行政院國家科學委員會專題研究計畫成果報告

廣義的機會式通訊：無線行動網路中之競爭、合作與感知

子計畫三：感知無線行動網路之協力式擇路協定及服務品質控制
期中報告

計畫編號：**NSC 96-2221-E-009-016-MY3**
執行期限：**96** 年 **8** 月 **1** 日 至 **99** 年 **7** 月 **31** 日
主持人：方凱田 教授 國立交通大學電信工程學系

針對部份可觀測感知網路所設計以POMDP為基礎的頻帶換手協定

近年的研究說明了靜態的頻帶分配是造成頻帶使用缺少效率的主因，為了增進頻帶使用率，可動態偵測且使用認證頻帶的感知無線電(Cognitive Radio, CR)因應而生。如何提供有效率的頻帶換手在感知無線電中是個很重要的議題。現存的頻帶換手(Spectrum Handoff)方法假定感知無線電使用者(CR User)可以正確的偵測每一個頻帶以便找到適合的頻帶進行換手。然而，這個假設在實際的情況下是不實際的，因為感知無線電使用者偵測頻帶所花費的時間將會太高而影響主要使用者(Primary User)的品質。

本研究團隊藉由部分可知的環境下的馬可夫決策過程(Partial Observable Markov Decision Process, POMDP) 之幫助，可以透過探測部分的頻帶來估測整個網路環境。研究團隊提出以 POMDP 為基準的頻帶換手機制(POMDP-based Spectrum Handoff, POSH)，其目的為藉由部分的通道狀態來找出最適合進行換手的頻帶。藉由 POSH 機制所選出的頻帶，可達到在每次換手時感知無線電使用者所需等待的時間(Waiting Time)最短，依據模擬結果顯示出此方法可有效率地讓感知無線電使用者在每次頻帶換手時達到最少的等待時間。

A POMDP-based Spectrum Handoff Protocol for Partially Observable Cognitive Radio Networks

Abstract

Recent studies have been conducted to indicate the ineffective usage of licensed bands due to the static spectrum allocation. In order to improve the spectrum utilization, the cognitive radio (CR) is therefore suggested to dynamically exploit the opportunistic primary frequency spectrums. How to provide efficient spectrum handoff has been considered a crucial issue in the CR networks. Existing spectrum handoff algorithms assume that all the channels within the network can be correctly sensed by the CR users in order to perform appropriate spectrum handoff process.

However, this assumption is considered impracticable in realistic circumstances primarily due to the excessive time required for the CR user to sense the entire spectrum space. In this paper, the partially observable Markov decision process (POMDP) is exploited to estimate the network information by partially sensing the frequency spectrums. A POMDP-based spectrum handoff (POSH) scheme is proposed to determine the optimal target channel for spectrum handoff according to the partially observable channel state information. By adopting the policy resulted from the POSH algorithm for target channel selection, minimal waiting time at each occurrence of spectrum handoff can be achieved. Numerical results illustrate that the proposed POSH scheme can effectively minimize the required waiting time for spectrum handoff in the CR networks.

Publication:

- Rui-Ting Ma, Yu-Pin Hsu, and Kai-Ten Feng, "A POMDP-based Spectrum Handoff Protocol for Partially Observable Cognitive Radio Networks," in *Proceedings of IEEE Wireless Communication and Networking Conference (WCNC 2009)*, Apr. 2009, Budapest, Hungary.

A POMDP-based Spectrum Handoff Protocol for Partially Observable Cognitive Radio Networks

Rui-Ting Ma, Yu-Pin Hsu, and Kai-Ten Feng
Department of Communication Engineering
National Chiao Tung University
Hsinchu, Taiwan
{scotie.cm96g, charles880353.ep90}@nctu.edu.tw and ktfeng@mail.nctu.edu.tw

Abstract*—Recent studies have been conducted to indicate the ineffective usage of licensed bands due to the static spectrum allocation. In order to improve the spectrum utilization, the cognitive radio (CR) is therefore suggested to dynamically exploit the opportunistic primary frequency spectrums. How to provide efficient spectrum handoff has been considered a crucial issue in the CR networks. Existing spectrum handoff algorithms assume that all the channels within the network can be correctly sensed by the CR users in order to perform appropriate spectrum handoff process. However, this assumption is considered impracticable in realistic circumstances primarily due to the excessive time required for the CR user to sense the entire spectrum space. In this paper, the partially observable Markov decision process (POMDP) is exploited to estimate the network information by partially sensing the frequency spectrums. A POMDP-based spectrum handoff (POSH) scheme is proposed to determine the optimal target channel for spectrum handoff according to the partially observable channel state information. By adopting the policy resulted from the POSH algorithm for target channel selection, minimal waiting time at each occurrence of spectrum handoff can be achieved. Numerical results illustrate that the proposed POSH scheme can effectively minimize the required waiting time for spectrum handoff in the CR networks.

