Age of Information Minimization for an Energy Harvesting Cognitive Radio

Shiyang Leng Aylin Yener
Wireless Communications and Networking Laboratory
Electrical Engineering Department, School of EECS
The Pennsylvania State University, University Park, PA 16802
sfl5154@psu.edu yener@ee.psu.edu

Abstract—Age of information (AoI) is a performance metric that measures the timeliness and freshness of information, and is particularly relevant in applications with time-sensitive data. This paper studies average AoI minimization in cognitive radio energy harvesting communications. More specifically, the system studied has a primary user with access rights to spectrum, and a secondary user who can utilize the spectrum only when it is left idle by the primary user. The secondary user is an energy harvesting sensor that harvests ambient energy with which it performs spectrum sensing and status updates of its sensing data to a destination. The status-updates are sent by opportunistically accessing the primary user’s spectrum. 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 with either perfect or imperfect spectrum sensing. The sequential decision problems are formulated as partially observable Markov decision processes and solved by dynamic programming for finite and infinite horizon. The properties of the optimal sensing and updating policies are investigated and shown to have threshold structure. Numerical results are presented to confirm the analytical findings.

Index Terms—Energy harvesting; cognitive radio; age of information; partially observable Markov decision processes.

I. INTRODUCTION

With the pervasive deployment of wireless nodes, timeliness of data delivery has become critical for various applications. Examples include 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. Maintaining information fresh in such scenarios brings about the need to consider a new network design metric. The novel concept of age of information (AoI) has been introduced to measure the freshness of information in [1], [2]. AoI quantifies the time elapsed since the generation of the latest successfully received update. Distinct from metrics of delay or latency, AoI thus captures the timeliness of the received information from the destination's perspective.

In early works on AoI, a queuing theoretic perspective has enabled the analysis and characterization of age [1]–[11]. 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) service is considered in [3] for an M/M/1 queue. Multiple sources are considered in [4]. Reference [5] has introduced peak age of information, where minimizing the peak age is of interest and managing packets by dropping or replacing is proposed in order to improve AoI. References [6] and [7] consider AoI in broadcast and multi-hop networks, respectively. Reference [8] introduces an AoI penalty function which generalizes linear and nonlinear models of age. Reference [9] considers nonlinear age. In [11], AoI of the primary user in a cognitive radio is investigated from a queuing theoretic perspective. In [12]–[14], user scheduling problems in broadcast networks are studied where the goal is to minimize the sum average AoI of the users. The problems are modeled as Markov decision processes (MDPs) in [12], [13]. In [15], a sampling decision for data generation is considered in addition to the updating decision, for which a constrained MDP is formulated to determine the optimal sampling and updating policy subject to an average energy cost constraint.

More recently, AoI has been investigated in energy harvesting systems, where each update consumes harvested energy [16]–[25]. 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 [16] has considered AoI minimization for point-to-point communication with energy harvesting constraints. By optimizing the inter-update time, it shows that waiting before updating improves AoI. Reference [17] considers that each update is generated, transmitted, and received instantly. The problem considered in [17] is to determine when to generate update with energy causality constraints. The offline knowledge of energy arrival is assumed and the optimal update policy is derived, which amounts to equalizing the inter-update time. Reference [18] studies the online setting for Poisson energy arrivals. A finite energy storage capacity and an erasure channel are considered. The long-term average age is analyzed by finding the age optimal threshold policy.

Copyright (C) 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. This work is supported by National Science Foundation Grants CNS-1526165 and ECCS-1748725. This paper is presented in part in IEEE WCNC 2019 and in IEEE ICC 2019.
recharge models for a finite battery. In the class of renewal policies, the optimal policy is proved to show an energy-dependent multi-threshold structure. Reference [23] studies the tradeoff between the achievable rate and AoI. References [24], [25] consider update failures due to the noisy channel.

The existing work on AoI minimization for energy harvesting communications assumes that the wireless channels for update transmissions are always available. By contrast, in cognitive-radio-based networks, a secondary user can only opportunistically access the primary spectrum when it is not occupied by a primary user. Energy harvesting cognitive radio networks (EH-CRN) have been studied previously with the throughput as the main metric [26]–[31]. In [27], [28], 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 partially observable Markov decision processes (POMDPs) subject to energy causality constraints. In particular, a collision constraint is also considered in [28]. In [29], the optimal detection threshold is derived to maximize the expected total throughput subject to the energy causality constraint and the collision constraint. Reference [30] characterizes the upper bound on the achievable throughput as a function of the energy arrival rate, the temporal correlation of the primary traffic, and the detection threshold for spectrum sensing. In [31], 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 [32], 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 wireless sensor networks, some cognitive nodes may not have exclusive spectrum available but may only operate as secondary users, whereas those sending data with higher priorities may be given unrestricted access to spectrum and be designated as primary users. For the secondary users that monitor the environment and opportunistically send time-sensitive status updates to the destination, AoI is of interest. Thus, in this paper, we consider an energy harvesting cognitive radio (EH-CR) with the objective of AoI minimization for the secondary user. The system consists of one primary user (PU) and an energy harvesting secondary user (SU). We consider that the harvested energy by SU is expended on spectrum sensing and status updating. Due to the randomness of energy harvesting and channel fading processes, the SU has to adaptively make sensing and updating 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 both perfect and imperfect spectrum sensing, we formulate POMDPs for sequential decision making to minimize the average AoI over a fixed time duration, and the long-term average AoI over an infinite horizon. The information state of the system is represented by the fully observable states of SU and the belief on PU, based on which sensing and update policies can be optimally determined using dynamic programming. We investigate the properties of the optimal policies and show that the optimal policy is one where the SU senses and updates when the harvested energy is larger than a threshold determined by battery and channel states, and the update policy has a threshold structure with respect to the state of age. In the numerical results, we demonstrate the threshold structure of the policy as well as the impact of system parameters on the system performance.

The remainder of the paper is organized as follows. In Section II, we introduce the system model of the EH-CR. In Section III, we formulate the finite-horizon POMDP, solve it by dynamic programming, and analyze the solution structure. In Section IV, we extend our investigation to the imperfect spectrum sensing case. In Section V, minimizing the long-term average AoI in an infinite horizon is considered for perfect and imperfect sensing. Section VI illustrates numerical results. Section VII concludes the paper.

II. SYSTEM MODEL

We consider a cognitive radio network with one primary user (PU) and one energy harvesting secondary user (SU) communicating with the corresponding primary receiver (PR) and secondary receiver (SR), respectively, as shown in Fig. 1. The SU has a wireless sensor that collects data and provides real-time status updates of this data to its destination. The SU sends the update data by accessing the primary spectrum opportunistically. We consider a time-slotted system with slots indexed by $t=0, 1, 2, \ldots$, normalized to the duration for one status-update data packet to be received.

A. Primary User Model

The PU has access rights to the spectrum. In each time slot, the PU either occupies the spectrum in an active (A) state or stays inactive (I), which forms a Markov chain [33]–[37]. The two-state (active/inactive) Markov chain model is commonly used for modeling PU activity [33], and has been verified to be an appropriate model to describe spectrum occupancy in the time domain [34]. In the time-slotted system considered in this paper, sensing decision is made at each slot, and the activity of PU can be captured with the unit the length of active/inactive period. Denote the state of the PU by $q_t \in \{A, I\}$. The transition probabilities of the two-state Markov chain are denoted by $p_{ii}$ and $p_{ai}$ for staying in inactive state and transitioning from the active state to inactive, respectively. That is, for $t=0,1,2,\ldots$, we have

\begin{align}
p_{ii} &\triangleq \mathbb{P}(q_{t+1} = I | q_t = I), \\
p_{ai} &\triangleq \mathbb{P}(q_{t+1} = I | q_t = A).
\end{align}
The transition probabilities are obtained by long-term measurements and are known to the SU.

