Dynamic Spectrum Access for RF-powered Ambient Backscatter Cognitive Radio Networks

Ahmed Y. Zakariya[1,2], Sherif I. Rabia[1,2] and Waheed K. Zahra[1,3]
[1]*Institute of Basic and Applied Sciences, Egypt-Japan University of Science and Technology (E-JUST), New Borg El-Arab City, Alexandria 21934, Egypt*
[2]*Department of Engineering Mathematics and Physics, Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt*
[3]*Department of Engineering Physics and Mathematics, Faculty of Engineering, Tanta University, Tanta, 31111, Egypt*

Keywords: Cognitive Radio, Markov Decision Process, Reinforcement Learning, Ambient Backscatter.

Abstract: In RF-powered backscatter cognitive radio networks, while the licensed channel is busy, the SU can utilize the primary user signal either to backscatter his data or to harvest energy. When the licensed channel becomes idle, the SU can use the harvested energy to actively transmit his data. However, it is crucial for the secondary user to determine the optimal action to do under the dynamic behavior of the primary users. In this paper, we formulate the decision problem as a Markov decision process in order to maximize the average throughput of the secondary user under the assumption of unknown environment parameters. A reinforcement learning algorithm is attributed to guide the secondary user in this decision process. Numerical results show that the reinforcement learning approach succeeds in providing a good approximation of the optimal value. Moreover, a comparison with the harvest-then-transmit and backscattering transmission modes is presented to investigate the superiority of the hybrid transmission mode in different network cases.

1 INTRODUCTION

Cognitive radio (CR) networks provide an effective solution for the issue of scarcity in spectrum bands (Gouda et al., 2018) by allowing an unlicensed (secondary) user (SU) to opportunistically access the unused licensed bands of the licensed (primary) user (PU) without making any harmful (Wang and Liu, 2010; Zakariya et al., 2019). At the beginning of each time slot, the SUs have to sense the license channels to check the existence of the PUs (Zakariya and Rabia, 2016; Zuo et al., 2018). In such network, the SU must evacuates the licensed channel when the PU appears in this channel (Fahim et al., 2018).

Energy efficiency is another important factor in wireless communication which needs to be considered in CR networks (Zakariya et al., 2021). Recently, radio frequency (RF) powered CR networks have been attributed to provide an innovative solution for both the spectrum scarcity and the energy limitation issues (Park et al., 2013). During the PU transmission period, the SU can harvest energy from the PU signal and store it until it is used in the active transmission when the channel is free from the PU signal. This transmission mode is called harvest-then-transmit (HTT) mode (Van Huynh et al., 2019). In (Lu et al., 2014; Niyato et al., 2014; Hoang et al., 2014), the HTT mode is studied for a single SU operating on multiple licensed channels. To maximize the throughput of the SU, an optimal channel selection policy is obtained by formulating a Markov decision process (MDP) problem. In (Lu et al., 2014), the tradeoff between data transmission and RF energy harvesting is studied assuming error-free sensing results. In (Niyato et al., 2014; Hoang et al., 2014), sensing errors are considered. In (Niyato et al., 2014), the model considered the known environment parameters scenario for both complete information and incomplete information cases. In (Hoang et al., 2014), in addition to the network cases considered in (Niyato et al., 2014), the unknown environment parameters scenario is considered. An online learning algorithm is attributed to determine the optimal policy for the SU.

The performance of the RF-powered CR networks strongly depends on the amount of the harvested energy. For channels with longer PU transmission periods, the SU will have lesser opportu-

