# Optimal Action for Cognitive Radio User under Constrained Energy and Data Traffic Uncertainty

Hiep Vu-Van, Young-Doo Lee and Insoo Koo
School of Electrical Engineering
University of Ulsan, Ulsan, South Korea
Email: vvhiep@gmail.com, leeyd1004@naver.com and iskoo@ulsan.ac.kr

***Abstract*—In cognitive radio network, cognitive radios should operate with a small battery. Operation schedule for cognitive radio user (CU) strongly affects performance of a device with a finite capacity battery. An energy harvester that harvests energy from the environment to recharge the battery can be utilized to extend the lifetime of the CU. While the CU consumes a similar energy for sensing and transmitting data in all active slots, its reward (i.e., throughput) may not be the same because of different data traffic (i.e., the amount of data that needs to be transmitted). Therefore, in order to maximize performance of the cognitive radio network, the CU needs to consider the current data traffic to determine its optimal action policy in terms of sleeping or active. In this paper, we formulate the problem of choosing an action by the CU by using a partially observable Markov decision process (POMDP) in which the CU's three states in terms of data traffic, belief and remaining energy are utilized as main factors to decide an optimal action of the CU in current time slot, and the effect of current action to future reward will be considered through POMDP. Simulation results prove the efficiency of the proposed scheme.**

***Keywords*—*cognitive radio; energy harvesting; energy efficiency; optimal action policy; POMDP.***

## I. Introduction

Cognitive radio (CR) technology can improve spectrum utilization by allowing cognitive radio users (CU) to share the frequency assigned to a licensed user, called the primary user (PU). In order to avoid interference with the operation of the licensed user, the CU is allowed to be active only when the frequency is free. Otherwise, when the presence of the PU is detected, the CU has to vacate their occupied frequency.

In the CR network, the CU often has a small battery that can maintain operations of the CU in a short time. Therefore, the performance of the CR network strongly depends on how effectively the CU uses its electric power. The problem of optimal energy management has been considered previously in [1], [2] where an optimal energy management scheme for a sensor node with an energy harvester to maximize throughput was proposed. In [3], [4], a scheme to find an optimal action policy including sleeping to save energy or active to take opportunity of transmitting data is proposed. Due to the energy-constrained of the CU, the proposed scheme applies the partially observable Markov decision process (POMDP) [5], [6] to determine that optimal action policy.

In most of the previous works, data traffic (i.e., the amount of data that needs to be transmitted) was not considered in order to decide the optimal action for energy efficiency. It is often assumed that the CU always has data to transmit. However, this assumption may be not real in practice. For example, in wireless sensor network, in order to save energy, only the change of data (e.g., the change of monitoring environment) is tracked and reported. In this case, the data traffic varies in time. Therefore, the performance of CR network can be strongly affected by the data traffic. In order to maximize performance of the cognitive radio network, the CU needs to consider the current data traffic to determine its optimal action policy in terms of sleeping or active. In sleeping mode, the CU is silent and waits until the next time slot for another action round. In active mode, the CU first determines the status of the PU signal, if the PU signal is absent, the CU is allowed to access the considered channel to transmit data. If the CU wants to switch to active mode, it must have enough energy for all operations of the mode (i.e., spectrum sensing and data transmitting). In both sleeping and active mode, the CU can harvest energy from the environment to recharge the battery.

In this paper, we formulate the problem of choosing an action by the CU by using a partially observable Markov decision process (POMDP) in which the CU's three states in terms of data traffic, belief and remaining energy are utilized as main factors to decide an optimal action of the CU in current time slot, and the effect of current action to future reward will be considered through POMDP. It is expected that the proposed scheme based on POMDP theory will provide the CR system an improved performance.

This paper is organized as follows. Section 2 describes the system model that we consider in the paper. Section 3 details the proposed optimal action decision scheme based on POMDP. Section 4 introduces simulation models and simulation results of the proposed scheme. Finally, Section 5 concludes this paper.