B. Secondary User Model

The SU is slot-synchronized with the PU. At the beginning of each slot, the SU decides to either sense the channel or not. If it stays idle, no action is needed. If it decides to sense, it takes a fixed fraction of the slot to sense the PU’s spectrum. Given the sensing result, the SU needs to further decide whether to update or not. If it decides to do so, it takes the remainder of the slot for the update to be sent (if successful) received at the destination. That is, spectrum sensing and one update transmission are completed in one slot. The SU aims to minimize the average age of information by making optimal sensing and update decisions over time $t = 0, 1, 2, \ldots, T - 1$. We will consider both finite and infinite horizon formulations. 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 based on SU's states and its statistical knowledge of the primary spectrum availability as introduced below.

1) Belief Model: The SU observes the availability of the primary spectrum by opportunistically sensing and accessing the spectrum. Based on its action and observation history, a sufficient statistic, the belief state, of the primary spectrum availability can be obtained. Specifically, at each slot, if the SU decides to sense, an observation of the state of PU $\hat{q}_t$ can be obtained, denoted by $\hat{q}_t \in \{A, I\}$. We shall consider both perfect and imperfect sensing scenarios, where for perfect sensing the observation reveals the true state of PU, i.e., $\hat{q}_t = q_t$. For imperfect sensing, false alarm, i.e., declaring PU active when it is not, and miss detection, i.e., declaring PU inactive when it in reality is transmitting, events occur. At the beginning of slot $t$, the SU forms a belief $\rho_t$, which is the conditional probability of PU being inactive, i.e., $q_t = I$, given the action and observation history.

2) Channel Model: We consider that the SU transmits data over a block fading channel with channel gain $h_t$ for slot $t$, $\forall t$. $h_t$ is a discrete random variable with distribution $p_{H_t}(h_t = h_{t})$ over a finite sample space $\mathcal{H}_t$. $h_t$ is independently and identically distributed (i.i.d.) over slots. The distribution is known apriori 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, $h_t$ is randomly assigned according to the distribution.

3) Energy Harvesting Model: The SU is able to harvest energy from ambient sources 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 $e_t$, whose distribution $p_{p}(e_t = e_{t})$ is known. Its realization is $e_t \in \mathcal{E}$ with $\mathcal{E}$ a finite set. The energy harvesting process is i.i.d. across slots. The harvested energy is used on sensing the spectrum and on updating over the wireless channel. Let $\sigma$ be the energy consumption on sensing. The energy consumption for an update includes a fixed cost for generating an update data packet and a time-varying cost for transmission. Specifically, the transmission energy depends on the channel gain. We denote the energy cost for an update in slot $t$ by $u(h_t)$. The function $u(\cdot)$ is nonincreasing [15], [26]. In general, we have a smaller energy cost on sensing than on transmitting updates [26]. Without loss of generality, let $e_t$, $\sigma$, and $u(h_t)$ be integer multiples of unit energy. We have the battery state $b_t \in B \triangleq \{0, 1, \ldots, \bar{b}\}$, which evolves as

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

The energy causality constraint [32] is given by

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

The harvested energy is first stored in the battery and then used at the next slot onwards. That is, we have a store-then-use model.

Remark 1: Note that for energy harvesting and channel fading, we assume that the distributions of the random processes are known. Under this assumption, the i.i.d. processes adopted here can be readily extended to any other time-correlated random processes without any conceptual modifications to the methodology and the analysis derived for this model. We also remark that in practice the available energy for storage in each time slot depends on the current battery state $b_t$ and the energy conversion efficiency $\eta_t$ [38], i.e., $e_t = f(b_t, \eta_t)$. The energy conversion efficiency is further a nonlinear function of the energy receiving power [39]. We consider $f(\cdot)$ is known such that the harvested energy $e_t$ is deterministic given $b_t$ and $\eta_t$. The characterization of function $f(\cdot)$ is out of the scope of this paper and the reader is referred to [38], [39] and the references therein.

4) Age of Information: We adopt a linear model for AoI [1], [2], where AoI is defined as the time elapsed since the time instant when the most recently received update is generated. Let $\alpha_t$ 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 [16], [17], [21], [22], i.e., the data packet is generated when update decision is made. The amount of data is small enough that it is generated and transmitted instantaneously when spectrum sensing is completed, and received by the end of the slot. Taking into account the time for sensing, it spends one slot to receive one update. If update is successfully received, the AoI decreases to 1; otherwise increases by 1.

Note that $\alpha_t$ is upper bounded by $\hat{a}$ such that $\alpha_t \leq A \triangleq \{1, 2, \ldots, \hat{a}\}$, where $\hat{a} = a_0 + T$ for $T$ finite. For infinite horizon, AoI approaching $\hat{a}$ indicates that the information received at the destination is expired so that there is no need for counting. A sample path of AoI is depicted in Fig. 2 with $a_0 = 1$. We consider an error-free channel over which the data can be received successfully if transmitted over an unoccupied spectrum. Therefore, only collision with the PU, i.e., $x_t = (1, 1)$ and $q_t = A$, 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
\[
J = \frac{1}{T} \sum_{t=0}^{T-1} \left( \frac{1}{2} + a_t \right).
\] (5)

In the next two sections, we formulate and solve the AoI minimization as finite-horizon POMDP problems for perfect and imperfect spectrum sensing, respectively.

III. FINITE HORIZON POMDP WITH PERFECT SPECTRUM SENSING

A. POMDP Formulation

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

- **Actions**: The SU first makes the 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 inactive, then it further makes an update decision based on its AoI, energy availability, and the channel state information\(^2\). The action for each slot is \( x_t = (w_t, z_t) \) \( \in \mathcal{X} = \{(0, 0), (1, 0), (1, 1) : b_t \geq w_t \sigma + z_t u(h_t)\} \), where \( w_t \in \Gamma_w = \{0, 1 : b_t \geq w_t \sigma\} \) and \( z_t \in \Gamma_z = \{0, 1 : b_t \geq \sigma + z_t u(h_t)\} \).

- **Observations and beliefs**: The observation of the PU’s state is \( \hat{q}_t \in \{A, I\} \). The belief \( \rho_t \in [0, 1] \) is a conditional probability on the availability of primary spectrum. Based on the action and observation history, the belief evolves over slots by \( \rho_{t+1} = \Lambda(\rho_t) \) specified as follows. 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
\[
\rho_{t+1} = \begin{cases} 
\Lambda_0(\rho_t) = \rho_t p_{h_1} + (1-\rho_t) p_{\bar{h}_1}, & \text{if } w_t = 0, \\
\Lambda_A(\rho_t) = \rho_t p_{a_1}, & \text{if } w_t = 1, \hat{q}_t = A, \\
\Lambda_I(\rho_t) = \rho_t p_{\bar{a}_1}, & \text{if } w_t = 1, \hat{q}_t = I.
\end{cases}
\] (6)

- **Cost**: Let \( C(s_t) \) be the immediate cost under state \( s_t \), which is the accumulated AoI of slot \( t \), i.e., the area under the age curve of slot \( t \),
\[
C(s_t) = \frac{1}{2} + a_t, \quad \forall t.
\] (11)

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 (6). Thus, for a finite time session of length \( T \), the belief space \( \mathcal{I} \) is a finite set.

