Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 129 (2018) 194–200

Procedia Computer Science

www.elsevier.com/locate/procedia

2017 International Conference on Identification, Information and Knowledge in the Internet of Things

Optimal Spectrum Sensing-Access Policy in Energy Harvesting Cognitive Radio Sensor Networks

Fan Zhang[a,b], Tao Jing[a], Yan Huo[a], Liran Ma[c]

[a]*School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China*
[b]*Science and Technology on Communication Networks Laboratory,Beijing,China*
[c]*Department of Computer Science, Texas Christian University, Fort Worth, TX, USA*

Abstract

In this paper, we introduce energy harvesting into a cognitive radio sensor network to power the network with renewable energy sources so as to achieve self-sustainability of energy-limited sensors. In our work, the cognitive radio technology enable sensors access to the underutilized spectrum for the purpose of coping with the spectrum-scarcity problem in the unlicensed band. Using centralized cooperative spectrum sensing, a set of cognitive sensors is chosen from candidate sensors with different received primary users' signal powers and energy-arrival rates. After detecting the state of a primary channel, we also need to determine which cognitive sensor can get to access the primary channel as well as the power level to be used upon the transmission. The above sensing-access design problem is formulated as an infinite-horizon partially observable Markov decision process, in which the primary goal is to maximize the long-term expected throughput. Through using a value iteration approach, we propose an optimal sensing-access policy. At last, numerical results are presented to verify the superiority of our proposed policy to the existing policy.

Selection and peer-review under responsibility of the scientific committee of the 2017 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2017).

Keywords: Energy harvesting; cognitive radio sensor networks; cooperative sensing; partially observable Markov decision process.

1. Introduction

A wireless sensor network (WSN), which is capable of performing event monitoring and data gathering, has been applied to various fields, including environment monitoring, military surveillance, patient monitoring and smart homes [1, 2]. Currently, most WSNs operate on unlicensed fixed spectrum for data transmission. Due to the coexistence of various emerging networking standards, particularly IEEE 802.11, Bluetooth (IEEE 802.15.1), and WSN itself, the

E-mail address: fanzhang2@bjtu.edu.cn

Selection and peer-review under responsibility of the scientific committee of the 2017 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2017).
10.1016/j.procs.2018.03.034

unlicensed spectrum has become saturated [3]. Cognitive Radio (CR) [4, 5, 6, 7, 8], which enables the secondary (unlicensed) users (SUs) to exploit the unused spectrum resource left by the primary (licensed) users (PUs), provides a promising path to realize dynamic spectrum access and enhance the wireless spectrum efficiency. It is therefore imperative to employ CR in WSNs to mitigate the spectrum-scarcity problem in the unlicensed band, which gives birth to CR sensor networks (CRSNs) [9, 10, 11].

In addition to bring above potential benefits by implementing CR in WSNs, many new issues have also been introduced. Specifically, in CRSNs, in order to utilize the idle spectrum and provide incumbent protection for the primary transmission, SUs need to continuously perform periodical spectrum sensing together with decision-makings on the availability of primary spectrum. Although battery replacing or recharging can extend the functioning time of SUs to a certain extent, such techniques usually afford high cost and involve inconvenient or even impossible operations in some cases [12]. One emerging technique targeting energy-constrained communication systems is energy harvesting (EH) [13, 14, 15]. Energy from renewable energy sources can recharge the sensor nodes' battery and enable the CRSNs to potentially operate perpetually without the need for external power cables or periodic battery replacements [16]. Therefore, incorporating EH in CRSN makes sustainable and environment-friendly sensor networks possible.

There have been many research studies on CR networks with energy harvesting capability. In [17], the authors introduce a stochastic formulation of the network utility optimization problem for EH CR sensor networks, and propose an online and low-complexity algorithm. [18] investigates an energy harvesting cognitive radio network with the save-then-transmit protocol, the authors mainly study the joint optimization of saving factor, sensing duration, sensing threshold and fusion rules to maximize the achievable throughput. In [19], for a single-user multichannel setting, jointly considering probabilistic arrival energy, channel conditions and the spectrum occupancy state of the primary network, the authors propose a channel selection criterion. In this research, we investigate EH CR sensor networks in which each cognitive sensor (i.e. SU) is equipped with a finite-capacity battery and powered by energy harvesting. The main contributions are summarized as follows:

1. Allowing for the sensing errors, channel condition variation as well as the fluctuation of harvested energy, we formulate the spectrum sensing-access problem as an infinite-horizon partially observable Markov decision process (POMDP), in which the primary goal is to maximize the long-term expected throughput.
2. We propose an optimal spectrum sensing-access policy through applying the value iteration approach in the POMDP, where the sensing strategy specifies the set of SUs for cooperative spectrum sensing, while the access strategy specifies which an SU gets to access the channel along with the transmission power.
3. We provide numerical results to evaluate the performance of our proposed policy and show that a significant gain is achieved by our proposed policy over the existing policy.