## II. System Model

We consider a CR network and a PU that is assumed to operate in a time slotted model. The status of the PU changes between two states of the Markov chain model, that is, Presence (P) and Absence (A) as shown in Figure 1. The transition probability of the PU from state P to state A and from state A to itself are defined as $P_{PA}$ and $P_{AA}$, respectively.

The data that needs to be transmitted is stored in a data buffer of the CU; the amount of data $D$ in the buffer is defined as data traffic of the CU. The buffer can store maximum $B_{max}$ units of data. At the time $t$, data traffic of the CU can be defined as,

 

$$D\left(t\right) \in \left\{c_1, c_2, ..., c_{\xi_d}\right\} \quad \text{(units of data)}, \qquad (1)$$

where $0 \leq c_1 < c_2 < ... < c_{\xi_d} \leq B_{max}$ and $\xi_d$ is the number of possible states of data traffic.

At each time slot, there are uncertainty amount of data $d^{in}$ coming the buffer. $d^{in}$ can take its values from the set of possible coming data as,

$$d^{in}\left(t\right) \in \left\{c_1^{in}, c_2^{in}, ..., c_{\xi_{in}}^{in}\right\} \quad \text{(units of data)}, \qquad (2)$$

where $0 \leq c_1^{in} < c_2^{in} < ... < c_{\xi_{in}}^{in}$ and $\xi_{in}$ is the number of possible states of coming data.

The probability mass function (PMF) of the coming data is given as follows:

$$P_{d^{in}}\left(k\right) = \Pr\left[d^{in}(t) = c_k^{in}\right], \; k = 1, 2, ..., \xi_{in}. \qquad (3)$$

We assume that the coming data follows the stochastic process that is marked by the Poisson process. Subsequently, $d^{in}(t)$ is a Poisson random variable with mean $d_{mean}^{in}$. The PMF in (3) can be equivalent to:

$$P_{d^{in}}\left(k\right) \approx \frac{e^{-d_{mean}^{in}}\left(d_{mean}^{in}\right)^k}{k!}, \; k = 1, 2, ..., \xi_{in}. \qquad (4)$$

In order to guarantee the security of the primary network, when the CU decides to utilize the channel of the PU, it firstly needs to check the status of the primary network by performing spectrum sensing. Only if no activity from the channel is captured, CU will be allowed to use the channel.

The CU utilizes an energy detector to monitor the activity of the PU. The Gaussian noise is considered in the sensing channel. Therefore, when the number of sensing samples is relatively large (e.g., $M \geq 200$), the received signal energy $xE$ from the detector can be closely approximated as a Gaussian random variable under both hypotheses of the PU signal [7]. So that, we have,

$$xE \sim \left\{ \begin{array}{ll} \mathrm{N}\left(M, 2M\right), & \mathrm{A} \\ \mathrm{N}\left(M\left(\gamma+1\right), 2M\left(2\gamma+1\right)\right), & \mathrm{P} \end{array} \right., \qquad (5)$$

where $\gamma$ is the signal-to-noise ratio (SNR) of the sensing channel between the CU and the PU.

According to the received signal energy $xE$, the decision on the PU status can be made as follows:

$$\left\{ \begin{array}{ll} G = 1\,\text{(the PU signal is present)}, & \text{if } xE \geq \lambda \\ G = 0\,\text{(the PU signal is absent)}, & \text{otherwise} \end{array} \right., \qquad (6)$$

where $\lambda$ is the energy threshold for a local decision.

The sensing performance of the CU can be evaluated by the probability of detection ($P_d$) and the probability of false alarm ($P_f$), which are given, respectively, as:

$$P_d = Q\left(\frac{\lambda - M\left(\gamma+1\right)}{\sqrt{2M\left(2\gamma+1\right)}}\right) \qquad (7)$$

and

$$P_f = Q\left(\frac{\lambda - M}{\sqrt{2M}}\right). \qquad (8)$$

The main tasks that consume energy of the CU are spectrum sensing and data transmitting. The energy is provided by a rechargeable battery with a finite capacity $E_{max}$ units of energy. The CU can be equipped with a separate energy harvester that can help the CU collect energy from ambient sources (e.g., solar, wind, thermal, vibration) concurrently with other operations. The harvested energy of the current time slot will recharge the battery for powering the CU in the next time slot.