Zakariya, A., Rabia, S. and Zahra, W.
Dynamic Spectrum Access for RF-powered Ambient Backscatter Cognitive Radio Networks.
DOI: 10.5220/0010511502240230
In *Proceedings of the 11th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2021)*, pages 224-230
ISBN: 978-989-758-528-9

nities for his transmission. With the need to improve the performance of the RF-powered CR networks, ambient backscatter technology is attributed due to its ability of transmission in a busy channel with low power consumption (Van Huynh et al., 2018a). Ambient backscattering (AB) is considered as an energy-efficient communication mechanism that enables communication devices to communicate by modulating and reflecting the signals from ambient RF sources (Liu et al., 2013; Zakariya et al., 2020a). In CR networks, the signals of PUs represent these ambient RF sources. Most importantly, this type of communication does not cause any harmful interference to the original RF signal (Van Huynh et al., 2018a). In (Van Huynh et al., 2019; Van Huynh et al., 2018b), a single SU working on a single channel using a hybrid HTT and backscattering transmission mode is assumed. In (Van Huynh et al., 2018b), only the unknown environment parameters case is considered, and in order to maximize the throughput of the SU a low-complexity online reinforcement learning algorithm is designed. In (Van Huynh et al., 2019), the known environment parameters network case is also studied. Under the dynamic behavior of the primary signal, an MDP framework is proposed to obtain the optimal policy that maximizes the throughput for SU. In (Anh et al., 2019), multiple SUs operating on a single channel in an unknown environment parameters network case is considered. Moreover, it is assumed that a gateway is responsible for coordinating and scheduling the backscattering time, the harvesting time, and the transmission time among the SUs. In order to maximize the total throughput, the authors propose a deep reinforcement learning algorithm to derive an optimal time scheduling policy for the gateway.

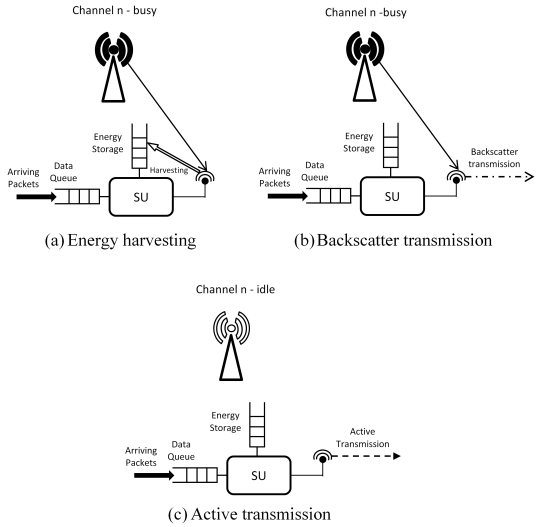
In (Van Huynh et al., 2019; Van Huynh et al., 2018b; Anh et al., 2019), only a single channel network case is considered. In the present work, we assume that the CR network consists of multiple licensed channels from which the SU can harvest energy or utilize for data transmission. Moreover, the incomplete information case (in which the SU does not know the current state of the licensed channels) with an unknown environment parameters scenario is considered. In each time slot, the SU has to select an operating channel and based on the sensing result, he has to choose between three possible operating modes: harvest energy, actively transmit data, or backscatter transmission. To maximize the average throughput for the SU, an MDP is attributed to represent and optimize the performance of such a dynamic environment. An online reinforcement learning algorithm is utilized to deal with the unknown environment parameters. Numerical results show that the performance of the proposed hybrid mode with the reinforcement learning approach outperforms both HTT mode and AB mode especially for high SU arrival rate or small PU idle probability.

The rest of this paper is organized as follows. System model and our assumptions are presented in Section II. Problem formulation and the online reinforcement learning algorithm are presented in Section III. Numerical results are shown in Section IV. Finally, concluding remarks are given in Section V.