The rest of the paper is organized as follows. In Section 2, we present the EH CR sensor network model. Section 3 formulates the design of sensing-access within the POMDP framework. In Section 4, we propose the optimal sensing-access policy. Numerical results are provided in Section 5. Finally, Section 6 concludes our work.

2. Network Model

In this paper, we consider the primary network is comprised of a licensed channel occupied by a PU transmitter, and employs the synchronous slotted communication protocol with duration T. For time slot i, we use $\theta_i \in \Theta \triangleq \{0(busy), 1(idle)\}$ denotes the status of the channel occupancy. The traffic of the primary network is modeled as a time-homogeneous discrete Markov process as assumed in [20], where the channel occupancy state randomly changes between idle and occupied according to a discrete Markov process. The state transition probability $Pr(\theta_i = c'|\theta_{i-1} = c)$ is denoted by $P_{c'c}$, where $c', c \in \Theta$. The stationary probabilities of channel being idle and busy are given by $\pi_1 = \frac{P_{01}}{P_{10}+P_{01}}$ and $\pi_0 = \frac{P_{10}}{P_{10}+P_{01}}$, respectively.

Assume there are N cognitive sensors, which are also called SUs, that cooperate in channel sensing and access. We consider that the SUs do not have a fixed power supply and is solely powered by energy scavenged from the ambient environment. The energy arrival process at the nth SU $\{E^h_{n,i}\}$ is assumed to be a sequence of independent random variables with Bernoulli distribution [19]. In each time slot, SUs can harvest energy e_h with probabilities $\mathbf{p_h} = [p_{h,1}, p_{h,2}, \cdots, p_{h,N}]$. Therefore, the average harvested energy rate of the nth SU is $P_{EH,n} = p_{h,n}e_h$. In order to

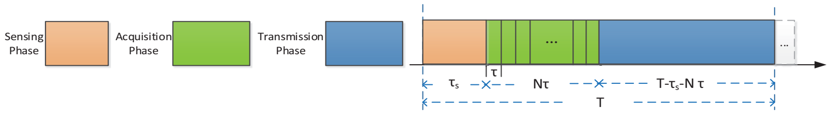

Fig. 1. Illustration of the transmission process.

opportunistically exploit the primary channel without interfering the primary transmission, each time slot is mainly comprised of sensing phase, the acquisition phase, and the transmission phase, as shown in Fig. 1. At the beginning of a time slot, the central controller (CC) broadcasts the information regarding the selected sensing SUs. In the sensing phase, each of the SU selected by the CC senses the status of the primary channel with duration τ_s, consuming energy $e_s = p_s \tau_s$ where p_s is the sensing power. In the acquisition phase, the sensing SUs send 1−bit sensing decisions and all the SUs send information respecting the harvested energy with 1−bit to the CC [21]. We assume each SU reports the information in a time-division multiple-access (TDMA) network, and the time duration cost by each SU is τ. In the transmission phase, CC chooses an SU to access the primary channel with a certain transmission power level.

At the beginning of time slot i, the available energy of SUs $\mathbf{E_i} = [E_{1,i}, E_{2,i}, \cdots, E_{N,i}]$ is known to the CC. Referring the available energy $\mathbf{E_i}$, the CC first makes a sensing decision $\mathbf{x_i} = [x_{1,i}, x_{2,i}, \cdots, x_{N,i}]$, where the available sensing action $x_{n,i} \in X_{n,i}(E_{n,i})$ of the nth SU depends on its available energy. Specifically, if $E_{n,i} \geq e_s$, then $X_{n,i}(E_{n,i}) = \{0, 1\}$ where 1 indicates the nth SU is chosen to sense the channel, while 0 indicates the nth SU stays in idle in the sensing phase; if $E_{n,i} < e_s$, then $X_{n,i}(E_{n,i}) = \{0\}$. The sensing SUs then send the sensing outcomes to the CC, which makes the final decision on the channel status based on the OR fusion rule [21]. Define the sensing result of the ith time slot as $O_i \in \{-1, 0, 1\}$, where "−1", "0", "1" represent no sensing, busy channel and idle channel, respectively. The false alarm probability (Q_{fi}) and the detection probability (Q_{di}) of the cooperative CR system are given by [22]:

$$Q_{fi}(\mathbf{x_i}) = 1 - \prod_{n=1}^{N}(1 - p_{fn,i}x_{n,i}), \quad Q_{di}(\mathbf{x_i}) = 1 - \prod_{n=1}^{N}(1 - p_{di}x_{n,i}), \tag{1}$$

where $p_{fn,i}$ and p_{di} are probability of false alarm and the probability of detection respecting the nth SU. The probability of the false alarm event by assuming the complex-valued PU signal and circularly symmetric complex Gaussian (CSCG) noise is given by [23]:

$$p_{fn,i} = Q(\sqrt{2\beta_n + 1}Q^{-1}(p_{di}) + \sqrt{\tau_s f_s}\beta_n), \tag{2}$$

where β_n means the received primary signal-to-noise ratio (SNR) at the nth SU, f_s is the sampling frequency and $Q(x) = (1/\sqrt{2\pi})\int_x^\infty exp(-t^2/2)dt$. After sensing the primary channel, the CC chooses an SU among the sensing SUs to access the primary channel according to an access decision $\phi_i \in \Phi_{n,i}(O_i, E_{n,i})$. Define $\Phi_{n,i}(O_i, E_{n,i}) = 0$ for $O_i \in \{-1, 0\}$, and $\Phi_{n,i}(O_i, E_{n,i}) = \{0, \{nI_{E_{n,i} \geq e_s} | x_{n,i} = 1\}\}$ for $O_i \in \{1\}$. For the nth SU, the available energy $E_{n,i}$ takes values from a finite set $\mathcal{B} \triangleq \{0, 1, 2, \cdots B\}$. Then the available energy evolves as $E_{n,i+1} = \min(E_{n,i} - E^c_{n,i} + E^h_{n,i}, B)$, where $E^c_{n,i} = e_s x_{n,i} + I_{\phi_i = n} E_t(p_{t,i})$. $E_t(p_{t,i}) = (T - \tau_s - N\tau)p_{t,i}$ represents the energy consumption in the data transmission phase, where $p_{t,i}$ indicates the transmission power. Let $G_{n,i}$ denote the channel power gain from the nth SU to the receiver at time slot i. The channel fading process of the nth SU $\{G_{n,i}\}$ is assumed to be a time-homogeneous Finite-State Markov Chain (FSMC) [24] with the one-step transition probability given by $P_{gg',n} = Pr(G_{n,i+1} = g' | G_{n,i} = g)$, $g, g' \in \mathcal{G}$. $\mathcal{G}$ is a finite set of discretized channel gains. It is assumed that at the beginning of slot i, the CC obtains the perfect knowledge of channel gains $\mathbf{G_i} = [G_{1,i}, G_{2,i}, \cdots, G_{N,i}]$, as assumed in [25]. We use N_0 denotes the noise power at the receiver, and $\tau_{tr} = T - \tau_s - n\tau$. If the channel state is $\theta_i = 1$ and the access decision is $\phi_i = n$, then the throughput of time slot i is

$$r(p_{t,i}, G_{n,i}) = \tau_{tr} \log(1 + \frac{p_{t,i}G_{n,i}}{N_0}). \tag{3}$$

3. Problem Formulation

We first introduce the state space, the action space, the observation space, the observation probability, the state transition probability, and the reward function. At the beginning of time slot i, the CC obtains the available energy $\mathbf{E_i}$ and the channel state information $\mathbf{G_i}$ which are fully observable and known exactly. However, the current spectrum occupancy state of the primary network θ_i cannot be directly observed due to the presence of sensing errors or a potential energy depletion. The CC can infer the spectrum occupancy state based on all its past actions and observation history $H_i \triangleq \{a_v, O_v\}_{v=1}^{i-1}$, where a_v indicates the action regarding the sensing decision as well as the access decision at slot v. As shown in [26], a sufficient statistic for the spectrum occupancy state θ_i is encapsulated by a belief $b_i \triangleq Pr(\theta_i = 1|H_i) \in \Omega$ which denotes the probability of the primary channel being idle in slot i, conditioned on the complete action and observation history H_i. Ω indicates the belief space. Therefore, the system state S_i of time slot i can be defined by a three tuple as $S_i = (b_i, \mathbf{E}_i, \mathbf{G}_i) \in \mathcal{S} = \Omega \times \mathcal{B}^N \times \mathcal{G}^N$. Based on the system state S_i, the CC first makes a spectrum access decision $\mathbf{x}_i \in \mathcal{X}_i = X_{1,i}(E_{1,i}) \times X_{2,i}(E_{2,i}) \times \cdots \times X_{N,i}(E_{N,i})$, where $\mathcal{X}_i$ denotes the admissible sensing action set. Based on the sensing result O_i, the CC then determines the access decision ϕ_i along with the transmission power $p_{t,i}$. Denote the $\mathbb{P}(\phi_i)$ as the action set of transmission power when the spectrum access decision is ϕ_i. If $\phi_i = 0$, we have $\mathbb{P}(0) = \{0\}$. If $\phi_i = n \in [1, N]$, then $\mathbb{P}(n) = \{0, \Delta, 2\Delta, \cdots, L_{n,i}^{\max}\Delta)\}$, where Δ denotes the step size of the transmission power, and $L_{n,i}^{\max}\Delta$ is the maximum transmission power of the nth SU. Considering the energy consumption for transmission should be no greater than the available energy, we have $L_{n,i}^{\max}\Delta \le \frac{[E_{n,i}-e_s]^+}{\tau_{tr}}$, where $[x]^+ = \max\{0, x\}$. In this paper, we set $\Delta = 1/\tau_{tr}$, corresponding to consuming one unit of the energy quantum during the transmission phase, and $L_{n,i}^{\max} = [E_{n,i} - e_s]^+$. The admissible action a_i can be represented as $a_i \in \mathcal{A} = \{(\mathbf{x}_i, \phi_i, p_{t,i})|\mathbf{x}_i \in \mathcal{X}_i, \phi_i \in [0, N], p_{t,i} \in \mathbb{P}(\phi_i)\}$. After performing the action $a_i = (\mathbf{x}_i, \phi_i, p_{t,i})$, the receiver broadcasts an acknowledge (ACK) ψ_i on the error-free common control channel. Define the observation at the ith time slot as $Z_i : (O_i, \psi_i)$, which is defined as follows: (1)$Z_i = 1 : (O_i = -1, \psi_i = 0)$: the CC does not carry out the spectrum sensing since the available energy is not sufficient to perform channel sensing or other reasons; (2)$Z_i = 2 : (O_i = 1, \psi_i = 1)$: the channel is sensed to be idle, the CC accesses the channel, and the ACK is received; (3)$Z_i = 3 : (O_i = 1, \psi_i = 0)$: the channel is sensed to be idle, but no ACK is received either because the CC does not access to the channel due to the energy shortage for transmission, or because of a miss detection; (4)$Z_i = 4 : (O_i = 0, \psi_i = 0)$: the channel is sensed to be busy, the CC does not access the channel, and received nothing. Given the S_i and a_i, the observation Z_i is:

$$Pr(Z_i|\theta_i = c, \mathbf{E}_i = \mathbf{E}, a_i) = \begin{cases} I_{\mathbf{1}\cdot\mathbf{x}_i=0}, \; Z_i = 1, \\ I_{\mathbf{1}\cdot\mathbf{x}_i\neq0} I_{\phi_i\neq0} \overline{Q}_{fi}(\mathbf{x_i})c, \quad Z_i = 2, \\ I_{\mathbf{1}\cdot\mathbf{x}_i\neq0}[\overline{Q}_{di}(\mathbf{x_i})\overline{c} + I_{\phi_i=0}\overline{Q}_{fi}(\mathbf{x_i})c], Z_i{=}3, \\ I_{\mathbf{1}\cdot\mathbf{x}_i\neq0}(Q_{fi}(\mathbf{x_i})c + Q_{di}(\mathbf{x_i})\overline{c}), Z_i = 4, \end{cases} \tag{4}$$

where $\overline{y} = (1-y)$. Based on the Bayes' rule, the CC then updates its belief $b_{i+1} = \mathcal{T}(b_i, Z_i, a_i)$ according to the current belief state b_i, the observation Z_i as well as the action a_i as follows

$$b_{i+1} = \begin{cases} \Gamma(b_i), \; Z_i = 1, \\ P_{11}, \quad Z_i = 2, \\ \Gamma(\dfrac{I_{\phi_i=0}\overline{Q}_{fi}(\mathbf{x_i})b_i}{\overline{Q}_{di}(\mathbf{x_i})(1-b_i) + I_{\phi_i=0}\overline{Q}_{fi}(\mathbf{x_i})b_i}), \quad Z_i = 3, \\ \Gamma(\dfrac{Q_{fi}(\mathbf{x_i})b_i}{Q_{fi}(\mathbf{x_i})b_i + Q_{di}(\mathbf{x_i})(1-b_i)}), Z_i = 4, \end{cases} \tag{5}$$

where the operator $\Gamma(\cdot)$ is defined as $\Gamma(y) = yP_{11} + (1-y)P_{01}$. At the end of time slot i, the system state then switches from S_i to a new state S_{i+1}. The state transition probability from $S_i = (b, \mathbf{E}, \mathbf{G})$ to the next system state $S_{i+1} = (\mathcal{T}(b, Z_i, a_i), \mathbf{E}', \mathbf{G}')$ is given by

$$Pr(S_{i+1} = (\mathcal{T}(b, Z_i, a_i), \mathbf{E}', \mathbf{G}')|S_i = (b, \mathbf{E}, \mathbf{G}), a_i, Z_i) = Pr(\mathbf{E}'|\mathbf{E}, a_i)Pr(\mathbf{G}'|\mathbf{G}), \tag{6}$$

where $Pr(\mathbf{E}'|\mathbf{E}, a_i) = \prod_{n=1}^{N} P_{E_nE'_n}$. $P_{E_nE'_n}$ indicates the transition of the available energy of the nth SU, and can be calculated by

$$P_{E_nE'_n} = \begin{cases} p_{hn}, \quad & E_n = \min(E_{n,i} - E_{n,i}^c + e_h, B), \\ 1 - p_{hn}, \quad & E_n = E_{n,i} - E_{n,i}^c. \end{cases} \tag{7}$$