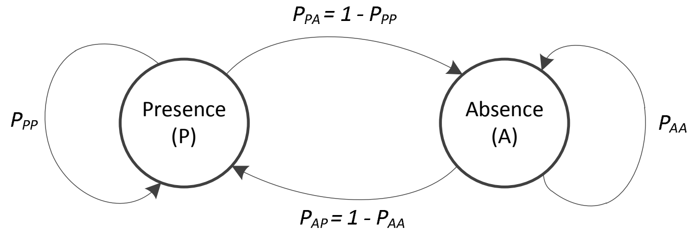

Fig. 1. Markov chain states of the PU.

At the time $t$, the harvested energy is either $e_h\left(t\right) = \left\{e | 0 < e \leq E_{\max}\right\}$ with probability $\tau_h$ or no energy ($e_h(t) = 0$) is harvested with probability $(1 - \tau_h)$. The PMF of the harvested energy can be given as,

$$P_{e_h} = \left\{ \begin{array}{ll} \tau_h, & \text{if } \mathrm{e}_h = \left\{e | 0 < e \leq E_{\max}\right\} \\ 1 - \tau_h, & \text{otherwise} \end{array} \right. . \qquad (9)$$

In this paper, we focus on energy efficiency of the CU under uncertain data traffic. There are three main factors affecting the CU throughput in terms of data traffic $D$, absence probability of the PU signal $p$ (that is defined as *belief* in this paper) and remaining energy $e_r$. Therefore, we use this information as main factors to decide the optimal action of the CU. At the beginning of the time slot, data traffic and remaining energy are available at the CU, and the information of belief $p$ can be estimated based on statistics of sensing results history of the CU. We define the state of the CU as $S = \{D, p, e_r\}$. Based on the state $S$, the CU decides its action including silent to save energy or carry out spectrum sensing to take opportunity of transmitting data.

## III. Action Decision Based on POMDP Under Uncertain Data Traffic of Cognitive Radio User

In this section, we propose a POMDP-based scheme to find an optimal action policy for the CU in order to maximize its throughout. The CU can take one of two actions as: $\Psi = \left\{\text{sleeping}\,(\mathrm{S}),\ \text{active}\,(\mathrm{Ac})\right\}$.

- Sleeping mode (S): as a normal device with limited energy resources, if the CU lacks energy for operations (i.e., spectrum sensing and data transmitting), it will keep sleeping and only harvest energy for operation in the next time slots.

- Active mode (Ac): the CU performs spectrum sensing to detect the state of the PU. If the state A of the PU is detected, the CU transmitter will send data to the CU receiver. At the same time, the harvester of the CU will also harvest energy from the environment.

In this paper, we define throughput $R$ (units of data) of the CU as the amount of data successfully transmitting from the CU, then $0 \leq R \leq S_{max}$, where $S_{max}$ is maximum transmission capacity per slot of the CU. The optimal mode

 ISBN: 978-1-61208-456-5 

decision policy in terms of sleeping or active is formulated as the framework of POMDP. For POMDP, we define the value function $V(D,p,e_r)$ as the maximum total discounted throughput from the current time slot when the current state of the CU is $S(k)=\{D(k),p(k),e_r(k)\}$ where $D(k)$, $p(k)$ and $e_r(k)$ are data traffic, belief and the remaining energy at the beginning of the $k^{th}$ time slot. The value function is given by:

$$V\left(S\left(k\right)\right)=\max_{a(k)}\mathrm{E}\left\{\sum_{t=k}^{\infty}\alpha^{t-k}R\left(S\left(t\right),a\left(t\right)\right)\left|S\left(k\right)\right.\right\}, \tag{10}$$

where $0\leq\alpha<1$ is the discount factor, $S\left(k\right)=\left\{D\left(k\right),p\left(k\right),e_r\left(k\right)\right\}$. $R\left(S(t),a(t)\right)$ is the throughput of the CU achieved at the $t^{th}$ time slot, which is mainly dependent on state $S(t)$ and action decision $a(t)$.