Keywords: Cognitive radio, partially observable Markov decision process (POMDP), spectrum handoff, channel selection.

I. Introduction

According to the research conducted by FCC [1], a large portion of the priced frequency spectrums remains idle at any given time and location. It has been indicated that the spectrum shortage problem is primarily resulted from the spectrum management policy rather than the physical scarcity of frequency spectrum. Consequently, a great amount of research is devoted to the study of cognitive radio (CR) in recent years [2; 3]. The CR user (i.e. the secondary user) is capable of sensing the channel condition and can adapt its internal parameters to access the licensed channels while these channels are not being utilized by the primary users. The IEEE 802.22 [4; 5], considered as a realistic implementation of the CR concept, is an emerging standard that allocates spectrums for TV broadcast services via a license-exempt basis. Since there is no promise for a CR user to finish its transmission on a certain spectrum, a mechanism called spectrum handoff has been introduced to allow the CR user to select another channel to maintain its data transmission. Consequently, the main objective for spectrum handoff is to select a feasible target channel such that the CR user can be switched into such as to retain its on-going transmission. The performance of spectrum handoff will primarily be dominated by the feasibility of conducting the channel selection.

Target channel selection can be categorized into two different types of schemes according to their sensing strategies [6], i.e. the pre-sensing and post-sensing methods. The pre-sensing scheme indicates that the secondary user will sense the frequency spectrums and consequently choose a sequence of selected target channels before the beginning of its data transmission. Once the secondary user is interrupted by the primary user, the secondary user will be switched to a channel which was determined in sequence from the pre-sensing phase. In general, pre-sensing techniques can reduce the waiting time for spectrum handoff since the target channel is selected based on a pre-determined channel list. However, since the stochastic characteristics of channel can vary drastically in realistic situations, the pre-determined channel list can possibly infeasible to be adopted in target channel selection for spectrum handoff. The channel reservation scheme as proposed in [7] conducts pre-sensing by exploiting the balance between blocking probability and forced termination in order to reserve idle channels for spectrum handoff. Analytical models have been studied in [8; 9] to illustrate the beneficial aspects of the pre-sensing strategy. However, those reserved idle channels can not be ensured available at the time for spectrum handoff. As a result, the performance of pre-sensing strategies can not be guaranteed especially under the fast fading channel environments.

On the other hand, the post-sensing techniques is implemented while the secondary user is forced to terminate its transmission by the primary user. The CR user will start to sense the spectrum in order to verify if there are available channels that can be accessed and consequently becomes its target channel. Compared to the pre-sensing approaches, a more feasible and accurate channel can be selected by exploiting the post-sensing schemes since the target channel is determined at the time while the secondary user is interrupted. Nevertheless, the post-sensing methods in general spend ex-

[1]This work was in part funded by the MOE ATU Program 95W803C, NSC 96-2221-E-009-016, MOEA 96-EC-17-A-01-S1-048, the MediaTek research center at National Chiao Tung University, and the Universal Scientific Industrial (USI) Co., Taiwan.

cessive time in spectrum sensing, especially under crowded network traffic. This situation is not permissible in spectrum handoff since the allowable time duration is considered limited for the CR user to vacate its current channel for the primary user.

Furthermore, it is assumed in most of the existing pre-sensing and post-sensing strategies that all the channels within the network can be correctly sensed, which is considered impracticable in realistic environments. In other words, the transition probabilities of all the channel states are not always served as the available information to the CR users within the network. An estimation algorithm has been proposed in [10] to estimate the transition probability by adopting the maximum likelihood function. However, the converging speed for the estimation can become intolerable while small value of transition probability has been occurred.

In this paper, without the necessity of obtaining all the correct channel information, the partially observable Markov decision process (POMDP) [11; 12] is utilized to reveal the network information by partially sensing the available frequency channels. A POMDP-based spectrum handoff (POSH) mechanism is proposed as a post-sensing strategy in order to acquire the policy such that optimal channel can be obtained with minimal waiting time at each occurrence of spectrum handoff. The transition probabilities for the channel states are derived in this paper by considering the channel as an M/G/1 system with given packet arrival rate and service rate. Numerical results are also presented to illustrate the effectiveness of the proposed POSH scheme, which can achieve minimized waiting time for spectrum handoff.