$Pr(\mathbf{G}'|\mathbf{G})$ indicates the evolvement of the channel states of SUs, which can be calculated by $Pr(\mathbf{G}'|\mathbf{G}) = \prod_{n=1}^{N} P_{G_nG'_n}$, where $P_{G_nG'_n}$ denotes the channel state transition from G_n to G'_n of the nth SU. The EH CR sensor network gains an immediate reward if one SU accesses the primary channel and successfully receives an acknowledgement from the receiver (i.e. $Z_i = 2$), otherwise no reward is received. The immediate reward is defined as the achieved throughput in a single time slot, which can be expressed as $R(a_i, Z_i) = r(p_{t,i}, G_{n,i})$ if $I_{\mathbf{1}\cdot\mathbf{x}_i\neq0}\phi_i p_{t,i} \neq 0, Z_i = 2$. Otherwise, $R(a_i, Z_i) = 0$

4. Optimal Policy

The optimal sensing-access policy for maximizing the expected total throughput of the EH CR sensor network can be formulated as a POMDP. Define a sensing-access policy $\Pi = [\pi_1, \pi_2, \cdots, \pi_T] : \mathcal{S} \to \mathcal{A}$, where π_i maps the system state S_i to the prescribed action $a_i \in \mathcal{A}$ in slot i. The objective is to develop the optimal stationary policy Π_{opt} to maximize the expected total throughput of the EH CR sensor network through cooperative sensing and access over an infinite horizon ($T \to \infty$), which is equivalently to maximize the expected total rewards over an infinite horizon. Given an arbitrary system state S_i and a policy Π, the expected total discounted rewards, also termed as the value function, is given by

$$V_{\Pi}(S_i) = \mathbb{E}_{\Pi}\Big\{\sum_{q=i}^{\infty} \lambda^q R(a_q, Z_q)|S_i\Big\}, S_i \in \mathcal{S}, a_q \in \mathcal{A}. \tag{8}$$

It is known that for an arbitrary system state $S \in \mathcal{S} = (b, \mathbf{E}, \mathbf{G})$, the optimal value function achieved by the optimal policy Π_{opt} satisfies the following Bellman optimality equation [27]:

$$V_{\Pi_{opt}}(b, \mathbf{E}, \mathbf{G}) = \max_{a\in\mathcal{A}} b \sum_{k=1}^{4} Pr(Z_k|1, \mathbf{E}, a)U(b, \mathbf{E}, a|Z_k) + (1-b) \sum_{k=1}^{4} Pr(Z_k|0, \mathbf{E}, a)U(b, \mathbf{E}, a|Z_k), \tag{9}$$

where

$$U(b, \mathbf{E}, a|Z_k) = \underbrace{R(a, Z_k)}_{\text{immediate reward}} \underbrace{+ \lambda \sum_{\mathbf{E}'} \sum_{\mathbf{G}'} Pr(\mathbf{E}'|\mathbf{E}, a)Pr(\mathbf{G}'|\mathbf{G})V_{\Pi_{opt}}(\mathcal{T}(b, Z_k, a), \mathbf{E}', \mathbf{G}')}_{\text{expected future reward}}. \tag{10}$$

$U(b, \mathbf{E}, a|Z_k)$ can be interpreted as the conditional maximum expected reward for a given observation Z_k, which is

Algorithm 1: Algorithm for the Optimal policy

Input: Error bound $\epsilon \to 0$
Output: Optimal policy Π_{opt}
Initialization: At $i = 0$, let $V_0(s) = 0$ for all $S \in \mathcal{S}$;
Repeat
for *each* $S \in \mathcal{S}$ **do**
  Compute:
  $V_{i+1}(S) = \max_{a\in\mathcal{A}_S} \Big\{b \sum_{k=1}^{4} Pr(Z_k|1, \mathbf{E}, a)U_i(b, \mathbf{E}, a|Z_k) + (1-b) \sum_{k=1}^{4} Pr(Z_k|0, \mathbf{E}, a)U_i(b, \mathbf{E}, a|Z_k)\Big\}$, where
  $U_i(b, \mathbf{E}, a|Z_k) = R(a, Z_k) + \lambda \sum_{\mathbf{E}'} \sum_{\mathbf{G}'} Pr(\mathbf{E}'|\mathbf{E}, a)Pr(\mathbf{G}'|\mathbf{G})V_i(\mathcal{T}(b, Z_k, a), \mathbf{E}', \mathbf{G}')$.
  Update:
  $i = i + 1$;
end
Until $\max_{S\in\mathcal{S}} |V_{i+1}(S) - V_i(S)| < \epsilon(1-\lambda)/2\lambda$.
$\Pi_{opt}(S) = \arg\max_{a\in\mathcal{A}} \Big\{b \sum_{k=1}^{4} Pr(Z_k|1, \mathbf{E}, a)U_{i+1}(b, \mathbf{E}, a|Z_k) + (1-b) \sum_{k=1}^{4} Pr(Z_k|0, \mathbf{E}, a)U_{i+1}(b, \mathbf{E}, a|Z_k)$.