### A. Sleeping Mode ($\phi_1$)

If the CU decides to remain sleeping, no throughput is achieved, then $R\left(S(t),S\left|\phi_1\right.\right)=0$.

State $S(t+1)=\{D(t+1),p(t+1),e_r(t+1)\}$ of the CU will be updated for the next time slot. Firstly, data traffic will be updated as,

$$D(t+1)=min\left(D\left(t\right)+d^{in}\left(t\right),S_{\max}\right), \tag{11}$$

with transition probability

$$\Pr\left(D(t)\to D(t+1)\left|\phi_1\right.\right)=\Pr\left[d^{in}\left(t\right)=c_k^{in}\right], \tag{12}$$

where $k=1,2,...,\xi_{in}$.

Secondly, the belief $p$ is updated as follows,

$$p(t+1)=p(t)P_{AA}+\left(1-p(t)\right)P_{PA}. \tag{13}$$

Finally, the remaining energy of the battery will be increased as,

$$e_r(t+1)=e_r(t)+e_h\left(t\right), \tag{14}$$

with transition probability

$$\Pr\left(e_r(t)\to e_r(t+1)\left|\phi_1\right.\right)=\Pr\left[e_h\left(t\right)\right]. \tag{15}$$

### B. Active mode

When the CU has enough energy for spectrum sensing and data transmitting (i.e., $e_r>e_s+e_t$), it may decide to be active. In this action mode, the achieved throughput of the system depends on the observations of the CU. In this paper, we define 3 observations for the active mode of the CU as,

***Observation 1*** *($\phi_2$):* The CU detects that the PU is present (the channel is used by the PU). Then, the CU is not allowed to access the channel and there is no achieved throughput $R\left(S(t),Ac\left|\phi_2\right.\right)=0$. The probability that $\phi_2$ happens is:

$$\Pr\left(\phi_2\right)=p(t)P_f+(1-p(t))P_d. \tag{16}$$

Data traffic for the next time slot will be updated similarly to the case of observation $\phi_1$.

The sensing result can be used to correct belief $p$ in the current time slot as,

$$p^u(t)=\frac{p(t)P_f}{p(t)P_f+(1-p(t))P_d}. \tag{17}$$

As a result, the updated belief for the next time slot is given by:

$$p(t+1)=p^u(t)P_{AA}+(1-p^u(t))P_{PA}. \tag{18}$$

The updated remaining energy is obtained as:

$$e(t+1)=e(t)+e_h\left(t\right)-e_s, \tag{19}$$

with transition probability

$$\Pr\left(e_r(t)\to e_r(t+1)\left|\phi_1\right.\right)=\Pr\left[e_h\left(t\right)\right]. \tag{20}$$

***Observation 2*** *($\phi_3$):* The CU does not detect any signal from the PU. The CU is allowed to use the channel to transmit data and receive an ACK message. This means that the sensing result is correct (the PU signal is absent) and the CU successfully transmits data. The throughput is achieved as:

$$R\left(S(t),Ac\left|\phi_3\right.\right)=\begin{cases}S_{\max}, & \text{if } D\left(t\right)\geq S_{\max}\\ D\left(t\right), & \text{otherwise.}\end{cases}. \tag{21}$$

The probability that $\phi_3$ happens is:

$$\Pr\left(\phi_3\right)=p\left(t\right)\left(1-P_f\right). \tag{22}$$

Data traffic for the next time slot will be updated as,

$$D(t+1)=min\left(D\left(t\right)+d^{in}\left(t\right)-R\left(S\left(t\right),Ac\left|\phi_3\right.\right),S_{\max}\right), \tag{23}$$

with transition probability

$$\Pr\left(D(t)\to D(t+1)\left|\phi_1\right.\right)=\Pr\left[d^{in}\left(t\right)=c_k^{in}\right]. \tag{24}$$

where $k=1,2,...,\xi_{in}$.