- **States**: The completely observable state of each slot consists of AoI state, battery state, energy harvesting state, and channel state, denoted by \( s_t = (a_t, b_t, e_t, h_t) \). Note that the state space, i.e., \( \mathcal{S} = \mathcal{A} \times \mathcal{B} \times \mathcal{E} \times \mathcal{H} \), is finite. Due to perfect sensing and transmission over an error-free channel, update is always successful when the sensing result is \( \hat{q}_t = I \) and update decision is \( z_t = 1 \). Thus, for \( t = 0, \ldots, T-1 \),
\[
a_{t+1} = \begin{cases} 
1, & \text{if } x_t = (1, 1) \\
a_t + 1, & \text{otherwise},
\end{cases}
\] (7)
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 \( p_t \). We denote the complete system state by \( s_t, (s_t, p_t) \), \( \forall t \in \mathcal{I} \). Since \( \mathcal{I} \) and \( \mathcal{S} \) are finite, the SU can only experience a finite number of possible system states \( (s_t, p_t) \in \mathcal{S} \times \mathcal{I} \).

- **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 \( p_x(s_{t+1}|s_t) \). Since the energy harvesting process and the channel fading process are i.i.d., we have
\[
p_x(s_{t+1}|s_t) = P(a_{t+1}|a_t, x_t) P(b_{t+1}|b_t, e_t, h_t, x_t) \cdot \quad p_{\mathcal{E}}(e_{t+1}) p_{\mathcal{H}}(h_{t+1}),
\] (8)
where
\[
P(a_{t+1}|a_t, x_t) = \begin{cases} 
1, & \text{if } a_{t+1} = (1 - z_t) a_t + 1, \\
0, & \text{otherwise},
\end{cases}
\] (9)
\[
P(b_{t+1}|b_t, e_t, h_t, x_t) = \begin{cases} 
1, & \text{if } b_{t+1} = \min\{b_t + e_t - w_t \sigma - z_t u(h_t)\}, \\
0, & \text{otherwise}.
\end{cases}
\] (10)

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

The POMDP can be written as a perfect state information problem by adopting the system state \((s, p) \in \mathcal{S} \times \mathcal{I} \) [40, 41]. Given SU’s initial state and belief, the finite-horizon average AoI under policy \( \pi \) is expressed as
\[
J^\pi(s_0, p_0) = \frac{1}{T} E \left[ \sum_{t=0}^{T-1} C(s_t)|s_0, p_0 \right],
\] (12)
where the expectation is taken over policy $\pi$. Finding the optimal sensing and update policy that minimizes the average AoI corresponds to solving
\[
\min_{\pi \in \Pi} J^\pi(s_0, \rho_0). \tag{13}
\]
Given $T$, (13) is a finite-state MDP with total cost.

### B. POMDP Solution

We use dynamic programming to solve the finite-horizon total cost minimization problem in (13) [41]. Let $V_t(s_t, \rho_t)$ denote the state-value function,
\[
V_t(s_t, \rho_t) = \min_{\pi_t \in \Pi_t} \mathbb{E} \left[ \sum_{i=t}^{T-1} C(s_i|s_t, \rho_t) \right], \tag{14}
\]
which is the minimum expected cost accumulated from slot $t$ to $T-1$ given $(s_t, \rho_t)$. Then, the minimum AoI in (13) is $J^* = V_0(s_0, \rho_0)/T$. 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 $(s_t, \rho_t)$ 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 $t = 0, 1, \ldots, T-1$,
\[
V_t(s_t, \rho_t) = \min_{w_t \in \Gamma_w} Q_t^{w_t}(s_t, \rho_t), \tag{15}
\]
where for $t = T-1$,
\[
\begin{align*}
Q_{T-1}^{0}(s_{T-1}, \rho_{T-1}) &= C(s_{T-1}) + C(s_T), \tag{16} \\
Q_{T-1}^{i}(s_{T-1}, \rho_{T-1}) &= (1 - \rho_{T-1})C(s_{T-1}) + \rho_{T-1} \min_{z_{T-1} \in \Gamma_z} C(s_{T-1}) + C(s_T), \tag{17}
\end{align*}
\]
and for $t = 0, \ldots, T-2$,
\[
\begin{align*}
Q_t^{0}(s_t, \rho_t) &= C(s_t) + \sum_{s_{t+1}} p_{0|s_t} v_{t+1}(s_{t+1}, \rho_0(\rho_t)), \tag{18} \\
Q_t^{1}(s_t, \rho_t) &= (1 - \rho_t)Q_t^{1A}(s_t, \rho_t) + \rho_t \min_{z_t \in \Gamma_z} Q_t^{1s}(s_t, \rho_t), \tag{19} \\
Q_t^{1A}(s_t, \rho_t) &= C(s_t) + \sum_{s_{t+1}} p_{10|s_t} v_{t+1}(s_{t+1}, \rho_1(\rho_t)), \tag{20} \\
Q_t^{1s}(s_t, \rho_t) &= C(s_t) + \sum_{s_{t+1}} p_{11|s_t} v_{t+1}(s_{t+1}, \rho_1(\rho_t)). \tag{21}
\end{align*}
\]
In particular, $Q_t^{1A}(s_t, \rho_t)$ in (20) denotes the conditional minimum expected cost given sensing result $\hat{q}_t = \Lambda$, i.e., adopting action $x_t = (1, 0)$. In (21) and (22), given sensing action $w_t = 1$ and sensing result $\hat{q}_t = I$, $Q_t^{1s}(s_t, \rho_t)$ and $Q_t^{1s}(s_t, \rho_t)$ characterize the conditional minimum expected costs by adopting update action $z_t = 0$ and $z_t = 1$, respectively. By recursion in (15)-(22), the optimal sensing and updating policies are obtained by
\[
\begin{align*}
w_t^*(s_t, \rho_t) &= \arg\min_{w_t \in \Gamma_w} Q_t^{w_t}(s_t, \rho_t), \tag{23} \\
z_t^*(s_t, \rho_t) &= \arg\min_{z_t \in \Gamma_z} Q_t^{1s}(s_t, \rho_t). \tag{24}
\end{align*}
\]

### C. Solution Structure

In this section, 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 system 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 $\rho_t$, for $p_{ii} \geq p_{ai}$.

The proof of Proposition 1 is provided in Appendix A.

It has been shown in [42] that for a finite and fixed time horizon POMDP, the value function is a piecewise linear, convex function 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 $\rho_t$, for $p_{ii} \geq p_{ai}$.

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

The proof of Theorem 1 is provided in 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., $z^*(a,b,e,h,\rho) = 1$.

The proof of Theorem 2 is provided in Appendix C.

**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 $a' \geq a$, $z^*(a',b,e,h,\rho) = 1$.

The proof of Theorem 3 is provided in Appendix D.

**Theorem 1** implies that 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 $b \geq \sigma + \bar{b} - e$, then an update decision is also made for any larger battery $b'$ as stated in **Theorem 2**. Similarly, the threshold structure for optimal update policy with respect to AoI state is stated in **Theorem 3**.

**Remark 2:** Note that accumulated AoI cost (11) is adopted in this paper. The solution derived here readily extends to the
formulation with a discrete-time AoI cost used in [12], [13], [15].

