Minimizing Age of Information for an Energy Harvesting Cognitive Radio

Shiyang Leng and Aylin Yener
Wireless Communications and Networking Laboratory (WCAN)
School of Electrical Engineering and Computer Science
The Pennsylvania State University, University Park, PA 16802.
sfl5154@psu.edu *yener@engr.psu.edu*

***Abstract*—This paper studies average age of information (AoI) minimization in cognitive radio energy harvesting communications. The secondary user is an energy harvesting sensor that harvests ambient energy with which it performs spectrum sensing and status updates. Status-update data is sent by opportunistically accessing the primary spectrum. Specifically, the secondary user aims to minimize the average AoI by adaptively making sensing and update decisions based on its energy availability and the availability of the primary spectrum. The sequential decision problem is formulated as a partially observable Markov decision process and solved by dynamic programming. The properties of the optimal sensing and updating policies are investigated and shown to have threshold structure. Numerical results confirm the analytical findings.**

I. Introduction

The timeliness of data delivery has become critical in wireless communications for various time-sensitive applications, for instance, vehicle-to-vehicle networking, unmanned vehicle tracking, and natural disaster monitoring, where the status of physical processes have to be updated in a timely manner. This highlights the issue of maintaining information fresh. The concept of *age of information* (AoI) has been introduced to measure the freshness of information [1], [2]. More specifically, AoI quantifies the time elapsed since the generation of the latest successfully received status update.

In early works on AoI, a queueing theoretic perspective has enabled the analysis and characterization of age [2], [3]. In [2], M/M/1, M/D/1, D/M/1 models and first-come-first-served (FCFS) queues are studied. Last-come-first-served (LCFS) is considered in [3] for an M/M/1 queue.

More recently, AoI has been investigated for energy harvesting systems, where each update consumes harvested energy [4]–[7]. Due to the randomness in the energy harvesting process, the information could become stale in these systems if energy shortage prevents updates. The main task thus is to optimally manage energy to keep updates fresh. Reference [4] considers AoI minimization for point-to-point communication with energy harvesting constraints, and shows that waiting before updating improves AoI when considering energy causality constraint. In [6] and [7], (asymptotically) optimal update policies for infinite, finite, and unit battery size are derived, where the optimal policy has a (multi-)threshold structure.

The existing work on AoI minimization for energy harvesting communications assumes that the wireless channel for update transmission is always available. By contrast, in cognitive-radio-based networks, a secondary user can only opportunistically access the primary spectrum when it is not occupied. Energy harvesting cognitive radio networks (EH-CRN) have been studied previously with the throughput as the main metric [8]–[11]. In [9], [12], accounting for the stochastic processes of primary spectrum availability and energy harvesting, decisions for sensing and/or transmitting are made by modeling the problem as a partially observable Markov decision process (POMDP) subject to energy causality constraints. In [11], a long-term average reward (throughput) for the secondary user is defined and upper bounded by a fixed fraction power allocation. While throughput continues to be a primary metric for energy harvesting communications at large, e.g., [13], in applications with energy harvesting cognitive radio sensors that have small data packets but a critical requirement on the freshness of information, AoI is a more appropriate metric. In [14], the AoI of the primary user in a cognitive radio network is characterized and minimized from a queuing theoretic perspective.

In this paper, we consider a cognitive radio (EH-CR) with one primary user (PU) and an energy harvesting cognitive secondary user (SU). We minimize the AoI for the SU, who monitors the environment and opportunistically sends status updates to the destination. The harvested energy is expended on spectrum sensing and statue update transmissions. A discrete-time system model for a finite horizon is adopted. Due to the randomness of energy harvesting and channel fading processes, the SU has to adaptively make sensing and update decisions in an online fashion. The primary user's state is modeled as a stationary two-state Markov chain whose state transition probabilities are known apriori to the SU. Considering reliable spectrum sensing, we formulate a POMDP for sequential decision making to minimize the average AoI over a fixed time duration. The information state of the system is represented by the fully observable states of SU and the belief on the state of PU, based on which sensing and update policies can be optimally determined using dynamic programming. We investigate the properties of the optimal policy, and verify that the SU senses and updates when the harvested energy is larger than a threshold determined

978-1-5386-7646-2/19/$31.00

by battery and channel states, and the update policy has a threshold structure with respect to the state of age. Numerical results are presented that confirm our analytical findings.

II. System Model

The SU sends the update by accessing the primary spectrum opportunistically. Consider a time-slotted system with slots indexed by $t = 0, 1, 2, \ldots$. The slot length is normalized and equal to the transmission time of one status-update packet.

The PU is licensed a legitimate spectrum. In each time slot, the PU either occupies the spectrum in an active (A) state or stays silent (S). Denote the state by $q_t \in \{\mathrm{A}, \mathrm{S}\}$. The transition probabilities of the two-state Markov chain are denoted by p_{ss} and p_{as} for staying in a silent state and transiting from an active state to silent, respectively. The transition probabilities are obtained by long-term measurements and known to the SU.