The belief and remaining energy for the next time slot can be updated, respectively, as:

$$p(t+1)=P_{AA} \tag{25}$$

and

$$e(t+1)=e(t)+e_h\left(t\right)-e_s-e_t, \tag{26}$$

with transition probability

$$\Pr\left(e_r(t)\to e_r(t+1)\left|\phi_1\right.\right)=\Pr\left[e_h\left(t\right)\right]. \tag{27}$$

***Observation 3*** *($\phi_4$):* This observation is similar to the observation $\phi_3$, state A of the PU is detected and the CU transmits its data. However, the CU can not receive ACK message. This means that the sensing result is incorrect (the PU signal is present), and the transmission data fails, no throughput is achieved, $R\left(S(t),Ac\left|\phi_4\right.\right)=0$. The probability that $\phi_4$ is obtained is:

$$\Pr\left(\phi_4\right)=\left(1-p\left(t\right)\right)\left(1-P_d\right). \tag{28}$$

Data traffic for the next time slot will be updated similarly to the case of observations $\phi_1$ and $\phi_2$.

The belief that the PU in state A at the next time slot is:

$$p(t+1)=P_{PA}. \tag{29}$$

$$\begin{aligned} V\left(S\left(k\right)\right) = \max_{a_k} \{ \sum_{t=k}^{\infty} \alpha^{t-k} \sum_{\phi_i \in a(t)} \Pr\left(\phi_i\right) \sum_{e(t+1)} \Pr\left(e\left(t\right) \rightarrow e\left(t+1\right) | \phi_i \right) \\ \sum_{D(t+1)} \Pr\left(D(t) \rightarrow D(t+1) \, | \phi_i \right) R\left(S\left(t\right), a\left(t\right) | \phi_i \right) | S\left(k\right) \} \, . \end{aligned} \tag{30}$$

TABLE I. SIMULATION PARAMETERS

| Symbol | Description | Value |
|---|---|---|
| $\gamma$ | SNR of the sensing channel | $-10$ dB |
| Pr($H_0$) | Average absence probability of the PU | 0.5 |
| $P_{AA}$ | Transition probability from state A to itself | 0.8 |
| $P_{PA}$ | Transition probability from state P to state A | 0.2 |
| $E_{ca}$ | Total capacity of battery | 10 units of energy |
| $e_h$ | Harvested energy | 1 units of energy |
| $\tau_h$ | Success probability of energy harvester | 0.8 |
| $e_t$ | Transmission energy | 2 units of energy |
| $e_s$ | Sensing energy | 1 units of energy |
| $B_{max}$ | Capacity of data buffer | 10 units of data |
| $s_{max}$ | Transmission capacity | 5 units of data |
| $d_{mean}^{in}$ | Mean value of coming data | 1 units of data |

The remaining energy of the CU for the next time slot can be updated similar to the case of observation $\phi_3$.

According to those observations, the value function in (10) can be expressed as (30). In order to find an optimal mode policy for maximizing throughput, the optimization problem in (30) will be solved by using the *value iterations* method [8].

## IV. SIMULATION RESULTS

In this section, we present simulation results to prove the efficiency of the proposed scheme. *Myopic* scheme only considers the current time slot for the *value function* (i.e., $\alpha = 0$) to decide the CU's action. This means that unless the CU has not enough energy for spectrum sensing and data transmitting or there is no data in the data buffer, *Myopic* scheme will always allow the CU to be active. Simulation results of *Myopic* will be provided for reference. The parameters for simulations are shown in Table I.

In order to evaluate the performance of the proposed scheme, we define *Reward* of the CU as the average number of units of data successfully transmitted in each time slot.

Figure 2 shows the *Reward* of the considered schemes according to the capacity of the CU battery $E_{max}$. It can be seen that the increase of battery capacity may help the CU achieve higher *Reward*. However, when the battery is big enough (i.e., it has enough space to store all harvested energy), the increase of battery will not affect the *Reward*.