2 SYSTEM MODEL

We consider an RF-powered backscatter CR network consisting of a single SU and N licensed channels working on a time-slotted manner. During each time slot, the SU receives a random batch of data packets to be stored in his data queue which has a finite capacity Q. The probability of receiving a batch of size i is denoted by $\alpha_i, i \in \{0, 1, ..., R\}$. If the available space in the data queue is not sufficient to store the incoming batch, the whole batch is blocked. The SU has also a finite energy storage of capacity E to store the harvested energy. The SU knows the licensed channels state (idle or busy) through sensing. Let η_n be the probability that channel n is being idle. At the beginning of each time slot, the SU has to select a channel to work on it. The selected channel is sensed and based on the sensing result, the SU has to perform a proper action. If the channel is sensed to be idle, the SU actively transmits a batch of R data packets which requires W energy units (see Figure 1c). If the SU has less than R packets or less than W energy units, then no transmission is performed. Let δ_n be the probability of successful active transmission in channel n. On the other hand, if the chosen channel is sensed to be busy, the SU has to choose either to harvest E_h energy units from the PU signal or use this signal to backscatter D ($D \leq R$) packets from his data queue (see Figure 1a and Figure 1b, respectively). If the SU has less than D packets, then no backscatter transmission is performed. The probability of successful harvesting and successful backscatter transmission are denoted by ν_n and β_n, respectively.

The SU has to take a decision based on the sensing result which suffers from both missed detection and false alarm errors with probability m_n and f_n in channel n, respectively. In missed detection, the channel is sensed to be idle while the actual channel state is busy (Wang and Liu, 2010). On the other hand, in the false alarm, the actual channel state is



(a) Energy harvesting (b) Backscatter transmission

(c) Active transmission

Figure 1: The different capabilities of the SU in the considered CR network.

Table 1: Environment parameters.

Symbol	Description
η_n	The probability that channel n is idle.
ν_n	The probability of successful energy harvesting from channel n.
δ_n	The probability of successful active transmission in channel n.
β_n	The probability of successful backscattring transmission in channel n.
m_n	The probability of missed detection in channel n.
f_n	The probability of false alarm in channel n.

idle, while the SU senses it to be busy (Zakariya et al., 2020b). The environment parameters are summarized in Table 1. We consider the unknown environment parameters case in which the SU doesn't know the value of these parameters in advance. Hence, we propose using a reinforcement learning approach to guide the SU decision process where the SU arrives at the optimal decision through interacting with the environment.

3 PROBLEM FORMULATION

In this section, to maximize the average throughput of the SU, the problem is formulated first as an MDP and then the applied online learning algorithm is introduced.

3.1 State Space

The state space of the SU is defined as $S = \{(q,e) : q \in \{0,...,Q\}; e \in \{0,...,E\}\}$, where q and e are the number of data packets buffered in the data queue and the number of energy units in the energy storage, respectively. Hence, the state of the SU is defined as a composite variable $s = (q,e) \in S$.

3.2 Action Space

At the beginning of each time slot, the SU has to select a channel to be sensed. If the channel is sensed to be idle, the SU selects the active transmission mode. On the other hand, if the channel is sensed to be busy, the SU selects between the energy harvesting mode and the backscattering transmission mode. This decision process is summarized in Figure 2. Hence, the action space of the SU can be defined as follows: $T = \{(n,a)|n \in \{1,2,...,N\}; a \in \{h,b\}\}$. Each action consists of a pair where the first element n is the selected channel and the second element a is the action to be taken (h: harvest, b: backscatter) when the channel is sensed to be busy. If the selected channel n is sensed to be idle, the SU will always perform active transmission if he has the sufficient data and energy.

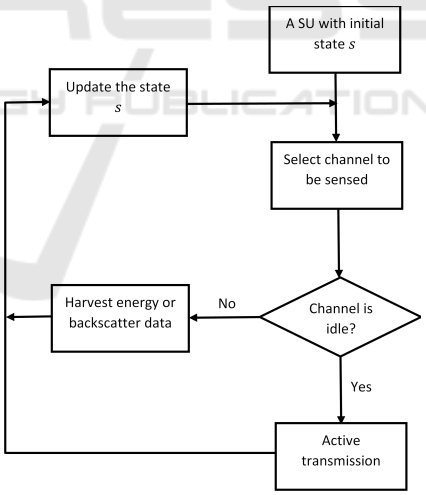


Figure 2: Typical decision process of the SU.