The SU is slot-synchronized with the PU. At the beginning of each slot, the SU decides its operation mode: idle or sensing. If it is staying idle, no action is needed. If it decides to sense, it further decides whether to update status or not. The SU aims to minimize the average AoI by making optimal sensing and update decisions over a finite horizon $t = 0, 1, 2, \ldots, T-1$. Let $x_t = (w_t, z_t)$ be the decision for slot t, where $w_t \in \{0 \text{ (idle)}, 1 \text{ (sense)}\}$ and $z_t \in \{0 \text{ (not update)}, 1 \text{ (update)}\}$ denote the sensing and update decisions, respectively. The decisions are made adaptively over $t = 0, 1, 2, \ldots, T-1$ based on SU's states and its statistical knowledge of the primary spectrum availability.

1) Belief Model: The SU cannot directly monitor the availability of primary spectrum, but only partially observe by opportunistically sensing and accessing. Based on its action and observation history, a sufficient statistic of the primary spectrum availability can be obtained, which is the belief. Specifically, at each slot, if the SU decides to sense, an observation of PU state can be obtained, denoted by $\hat{q}_t \in \{\mathrm{A}, \mathrm{S}\}$. For reliable spectrum sensing as we assume in this paper, $\hat{q}_t = q_t$, $\forall t$. Given action and observation history, the SU forms a belief ρ_t, which represents the conditional probability of the PU being silent, i.e., $q_t = \mathrm{S}$.

2) Channel Model: We consider that the SU transmits data over a block fading channel with channel gain h_t for slot t, $\forall t$. $\mathbf{h}_t$ is a discrete random variable with distribution $p_{\mathcal{H}}(\mathbf{h}_t = h_t)$ over a finite sample space $\mathcal{H}$. This is mainly for mathematical tractability and can be interpreted as quantization of channel gains. Assume $\mathbf{h}_t$ is independently and identically distributed (i.i.d.) over slots. The distribution is known a prior by the SU. At the beginning of slot t, if the SU senses, it obtains the channel gain h_t causally when the spectrum is unoccupied; otherwise, it keeps the old channel information, i.e., $h_t = h_{t-1}$.

3) Energy Harvesting Model: The SU is able to harvest energy from its ambient environment and store it in the battery before use. The battery capacity is $\bar{b}$. The energy harvested at slot t is a discrete random variable $\mathbf{e}_t$, whose distribution $p_{\mathcal{E}}(\mathbf{e}_t = e_t)$ is known apriori and realization is $e_t \in \mathcal{E}$ with $\mathcal{E}$ finite. The energy harvesting process is assumed to be i.i.d. across slots. Energy is mainly consumed on sensing the spectrum and transmitting the status updates over the wireless channel. Other operations of the SU is assumed to expend negligible energy. Let σ be the energy consumption on sensing. The transmission energy depends on the channel gain, which is given by $u(h_t)$. The function $u(\cdot)$ is nonincreasing and convex, for instance, $u(h_t) = \frac{\tau}{h_t}(\exp(R/\tau) - 1)$ for Gaussian channel with transmission rate R and time τ. Let e_t, σ, and $u(h_t)$ be integer multiples of unit energy. We have the battery state $b_t \in \mathcal{B} \triangleq \{0, 1, \ldots, \bar{b}\}$, which evolves as

$$b_{t+1} = \min\{b_t + e_t - w_t\sigma - z_t u(h_t), \bar{b}\}. \quad (1)$$

The energy causality constraint has to be satisfied, which is:

$$w_t\sigma + z_t u(h_t) \leq b_t. \quad (2)$$

Note that, here, we harvest the energy first and then use at the next slot onwards. That is, we have a store-then-use model.

4) Age of Information: We adopt a linear model for AoI, where AoI is defined as the time elapsed since the time instant when the most recently received update is generated. Let a_t denote the AoI of slot t. Once the SU decides to update status, it generates and transmits a data packet. We consider the generate-at-will scheme [4]–[7], that the data packet is generated when update decision is made. Assume that the data is small such that it is generated and transmitted instantaneously, and received at the end of the slot, i.e., transmission time is one slot. If update is successfully received, the AoI decreases to 1; otherwise increases by 1. Note that a_t is upper bounded by $a_0 + T$ and $a_t \in \mathcal{A} \triangleq \{1, 2, \ldots, a_0 + T\}$ for a finite T. A sample path of AoI is depicted in Fig. 1 with $a_0 = 1$. Consider an error-free channel over which the data can be received successfully once transmitted over an unoccupied spectrum. Therefore, only transmission collision, i.e., $x_t = (1, 1)$ and $q_t = \mathrm{A}$, leads to an update failure. The average AoI is the cumulative AoI (the area under the age curve) averaged over time. For an interval of T slots, the average AoI can be represented as the average sum of disjoint geometric parts of each slot,

$$J = \frac{1}{T}\sum_{t=0}^{T-1}\Big(\frac{1}{2} + a_t\Big). \quad (3)$$

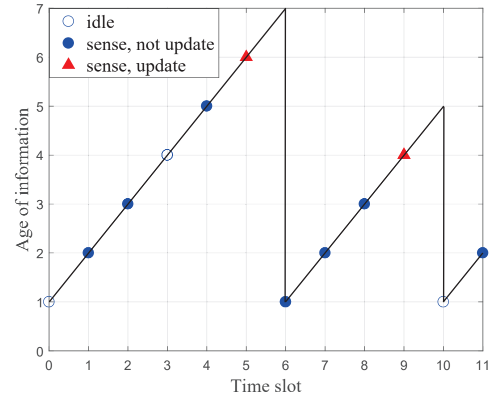

Fig. 1. A sample path of AoI.