IV. FINITE HORIZON POMDP WITH IMPERFECT SPECTRUM SENSING

In this section, we consider the SU can make erroneous decisions when sensing the PU activity. Let \( p_t \) denote the probability of false alarm, namely, the probability of deciding the PU is active when it is not. The probability of detecting a PU when it is active, i.e., the probability of detection, is denoted by \( p_d \).

\[
p_t \triangleq \mathbb{P}(\hat{q}_t = 1 | q_t = 0), \quad \forall t,
\]
\[
p_d \triangleq \mathbb{P}(\hat{q}_t = 1 | q_t = 1), \quad \forall t.
\]

Based on the observation the SU gets from sensing the PU activity, it will take one of two actions. If the PU is sensed to be active, the SU will not update its belief, namely, this is indeed the case or in the event of a false alarm. Then, the belief is updated solely based on this sensing decision. When the PU is sensed to be inactive, the SU needs to make a decision whether to update. If an update is transmitted, the SU will receive a 1-bit feedback signal from the destination as to whether the update is successful or not, which is energy-cost negligible. The update is successful when the sensing result \( q_t = 1 \) is correct; this happens with probability \( 1 - p_t \). Update failure occurs if the PU is active despite SU declaring it inactive, which leads to a transmission collision between the PU and the SU; this happens with probability \( 1 - p_d \). In particular, the belief is updated according to the following cases.

- **Case 1:** If the SU stays idle without sensing, the belief is updated as
  \[
  \rho_{t+1} = \Lambda_0(\rho_t) = \rho_t p_{ii} + (1 - \rho_t) p_{ai}.
  \]
- **Case 2:** If the PU is sensed to be active, the SU does not update. The new belief is
  \[
  \rho_{t+1} = \Lambda_1(\rho_t) = \theta_t p_{ii} + (1 - \theta_t) p_{ai}, \quad \text{where}
  \theta_t \triangleq \mathbb{P}(q_t = 1 | \hat{q}_t = 1) = \frac{\rho_t p_{ii}}{\rho_t p_{ii} + (1 - \rho_t) p_{ai}}.
  \]
- **Case 3:** When the PU is sensed to be inactive and the SU decides not to update. The new belief is given by
  \[
  \rho_{t+1} = \Lambda_{\bar{1}}(\rho_t) = \bar{\theta}_t p_{ii} + (1 - \bar{\theta}_t) p_{ai}, \quad \text{where}
  \bar{\theta}_t \triangleq \mathbb{P}(q_t = 1 | \hat{q}_t = 0) = \frac{\rho_t (1 - p_t)}{\rho_t (1 - p_t) + (1 - \rho_t) (1 - p_d)}.
  \]
- **Case 4:** If the PU is sensed to be inactive, i.e., \( \hat{q}_t = 1 \), and the SU updates successfully, the sensing result correctly indicates the true state of PU, i.e., \( q_t = 1 \). We have
  \[
  \rho_{t+1} = \Lambda_I(\rho_t) = p_{ii}.
  \]
- **Case 5:** If the PU is sensed to be inactive, i.e., \( \hat{q}_t = 1 \), and the SU fails to update, a miss event has occurred in spectrum sensing and the true state is \( q_t = 1 \). We have
  \[
  \rho_{t+1} = \Lambda_A(\rho_t) = p_{ai}.
  \]

The transition probabilities for taking actions other than \( x_t = (1, 1) \) are given as the same in (8)-(10). For \( x_t = (1, 1) \), the transition probability is obtained by taking into account imperfect sensing results.

\[
p_{x_t}(s_{t+1}|s_t, \hat{q}_t, q_t) = \mathbb{P}(a_{t+1}|a_t, x_t, \hat{q}_t, q_t) \mathbb{P}(h_{t+1}|b_t, e_t, h_t, x_t) \cdot p_{\rho}(e_{t+1}|p_{\rho})(h_{t+1}),
\]

where \( \mathbb{P}(h_{t+1}|b_t, e_t, h_t, x_t) \) is given in (10) and \( \mathbb{P}(a_{t+1}|a_t, x_t, \hat{q}_t, q_t) \) is specified as follows.

\[
\mathbb{P}(a_{t+1}|a_t, (1, 1), \hat{q}_t, q_t) = \begin{cases} \hat{\theta}_t, & \text{if } \hat{q}_t = q_t, a_{t+1} = 1, \\ 1 - \hat{\theta}_t, & \text{if } \hat{q}_t \neq q_t, a_{t+1} = a_t + 1, \\ 0, & \text{otherwise.} \end{cases}
\]

where \( \hat{\theta}_t \) is given in (29).

Similar to perfect spectrum sensing, the POMDP can be written as a perfect state information problem with finite states since at each slot the belief can only transit to a finite number of possible new beliefs such that \( \mathcal{I} \) is finite. Again, dynamic programming is used to solve (13). We have the recursion equations as follows. For \( t = 0, 1, \ldots, T - 1 \),

\[
V_i(s_t, \rho_t) = \min_{w_i \in \mathcal{I}_i} Q_{\nu_i}^w(s_t, \rho_t),
\]

where for \( t = T - 1 \),

\[
Q^0_{T-1}(s_{T-1}, \rho_{T-1}) = C(s_{T-1}) + C(s_T),
\]

\[
Q^1_{T-1}(s_{T-1}, \rho_{T-1}) = (1 - \eta_T) C(s_{T-1}) + \eta_T \min_{z_{T-1} \in \mathcal{Z}_{T-1}} C(s_{T-1}) + C(s_T),
\]

and for \( t = 0, \ldots, T - 2 \),

\[
Q^0_t(s_t, \rho_t) = C(s_t) + \sum_{s_{t+1}} p_{00}(s_{t+1}|s_t) V_{t+1}(s_{t+1}, \Lambda_0(\rho_t)),
\]

\[
Q^1_t(s_t, \rho_t) = (1 - \eta_t) Q^1_{t+1}(s_t, \rho_t) + \eta_t \min_{z_t \in \mathcal{Z}_t} Q^1_{t+1}(s_t, \rho_t),
\]

\[
Q^{1A}_t(s_t, \rho_t) = C(s_t) + \sum_{s_{t+1}} p_{10}(s_{t+1}|s_t) V_{t+1}(s_{t+1}, \Lambda_1(\rho_t)),
\]

\[
Q^{10}_t(s_t, \rho_t) = C(s_t) + \sum_{s_{t+1}} p_{11}(s_{t+1}|s_t) V_{t+1}(s_{t+1}, \Lambda_1(\rho_t)),
\]

\[
Q^{11}_t(s_t, \rho_t) = C(s_t) + \sum_{s_{t+1}} p_{11}(s_{t+1}|s_t, \hat{q}_t = q_t) V_{t+1}(s_{t+1}, \Lambda_1(\rho_t))
\]

\[
+ \sum_{s_{t+1}} p_{11}(s_{t+1}|s_t, \hat{q}_t \neq q_t) V_{t+1}(s_{t+1}, \Lambda_A(\rho_t)),
\]

where \( \eta_t \) in (36) and (38) denotes the probability of observing PU inactive, that is,

\[
\eta_t \triangleq \mathbb{P}(\hat{q}_t = 1) = \rho_t (1 - p_t) + (1 - \rho_t) (1 - p_d).
\]

The optimal sensing and updating policies are given by (23) and (24).