3.3 Immediate Reward

The SU receives an immediate reward (throughput) when he successfully transmits his data either by active or backscattering transmission. Let $\tau(s,c)$ be the average throughput assuming the SU state $s \in S$ and the taken action $c \in T$. The expected value of $\tau(s,c)$ can be expressed as follows:

$$\mathrm{E}[\tau(s,c)] = \begin{cases} \eta_n f_n^o \delta_n R & ,e \geq W, q \geq R\ ,\ c=(n,h) \\ \eta_n^o m_n^o \beta_n D & ,D \leq q < R\ ,\ c=(n,b) \\ \eta_n^o m_n^o \beta_n D & ,e < W, q \geq R\ ,\ c=(n,b) \\ \eta_n f_n^o \delta_n R + & \\ \eta_n^o m_n^o \beta_n D & ,e \geq W, q \geq R\ ,\ c=(n,b) \\ 0 & ,\text{otherwise} \end{cases}, \tag{1}$$

where we use the notation z^o to denote $1-z$ throughout this work.

3.4 Learning Algorithm

For the unknown environment parameters case, the transition probability matrix for the MDP cannot be derived. Hence, a reinforcement learning approach is attributed to solve this issue. A parameterized policy is considered and a softmax action selection rule is applied to find the SU decisions (Sigaud and Buffet, 2013). In this policy, the probability of executing action c at state s can be calculated as follows:

$$q(c|s;\Theta) = \frac{e^{\theta_{s,c}}}{\sum_{g\in T} e^{\theta_{s,g}}}, \tag{2}$$

where $\Theta = [\theta_{s,c}]; s \in S, c \in T$ is the parameter vector (also called preference vector) of the learning algorithm which will be updated iteratively by interacting with the environment to maximize the throughput of the SU. The parameterized immediate throughput of the SU in state s is $\tau_\Theta(s) = \sum_{c\in T} q(c|s;\Theta)\tau(s,c)$. Finally, the average throughput of the SU can be parameterized as follows:

$$\Re(\Theta) = \lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} \mathrm{E}[\tau_\Theta(s_k)], \tag{3}$$

where $s_k \in S$ is the state at time step k and our goal is to maximize $\Re(\Theta)$ by updating the parameter vector Θ. The OLGARB algorithm introduced in (Weaver and Tao, 2001) is selected to be implemented in our work and it is outlined in Algorithm 1.

4 NUMERICAL RESULTS

A two-channel ($N = 2$) CR network is considered. The assumed environment parameters are summarized in Table 2. A batch of $i \in \{0,1,2\}$ data packet arrives in each time slot with probability α_i. In Figure 3, the convergence of the learning algorithm is investigated. At the beginning of the learning process, the throughput of the SU is fluctuating because the SU is still in the starting process of adjusting the parameter Θ. As the number of iterations increases the performance of the SU starts to stabilize and the throughput approximately converges to 0.92 after 6×10^5 iterations. This value is close to the optimal policy (0.94)

Algorithm 1: OLGARB Algorithm.

Input: Θ_0, ε, γ ▷ Θ_0: The initial value for the preference vector.
▷ ε: The learning step-size.
▷ γ: The discount factor $\in [0,1]$.
Output: Θ ▷ Θ: The optimal preference vector.

1: $z \leftarrow 0$ ▷ Initialization for the eligibility trace vector.
2: $B \leftarrow 0$ ▷ Initialization for the baseline (estimated average reward).
3: Get initial state: s_1 ▷ Randomly select an initial state
4: **for** t from 1 to M **do**
5: Get c_t from $q(.|s_t;\Theta)$
6: Execute c_t
7: Get the new state s_{t+1} and the reward $\tau(s_t,c_t)$
8: $B \leftarrow B + \frac{\tau(s_t,c_t)-B}{t}$
9: $z \leftarrow \gamma z + \frac{\nabla q(c_t|s_t;\theta)}{q(c_t|s_t;\Theta)}$
10: $\Theta \leftarrow \Theta + \varepsilon(\tau(s_t,c_t)-B)z$
11: **end for**

Table 2: Parameters setting.