III. POMDP Formulation

The sequential sensing and update decisions for average AoI minimization is formulated as a POMDP. We describe the components of the POMDP as follows.

Actions: The SU first makes sensing decision. If the SU senses and observes PU to be active, then it does not update, i.e., $x_t = (1,0)$; if the SU senses and observes PU silent, then it further makes an update decision based on its AoI, energy availability, and the channel gain. The action for each slot is $x_t = (w_t, z_t) \in \mathcal{X} \triangleq \{(0,0),(1,0),(1,1) : b_t \geq w_t\sigma + z_t u(h_t)\}$, where $w_t \in \Gamma_w \triangleq \{0,1 : b_t \geq w_t\sigma\}$ and $z_t \in \Gamma_z \triangleq \{0,1 : b_t \geq \sigma + z_t u(h_t)\}$.

Observations and beliefs: The observation of the PU's state is $\hat{q}_t \in \{\mathrm{A},\mathrm{S}\}$. The belief $\rho_t \in [0,1]$ is a conditional probability indicating the availability of primary spectrum. Based on the action and observation history, the belief evolves over slot by $\rho_{t+1} = I(\rho_t)$. If the SU stays idle without sensing, the new belief is updated solely based on the underlying Markov chain of the PU state. Otherwise, the sensing result shows the true state. Specifically, we have

$$\rho_{t+1} = \begin{cases} I_0(\rho_t) = \rho_t p_{\mathrm{ss}} + (1-\rho_t)p_{\mathrm{as}}, & \text{if } w_t = 0 \\ I_A(\rho_t) = p_{\mathrm{as}}, & \text{if } w_t = 1, \hat{q}_t = \mathrm{A} \\ I_S(\rho_t) = p_{\mathrm{ss}}, & \text{if } w_t = 1, \hat{q}_t = \mathrm{S}. \end{cases} \quad (4)$$

Given initial belief, the number of possible beliefs over T slots is finite, since from the current belief, the SU can only transit to three beliefs by (4). Thus, for a finite time T, the belief space $\mathcal{I}$ is a finite set.

States: The completely observable states of each slot consists of AoI state, battery state, energy harvesting state, and channel state, denoted by $s_t \triangleq (a_t, b_t, e_t, h_t)$. Note the state space, i.e., $\mathcal{S} \triangleq \mathcal{A} \times \mathcal{B} \times \mathcal{E} \times \mathcal{H}$, is finite. In particular, over an error-free channel, update is always successful when the sensing result is $\hat{q}_t = \mathrm{S}$ and update decision is $z_t = 1$. Thus, for $t = 0, \ldots, T-1$,

$$a_{t+1} = \begin{cases} 1, & \text{if } x_t = (1,1) \\ a_t + 1, & \text{otherwise}, \end{cases} \quad (5)$$

or more compactly, $a_{t+1} = (1 - z_t)a_t + 1$. Additionally, the spectrum state is only partially observable and is described by the sufficient statistic, i.e., belief ρ_t. We denote the complete information state by (s_t, ρ_t), $\forall t$. Since $\mathcal{S}$ and $\mathcal{I}$ are finite, the SU can only experience a finite number of possible information states $(s_t, \rho_t) \in \mathcal{S} \times \mathcal{I}$.

Transition probabilities: Given current state $s_t = (a_t, b_t, e_t, h_t)$ and action $x_t = (w_t, z_t)$, the transition probability to state $s_{t+1} = (a_{t+1}, b_{t+1}, e_{t+1}, h_{t+1})$ is denoted by $p_{x_t}(s_{t+1}|s_t)$. Due to the independency of energy harvesting and channel fading and their being i.i.d., we have

$$\begin{aligned} p_{x_t}(s_{t+1}|s_t) = \; & \mathbb{P}(a_{t+1}|a_t, x_t)\mathbb{P}(b_{t+1}|b_t, e_t, h_t, x_t) \cdot \\ & p_{\mathcal{E}}(e_{t+1}) p_{\mathcal{H}}(h_{t+1}), \end{aligned} \quad (6)$$

where $\mathbb{P}(a_{t+1}|a_t, x_t) = 1$ if $a_{t+1} = (1 - z_t)a_t + 1$ and $\mathbb{P}(b_{t+1}|b_t, e_t, h_t, x_t) = 1$ if $b_{t+1} = \min\{b_t + e_t - w_t\sigma - z_t u(h_t), \bar{b}\}$.

Cost: Let $C(s_t)$ be the immediate cost taken under state s_t, which is given by

$$C(s_t) = \frac{1}{2} + a_t, \; t = 0, \ldots, T. \quad (7)$$

Policy: Denote the policy $\pi = \{\mu_0, \ldots, \mu_{T-1}\}$, where μ_t is a deterministic decision rule that maps an information state $(s_t, \rho_t) \in \mathcal{S} \times \mathcal{I}$ into an action $x_t \in \mathcal{X}$, i.e, $x_t = \mu_t(s_t, \rho_t)$. Let Π denotes the set of all deterministic policies.