comprised of two parts: the first term represents the immediate reward of the current slot defined in (7), while the second term is the (discounted) expected future reward accrued starting from the current time slot with the updated system state. The optimal policy can be found by the value iteration method shown in Algorithm 1. The algorithm utilizes the fixed-point iteration method to solve the Bellman optimality equation with a stopping criteria. If we let $\epsilon \to 0$, then the algorithm returns the optimal policy Π_{opt} [27].

5. Numerical Results

We present numerical results to evaluate the performance of the proposed policy in this section. The unit of energy quantum is $e_u = 1$ mJ, the battery capacity is 6 mJ. The channel state can be "B=Bad", or "G=Good", and the state transition probabilities as follows: $P_{gg'} = \{P_{B,B} = 0.3, P_{B,G} = 0.7, P_{G,B} = 0.7, P_{G,G} = 0.3\}$. The channel power gains at the "Bad", and "Good" states equal to 0.2 and 1.5, respectively. The PU channel state transition probabilities are $P_{11} = 0.8$ and $P_{01} = 0.4$. The probable arrival energy is $e_h = 3$mJ. We assume there are three SUs, and the energy harvesting probabilities are $\mathbf{p_h} = [p_{h,1} = 0.8, p_{h,2} = 0.6, p_{h,3} = 0.4]$. The primary SNRs received at the SUs are $\beta_1 = -17dB, \beta_2 = -16dB, \beta_3 = -15dB$. The initial channel state is $\mathbf{G} =$ ["Good","Good","Bad"], and the initial belief is $b_i = 0.4$. We set $T = 100ms$, $\tau_s = 10ms$, and the sampling frequency is $1MHz$. The information transmission duration is $\tau = 1us$. We assume te default sensing power is $p_s = 100mW$, and $p_{di} = 0.9$. We set $\lambda = 0.99$, and the normalized SNR by $1mw$ is $1/N_0 = -5$ dB. We compare the proposed optimal spectrum sensing-access policy with a benchmark policy proposed in [21] to evaluate the performance. The benchmark policy allows the SU to explore the channel diversity, but it adopts a fixed transmission power. We assume the fixed transmission power is $\Delta = e_u/\tau_{tr}$.

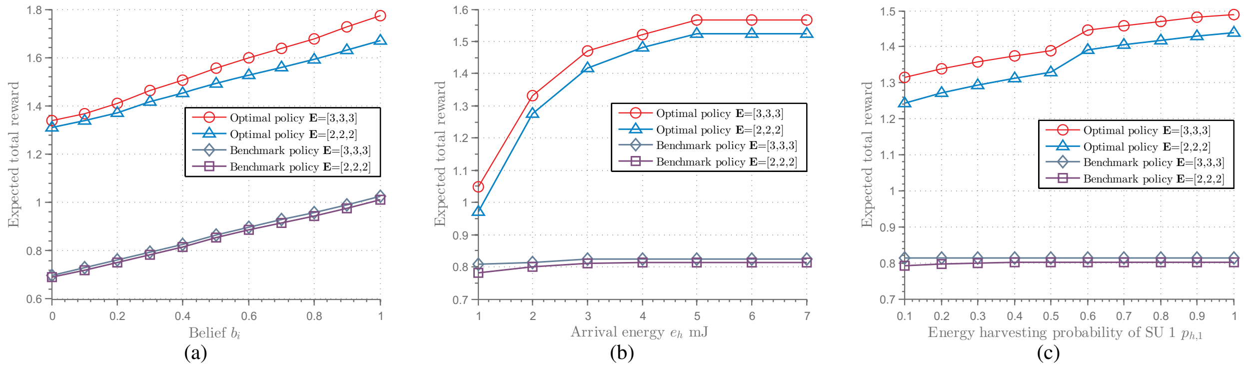

Fig. 2. (a) Expected total reward vs. channel idle belief; (b) Expected total reward vs. probable arrival energy;(c)Expected total reward vs. Energy harvesting probability of SU 1.