The monotonicity of the value function with respect to state components \( a_t, b_t, e_t, \) and \( h_t, \forall t \), stated in Proposition 1 holds for imperfect sensing as well. In particular, the value function
is nonincreasing in belief $\rho_t$, $\forall t$, if the transition probabilities of the state of PU given in (1), (2) and the probabilities of false alarm and detection events given in (25), (26) satisfy $\frac{\lambda_{ui}}{p_{ai}} \geq \frac{\lambda_{di}}{p_{ai}} \geq 1$. The value function with respect to the belief is concave. The optimal policy structure stated in Theorem 1 and Theorem 2 also hold. The proofs are provided in Appendix E.

V. INFINITE HORIZON POMDP

In this section, we consider an infinite-horizon POMDP for the long-term average AoI minimization. For the same setting for perfect and imperfect spectrum sensing described in Section III and IV, respectively, the long-term average AoI under policy $\pi$ is given by

$$J^\pi(s_0, \rho_0) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} C(s_t)|s_0, \rho_0 \right].$$ (43)

For an infinite horizon, we focus on the set of deterministic stationary policies $\Pi^s$ that satisfies the energy causality constraint, where $\pi = \{\mu_0, \mu_1, \ldots\} \in \Pi^s$ such that $\mu_{t_1} = \mu_{t_2}$ when $(s_{t_1}, \rho_{t_1}) = (s_{t_2}, \rho_{t_2})$ for any $t_1, t_2$. Thus, we omit the time index in the sequel. The goal is to find an optimal sensing and update policy that solves the long-term average AoI minimization problem as follows.

$$\min_{\pi \in \Pi^s} J^\pi(s_0, \rho_0).$$ (44)

Distinct from the finite-horizon formulation, where a finite set of system states can be used to fully characterize the decision making process, the infinite-horizon POMDP has a countably infinite set of beliefs $\mathcal{I}$ leading to a countably infinite set of system states. Based on [43, Theorem 4.2], next we prove that a solution exists for the POMDP with average cost formulated in (44).

**Theorem 4:** There exists $(J^*, G(s, \rho))$ that satisfies the Bellman equation

$$J^* + G(s, \rho) = \min_{w \in \mathcal{W}} Q^w(s, \rho), \forall (s, \rho) \in \mathcal{S} \times \mathcal{I},$$ (45)

where $J^*$ is the optimal average cost which is a constant for all $(s, \rho) \in \mathcal{S} \times \mathcal{I}$, $G(s, \rho)$ is the relative value function defined in (49), and $Q^w(s, \rho)$ is the Q-function for taking sensing action $w$, which is given in (51) and (52). The optimal policy $\pi^*$ exists and is obtained by

$$w^*(s, \rho) \in \arg\min_{w \in \mathcal{W}} Q^w(s, \rho),$$ (46)

$$z^*(s, \rho) \in \arg\min_{z \in \mathcal{I}} Q^z(s, \rho),$$ (47)

where $Q^z(s, \rho)$ is the Q-function for taking action $(1, z)$ as given in (51) and (52). Furthermore, for $\beta \in (0, 1)$, we have $(1 - \beta) V_{\beta}(s, \rho) \geq J^\pi$, where $V_{\beta}(s, \rho)$ is the value function of the corresponding discounted cost problem with objective

$$J^\beta(s_0, \rho_0) = \lim_{T \to \infty} \mathbb{E} \left[ \sum_{t=0}^{T-1} \beta^t C(s_t)|s_0, \rho_0 \right],$$ (48)

and $V_{\beta}(s, \rho) \leq \min_{\pi \in \Pi^s} J^\pi(s, \rho)$.

The relative value function is defined as

$$G(s, \rho) \triangleq \hat{V}(s, \rho) - \tilde{V}(s^0, \rho^0),$$ (49)

where $(s^0, \rho^0) \in \mathcal{S} \times \mathcal{I}$ is a reference system state, and $\hat{V}(s, \rho)$ is computed as

$$\hat{V}(s, \rho) = \min_{w \in \mathcal{W}} Q^w(s, \rho).$$ (50)

For perfect spectrum sensing, the Q-functions are given by

$$Q^0(s, \rho) = C(s) + \sum_{s'} p_{00}(s'|s) G(s', \Lambda_0(\rho)),$$ (51a)

$$Q^1(s, \rho) = (1 - \eta(\rho))Q^1A(s, \rho) + \eta(\rho) \min_{z \in \mathcal{I}} Q^1z(s, \rho),$$ (51b)

$$Q^1A(s, \rho) = C(s) + \sum_{s'} p_{10}(s'|s) G(s', \Lambda_A(\rho)),$$ (51c)

$$Q^0(s, \rho) = C(s) + \sum_{s'} p_{10}(s'|s) G(s', \Lambda_1(\rho)),$$ (51d)

$$Q^1(s, \rho) = C(s) + \sum_{s'} p_{11}(s'|s, \tilde{q} = q) G(s', \Lambda_1(\rho)),$$ (51e)

For imperfect spectrum sensing, the Q-functions are listed as follows.

$$Q^0(s, \rho) = C(s) + \sum_{s'} p_{00}(s'|s) G(s', \Lambda_0(\rho)),$$ (52a)

$$Q^1(s, \rho) = (1 - \eta(\rho))Q^1A(s, \rho) + \eta(\rho) \min_{z \in \mathcal{I}} Q^1z(s, \rho),$$ (52b)

$$Q^1A(s, \rho) = C(s) + \sum_{s'} p_{10}(s'|s) G(s', \Lambda_A(\rho)),$$ (52c)

$$Q^0(s, \rho) = C(s) + \sum_{s'} p_{10}(s'|s) G(s', \Lambda_1(\rho)),$$ (52d)

$$Q^1(s, \rho) = C(s) + \sum_{s'} p_{11}(s'|s, \tilde{q} = q) G(s', \Lambda_1(\rho)),$$ (52e)

where $\eta(\rho) = \rho (1 - p_f) + (1 - \rho)(1 - p_a)$.

To solve (45) for $(J^*, G(s, \rho))$ and the optimal policy, we apply the relative value iteration algorithm [41]. Since the set of belief states $\mathcal{I}$ is countably infinite, we approximate it by a finite set $\bar{\mathcal{I}}$ for a given initial belief $\rho_0$. The algorithm is summarized in Algorithm 1. For perfect sensing, beginning with initial belief $\rho_0$, the belief can be updated to $p_{ai}, p_{ai}$, and $\Lambda_0(\rho_i)$ for slot $t + 1$. As $t \to \infty$, $\Lambda_0(\rho_i)$ converges to the stationary probability of the spectrum availability, i.e., $\bar{p} = p_{ai}/(1 - p_{ai} + p_{ai})$. Thus, $\bar{\mathcal{I}}$ can be obtained by including $p_{ai}, p_{ai}, \rho_0, \bar{p}$ and all the intermediate beliefs evolved from $\rho_0, p_{ai}, p_{ai}$ to $\bar{p}$. Similarly, $\bar{\mathcal{I}}$ can be obtained for imperfect sensing.

The optimal policies for perfect and imperfect spectrum sensing infinite-horizon POMDPs demonstrate the same properties as finite-horizon settings for perfect and imperfect spectrum sensing, respectively. Considering the POMDP with discounted cost (48), the value function $V_{\beta}(s, \rho)$ satisfies the following Bellman equation [43, Theorem 2.1]

$$V_{\beta}(s, \rho) = \min_{w \in \mathcal{W}} Q^w_{\beta}(s, \rho),$$ (53)

where $Q^{\beta}_{\beta}(s, \rho)$ is the corresponding Q-function with discount $\beta$, which is the sum of the immediate cost and the $\beta$-
Algorithm 1: Relative value iteration algorithm

1: For all \((s, \rho) \in S \times \tilde{I}\), initialize \(\tilde{V}_0(s, \rho) = 0\), choose \((s^0, \rho^0)\), set \(G_0(s, \rho) = \tilde{V}_0(s, \rho) - \tilde{V}_0(s^0, \rho^0)\) and \(k = 0\).