The POMDP can be reformulated as a perfect state information problem by adopting the information state (s, ρ) in $\mathcal{S} \times \mathcal{I}$ [15]. Given SU's initial state and belief, the finite-horizon average AoI under policy π is expressed as

$$J^\pi(s_0, \rho_0) = \frac{1}{T}\mathbb{E}\Big[\sum_{t=0}^{T-1} C(s_t)|s_0, \rho_0\Big], \quad (8)$$

where the expectation is taken over policy π. Finding the optimal sensing and update policy that minimizes the average AoI corresponds to solve the optimization problem

$$\min_{\pi \in \Pi} J^\pi(s_0, \rho_0). \quad (9)$$

For a fixed T, (9) is a finite-state MDP with total cost.

IV. POMDP Solution

We use dynamic programming to solve the finite-horizon total cost minimization problem in (9) [15]. Let $V_t(s_t, \rho_t)$ denote the value function,

$$V_t(s_t, \rho_t) \triangleq \min_{\{x_i\}_{i=t}^{T-1}} \mathbb{E}\Big[\sum_{i=t}^{T-1} C(s_i)|s_t, \rho_t\Big], \quad (10)$$

which is the minimum expected cost accumulated from slot t to $T-1$ given information state (s_t, ρ_t). Then, the minimum AoI in (9) is $J^* = V_0(s_0, \rho_0)/T$ for fixed T. Let $Q_t^{w_t}(s_t, \rho_t)$ denote the action-value function or Q-function, which represents the minimum expected cost for taking sensing action w_t in state (s_t, ρ_t) that is accumulated since t. The Q-function consists of two parts: the immediate cost obtained under current state and the expected sum of value functions for the next slot. The finite-horizon MDP problem can be solved via dynamic programming recursion as follows. For $t = 0, 1, \ldots, T-1$,

$$V_t(s_t, \rho_t) = \min_{w_t \in \Gamma_w} Q_t^{w_t}(s_t, \rho_t), \quad (11)$$

where for $t = T-1$,

$$Q_{T-1}^0(s_{T-1}, \rho_{T-1}) = C(s_{T-1}) + C(s_T), \quad (12)$$

$$\begin{aligned} Q_{T-1}^1(s_{T-1}, \rho_{T-1}) = \; & (1 - \rho_{T-1})C(s_{T-1}) + \\ & \rho_{T-1} \min_{z_{T-1} \in \Gamma_z} C(s_{T-1}) + C(s_T), \end{aligned} \quad (13)$$

and for $t = 0, \ldots, T-2$,

$$Q_t^0(s_t,\rho_t) = C(s_t) + \sum_{s_{t+1}} p_{00}(s_{t+1}|s_t) V_{t+1}(s_{t+1}, I_0(\rho_t)), \quad (14)$$

$$Q_t^1(s_t,\rho_t) = (1-\rho_t) Q_t^{1A}(s_t,\rho_t) + \rho_t \min_{z_t \in \Gamma_z} Q_t^{1z_t}(s_t,\rho_t), \quad (15)$$

$$Q_t^{1A}(s_t,\rho_t) = C(s_t) + \sum_{s_{t+1}} p_{10}(s_{t+1}|s_t) V_{t+1}(s_{t+1}, I_A(\rho_t)), (16)$$

$$Q_t^{10}(s_t,\rho_t) = C(s_t) + \sum_{s_{t+1}} p_{10}(s_{t+1}|s_t) V_{t+1}(s_{t+1}, I_S(\rho_t)), \ (17)$$

$$Q_t^{11}(s_t,\rho_t) = C(s_t) + \sum_{s_{t+1}} p_{11}(s_{t+1}|s_t) V_{t+1}(s_{t+1}, I_S(\rho_t)), \ (18)$$

In particular, $Q_t^{1A}(s_t, \rho_t)$ in (16) denotes the conditional minimum expected cost given sensing result $\hat{q}_t = \text{A}$, i.e., adopting action $x_t = (1, 0)$. In (17) and (18), given sensing result $\hat{q}_t = \text{S}$, $Q_t^{10}(s_t, \rho_t)$ and $Q_t^{11}(s_t, \rho_t)$ characterize the conditional minimum expected costs by adopting update action $z_t = 0$ and $z_t = 1$, respectively. By recursion in (11)-(18), the optimal sensing and updating policies are obtained by

$$w_t^*(s_t, \rho_t) \in \arg\min_{w_t \in \Gamma_w} Q_t^{w_t}(s_t, \rho_t), \quad (19)$$

$$z_t^*(s_t, \rho_t) \in \arg\min_{z_t \in \Gamma_z} Q_t^{1z_t}(s_t, \rho_t). \quad (20)$$

Next, we analyze the structure of the optimal policy to gain insights for optimum sequential decision making in EH-CR with the objective of AoI minimization. We first show the monotonicity of the value function with respect to each component of the information state.

Proposition 1: For $t = 0, \ldots, T-1$,

1) $V_t(s_t, \rho_t)$ is nondecreasing with respect to the AoI state a_t.
2) $V_t(s_t, \rho_t)$ is nonincreasing with respect to battery state b_t, energy harvesting state e_t, and channel state h_t.
3) $V_t(s_t, \rho_t)$ is nonincreasing with respect to belief ρ_t if $p_{\text{ss}} \geq p_{\text{as}}$.