Fig. 2(a) compares the expected total reward of the optimal policy and the benchmark policy with different settings of belief. First, it can be seen that the performance of two policies strictly increases with b_i. This is because a higher b_i indicates the CC believes the primary channel is idle with a higher probability, therefore the expected total reward can be improved. Besides, we can also observe that the performance of the optimal policy outperforms the benchmark policy for all settings of b_i. This is because the optimal policy jointly considering the spectrum sensing as well as the transmission power allocation, while the benchmark policy unable to adjust the transmission power to the belief. Fig. 2(b) compares the expected total reward of the optimal policy and the benchmark policy with different settings of probable arrival energy. First, it can be seen that the performance of the optimal policy monotonically increasing with the e_h, and finally achieves a saturation effect. This is because as to the lower region of e_h, the available energy increases as e_h grows, and the SU is able to allocate more energy for channel sensing as well as the data transmission. When e_h is sufficient high, the total available energy is limited by the finite energy capacity B. Besides, we observe that the performance of the benchmark policy slightly increases as e_h grows. This is owing to that the benchmark policy employs a fixed transmission power, and failed to efficiently utilize the higher available energy. Fig. 2(c) shows the expected total reward of the optimal and the benchmark policy with different settings of energy harvesting probability of SU 1. It can be seen that the performance of the optimal policy strictly increases with $p_{h,1}$. This is because as $p_{h,1}$ grows, the optimal policy can exploit more energy for either spectrum sensing or data transmission to improve performance. However, as to the benchmark policy, the performance is almost remains unchanged as $p_{h,1}$ increases. This is owing to that the benchmark policy can not utilize the adaptive transmission power.

6. Conclusion

In this research, we have investigated the optimal sensing-access policy for an EH CR sensor network, where the sensing strategy specified whether to sense the spectrum as well as the set of sensing SUs, while the access strategy specified which SU should access the spectrum along with the transmission power. Aiming to maximize the long-term expected throughput, the design of the optimal sensing-access policy was formulated as an infinite-horizon POMDP, in which the central controller dynamically adapts the sensing strategy together with the access strategy to the system states comprised of the channel idle belief, available energy and channel quality. Through using the value iteration approach in the POMDP, we derived the optimal sensing-access policy. Finally, numerical results were provided to evaluate the performance of our proposed policies and it was shown that a significant gain was achieved by our proposed policy over the existing policy.

7. Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 61471028, 61571010, and 61572070), the Fundamental Research Funds for the Central Universities (Grant No. 2017JBM004 and

2016JBZ003), and the Open Project Program of the Science and Technology on Communication Networks Laboratory (KX162600033).

References

[1] L. M. Borges, Fernando J. Velez, and A. S. Lebres. (2014) "Survey on the characterization and classification of wireless sensor network applications." *IEEE Communications Surveys Tutorials* **16** (4): 1860–1890.

[2] Siyao Cheng, Zhipeng Cai, Jianzhong Li, and Xiaolin Fang. (2015) "Drawing dominant dataset from big sensory data in wireless sensor networks." *The 34th Annual IEEE International Conference on Computer Communications (INFOCOM 2015)* : 531–539.

[3] Ghalib A. Shah, and Ozgur B. Akan. (2014) "Cognitive Adaptive Medium Access Control in Cognitive Radio Sensor Networks." *IEEE Transactions on Vehicular Technology* **64** (2): 757–767.

[4] Jianhui Huang, Shengling Wang, Xiuzhen Cheng, Min Liu, Zhongcheng Li, and Biao Chen. (2014) "Mobility-Assisted Routing in Intermittently Connected Mobile Cognitive Radio Networks." *IEEE Transactions on Parallel and Distributed Systems* **25** (11): 2956–2968.

[5] Jianhui Huang, Shengling Wang, Xiuzhen Cheng, and Jingping Bi. (2016) "Big data routing in D2D communications with cognitive radio capability." *IEEE Wireless Communications* **23** (4): 45–51.

[6] Zhipeng Cai, Shouling Ji, Jing He, Lin Wei, and Anu G. Bourgeois. (2014) "Distributed and Asynchronous Data Collection in Cognitive Radio Networks with Fairness Consideration." *IEEE Transactions on Parallel and Distributed Systems* **25** (8): 2020–2029.

[7] Zhipeng Cai, Shouling Ji, Jing He, and Anu G. Bourgeois. (2012) "Optimal Distributed Data Collection for Asynchronous Cognitive Radio Networks." *The 32nd Proceeding. of International Conference on Distributed Computing Systems 2012 (ICDCS 2012)* : 245–254.

[8] Zhipeng Cai, Yueming Duan, Jing He, and Anu G. Bourgeois. (2015) "Delay efficient opportunistic routing in asynchronous multi-channel cognitive radio networks." *Journal of Combinatorial Optimization* **29** (4): 815–835.

[9] Ghalib A. Shah, and Ozgur B. Akan. (2015) "Cognitive Adaptive Medium Access Control in Cognitive Radio Sensor Networks." *IEEE Transactions on Vehicular Technology* **64** (2): 757–767.