Figure 3 presents the relation between *Reward* and harvested energy $e_h$ of the CU. Higher amount of $e_h$ provide more energy for active mode of the CU, so that the CU can get more *Reward*. However, when $e_h$ is high enough for active mode of the CU in all time, $e_h$ has no more effect on the CU's *Reward*. In this case, the energy constraint will disappear; and then the action of the proposed scheme and *Myopic* scheme will be the same. That is the reason why they have the same performance when the harvested energy is high.

Transmission energy $e_t$ may give strong effect to the *Reward* of the proposed scheme, as shown in Figure 4. More energy is consumed by transmitting data, less *Reward* the CU achieves.

Figures 5 and 6 illustrate the *Reward* according to the change of maximum transmission capacity $s_{max}$ of the CU (i.e., the maximum amount of data that the CU can transmit in whole duration of a time slot) and the change of SNR in the sensing channel, respectively. These figures show that better transmission capacity or better SNR will improve the performance of the proposed scheme.

The simulations results shown in all figures prove that the proposed scheme can offer the CU better performance than the conventional *Myopic* scheme. That benefit of the proposed scheme is achieved by considering future *Reward* on deciding current action.

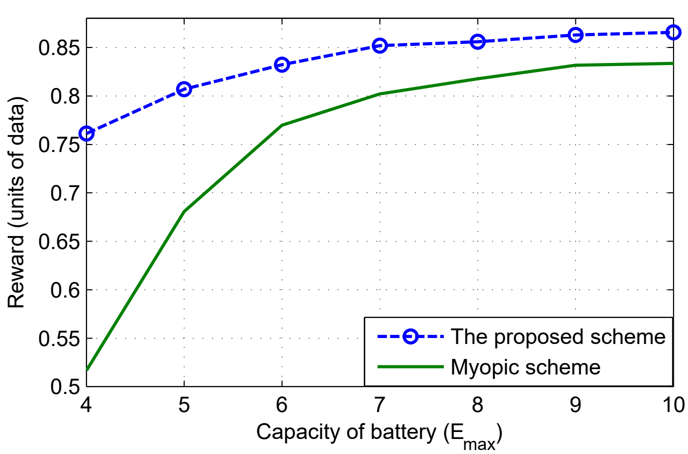


Fig. 2. *Reward* versus battery capacity $E_{max}$.

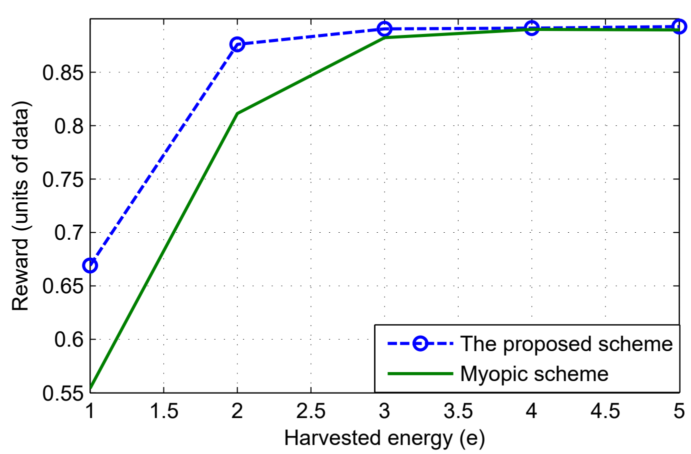


Fig. 3. *Reward* versus harvested energy $e_h$.

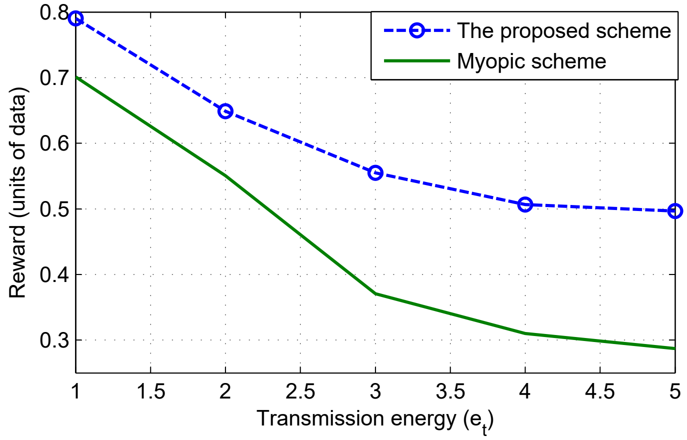


Fig. 4. *Reward* versus transmission energy $e_t$.