Symbol	Value	Symbol	Value
Q	10	E	10
W	1	R	2
β_1, β_2	0.95	N	2
δ_1, δ_2	0.95	E_h	1
ν_1, ν_2	0.95	D	1
m_1, m_2	0.01	ε	0.00005
f_1, f_2	0.01	γ	0.99

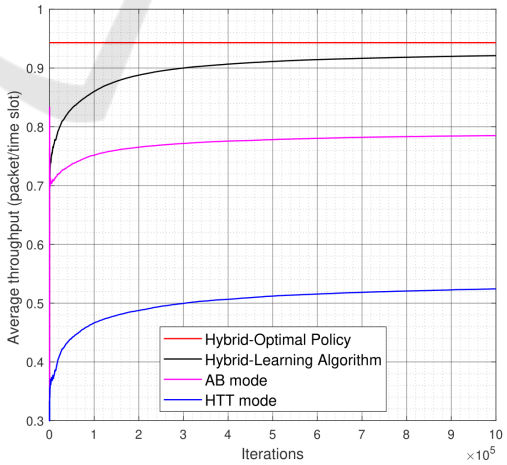
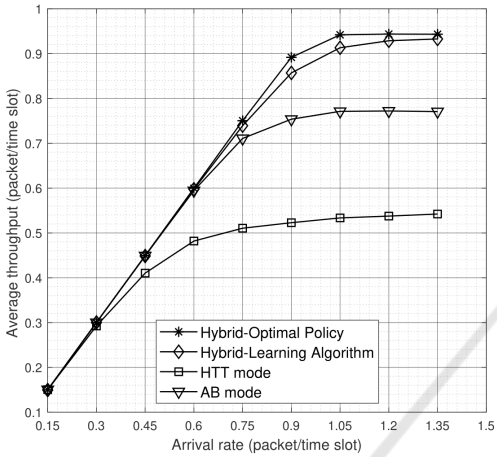


Figure 3: The convergence of the learning algorithm.

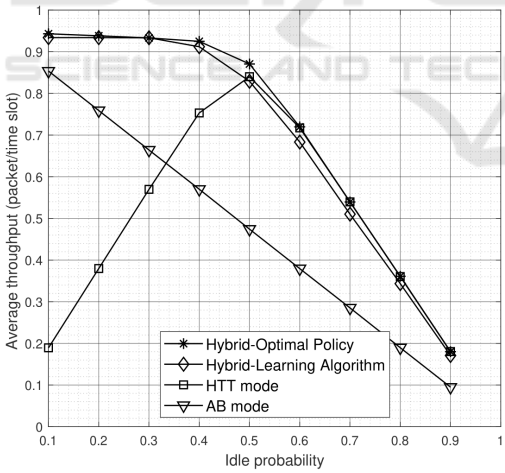
obtained through value iteration algorithm.

In Figures 4 and 5, the performance of the proposed hybrid transmission mode is investigated and compared with HTT mode and AB mode in terms

of throughput and blocking probability, respectively. Figures 4*a* and 4*b* show the impact of the arrival rate of the SU and the idle probability of the licensed channels on the achievable throughput for different policies. The arrival rate is simply calculated as $\sum_{i=0}^{R} i\alpha_i$. Moreover, we assume that $\alpha_1 = \alpha_2 = \alpha$ where α is changed between 0.05 and 0.45 which allows the arrival rate to vary between 0.15 to 1.35.



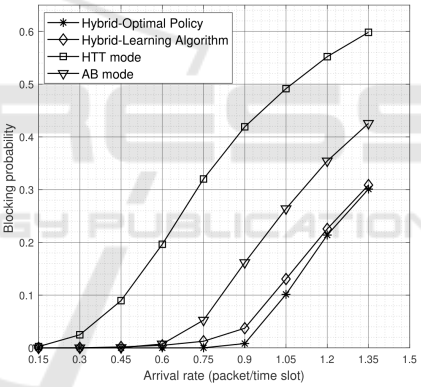
(a) The effect of the SU arrival rate for $\eta_1 = 0.1$ and $\eta_2 = 0.3$