[10] Fan Zhang, Tao Jing, Yan Huo, and Kaiwei Jiang (2017) "Outage Probability Minimization for Energy Harvesting Cognitive Radio Sensor Networks." *Sensors*, **17** (2)

[11] Lichen Zhang, Zhipeng Cai, Peng Li, Liang Wang, and Xiaoming Wang (2017) "Spectrum-Availability Based Routing for Cognitive Sensor Networks." *IEEE Access*, **5**: 4448–4457

[12] Feng Zhao, Lina Wei, and Hongbin Chen. (2016) "Optimal Time Allocation for Wireless Information and Power Transfer in Wireless Powered Communication Systems." *IEEE Transactions on Vehicular Technology* **65** (3): 1830–1835.

[13] Q. Chen, H. Gao, Z. Cai, L. Chen, and J. Li. (2018) "Energy-Collision Aware Data Aggregation Scheduling for Energy Harvesting Sensor Networks." *The 37th Annual IEEE International Conference on Computer Communications (INFOCOM 2018)* .

[14] Tuo Shi, Siyao Cheng, Zhipeng Cai, Yingshu Li, and Jianzhong Li. (2017) "Exploring Connected Dominating Sets in Energy Harvest Networks." *IEEE/ACM Transactions on Networking* **25** (3): 1803–1817.

[15] Tuo Shi, Siyao Cheng, Zhipeng Cai, and Jianzhong Li. (2016) "Adaptive connected dominating set discovering algorithm in energy-harvest sensor networks." *The 35th Annual IEEE International Conference on Computer Communications (INFOCOM 2016)*: 1-9.

[16] Sungsoo Park, and Daesik Hong. (2013) "Optimal Spectrum Access for Energy Harvesting Cognitive Radio Networks." *IEEE Transactions on Wireless Communications* **12** (12): 6166–6179.

[17] Deyu Zhang, Zhigang Chen, Mohamad Khattar Awad, Ning Zhang, Haibo Zhou, and Xuemin Sherman Shen. (2016) "Utility-Optimal Resource Management and Allocation Algorithm for Energy Harvesting Cognitive Radio Sensor Networks." *IEEE Journal on Selected Areas in Communications* **34** (12): 3552–3565.

[18] Sixing Yin, Zhaowei Qu, and Shufang Li. (2015) "Achievable Throughput Optimization in Energy Harvesting Cognitive Radio Systems." *IEEE Journal on Selected Areas in Communications* **33** (3): 407–422.

[19] Jeya Pradha J, Sanket S. Kalamkar, and Adrish Banerjee. (2014) "Energy Harvesting Cognitive Radio With Channel-Aware Sensing Strategy." *IEEE Communications Letters* **18** (7): 1171–1174.

[20] Sungsoo Park, and Daesik Hong. (2013) "Optimal Spectrum Access for Energy Harvesting Cognitive Radio Networks." *IEEE Transactions on Wireless Communications* **12** (12): 6616–6179.

[21] Pratibha, Kwok Hung Li, and Kah Chan Teh. (2016) "Dynamic Cooperative Sensing-Access Policy for Energy-Harvesting Cognitive Radio Systems." *IEEE Transactions on Vehicular Technology* **65** (12): 10137–10141.

[22] Ataollah Ebrahimzadeh, Maryam Najimi, Seyed Mehdi Hosseini Andargoli, and Afshin Fallahi. (2015) "Sensor Selection and Optimal Energy Detection Threshold for Efficient Cooperative Spectrum Sensing." *IEEE Transactions on Vehicular Technology* **64** (4): 1565–1577.

[23] Ying-Chang Liang, Yonghong Zeng, Edward C. Y. Peh, and Anh Tuan Hoang. (2008) "Sensing-Throughput Tradeoff for Cognitive Radio Networks." *IEEE Transactions on Wireless Communications* **7** (4): 1326–1337.

[24] Meng-Lin Ku, Yan Chen, K. J. Ray Liu. (2015) "Data-Driven Stochastic Models and Policies for Energy Harvesting Sensor Communications." *IEEE Journal on Selected Areas in Communications* **33** (8): 1505–1520.

[25] Yuan Wu, Vincent K. N. Lau, Danny H. K. Tsang, and Li Ping Qian. (2012) "Energy-Efficient Delay-Constrained Transmission and Sensing for Cognitive Radio Systems." *IEEE Transactions on Vehicular Technology* **61** (7): 3100–3113.

[26] Yunxia Chen, Qing Zhao, and Ananthram Swami. (2009) "Distributed Spectrum Sensing and Access in Cognitive Radio Networks With Energy Constraint." *IEEE Transactions on Signal Processing* **57** (2): 783–797.

[27] Martin L. Puterman. (2005) "Markov Decision Processes: Discrete Stochastic Dynamic Programming." *John Wiley & Sons, Inc*.