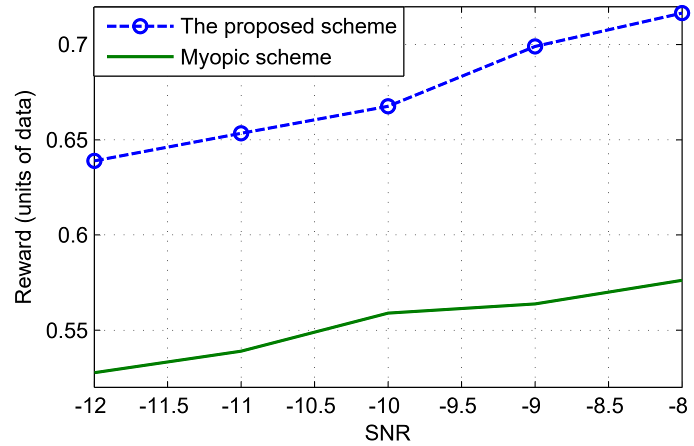


Fig. 6. *Reward* versus SNR of the sensing channel $\gamma$.

## V. Conclusion

In this paper, we proposed a scheme to decide an optimal action to maximize *Reward* of the CU on energy-constrained and uncertainty data traffic manner. By focusing on uncertainty data traffic, the proposed scheme is more practical than in the previous studies. On the other hand, the proposed scheme can take consideration into the effect of the future *Reward* on current action of the CU by applying POMDP theory. Simulation results show that the proposed scheme can obtain better performance than conventional *Myopic* scheme.

## Acknowledgement

This work was supported by the KRF funded by the MEST (NRF-2014R1A1A2005378).

## References

[1] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Transactions on Wireless Communications,*, vol. 9, no. 4, pp. 1326–1336, 2010.

[2] S. Mao, M. H. Cheung, and V. Wong, "An optimal energy allocation algorithm for energy harvesting wireless sensor networks," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 265–270.

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

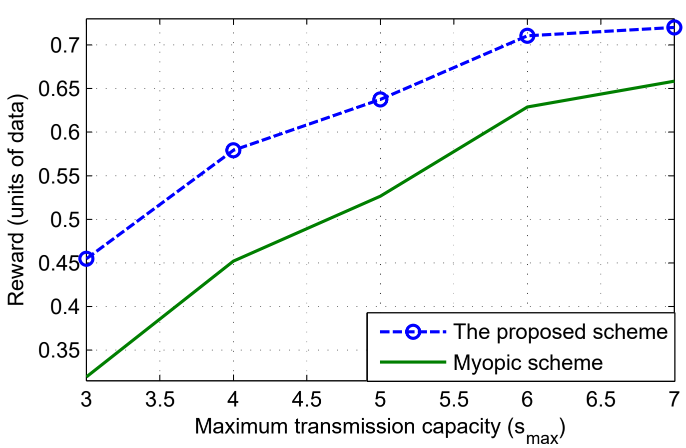


Fig. 5. *Reward* versus maximum transmission capacity $s_{max}$.

[4] S. Park, J. Heo, B. Kim, W. Chung, H. Wang, and D. Hong, "Optimal mode selection for cognitive radio sensor networks with rf energy harvesting," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, 2012, pp. 2155–2159.

[5] X. Cao and X. Guo, "Partially observable markov decision processes with reward information," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 4, 2004, pp. 4393–4398.

[6] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *ARTIFICIAL INTELLIGENCE*, vol. 101, pp. 99–134, 1998.

[7] J. Ma and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2007, pp. 3139–3143.

[8] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientic, 2nd edition, 2001, vol. 1 and 2.

 ISBN: 978-1-61208-456-5 