} | \vec{B}, \mathcal{K}}\left[R(\vec{B}, \hat{\mathbb{B}}(\beta, \mathcal{K}, \vec{Y})) + \gamma V_t(\mathbb{B}(\hat{\mathbb{B}}(\beta, \mathcal{K}, \vec{Y})))\right]$$

The Parameter Estimator[15]

- Goal: Determine the time-frequency occupancy correlation structure of the PUs – parameterized by $\vec{\theta}$
- Maximum Likelihood Estimation (MLE) problem:

$$\vec{\theta}^* = \arg\max_{\vec{\theta}} \log\Big(\sum_{\mathsf{B}} \mathbb{P}(\mathsf{B},\mathsf{Y}|\vec{\theta})\Big)$$

- Baum-Welch (Expectation-Maximization (EM) for HMMs):

$$\text{E-step: } Q(\vec{\theta}|\vec{\theta}^{(t)}) = \mathbb{E}_{\mathsf{B}|\mathsf{Y},\vec{\theta}^{(t)}}\Big[\log\Big(\mathbb{P}(\mathsf{B},\mathsf{Y}|\vec{\theta}^{(t)})\Big)\Big]$$

$$\text{M-step: } \vec{\theta}^{(t+1)} = \arg\max_{\vec{\theta}} Q(\vec{\theta}|\vec{\theta}^{(t)})$$

[15] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE 77, no. 2, Feb 1989

PERSEUS[16] (1/3)

- Initial exploration: $\tilde{\mathcal{B}}$
- Goal: Improve the value of all the belief points in $\tilde{\mathcal{B}}$ by updating the value of only a subset of these belief points, chosen iteratively at random
- $$V_t(\beta) \approx \beta \cdot \vec{\alpha}_t^{u^*}, \; u^* = \underset{u \in \{1,2,\ldots,|\tilde{\mathcal{B}}|\}}{\arg\max} \; \beta \cdot \vec{\alpha}_t^{u}, \; \beta \cdot \vec{\alpha} = \sum_{\vec{B}} \beta(\vec{B}) \vec{\alpha}(\vec{B})$$

[16] T.J. Spaan, et. al., "Perseus: Randomized Point-based Value Iteration for POMDPs", Journal of Artificial Intelligence Research, 2005

PERSEUS (2/3)

- Initialization: $\tilde{\mathcal{U}}=\tilde{\mathcal{B}}$
- Backup:
  - Find a new hyperplane associated with randomly chosen β_u:

$$\vec{\alpha}^u_{t+1} = \Xi^u_{\mathcal{K}^u_{t+1}},\ \ \mathcal{K}^u_{t+1} = \arg\max_{\mathcal{K}\in\mathcal{A}} \beta_u \cdot \Xi^u_{\mathcal{K}}$$

$$\Xi^u_{\mathcal{K}}(\vec{B}) = \mathbb{E}_{\vec{Y}|\vec{B},\mathcal{K}}\Big[R(\vec{B},\hat{\mathbb{B}}(\beta_u,\mathcal{K},\vec{Y}))+ \gamma\sum_{\vec{B}'}\mathbb{P}(\vec{B}(i+1){=}\vec{B}'|\vec{B}(i){=}\vec{B})\Xi^u_{\mathcal{K},\vec{Y}}(\vec{B}')\Big]$$

Future value function: $\Xi^u_{\mathcal{K},\vec{Y}} = \arg\max_{\alpha^{u'}_t,u'\in\{1,2,\ldots,|\tilde{\mathcal{B}}|\}} \mathbb{B}(\hat{\mathbb{B}}(\beta_u,\mathcal{K},\vec{Y}))\cdot\alpha^{u'}_t$

  - $$\tilde{\mathcal{U}} \leftarrow \tilde{\mathcal{U}} \setminus \{\beta_u\} \setminus \{\beta' \in \tilde{\mathcal{U}} : \beta'\cdot\vec{\alpha}^u_{t+1} \geq V_t(\beta')\}$$
  - Backup termination: $\tilde{U}{=}\phi$
- PERSEUS Termination: $|V_{t+1}(\beta){-}V_t(\beta)|{<}\epsilon,\ \forall\beta{\in}\tilde{\mathcal{B}},\ \epsilon{>}0$

PERSEUS (3/3): Heuristics

- Fragmentation: Smaller, independent sets of correlated channels governed by a dominant PU

$$(\mathcal{B}_{\Delta}, \mathcal{A}_{\Delta}, \mathcal{Y}_{\Delta}, \mathsf{A}_{\Delta}, \mathsf{M})$$

- Belief Update Simplification: Hamming distance filter to avoid iterating over all possible states

$$\mathcal{B}_{\delta}(\vec{B}) \equiv \{\vec{B}' \in \mathcal{B} : \psi(\vec{B}, \vec{B}') \leq \delta\}$$

Numerical Evaluations (1/2)

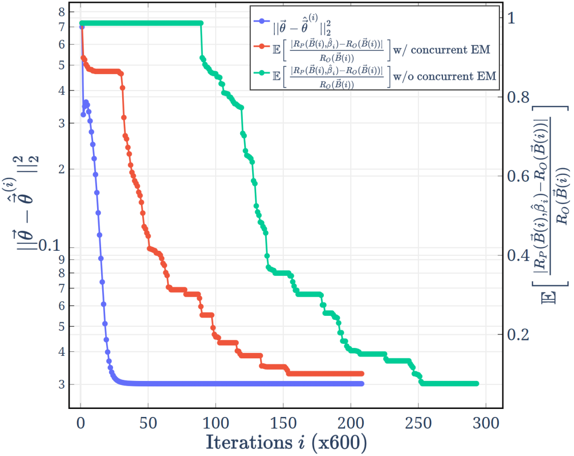

Figure 5: Convergence of: MSE of the EM algorithm to estimate $\vec{\theta}$ | Normalized sub-optimality gap of fragmented PERSEUS with belief update simplification to concurrently find the policy

Numerical Evaluations (2/2)

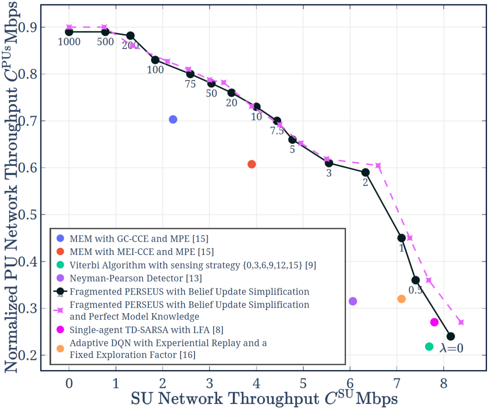

Figure 6: C^{SU} v C^{PUs} for different values of the penalty term λ | Comparison with state-of-the-art

[15]: M. Gao, et. al., "Fast Spectrum Sensing...", 2014 IEEE MilCom
[9]: C. Park, et. al., "HMM Based Channel Status Predictor for Cognitive Radio", 2007 Asia-Pacific MW Conf
[13]: S. Mosleh, et. al., "Performance analysis of the Neyman-Pearson fusion center...", 2009 IEEE EUROCON
[8]: J. Lundén, et. al., "Multiagent Reinforcement Learning...", IEEE Journal of Selected Topics in SigProc, 2013
[16]: S. Wang, et. al., "Deep Reinforcement Learning for Dynamic Multichannel Access...", IEEE TCCN, 2018

Extensions (1/2): Distributed Multi-agent Deployment[18]

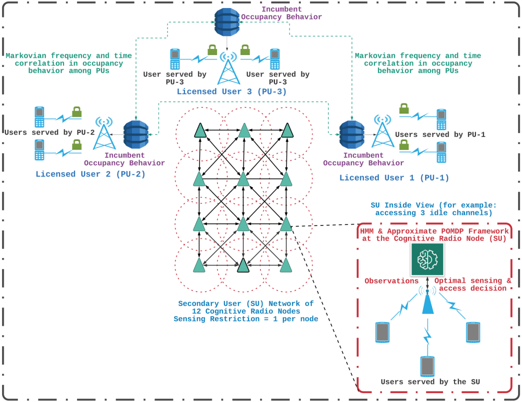

Figure 7: Distributed multi-agent example: 18 channels, 3 PUs, 12 SUs

Also, our multi-agent approximate POMDP solution is evaluated in centralized settings via emulations in the DARPA SC2 Active Incumbent scenario[17].

[17] Active Incumbent, DARPA SC2, [Online], Aug 28, 2019

[18] B. Keshavamurthy and N. Michelusi, "Learning-based Spectrum Sensing in Cognitive Radio Networks via Approximate POMDPs", Under review at IEEE TCCN, 2021

Extensions (2/2): Multi-agent POMDP Model[19]

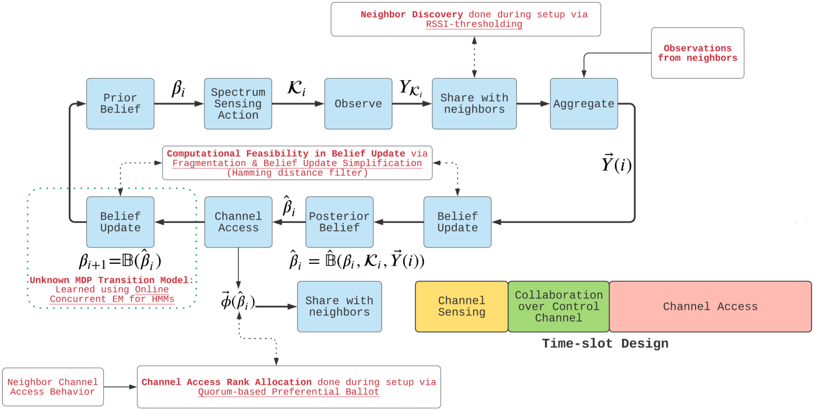

Figure 8: The POMDP model for multi-agent deployments

[19] B. Keshavamurthy and N. Michelusi, “Learning-based Spectrum Sensing in Cognitive Radio Networks via Approximate POMDPs”, Under review at IEEE TCCN, 2021

Future View

- NSF PAWR POWDER[20] testbed:
  - Over-the-Air (OTA) [CBRS: 3.4–3.8 GHz] impl of this approximate POMDP solution in multi-agent settings

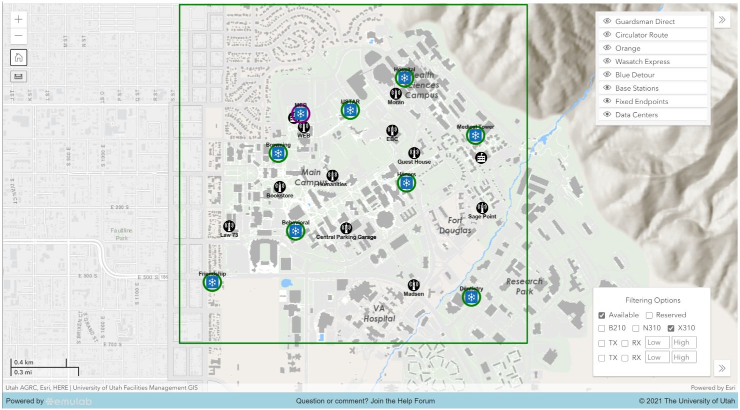

Figure 9: POWDER at the University of Utah: our OTA arena

[20] NSF PAWR POWDER, University of Utah, [Online]

Conclusion

We have addressed the challenges & proved superior performance:

.

- Time-freq correlated PU occupancies | SU with sensing limits
- Fully online Baum-Welch (HMM EM) | Concurrent PERSEUS
- Fragmentation | Hamming distance state filters
- Superior performance over TD-SARSA, DQN, Viterbi, etc.
- Regulate the trade-off between SU and PU throughputs