2: repeat
3: for \((s, \rho) \in S \times \tilde{I}\) do
4: compute \(\tilde{V}_{k+1}(s, \rho)\) by (50) and (51)/(52) using \(G_k(s, \rho)\).
5: Let \(G_{k+1}(s, \rho) = \tilde{V}_{k+1}(s, \rho) - \tilde{V}_{k+1}(s^0, \rho^0)\).
6: end for
7: until \(G_k(s, \rho) \rightarrow G(s, \rho)\) for all \((s, \rho)\), otherwise increase \(k\) by 1.
8: \(J^* = \tilde{V}(s^0, \rho^0)\), optimal policy is obtained by (46).

In this section, we present numerical results for perfect and imperfect spectrum sensing to verify our findings. The PU has state transition probabilities \(p_{ai} = 0.8\) and \(p_{ai} = 0.2\). The probability of detecting an active PU is \(p_d = 0.8\) for the imperfect sensing scenario. 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 gains, we set \(\mathcal{H} = \{h_1, h_2\}\) with \(p_h\) as the probability of observing \(h_1\), where each level corresponds to an energy cost on update. We set \(u(h_1) = 2\) and \(u(h_2) = 4\), i.e., \(h_1\) is a better channel than \(h_2\) and requires half the energy to transmit an update. 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. 3, we plot two sample paths of AoI by optimal policy for perfect sensing with \(b = 5\), \(p_e = 0.3\), \(p_h = 0.5\).

VI. NUMERICAL RESULTS

In this section, we present numerical results for perfect and imperfect spectrum sensing to verify our findings. The PU has state transition probabilities \(p_{ai} = 0.8\) and \(p_{ai} = 0.2\). The probability of detecting an active PU is \(p_d = 0.8\) for the imperfect sensing scenario. 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 gains, we set \(\mathcal{H} = \{h_1, h_2\}\) with \(p_h\) as the probability of observing \(h_1\), where each level corresponds to an energy cost on update. We set \(u(h_1) = 2\) and \(u(h_2) = 4\), i.e., \(h_1\) is a better channel than \(h_2\) and requires half the energy to transmit an update. 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. 3, we plot two sample paths of AoI by the optimal policy in perfect sensing case with battery capacity \(b = 5\) and initial belief \(\rho_0 = p_{ai}\). Table I provides the corresponding system 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 fill up the battery at the 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. 4 and 5 show the average AoI versus the probability of energy arrival \(p_e\) for \(T = 50\), \(p_h = 0.5\), \(a_0 = 1\).
of randomness in the energy state by various sensing and update decisions. We also observe that, as expected, a larger battery is present, the average AoI is lower.

Similar performance insights hold for the impact of the channel quality on average AoI, as shown in Fig. 6 and Fig. 7 for finite horizon ($T = 50$) and infinite horizon ($T = 10000$), respectively. As the probability of experiencing the better channel gain increases, the average AoI decreases since less energy is consumed for each update.

For imperfect spectrum sensing, Fig. 8 presents the average AoI versus the probability of false alarm for finite horizon and infinite horizon. As the probability of false alarm increases, the average AoI becomes larger since the SU observes the spectrum to be occupied and decides not to update due to this sensing error. The optimal policy outperforms the myopic policy significantly.

VII. CONCLUSION

In this paper, we have investigated an energy harvesting cognitive secondary user, e.g., an energy harvesting 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 finite and infinite horizon are considered. Taking into account the partially observable state of the primary user, POMDP is adopted to formulate the average AoI minimization problem subject to the energy causality constraint. For perfect and imperfect spectrum sensing, the POMDPs are formulated as perfect state information problems, which are solved by dynamic programming. The monotonicity of the value function and the threshold structure of the optimal policy are shown. Numerical results illustrate the policy structures, highlight the impact of energy harvesting system parameters, and demonstrate that optimal policies significantly outperform myopic policies. Future work includes continuous time system models and optimizing general forms of age of information, as well as systems with multiple secondary and primary users. Considering models where secondary users harvest energy from the primary signals is an interesting future direction, as in these, the transmit power of the primary users will impact the energy state of the secondary users and their age of information. Another interesting direction is when update data availability is stochastic in nature.
Appendix A

Proof of Proposition 1

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

(1) Nondecreasing in $a$: We show that $V_t(a') \geq V_t(a_i)$ for $a'_t \geq a_i$ by induction according to the recursion in (15)-(22).

For $t = T - 1$, by (16) and (17), $Q^{Q_T-1}(a_{T-1}) = C(a_{T-1}) + C(a'_T) = C(a_{T-1}) + C(1 - z_{T-1})a_{T-1} + 1 \geq Q^{Q_T-1}(a_{T-1})$ for $a'_{T-1} \geq a_{T-1}$. Since $\pi$ preserves the monotonicity, $V_{T-1}(a'_{T-1}) \geq V_{T-1}(a_{T-1})$ from (15).

Suppose $V_{t+1}(a'_{t+1}) \geq V_{t+1}(a_{t+1})$ for some $t$, we next show $V_t(a') \geq V_t(a_i)$. From (18), $Q^0_t(a'_t) \geq Q^0_t(a_i)$ holds as $C(a'_t) \geq C(a_i)$ and $V_{t+1}(a'_{t+1}) \geq V_{t+1}(a_{t+1})$. Similarly, we have $Q^1_t(a'_t) \geq Q^1_t(a_i)$, $Q^{10}_t(a'_t) \geq Q^{10}_t(a_i)$, and $Q^{11}_t(a'_t) \geq Q^{11}_t(a_i)$. Then, $Q^1_t(a'_t) \geq Q^1_t(a_i)$ from (19). Consequently, $V_t(a'_t) \geq V_t(a_i)$ from (15).

(2) Nonincreasing in $b, c,$ 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, residual energy can be kept in the battery to provide a lower value function.

(3) Nonincreasing in $\rho$: We show that $V_t(\rho'_t) \leq V_t(\rho_t)$ for $\rho'_t \leq \rho_t$ by induction according to the recursion in (15)-(22).

For $t = T - 1$, if the secondary user stays idle, we have $Q^{Q_T-1}(\rho'_{T-1}) = Q^{Q_T-1}(\rho_{T-1})$ from (16). If sensing, from (17),

$$Q^1_{T-1}(\rho'_{T-1}) = \begin{cases} C(a_{T-1}) + C(a_{T-1} + 1), & \text{if } z_{T-1} = 0 \\ C(a_{T-1}) + C(1), & \text{if } z_{T-1} = 1. \end{cases}$$

(Since the update policy does not depend on the belief, $Q^1_{T-1}(\rho'_{T-1}) \leq Q^1_{T-1}(\rho_{T-1})$. Thus, $V_{T-1}(\rho'_{T-1}) \leq V_{T-1}(\rho_{T-1})$ by (15).