The rest of this paper is organized as follows. Section II briefly summarizes the concept of the POMDP approach. The proposed POSH scheme is modeled and derived in Section III. Section IV illustrates the performance evaluation for the proposed POSH mechanism; while the conclusions are drawn in Section V.

II. POMDP FRAMEWORK

A Markov decision process (MDP) refers to a discrete time stochastic control process that conducts decision-making based on the present state information, i.e. $s_k \in \boldsymbol{S}_k$. It is noted that the subscript k is utilized to denote the kth time slot in consideration; while $\boldsymbol{S}_k$ represents the entire state space at the kth time slot. Considering the realistic case that not all the current states are obtainable, the partially observable Markov decision process (POMDP) [11] is utilized to determine the decision policy based on the partially available information and the observations from the environment. The schematic diagram of the POMDP framework is illustrated in Fig. 1. In general, optimization techniques are required to be exploited in order to obtain the solution for the POMDP-based problem.

A. Observation

Since not all the states are directly observable within the POMDP setting, a set of observations $z_k \in \boldsymbol{Z}_k$ is essential to provide an indication about which state the environment should be located. The observations can be considered with probabilistic nature, where an observation function o is defined as a probability distribution over all possible observations z_k for each action a_k and resulting states s_{k+1}, i.e.

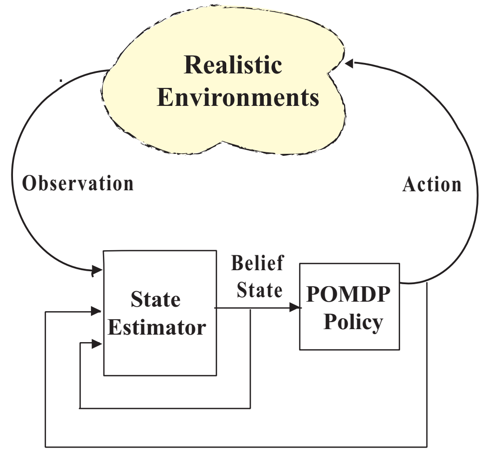

Fig. 1. The schematic diagram of the POMDP framework.

$$o(s_{k+1}, a_k, z_k) = Pr(z_k|a_k, s_{k+1}) \quad (1)$$

$\forall z_k \in \boldsymbol{Z}_k, a_k \in \boldsymbol{A}_k, s_{k+1} \in \boldsymbol{S}_{k+1}$ where $\boldsymbol{A}_k$ stands for the action set at the kth time slot. The parameter a_k denotes the action chosen by the POMDP formulation and s_{k+1} is the resulting state after executing action a_k. Considering the MDP case, a policy is determined to map from the current state to the corresponding action since the present state of MDP is fully observable. On the other hand, the POMDP can only map from the latest observation to the corresponding action (as in Fig. 1), which is in general considered insufficient to represent the history of the process. Therefore, the belief state is utilized to reveal the statistic distribution of the current state information, which will be explained in the next subsection.

B. Belief State

The concept of the belief state, i.e. the information state, is developed to reveal the state of environment and help to behave truly and effectively in a partially observable world. The belief state $b(\boldsymbol{S}_k)$ is a statistic distribution over the state space $\boldsymbol{S}_k$; while $b(s_k)$ corresponds to the probability of state s_k with $\sum_{s_k \in \mathbf{S}_k} b(s_k) \leq 1$. It is noticed that the belief state comprises a sufficient statistical information for the past history, including all the actions and observations that can provide a basis for decision-making under environmental uncertainties. Furthermore, the essential part of belief state is that it can be updated after each corresponding action in order to incorporate one additional step of information into the history. It is considered beneficial to capture the variations from dynamic environment and consequently obtain a more accurate information of the environment. As shown in the Fig. 1, the updated belief state is acquired as the outcome of the state estimator, which is consisted by the inputs of observation, action, and the former belief state. Consequently, the resulting belief state $b(s_{k+1})$ w.r.t. the state s_{k+1} can be obtained as

$$b(s_{k+1}) = Pr(s_{k+1}|b(\boldsymbol{S}_k), a_k, z_k) = \frac{o(s_{k+1}, a_k, z_k) \sum_{s_k \in \boldsymbol{S}_k} \Gamma(s_k, a_k, s_{k+1}) b(s_k)}{Pr(z_k|b(\boldsymbol{S}_k), a_k)} \quad (2)$$

where $b(s_k)$ indicates the former belief state of s_k. The parameter $\Gamma(s_k, a_k, s_{k+1})$ represents the state transition probability from s_k to s_{k+1} according to the action a_k, i.e. $\Gamma(s_k, a_k, s_{k+1}) = Pr(s_{k+1}|a_k, s_k)$. The denominator of (2) can be considered as a normalizing factor, which is obtained as

$$Pr(z_k|b(\boldsymbol{S}_k), a_k) = \sum_{s_k \in \boldsymbol{S}_k, } \sum_{s_{k+1} \in \mathbf{S}_{k+1}} o(s_{k+1}, a_k, z_k) \cdot \Gamma(s_k, a_k, s_{k+1}) b(s_k) \quad (3)$$

C. Reward and Value Functions

In order to ensure optimal decision is made by adopting the POMDP, it is necessary to provide a measurement such as to evaluate the cost or to reward the update from each state. An immediate reward function $r(s_k, a_k, s_{k+1}, z_k)$ is defined to represent the reward by executing action a_k to turn from state s_k to s_{k+1} associated with the observation z_k. Since both the state transition and observation function are probabilistic, the expected reward $R(s_k, a_k)$ can be obtained as

$$R(s_k, a_k) = \sum_{s_{k+1} \in \boldsymbol{S}_{k+1},} \sum_{z_k \in \boldsymbol{Z}_k} \Gamma(s_k, a_k, s_{k+1}) \cdot o(s_{k+1}, a_k, z_k) \cdot r(s_k, a_k, s_{k+1}, z_k) \quad (4)$$

It is noticed that the immediate reward function is denoted as the one-step value function since only the present reward is the major concern. The optimal policy can be directly determined by adopting the reward function as in (4). However, certain period of time is in general considered for evaluating the value of the reward. Therefore, the decision policy by exploiting the POMDP is determined by optimizing the L-step value function with $L \geq 1$.

III. Proposed POSH Scheme

In this section, the POMDP framework will be utilized to model the spectrum handoff problem in a slotted overlay CR network. The value function will consequently be formulated in order to obtain the optimal policy for spectrum handoff.

A. System Model for POSH Scheme

In the considered CR network, there are N channels that are available to be accessed by both the primary and the secondary users. Based on the secondary user's point of view, each channel is assumed to be either in the busy state, i.e. occupied by the primary user, or in the idle state, i.e. free to be accessed. Considering that $c_{i,k}$ denotes the state of the ith channel in time slot k (i.e. $c_{i,k} = 0$ indicates the idle state and $c_{i,k} = 1$ represents the busy state), the state of the entire network in the kth time slot can be written as

$$s_k = [c_{1,k}, \dots c_{i,k}, \dots c_{N,k}], c_{i,k} \in \{0, 1\}, \forall s_k \in \boldsymbol{S}_k \quad (5)$$

The most essential part in spectrum handoff is the target channel selection, which is defined as the action set within the POMDP framework. In other words, an action is to appropriately choose the target handoff channel from the entire N channels within the CR network. After the execution of an action, the channel state can consequently be observed. The set of observations $z_k \in \{0, 1\}$ can be defined as the sensing outcome, where 0 represents the idle state and 1 stand for the busy state.

Furthermore, the transition probability can be determined by modeling a channel as an M/G/1 system with arrival rate λ and service rate μ. By assuming poisson traffic for the arrival packets, the probability distribution of arriving packets can be represented as

$$Pr(n_{\lambda,k} = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (6)$$

where $n_{\lambda,k}$ is denoted as the number of arriving packets in the kth time slot. With the execution of action a_k, the channel transition probability $\tau(c_{i,k}, a_k, c_{i,k+1})$ represents the transition from the present channel state $c_{i,k}$ to the channel state $c_{i,k+1}$ at the next time slot. By adopting the result from (6), the transition probability from idle to idle state for a specific channel $c_{i,k}$ can be acquired as

$$\tau(c_{i,k} = 0, a_k, c_{i,k+1} = 0) = \frac{Pr(n_{\lambda,k} = 0, c_{i,k} = 0)}{Pr(c_{i,k} = 0)} = \frac{Pr(n_{\lambda,k} = 0) \cdot Pr(c_{i,k} = 0)}{Pr(c_{i,k} = 0)} = e^{-\lambda} \quad (7)$$

It is noted that the second equality in (7) is attributed to the fact that the availability of a channel in any kth time slot is independent to the total number of arrival packets within the same slot. On the other hand, the transition probability for the channel $c_{i,k}$ coming from busy to idle state can be represented as

$$\tau(c_{i,k} = 1, a_k, c_{i,k+1} = 0) = \frac{Pr(n_{\lambda,k} = 0, n_{\lambda,k-1} > 0, T_{s,k} \leq 1)}{Pr(c_{i,k} = 1)} = \frac{Pr(n_{\lambda,k} = 0) \cdot Pr(n_{\lambda,k-1} > 0) \cdot Pr(T_{s,k} \leq 1)}{Pr(c_{i,k} = 1)} \quad (8)$$

where the second equality is also contributed to the independency of the three probabilities within the numerator of (8). The parameter $T_{s,k}$ is the total service time in time slot k which includes both the time durations for serving packets coming into this kth time slot and the remaining packets acquired from the previous $(k-1)$th slot. It is assumed that $\gamma_1, \gamma_2, \dots \gamma_\alpha$ are the random variables of a packet service time with mean value of $1/\mu$, where α represents the number of packets arrived from the previous $(k-1)$th slot, i.e. $n_{\lambda,k-1} = \alpha$. The time server take for serving these α packets within the kth slot is denoted as $T_{\alpha,k} = \sum_{j=0}^{\alpha} \gamma_j Pr(n_{\lambda,k-1} = \alpha)$. Therefore, the third term in the numerator of (8) can be rewritten as

$$Pr(T_{s,k} \le 1) = Pr(T_{s,k-1} \le 1) \cdot Pr(T_{\alpha,k} \le 1) + Pr(T_{s,k-1} > 1) \cdot Pr(T_{s,k-1} - 1 + T_{\alpha,k} \le 1) \quad (9)$$

which is the combination of two cases as follows: (a) the packets can be served in both previous and this time slots; and (b) the packets have not been entirely served in the previous slot but are able to be served in this time slot. Furthermore, the denominator of (8) that representing the probability for a busy channel can be expressed as

$$\Pr(c_{i,k}{=}1){=} Pr(n_{\lambda,k-1}{>}0){+}Pr(n_{\lambda,k-1}{=}0){\cdot}Pr(T_{s,k-1}{>}1) \quad (10)$$

Based on (7) and (8), the transition probabilities from idle to busy state and from busy to busy state can be respectively obtained as $\tau(c_{i,k} = 0, a_k, c_{i,k+1} = 1) = 1 - \tau(c_{i,k} = 0, a_k, c_{i,k+1} = 0)$ and $\tau(c_{i,k} = 1, a_k, c_{i,k+1} = 1) = 1 - \tau(c_{i,k} = 1, a_k, c_{i,k+1} = 0)$. As a result, by assuming that each channel $c_{i,k}$ is independent with each other for $i = 1$ to N, the transition probability for the entire network $\Gamma(s_k, a_k, s_{k+1})$ can be obtained as

$$\Gamma(s_k, a_k, s_{k+1}) = \prod_{i=1}^{N} \tau(c_{i,k} = \varsigma_1, a_k, c_{i,k+1} = \varsigma_2) \quad (11)$$

where $c_{i,k} \in s_k$ and $c_{i,k+1} \in s_{k+1}$, the time required for spectrum handoff of a CR user is defined as the time duration from the termination of packet transmission in one channel to the starting time of retransmission in another channel. Three factors are considered for the spectrum handoff time including the switching time, the handshaking time, and the waiting time. In general, both the switching and the handshaking time intervals are assumed fixed with comparably smaller values. Therefore, the main objective of the proposed POSH scheme is to select a target channel that has the minimum waiting time, i.e. the smallest number of waiting slots required for the CR user in the case that the target channel is still occupied by the primary user.

The waiting time, which is served as the reward function within the POMDP framework, will be minimized by the POSH scheme with the selection of an optimal channel in the spectrum handoff process. The immediate reward is considered as the total number of waiting slots n_w required by the secondary user while executing a certain action for spectrum handoff. Consequently, the expected reward $R(s_k, a_k)$ can be written as

$$\begin{aligned} R(s_k,a_k) &= E[n_w{=}\ell|a_k] = \sum_{\ell=0}^{\infty} \ell \cdot Pr(n_w{=}\ell|a_k) \\ &= \sum_{\ell=1}^{\infty} \ell \cdot Pr\left(c_{i,\ell+k}{=}0, \bigcap_{p=1}^{\ell} c_{i,k+p-1}{=}1\right) \\ &= \sum_{\ell=1}^{\infty} \ell \cdot \tau(c_{i,\ell+k-1}{=}1, a_k, c_{i,\ell+k}{=}0) \cdot Pr\left(\bigcap_{p=1}^{\ell} c_{i,k+p-1}{=}1\right) \end{aligned} \quad (12)$$

It is noted that the intersecting probability of the last equality in (12) can further be simplified as

$$\begin{aligned} & Pr(\bigcap_{p=1}^{\ell} c_{i,k+p-1} = 1) \\ &= Pr(\bigcap_{p=1}^{\ell-1} c_{i,k+p-1} = 1) \cdot \tau(c_{i,k+\ell} = 1, a_k, c_{i,k+\ell-1} = 1) \\ &= c_{i,k} \cdot [\tau(c_{i,k} = 1, a_k, c_{i,k+1} = 1)]^{\ell-1} \end{aligned} \quad (13)$$

B. Protocol Implementation of POSH Scheme

An overlay slotted CR network with partially observable information is considered for the POSH scheme, which indicates that the secondary user is not allow to coexist with the primary user while the time duration for packet transmission is divided into time slots. As shown in Fig. 1, partial channel information $o(s_{k+1}, a_k, z_k)$ is assumed available to be observed by the secondary users, which will be exploited for the update of the belief state $b(s_{k+1})$ as in (2). The secondary user utilizes the updated belief state in order to estimate the channel state of the CR network.

According to the POSH scheme, an L-step value function will be adopted to obtain the corresponding action that results in the minimal waiting time after the handoff process. In other words, based on the L-step value function $V_L^*[b(s_k)]$, which is mapped from the belief state space, the CR user will determine the feasible action to take in order to achieve the highest reward. The L-step value function for the CR user can be obtained as

$$\begin{aligned} V_L^*[b(s_k)] = \max_{a_k \in \boldsymbol{A}_k} & \sum_{s_k \in \boldsymbol{S}_k} b(s_k) \cdot R(s_k, a_k) + \\ & \rho \sum_{z_k \in \boldsymbol{Z}_k} Pr(z_k | b(\mathbf{S}_k), a_k) \cdot V_{L+1}^*[b(s_{k+1})] \end{aligned} \quad (14)$$

where ρ is denoted as a discount factor for convergence control of the value function. The probability $Pr(z_k | b(\mathbf{S}_k), a_k)$ is defined as in (3). At the beginning of the time slot where the spectrum handoff occurred, the CR user will choose a target channel that possesses the minimum waiting slots based on the results obtained from the L-step value function as in (14). After switching to the target channel, the CR user will conduct the sensing task for observing the newly updated channel state even though only partial state information is obtainable. After waiting for the required time slots that are determined by the POSH scheme, the secondary user can start to conduct its packet transmission within the target channel.

Moreover, it is noticed that the computation of L-step value function (in (14)) is considered complex and in general difficult to solve. The dimension of the belief state can grow exponentially as the number of channel is augmented, which makes it difficult to be adopted for practical implementation. A reduced state strategy has been proposed in [13] to establish an approximated linear state vector, which can effectively decrease the computation complexity of the value function. The complex optimization problem associated with (14) can therefore be resolved in an efficient manner.

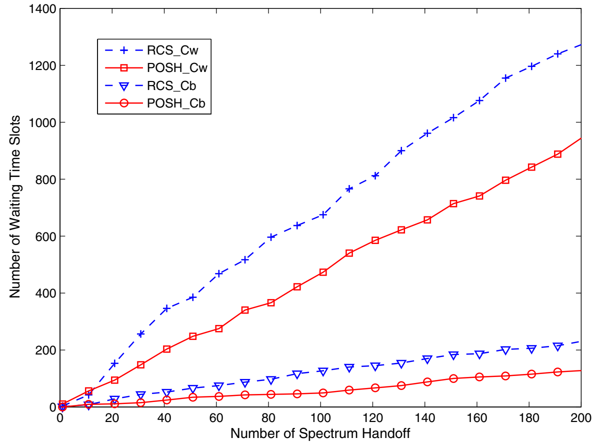

Fig. 2. Performance comparison: the number of waiting time slots versus the number of spectrum handoff.

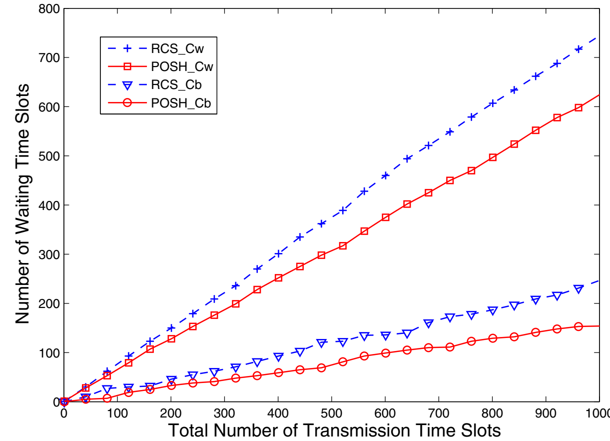

Fig. 3. Performance comparison: the number of waiting time slots versus the total number of transmission time slots.

IV. Performance Evaluation

In this section, simulations are presented to demonstrate the performance of the proposed POSH scheme. The major concern in the simulations is to obtain the required waiting time slots for the secondary user while it has been directed to the target channel. Since full channel state information is required by all of the existing spectrum handoff algorithms, it is considered unfair to compare the existing schemes within the environment adopted by the proposed POSH algorithm, where only partial channel information is observable. Therefore, the proposed POSH scheme will be compared with another two different cases as follows: (a) the random channel selection (RCS) scheme is adopted by the CR user to perform spectrum handoff by randomly selecting the target channel; and (b) no spectrum handoff (NSH) scheme is performed for the CR user under any circumstance. The traffic of the primary user follows the poison distribution, and the service time is assumed to be a uniform distribution with mean $1/\mu = 1$. Three channels are considered in the simulations, i.e. $s_k = [c_{1,k}, c_{2,k}, c_{3,k}]$; while the discount factor in (14) is selected as $\rho = 1$. The reduced state strategy [13] is utilized in the simulations to obtain the numerical results of the POMDP-based optimization problem.

Fig. 2 shows the performance comparison of the number of waiting time slots versus the total number of spectrum handoff for both the proposed POSH scheme and the RCS method. Two different channel conditions are considered for comparison purpose as follows. A better channel condition C_b is chosen with the transition probability from idle to idle state for each channel as $\tau(c_{i,k} = 0, a_k, c_{i,k+1} = 0) = 0.8, 0.7, 0.65$ for $i = 1, 2, 3$; while that from busy to idle state is selected as $\tau(c_{i,k} = 1, a_k, c_{i,k+1} = 0) = 0.4, 0.5, 0.55$ for $i = 1, 2, 3$. On the other hand, a worse channel condition C_w is determined with the transition probability from idle to idle state as $\tau(c_{i,k} = 0, a_k, c_{i,k+1} = 0) = 0.4, 0.3, 0.35$, and that from busy to idle state is set as $\tau(c_{i,k} = 1, a_k, c_{i,k+1} = 0) = 0.1, 0.2, 0.15$. For fair comparison, the NSH method is not implemented in this case since the CR user can always stay at the channel with better condition. It is intuitive to see that the total number of waiting time slots is increased as the number of spectrum handoff is augmented. Furthermore, the secondary user has to wait for comparably more time slots in the worse channel case under both schemes. Nevertheless, the total waiting time slots acquired from the proposed POSH scheme is comparably smaller than that from the RCS method under both channel conditions. It is also observed that the POSH scheme performs better as the number of spectrum handoff is increased. The reason can be attributed to the situation that more updated belief states are acquired by the POSH scheme as the number of handoff is augmented.

Fig. 2 illustrates the performance comparison between the number of waiting time slots and the total number of transmission time slots. It is noticed that different numbers of waiting time slots and handoff numbers will be resulted by each scheme at every specific number of transmission time slots. In other words, the combining effects from both the waiting time slots and the handoff numbers will be revealed in Fig. 2 at each horizontal data point. It can be observed that the proposed POSH algorithm still outperforms the RCS scheme under both the C_b and C_w channel conditions. Even though the effect from the total number of spectrum handoff has not been considered in the value function (in (14)), the POSH scheme can still provide smaller waiting time comparing with the RCS method.

Figs. 4 illustrate the performance comparison between the POSH, the RCS, and the NSH schemes under different values of packet arrival rate λ of the primary user. The left subplot of Fig. 4 shows the comparison under fix numbers of spectrum handoff $= 250$ and the right subplot is performed under fixed number of transmission time slots $= 1200$. The percentage of waiting time slots as shown in Fig. 4 is defined as

$$\eta(\%) = \frac{n_{w,strategy} - n_{w,posh}}{n_{w,posh}} \times 100 \qquad (15)$$

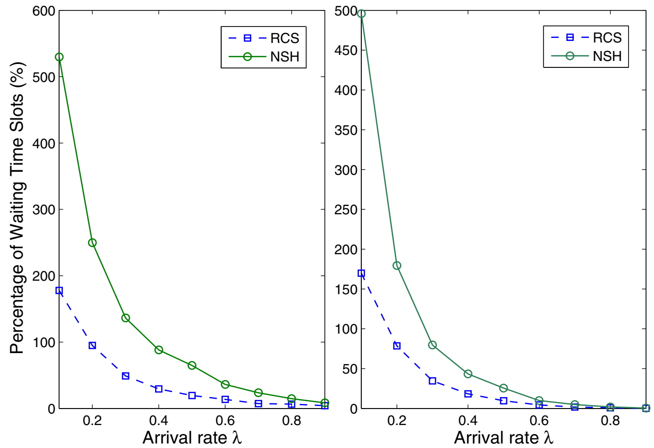

Fig. 4. Performance comparison: the percentage of waiting time slots versus the arrival rate (left: with fixed handoff number = 250; right: with fixed number of transmission time slots = 1200).

where $n_{w,strategy}$ is the total number of waiting slots resulting from either the RCS or the NSH scheme; and $n_{w,posh}$ indicates that from the proposed POSH algorithm. It can be observed that the proposed POSH scheme can outperform the other two methods under different packet arrival rates. The benefits from the adoption of the POSH algorithm is especially revealed at smaller values of packet arrival rate since there can be more opportunity for the POSH scheme to select a feasible target channel.

V. Conclusion

This paper proposes a strategy for post-sensing spectrum handoff based on the partially observable Markov decision process (POMDP) in the overlay cognitive radio (CR) networks. With only partially observable state information, the proposed POMDP-based spectrum handoff (POSH) scheme selects the optimal target channel in order to achieve the minimal waiting time for packet transmission. Simulation results show that the proposed POSH scheme can effectively reduce the waiting time for spectrum handoff for a partially observable CR network.

References

[1] F. C. Commission, “Unlicensed operation in the TV broadcast bands,” *ET Docket*, pp. 4–186, 2006.

[2] V. K. Bhargava, “Research challenges in cognitive radio networks,” *Proc. IEEE ICCCN*, pp. 1–6, 2007.

[3] Q. Zhao, L. Tong, A. Swam, and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework,” *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 589–600, 2007.

[4] S. Sengupta, S. Brahma, M. Chatterjee, and N. S. Shankar, “Enhancements to cognitive radio based IEEE 802.22 air-interface,” *Proc. IEEE ICC*, 2007.

[5] C. Cordeiro, K. Challapali, D. Birru, and N. S. Shankar, “IEEE 802.22: The first worldwide wireless standard based on cognitive radios,” *Proc. IEEE DySPAN*, pp. 328–337, 2005.

[6] L. C. Wang and A. Chen, “On the performance of spectrum handoff for link maintenance in cognitive radio,” *Proc. IEEE ISWPC*, pp. 670–674, 2008.

[7] X. Zhu, L. Shen, and T. S. P. Yum, “Analysis of cognitive radio spectrum access with optimal channel reservation,” *IEEE Communications Letters*, vol. 11, pp. 304–306, 2007.

[8] L.-C. Wang and C. W. Wang, “Modeling and analysis for proactive-decision spectrum handoff in cognitive radio networks,” *Proc. IEEE ICC*, 2009.

[9] L. C. Wang and C. W. Wang, “Spectrum handoff for cognitive radio networks: Reactive-sensing or proactive-sensing?” *Proc. IEEE IPCCC*, 2008.

[10] X. Long, X. Gan, Y. Xu, J. Liu, and M. Tao, “An estimation algorithm of channel state transition probabilities for cognitive radio systems,” *Proc. IEEE CROWNCOM*, 2008.

[11] A. R. Cassandra, “Exact and approximate algorithms for partially observable markov decision processes,” Ph.D. dissertation, Brown University, 1994.

[12] 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.

[13] Q. Zhao, L. Tong, and A. Swami, “Decentralized cognitive mac for dynamic spectrum access,” *Proc. IEEE DYSPAN*, pp. 224–232, 2005.

[14] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed. Wiley, 1998.

[15] W. Hu, D. Willkomm, M. Abusubaih, J. Gross, G. Vlantis, M. Gerlaand, and A. Wolisz, “Cognitive radios for dynamic spectrum access - dynamic frequency hopping communities for efficient IEEE 802.22 operation,” *IEEE Communications Magazine*, vol. 45, pp. 80–87, 2007.

[16] Z. Kostic and I. Maric, “Dynamic frequency hopping in wireless cellular systems-simulations of full-replacement and reduced-overhead methods,” *Proc. IEEE VTC*, pp. 914–918, 1999.