Proof: See Appendix A. ■

In [16], it is proved that for a finite and fixed time horizon POMDP, the value function is piecewise linear and convex with respect to the belief state for a reward maximization problem. Applying the theory developed there, we can verify that $V_t(s_t, \rho_t)$ of our total cost minimization problem is piecewise linear and concave with respect to belief ρ_t, $\forall t$. The monotonicity and concavity of the value function establish the basis for the following analysis on the solution structure of the optimal policy, which holds for all t, (we will be omitting the time index without loss of generality).

Theorem 1: For the optimal sensing policy, the SU senses, i.e., $w^*(a, b, e, h, \rho) = 1$, if $e \geq \sigma + \bar{b} - b$.

Proof: See Appendix B. ■

Theorem 2: For the optimal update policy, if $e \geq \sigma + \bar{b} - b$, and $z^*(a, b, e, h, \rho) = 1$, then for any $b' \geq b$, $z^*(a, b', e, h, \rho) = 1$; if $e \geq \sigma + u(h) + \bar{b} - b$, the SU updates, i.e., $z^*(a, b, e, h, \rho) = 1$.

Proof: See Appendix C. ■

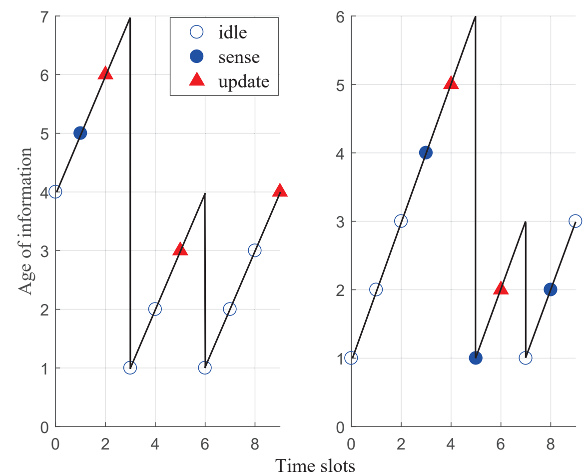

Fig. 2. Two sample paths of AoI by the optimal policy for $\bar{b} = 5$, $p_e = 0.3$, $p_h = 0.5$.

Theorem 3: The optimal update policy has a threshold structure with respect to the AoI state: if $z^*(a, b, e, h, \rho) = 1$, then for any $a' \geq a$, $z^*(a', b, e, h, \rho) = 1$.

Proof: See Appendix D. ■

Theorem 1 implies if the harvested energy is large enough such that the battery is full at the beginning of the next slot, then the SU always decides to sense. A similar result can be concluded for the update policy from Theorem 2, that if the harvested energy is large enough such that the battery can be fully charged taking account the sensing and update cost, then the SU always decides to sense and update, i.e., $x^* = (1, 1)$. Furthermore, if the update is transmitted at battery state b, then update is also decided for any larger battery b' as stated in Theorem 2. Similar threshold structure for the optimal update policy with respect to AoI state is stated in Theorem 3. The computation of value iteration can be reduced by exploiting the threshold structure of the optimal policy.

V. Numerical Results

In this section, we present numerical results to verify our findings. The PU has state transition probabilities $p_{\text{ss}} = 0.8$ and $p_{\text{sa}} = p_{\text{as}} = 0.2$. The initial belief is set to be $\rho_0 = p_{\text{ss}}$. The energy consumption for sensing is $\sigma = 1$. The energy harvesting process is i.i.d. Bernoulli with probability p_e for harvesting $e = 3$ and probability $1 - p_e$ for $e = 0$. For the channel fading level, we set $\mathcal{H} = \{h_1, h_2\}$ with p_h for h_1, where each level corresponds to an energy cost on update. Set $u(h_1) = 2$ and $u(h_2) = 4$. We compare the proposed optimal policy with a myopic policy. In particular, in a myopic policy, the SU senses the primary spectrum whenever it has enough energy for sensing. If the primary spectrum is sensed to be unoccupied, the update takes place if the residual energy is sufficient for an update.

In Fig. 2, we plot a sample path of AoI by the optimal policy for battery capacity $\bar{b} = 5$. Table I provides the corresponding information states and actions. By comparing the states of $t = 1, 2$ in the first sample path, we can observe that when the harvested energy is large enough to make the battery full at the

TABLE I
TWO SAMPLE PATHS OF AOI BY THE OPTIMAL POLICY

Time t	a	b	e	$u(h)$	ρ	Action x
1	5	3	3	4	0.5389	$(1,0)$
2	6	5	3	2	0.8	$(1,1)$
3	1	5	0	4	0.8	$(0,0)$
5	3	5	0	2	0.6080	$(1,1)$
9	4	5	0	2	0.5648	$(1,1)$
3	4	3	3	4	0.5648	$(1,0)$
4	5	5	3	4	0.8	$(1,1)$
5	1	3	3	4	0.8	$(1,0)$
6	2	5	3	4	0.8	$(1,1)$
7	1	3	0	4	0.8	$(0,0)$
8	2	3	3	2	0.68	$(1,0)$

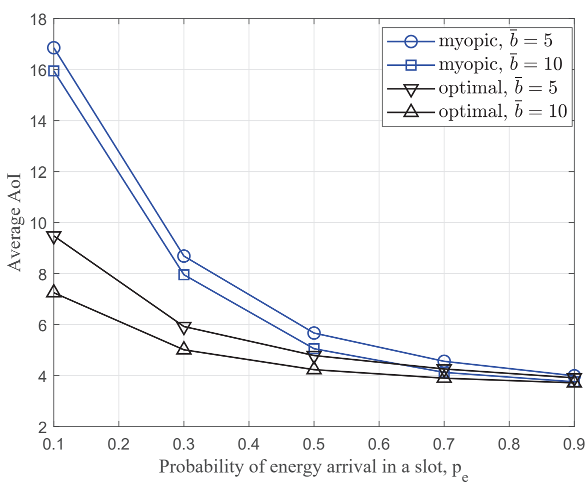

Fig. 3. Average AoI versus the probability of energy harvesting p_e for $T = 50$, $p_h = 0.5$, $a_0 = 1$.

beginning of next slot, sensing and update are implemented. At $t = 5, 9$, the threshold structure of the optimal update policy in terms of AoI state is shown. Similar results can be observed in the second sample path at the sensing instants $t = 3, 5$ and the update instants $t = 4, 6$.

Fig. 3 shows the average AoI versus the probability of energy harvesting p_e for $T = 50$. It can be observed that the average AoI decreases as the probability of energy harvesting grows. The optimal policy performs better than myopic policy, especially when energy is scarce since the optimal policy mitigates the randomness of energy harvesting by decisions. The same reason is for the performance of cases with small and large battery capacities, i.e, $\bar{b} = 5, 10$. When a larger battery is present, the average AoI is lower.

VI. CONCLUSION

In this paper, we have investigated an energy harvesting cognitive secondary user sensor with the aim of AoI minimization. For the energy harvesting cognitive radio who needs to keep the information at its destination as fresh as possible, optimal sensing and update decisions that minimize the average AoI over a finite horizon are considered. Taking into account the partially observable state of the primary user, POMDP is adopted to formulate the total cost minimization problem subject to the energy causality constraint. The POMDP is formulated as a perfect state information problem, which is solved by dynamic programming. The monotonicity of the value function and a threshold structure for the optimal policy are shown. The numerical results illustrate the policy structures and the impact of system parameters. Future work includes the AoI minimization for the secondary user with imperfect spectrum sensing as well as the infinite horizon problem.

APPENDIX A
PROOF OF PROPOSITION 1

For the ease of notation, we omit the notation for irrelevant state components in the sequel.

(1) Nondecreasing in a: We show that $V_t(a_t') \geq V_t(a_t)$ for $a_t' \geq a_t$ by induction according to the recursion in (11)-(18).

For $t = T-1$, by (12) and (13), $Q_{T-1}^{w_{T-1}}(a_{T-1}') \geq C(a_{T-1}') + C(a_T') = C(a_{T-1}') + C((1-z_{T-1})a_{T-1}'+1) \geq Q_{T-1}^{w_{T-1}}(a_{T-1})$ for $a_{T-1}' \geq a_{T-1}$. Since min preserves the monotonicity, $V_{T-1}(a_{T-1}') \geq V_{T-1}(a_{T-1})$ from (11).

Suppose $V_{t+1}(a_{t+1}') \geq V_{t+1}(a_{t+1})$ for some t, we next show $V_t(a_t') \geq V_t(a_t)$. From (14), $Q_t^0(a_t') \geq Q_t^0(a_t)$ holds as $C(a_t') \geq C(a_t)$ and $V_{t+1}(a_{t+1}') \geq V_{t+1}(a_{t+1})$. Similarly, we have $Q_t^{1A}(a_t') \geq Q_t^{1A}(a_t)$, $Q_t^{10}(a_t') \geq Q_t^{10}(a_t)$ and $Q_t^{11}(a_t') \geq Q_t^{11}(a_t)$. Then, $Q_t^1(a_t') \geq Q_t^1(a_t)$ from (15). Consequently, $V_t(a_t') \geq V_t(a_t)$ from (11).

(2) Nonincreasing in b, e, and h: Same induction procedure as for state a follows for verifying the nonincreasing in b. Note that if $b_t' \geq b_t$ for any $t = 0, \ldots, T-1$ with other states the same, the SU with b_t' can sense and update no less times from slot t to the end than with b_t, which leads to no larger cost. Considering energy harvesting state e_t, a larger e_t' results in battery state b_{t+1}' no less than b_{t+1}, which implies a lower value function. Similarly, a higher channel state h_t' leads to a smaller transmission cost $u(h_t')$ due to the nonincreasing function $u(\cdot)$, thus, more residual energy can be kept in the battery to provide a lower value function.

(3) Nonincreasing in ρ: We show that $V_t(\rho_t') \leq V_t(\rho_t)$ for $\rho_t' \geq \rho_t$ by induction according to the recursion in (11)-(18).

For $t = T-1$, if the secondary user stays idle, we have $Q_{T-1}^0(\rho_{T-1}') = Q_{T-1}^0(\rho_{T-1})$ from (12). If sensing, from (13),

$$Q_{T-1}^1(\rho_{T-1}') = \begin{cases} C(a_{T-1}) + C(a_{T-1}+1), & \text{if } z_{T-1} = 0 \\ C(a_{T-1}) + C(1), & \text{if } z_{T-1} = 1. \end{cases} \quad (21)$$

Since the update policy does not depend on the belief, $Q_{T-1}^1(\rho_{T-1}') \leq Q_{T-1}^1(\rho_{T-1})$. Thus, $V_{T-1}(\rho_{T-1}') \leq V_{T-1}(\rho_{T-1})$ by (11).

Suppose $V_{t+1}(\rho_{t+1}') \leq V_{t+1}(\rho_{t+1})$ for some t, we next show $V_t(\rho_t') \leq V_t(\rho_t)$. From (4), $I_0(\rho_t)$ is nondecreasing in ρ_t as $p_{\text{ss}} \geq p_{\text{as}}$. Then, $V_{t+1}(I_0(\rho_t')) \leq V_{t+1}(I_0(\rho_t))$ for $\rho_t' \geq \rho_t$ by assumption. This implies $Q_t^0(\rho_t') \leq Q_t^0(\rho_t)$ according to (14). By similar argument, it can be verified that $Q_t^{1A}(\rho_t') \leq$

$Q_t^{1A}(\rho_t)$, $Q_t^{10}(\rho_t') \leq Q_t^{10}(\rho_t)$, and $Q_t^{11}(\rho_t') \leq Q_t^{11}(\rho_t)$. Then, for $Q_t^1(\rho_t')$ given in (15),

$$\begin{aligned} Q_t^1(\rho_t') &\leq (1-\rho_t')Q_t^{1A}(\rho_t) + \rho_t' \min_{z_t \in \Gamma_z} \{Q_t^{10}(\rho_t), Q_t^{11}(\rho_t)\} \\ &= Q_t^{1A}(\rho_t) + \rho_t' \Delta Q_t(\rho_t), \end{aligned} \tag{22}$$

where $\Delta Q_t(\rho_t) \triangleq \min_{z_t \in \Gamma_z} \{Q_t^{10}(\rho_t), Q_t^{11}(\rho_t)\} - Q_t^{1A}(\rho_t) \leq Q_t^{10}(\rho_t) - Q_t^{1A}(\rho_t) \leq 0$. The nonpositivity is by (17) and (16), where $V_{t+1}(I_S(\rho_t)) \leq V_{t+1}(I_A(\rho_t))$ holds by assumption for $I_S(\rho_t) = p_{\text{ss}} \geq I_A(\rho_t) = p_{\text{as}}$. Therefore, from (22) $Q_t^1(\rho_t') \leq Q_t^{1A}(\rho_t) + \rho_t' \Delta Q_t(\rho_t) \leq Q_t^{1A}(\rho_t) + \rho_t \Delta Q_t(\rho_t) = Q_t^1(\rho_t)$. By (11), we conclude $V_t(\rho_t') \leq V_t(\rho_t)$.

Appendix B
Proof of Theorem 1

Let C denote $C(s)$ if a is not changed. To prove $w(b,\rho) = 1$, we need to show that $Q^0(b,\rho) \geq Q^1(b,\rho)$. Since $e \geq \sigma + \bar{b} - b$, the new battery state becomes $b' = \min\{b+e-\sigma, \bar{b}\} = \bar{b}$ if sensing, and $b'' = \min\{b+e, \bar{b}\} = \bar{b}$ if not sensing. By (15),

$$\begin{aligned} Q^1(b,\rho) &\leq (1-\rho)Q^{1A}(b,\rho) + \rho Q^{10}(b,\rho) \\ &= (1-\rho)\big[C + \textstyle\sum_{s'} p_{10}(s'|s)V(b', I_A(\rho))\big] \\ &\quad + \rho\big[C + \textstyle\sum_{s'} p_{10}(s'|s)V(b', I_S(\rho))\big] \\ &\overset{(1)}{\leq} C + \textstyle\sum_{s'} p_{10}(s'|s)V(\bar{b}, (1-\rho)I_A(\rho) + I_S(\rho)) \\ &\overset{(2)}{=} C + \textstyle\sum_{s'} p_{10}(s'|s)V(\bar{b}, I_0(\rho)) \\ &= C + \textstyle\sum_{s''} p_{00}(s''|s)V(b'', I_0(\rho)) = Q^0(b,\rho) \end{aligned} \tag{23}$$

where (1) is by the concavity of value function with respect to the belief, and (2) is from the belief update equation in (4).

Appendix C
Proof of Theorem 2

First we show that when $e \geq \sigma + \bar{b} - b$, for any larger battery state $b' \geq b$, if $z^*(a,b,e,h,\rho) = 1$, then $z^*(a,b',e,h,\rho) = 1$. We need to show that $Q^{11}(b',\rho) \leq Q^{10}(b',\rho)$. Since $e \geq \sigma + \bar{b} - b$, the new battery state becomes $\tilde{b} = \min\{b - \sigma + e, \bar{b}\} = \bar{b}$ if solely sensing. By (18),

$$\begin{aligned} Q^{11}(b',\rho) &= C + \textstyle\sum_{\tilde{s}'} p_{11}(\tilde{s}'|s)V(\min\{b' - \sigma - u(h) + e, \bar{b}\}, I_S(\rho)) \\ &\overset{(1)}{\leq} C + \textstyle\sum_{\tilde{s}} p_{11}(\tilde{s}|s)V(\min\{b - \sigma - u(h) + e, \bar{b}\}, I_S(\rho)) \\ &\overset{(2)}{\leq} C + \textstyle\sum_{\tilde{s}} p_{10}(\tilde{s}|s)V(\tilde{b}, I_S(\rho)) = Q^{10}(b',\rho), \end{aligned} \tag{24}$$

where (1) is by the monotonicity of value function with respect to the battery state, and (2) is due to $z^*(a,b,e,h,\rho) = 1$ implying $Q^{11}(b) \leq Q^{10}(b)$.

Next, we prove that if the battery state satisfies $e \geq \sigma + u(h) + \bar{b} - b$, $z^*(a,b,e,h,\rho) = 1$. By Theorem 1, sensing is carried out, i.e., $w^*(a,b,e,h,\rho) = 1$. Thus, we only need to show $Q^{10}(a,b,e,h,\rho) \geq Q^{11}(a,b,e,h,\rho)$. Since $e \geq \sigma + u(h) + \bar{b} - b$, the new battery state becomes $b' = \bar{b}$ if update is transmitted. By (18),

$$\begin{aligned} Q^{11}(a,b,e,h,\rho) &= C + \textstyle\sum_{s'} p_{11}(s'|s)V(1, \bar{b}, e', h', I_S(\rho)) \\ &\leq C + \textstyle\sum_{s''} p_{10}(s''|s)V(a+1, \bar{b}, e', h', I_S(\rho)) \\ &= Q^{10}(a,b,e,h,\rho), \end{aligned} \tag{25}$$

where the inequality is due to the monotonicity of the value function with respect to the AoI state.

Appendix D
Proof of Theorem 3

To prove $z(a',b,e,h,\rho) = 1$, all need to show is $Q^{10}(a',b,e,h,\rho) \geq Q^{11}(a',b,e,h,\rho)$. By (20), $z(a,b,e,h,\rho) = 1$ implies that $Q^{11}(a,b,e,h,\rho) \leq Q^{10}(a,b,e,h,\rho)$. That is, by (17) and (18),

$$\begin{aligned} &C + \textstyle\sum_{\tilde{s}} p_{11}(\tilde{s}|s)V(1, b_{11}, \tilde{e}, \tilde{h}, I_S(\rho)) \\ &\qquad \leq C + \textstyle\sum_{\tilde{s}} p_{10}(\tilde{s}|s)V(a+1, b_{10}, \tilde{e}, \tilde{h}, I_S(\rho)), \end{aligned} \tag{26}$$

where $b_{11} = \min\{b - \sigma - u(h) + e, \bar{b}\}$, $b_{10} = \min\{b - \sigma + e, \bar{b}\}$, $\tilde{e}$ and $\tilde{h}$ are the energy harvesting and channel states of the next slot. Thus, $V(1, b_{11}, \tilde{e}, \tilde{h}, I_S(\rho)) \leq V(a+1, b_{10}, \tilde{e}, \tilde{h}, I_S(\rho)) \leq V(a'+1, b_{10}, \tilde{e}, \tilde{h}, I_S(\rho))$, where the last inequality is due to the monotonicity of value function with respect to AoI state. Then, $Q^{11}(a',b,e,h,\rho) \leq Q^{10}(a',b,e,h,\rho)$ again by (17) and (18).

References

[1] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2011, pp. 350–358.

[2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, 2012, pp. 2731–2735.

[3] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Proc. 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012.

[4] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE Intern. Sympos. on Inf. Theory*, 2015, pp. 3008–3012.

[5] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. IEEE Inf. Theory and Application Workshop*, 2015, pp. 25–31.

[6] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. Green Commun. and Networking*, vol. 2, no. 1, pp. 193–204, 2018.

[7] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *arXiv:1806.07271*, 2018.

[8] Y. Chen, Q. Zhao, and A. Swami, "Distributed spectrum sensing and access in cognitive radio networks with energy constraint," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 783–797, 2009.

[9] A. Sultan, "Sensing and transmit energy optimization for an energy harvesting cognitive radio," *IEEE Commun. Lett.*, vol. 1, no. 5, pp. 500–503, 2012.

[10] S. Park, H. Kim, and D. Hong, "Cognitive radio networks with energy harvesting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1386–1397, 2013.

[11] B. Varan and A. Yener, "Online transmission policies for cognitive radio networks with energy harvesting secondary users," in *Proc. IEEE Wireless Commun. and Networking Conf.*, 2017.

[12] S. Park and D. Hong, "Optimal spectrum access for energy harvesting cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6166–6179, 2013.

[13] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1180–1189, 2012.

[14] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age of information and throughput in a shared access network with heterogeneous traffic," *arXiv preprint arXiv:1806.08776*, 2018.

[15] D. P. Bertsekas, *Dynamic programming and optimal control, Vol. I.* Athena scientific, 2005.

[16] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations research*, vol. 21, no. 5, pp. 1071–1088, 1973.