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 (6), $A_0(\rho_t)$ is nondecreasing in $\rho_t$ as $p_{ii} \geq p_{ai}$. Then, $V_{t+1}(A_0(\rho_t')) \leq V_{t+1}(A_0(\rho_t))$ for $\rho'_t \geq \rho_t$ by assumption. This implies $Q_t^0(\rho'_t) \leq Q_t^0(\rho_t)$ according to (18). By similar argument, it can be verified that $Q_t^1(\rho'_t) \leq Q_t^1(\rho_t)$, $Q_t^{10}(\rho'_t) \leq Q_t^{10}(\rho_t)$, and $Q_t^{11}(\rho'_t) \leq Q_t^{11}(\rho_t)$. Then, for $Q_t^1(\rho'_t)$ given in (19),

$$Q^1_t(\rho'_t) \leq (1 - \rho_t)Q^1_t(\rho_t) + \rho_t \min_{z_{t+1} \in z} \left\{ Q_t^{10}(\rho_t), Q_t^{11}(\rho_t) \right\} = Q^1_t(\rho_t) + \rho_t \Delta Q_t(\rho_t).$$

$\Delta Q_t(\rho_t) \triangleq \min_{z_{t+1} \in z} \left\{ Q_t^{10}(\rho_t), Q_t^{11}(\rho_t) \right\} - Q^1_t(\rho_t) \leq Q_t^{10}(\rho_t) - Q^1_t(\rho_t) \leq 0.\quad (56)$

The nonpositivity is by (21) and (20), where $V_{t+1}(\Lambda_1(\rho_t)) \leq V_{t+1}(A_0(\rho_t))$ holds by assumption for $\Lambda_1(\rho_t) = \rho_t \geq \Lambda_0(\rho_t) = \rho_{ei}$. Therefore, from (55), $Q^1_t(\rho'_t) \leq Q^1_t(\rho_t) + \rho_t \Delta Q_t(\rho_t) \leq Q^1_t(\rho_t) + \rho_t \Delta Q_t(\rho_t) = Q^1_t(\rho_t)$. By (15), we conclude $V_t(\rho'_t) \leq V_t(\rho_t)$.

Appendix B

Proof of Theorem 1

Let $C$ denote $C(s)$ if $s$ is not changed. To prove $w(b, \rho) = 1$, we need to show that $Q^0(b, \rho) \geq Q^1(b, \rho)$. Since $e \geq e + b - b$, 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 (19),

$$Q^1(b, \rho) \leq (1 - \rho)Q^A(b, \rho) + \rho Q^{10}(b, \rho)$$

$$= (1 - \rho)[C + \sum_{s'} p_{10}(s'|s)V(b', \Lambda_{1}(\rho))] + \rho[C + \sum_{s'} p_{10}(s'|s)V(b', \Lambda_{1}(\rho))]$$

$$\leq C + \sum_{s'} p_{10}(s'|s)V(\bar{b}, (1 - \rho)\Lambda_{1}(\rho) + \Lambda_{1}(\rho))$$

$$= C + \sum_{s'} p_{10}(s'|s)V(\bar{b}, \Lambda_{0}(\rho))$$

$$= C + \sum_{s''} p_{00}(s''|s)V(b'', \Lambda_{0}(\rho))$$

$$= Q^0(b, \rho),$$

(57)

where (1) is by the concavity of value function with respect to the belief, and (2) is from the belief update equation in (6).

Appendix C

Proof of Theorem 2

First we show that when $e \geq \sigma + b - b$, 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 $e \geq \sigma + b - b$, the new battery state becomes $b = \min\{b - \sigma + e, \bar{b}\} = \bar{b}$ if solely sensing. By (22),

$$Q^{11}(b', \rho) = C + \sum_{s'} p_{11}(s'|s)V(\min\{b' - \sigma - u(h) + e, \bar{b}\}, \Lambda_{1}(\rho))$$

$$\leq C + \sum_{s'} p_{11}(s'|s)V(\min\{b - \sigma - u(h) + e, \bar{b}\}, \Lambda_{1}(\rho))$$

(1)

$$= C + \sum_{s'} p_{10}(s'|s)V(\bar{b}, \Lambda_{1}(\rho))$$

(2)

$$= Q^{10}(b', \rho),$$

(58)

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 $e \geq \sigma + u(h) + b - b$, $z^*(a, b, e, h, \rho) = 1$. By Theorem 1, sensing action is taken, i.e., $w^*(a, b, e, h, \rho) = 1$. Thus, we only need to show $Q^{10}(a, b, e, h, \rho) \geq Q^{11}(a, b, e, h, \rho)$. The new battery state becomes $b' = \bar{b}$ if update is transmitted. By (22),

$$Q^{11}(a, b, e, h, \rho) = C + \sum_{s'} p_{11}(s'|s)V(\bar{b}, e', h', \Lambda_{1}(\rho))$$

$$\leq C + \sum_{s''} p_{10}(s''|s)V(a + 1, \bar{b}, e', h', \Lambda_{1}(\rho))$$

$$= Q^{10}(a, b, e, h, \rho),$$

(59)
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 (24), \( z(a, b, e, h, \rho) = 1 \) implies that \( Q^{11}(a, b, e, h, \rho) \leq Q^{10}(a, b, e, h, \rho) \). That is, by (21) and (22),

\[
C + \sum_{\tilde{s}} p_{11}(\tilde{s}|s)V(1, b_{11}, \tilde{e}, \tilde{h}, \Lambda_I(\rho)) \\
\leq C + \sum_{\tilde{s}} p_{10}(\tilde{s}|s)V(a + 1, b_{10}, \tilde{e}, \tilde{h}, \Lambda_I(\rho)), \quad (60)
\]

where \( b_{11} = \min\{b - \sigma + u(h) + e, \tilde{b}\} \), \( b_{10} = \min\{b - \sigma + e, \tilde{b}\} \), and \( \tilde{e} \) and \( \tilde{h} \) are the energy harvesting and channel states of the next slot. Thus, \( V(1, b_{11}, \tilde{e}, \tilde{h}, \Lambda_I(\rho)) \leq V(a + 1, b_{10}, \tilde{e}, \tilde{h}, \Lambda_I(\rho)) \), 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 (21) and (22).

APPENDIX E
PROOFS FOR IMPERFECT SPECTRUM SENSING

We prove that Proposition 1, Theorem 1, and Theorem 2 hold for imperfect spectrum sensing.

**Proposition 1:** The monotonicity of the value function with respect to \( a, b, e, h \) can be proved by the same argument as in Appendix A. Here, by induction we show that the nonincreasing in belief \( \rho \) holds, i.e., \( V_t(\rho'_t) \leq V_t(\rho_t) \) for \( \rho'_t \geq \rho_t \) if \( \frac{p_{a_i}}{p_{a_i}} \geq \frac{p_{a_i}}{p_{a_i}} \geq 1 \).

For \( t = T - 1 \), the same argument in Appendix A holds for \( V_{T-1}(\rho'_{T-1}) \leq V_{T-1}(\rho_{T-1}) \). Suppose \( V_{t+1}(\rho'_{t+1}) \leq V_{t+1}(\rho_{t+1}) \) for \( \rho'_t \geq \rho_t \), we next show \( V_t(\rho'_t) \leq V_t(\rho_t) \). From (27), \( \Lambda_0(\rho) \) is nondecreasing in \( \rho \) as \( p_{a_i} \geq p_{a_i} \). Then, \( V_{t+1}(\Lambda_0(\rho'_t)) \leq V_{t+1}(\Lambda_0(\rho_t)) \) for \( \rho'_t \geq \rho_t \) by assumption. This implies \( Q^1(\rho'_t) \leq Q^1(\rho_t) \) according to (37). Similarly, it can be easily verify from (28) and (29) that \( \Lambda_1(\rho) \) and \( \Lambda_1(\rho) \) are nondecreasing in \( \rho \), as well as \( \Lambda_1(\rho) \geq \Lambda_1(\rho) \) due to \( \frac{p_{a_i}}{p_{a_i}} \geq \frac{p_{a_i}}{p_{a_i}} \geq 1 \). Then from (39)-(41), \( Q^1(\rho'_t) \leq Q^1(\rho_t) \), \( Q^1(\rho'_t) \leq Q^1(\rho_t) \), and \( Q^1(\rho'_t) \leq Q^1(\rho_t) \). Thus, the same argument as in Appendix A follows.

**Theorem 1:** Applying similar argument as in Appendix B, we need to show \( Q^1(\rho, \beta) \leq Q^0(\rho, \beta) \).

\[
Q^1(\rho, \beta) \\
\leq (1 - \eta(\rho))Q^1(\rho, \beta) + \eta(\rho)Q^0(\beta, \rho) \\
\leq C + \sum_{s'} p_{10}(s'|s)V(b, (1 - \eta(\rho))\Lambda_1(\rho) + \eta(\rho)\Lambda_1(\rho)) \\
\leq C + \sum_{s'} p_{10}(s'|s)V(b, \Lambda_0(\rho)) \\
= C + \sum_{s''} p_{00}(s''|s)V(b, \Lambda_0(\rho)) \\
= Q^0(\rho, \beta), \quad (61)
\]

where (1) is by the concavity of value function with respect to the belief, and (2) is from the belief update equations in (27)-(29).

**Theorem 2:** Based on the recursion equations in (40) and (41), the threshold structure of update policy for \( e \geq \sigma + b - b \) can be proved by the same procedure as in Appendix C. Now, we prove that when \( e \geq \sigma + u(h) + b - b, z^*(a, b, e, h, \rho) = 1 \). By the recursion equations in (40) and (41),

\[
Q^{11}(a, b, e, h, \rho) \\
= C + \sum_{s''} p_e(e'|s')V(h') \left[ \tilde{h}V(a, e', h', \Lambda_1(\rho)) \right] \\
\leq C + \sum_{s''} p_e(e'|s')V(h') \left[ \tilde{h}V(a, e', h', \Lambda_1(\rho)) \right] \\
\leq C + \sum_{s''} p_e(e'|s')V(h')V(a + 1, e', h', \Lambda_1(\rho)) \\
\leq C + \sum_{s''} p_e(e'|s')V(h')V(a + 1, e', h', \Lambda_1(\rho)) \\
= C + \sum_{s''} p_{00}(s''|s)V(a + 1, h', \Lambda_1(\rho)) \\
\leq Q^0(\rho, \beta), \quad (62)
\]

where (1) is since the value function is nondecreasing in AoI state, (2) is due to the concavity of value function, and (3) is from (29)-(31).

APPENDIX F
PROOF OF THEOREM 4

According to [43, Theorem 4.2], it suffices to show that the following two conditions are satisfied: (i) \( \Lambda^{-1}_i(\rho), \forall i \in \{0, 1, 4, 1, 5, 6, 7, 8\} \) is a countable set; (ii) there is a constant \( L \geq 0 \) such that \( |V_\beta(s, \rho) - V_\beta(s', \rho')| \leq L, \forall 0 < \beta < 1, \forall (s, \rho), (s', \rho') \in S \times \mathcal{I} \).

For (i), the condition holds if \( \Lambda^{-1}_i(\rho) \) is an injective map. Since \( \frac{p_{a_i}}{p_{a_i}} \geq \frac{p_{a_i}}{p_{a_i}} \geq 1 \), the matrices \( \left( \frac{p_{a_i}}{p_{a_i}} \right) \) are nonsingular. Thus, \( \Lambda^{-1}_i(\rho) \) is an injective map based on [43, Lemma 4.2]. For (ii), consider a system state \( (\hat{s}, \hat{\rho}) = (\hat{a}, 0, 0, \hat{h}_{\min}, 0) \), where \( \hat{a} \) is the upper bound of AoI and \( h_{\min} \in \mathcal{H} \) is the worst channel level. Due to the monotonicity of \( V_\beta(s, \rho), 0 \leq V_\beta(s, \rho) \leq V_\beta(\hat{s}, \hat{\rho}) \) for any \( (s, \rho) \in S \times \mathcal{I} \). Then, it suffices to show that \( V_\beta(s, \rho) \) is no larger than a constant \( L \). For both perfect and imperfect sensing, \( \Lambda_0(0) = p_{a_i} \), then the Q-function with discount \( \beta \) can be written according to (51)-(52),

\[
Q^0(\rho, \beta) = C(\hat{a}) + \beta \sum_{s'} p_{00}(s'|s)V_\beta(s', \rho_{a_i}). \quad (63)
\]

Thus, \( V_\beta(s, \rho) \leq Q^0(\rho, \beta) \leq C(\hat{a}) + \beta V_\beta(\hat{s}, \hat{\rho}) \), which results in \( V_\beta(s, \rho) \leq C(\hat{a})/(1 - \beta) \leq (1 + \hat{a})/(1 - \beta) \). Then, \( L = (1 + \hat{a})/(1 - \beta) \).

REFERENCES

A. Kosta, N. Pappas, V. Angelakis, B. T. Bacinoglu and E. Uysal-Biyikoglu, “Scheduling status updates to
B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, “Age of informa-
I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and N. B. Shroff, “Age-optimal information
Aylin Yener (S’91–M’01–SM’14–F’15) received the B.Sc. degree in electrical and electronics engineering and the B.Sc. degree in physics from Bogazici University, Istanbul, Turkey, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, NJ, USA. She is a Distinguished Professor of Electrical Engineering at The Pennsylvania State University, University Park, PA, USA, where she joined the faculty as an assistant professor in 2002. Since 2017, she is also a Dean’s Fellow in the College of Engineering at The Pennsylvania State University. She was a visiting professor of Electrical Engineering at Stanford University in 2016-2018 and a visiting associate professor in the same department in 2008-2009. Her current research interests are in information security, green communications, caching systems, and more generally in the fields of information theory, communication theory and networked systems. She received the NSF CAREER Award in 2003, the Best Paper Award in Communication Theory from the IEEE International Conference on Communications in 2010, the Penn State Engineering Alumni Society (PSEAS) Outstanding Research Award in 2010, the IEEE Marconi Prize Paper Award in 2014, the PSEAS Premier Research Award in 2014, the Leonard A. Doggett Award for Outstanding Writing in Electrical Engineering at Penn State in 2014, and the IEEE Women in Communications Engineering Outstanding Achievement Award in 2018. She is a distinguished lecturer for the IEEE Information Theory Society (2019-2020), the IEEE Communications Society (2018-2020) and the IEEE Vehicular Technology Society (2017-2019).

Dr. Yener is serving as the vice president of the IEEE Information Theory Society in 2019. Previously she was the second vice president (2018), member of the Board of Governors (2015-2018) and the treasurer (2012-2014) of the IEEE Information Theory Society. She served as the Student Committee Chair for the IEEE Information Theory Society (2007-2011), and was the co-Founder of the Annual School of Information Theory in North America in 2008. She was a Technical (Co)-Chair for various symposia/tracks at the IEEE ICC, PIMRC, VTC, WCNC, and Asilomar in 2005, 2008-2014 and 2018. Previously, she served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS (2009-2012), an Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING (2017-2018), and an Editor and an Editorial Advisory Board Member for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2001-2012). She also served a Guest Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY in 2011, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2015. Currently, she serves as a Senior Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.