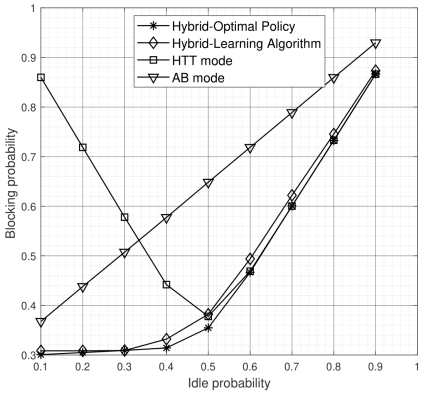
(b) The effect of the idle channel probability for $\alpha = 0.5$

Figure 4: Average throughput performance for different transmission modes.

Figure 4*a* shows that for low arrival rate, all modes almost give the same average throughput. That is because, the amount of data units stored in the data queue is relatively low. Hence, an inconsiderable opportunity is sufficient to transmit the existing data packets. When the arrival rate increases, the difference in performance between the modes appears and the proposed hybrid mode achieves a higher average throughput. When the arrival rate is relatively high, the system reaches to the saturated state and thus the average throughput is steady. Figure 4*b* shows that for low idle probability, the proposed hybrid mode achieves significantly higher throughput compared to the HTT policy. This is because, for low idle probability, the SU in the HTT mode almost has no opportunity to transmit his data. The performance of the HTT mode enhances as the idle probability increases because the opportunity of transmission increases. On the other hand, as the idle probability increases, the performance of the AB decreases because the availability of the RF signal of the PU decreases. For higher idle probability (greater than 0.5) the performance of all modes starts to decrease because the availability of the RF signal of the PU which is used for backscattering transmission and for energy harvesting decreases. Moreover, Figures 4*a*, 4*b* show that the learning algorithm always achieves performance close to that of the optimal policy.



(a) The effect of the SU arrival rate for $\eta_1 = 0.1$ and $\eta_2 = 0.3$



(b) The effect of the idle channel probability for $\alpha = 0.5$

Figure 5: Blocking probability performance for different transmission modes.

In Figure 5, the blocking probability is investigated. From Figure 5*a*, as the arrival rate increases the blocking probability for the arrival packets increases. That is because, when the number of arriving packets increases, the probability that the finite queue size reaches its maximum value increases which in turns increases the blocking probability. From Figure 5*b*, for low idle channel probabilities, the HTT mode gives the highest blocking probability because there is almost no opportunity to transmit any packets which consequently accumulate the data packets in the queue till it reaches its maximum capacity and blocks any further arrival packets. The blocking probability of the HTT mode decreases as the idle probability increases. However, when the idle probability is greater than 0.5, the blocking probability of the HTT mode starts to increase in a pattern similar to the other modes.

5 CONCLUSIONS

In this paper, we applied a reinforcment learning approach to study a hybrid HTT/backscattering transmission mode of an RF-powered CR network. The average throughput and the blocking probability for the SU are investigated for the incomplete information channel case under the unknown environment parameters assumption. Numerical results showed that the performance of the proposed hybrid mode is better than that of using HTT and backscattering transmission modes especially for the case of heavy SU loads or small PU idle probability. Finally, the proposed model can be extended by considering the mode and channel selection problem in a multi-channel RF-powered cognitive radio networks composed of multiple SUs, where the optimization problem is more challenging.

ACKNOWLEDGEMENTS

This work is supported by the National Telecom Regulatory Authority (NTRA) of Egypt under the project entitled "Security-Reliability Tradeoff in Spectrum Sharing Networks with Energy Harvesting". Ahmed Y. Zakariya acknowledges also the support from the Missions Sector of the Higher Education Ministry in Egypt through Ph.D. scholarship.

REFERENCES

Anh, T. T., Luong, N. C., Niyato, D., Liang, Y.-C., and Kim, D. I. (2019). Deep reinforcement learning for time scheduling in RF-powered backscatter cognitive radio networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–7.

Fahim, T. E., Zakariya, A. Y., and Rabia, S. I. (2018). A novel hybrid priority discipline for multi-class secondary users in cognitive radio networks. *Simulation Modelling Practice and Theory*, 84:69–82.

Gouda, A. E., Rabia, S. I., Zakariya, A. Y., and Omar, M. (2018). Reactive spectrum handoff combined with random target channel selection in cognitive radio networks with prioritized secondary users. *Alexandria engineering journal*, 57(4):3219–3225.

Hoang, D. T., Niyato, D., Wang, P., and Kim, D. I. (2014). Opportunistic channel access and RF energy harvesting in cognitive radio networks. *IEEE Journal on Selected Areas in Communications*, 32(11):2039–2052.

Liu, V., Parks, A., Talla, V., Gollakota, S., Wetherall, D., and Smith, J. R. (2013). Ambient backscatter: Wireless communication out of thin air. *ACM SIGCOMM Computer Communication Review*, 43(4):39–50.

Lu, X., Wang, P., Niyato, D., and Hossain, E. (2014). Dynamic spectrum access in cognitive radio networks with RF energy harvesting. *IEEE Wireless Communications*, 21(3):102–110.

Niyato, D., Wang, P., and Kim, D. I. (2014). Channel selection in cognitive radio networks with opportunistic RF energy harvesting. In *IEEE International Conference on Communications (ICC)*, pages 1555–1560.

Park, S., Kim, H., and Hong, D. (2013). Cognitive radio networks with energy harvesting. *IEEE Transactions on Wireless Communications*, 12(3):1386–1397.

Sigaud, O. and Buffet, O. (2013). *Markov decision processes in artificial intelligence*. John Wiley & Sons.

Van Huynh, N., Hoang, D. T., Lu, X., Niyato, D., Wang, P., and Kim, D. I. (2018a). Ambient backscatter communications: A contemporary survey. *IEEE Communications Surveys & Tutorials*, 20(4):2889–2922.

Van Huynh, N., Hoang, D. T., Nguyen, D. N., Dutkiewicz, E., Niyato, D., and Wang, P. (2018b). Reinforcement learning approach for RF-powered cognitive radio network with ambient backscatter. In *IEEE Global Communications Conference (GLOBECOM)*, pages 1–6.

Van Huynh, N., Hoang, D. T., Nguyen, D. N., Dutkiewicz, E., Niyato, D., and Wang, P. (2019). Optimal and low-complexity dynamic spectrum access for RF-powered ambient backscatter system with online reinforcement learning. *IEEE Transactions on Communications*, 67(8):5736–5752.

Wang, B. and Liu, K. R. (2010). Advances in cognitive radio networks: A survey. *IEEE Journal of Selected Topics in Signal Processing*, 5(1):5–23.

Weaver, L. and Tao, N. (2001). The optimal reward baseline for gradient-based reinforcement learning. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 538–545.

Zakariya, A. Y. and Rabia, S. I. (2016). Analysis of an interruption-based priority for multi-class secondary users in cognitive radio networks. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE.

Zakariya, A. Y., Rabia, S. I., and Abouelseoud, Y. (2019). An optimized general target channel sequence for prioritized cognitive radio networks. *Computer Networks*, 155:98–109.

Zakariya, A. Y., Rabia, S. I., and Zahra, W. (2020a). Analysis of a hybrid overlay/semi-passive backscattering spectrum access in cognitive radio networks. In *2020 24th International Conference on Circuits, Systems, Communications and Computers (CSCC)*, pages 252–255. IEEE.

Zakariya, A. Y., Rabia, S. I., and Zahra, W. (2021). Optimal decision making in multi-channel rf-powered cognitive radio networks with ambient backscatter capability. *Computer Networks*, page 107907.

Zakariya, A. Y., Tayel, A. F., Rabia, S. I., and Mansour, A. (2020b). Modeling and analysis of cognitive radio networks with different channel access capabilities of secondary users. *Simulation Modelling Practice and Theory*, 103:102096.

Zuo, P., Wang, X., Linghu, W., Sun, R., Peng, T., and Wang, W. (2018). Prediction-based spectrum access optimization in cognitive radio networks. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE.