Detection of Spectral Resources in Cognitive Radios Using Reinforcement Learning

Ulrich Berthold*, Fangwen Fu**, Mihaela van der Schaar**, Friedrich K. Jondral*
*Universität Karlsruhe (TH), Institut für Nachrichtentechnik, Germany
**University of California Los Angeles (UCLA), Dept. of Electrical Engineering
berthold@int.uni-karlsruhe.de, {fwfu, mihaela}@ee.ucla.edu

(Short Paper)

***Abstract*—Available spectrum for wireless communications is a limited resource which gains importance with the increasing demand for mobile communication services with high data rates. Measurements show, that assigned frequency bands (FBs) are not used efficiently. One approach for increasing the efficiency in spectrum use is the concept of overlay systems, which can be seen as an enabling technology for cognitive radio and dynamic spectrum access by providing frequency agility. In this paper, we propose an approach for the detection of spectral resources based on reinforcement learning, allowing the cognitive radio to select the FBs with the most available resources.**

I. Introduction

Recent developments in communication hardware allow the design of small mobile devices which are capable of processing and transmitting high bit rate data at affordable costs, resulting in a growing mass market for mobile multimedia communications. The main technological resource for providing mobile multimedia services is the available spectrum suitable for wireless communications. This resource is limited due to the physical properties of electro-magnetical wave propagation. When looking at the frequency assignments managed by the regulatory body (e. g. the FCC) and prices recently paid in FB auctions, it is very difficult and expensive for service providers to obtain new FBs for deploying new wireless communication systems or extending existing ones.

Nevertheless, measurements show [1] that although nearly all FBs are assigned, most of them are not used in a very efficient way, resulting in a very fragmented allocation in the time-frequency plane. Unused parts of the time-frequency plane are also called spectrum holes. One approach to increase the efficiency in spectrum use, is the concept of overlay systems which is the focus of this paper.

We assume a scenario where there already is a system deployed in a given FB and an independent overlay system dynamically fills the spectrum holes [2]. An efficient design, configuration and operation of the overlay system, and especially of its detection component, is vital to a coexistence of both systems in the same FB. To avoid collisions with the PU system, the detection subsystem has to periodically perform allocation measurements in the FB which is currently used for transmitting data. Additionally, it has to observe the spectral resource situation in all other available FBs to be able to switch the active FB if necessary.

The remainder of this paper is organized as follows. In Section II we give an overview on the assumed overlay scenario and discuss the difference between detection of the PU system's allocation and the detection of spectral resources. In Section III the detection of spectral resources is formulated as a Markov decision process and a solution strategy based on a actor-critic method is proposed. Simulation results are presented in Section IV and the paper is finally concluded in Section V.

II. System Model

We assume a scenario where a primary user (PU) system and a secondary user (SU) system are being operated in the same FB. Two assumptions are made for the PU system [3]: The PU system has priority and must not be affected by the SU system and the PU system must not be modified. Thus, all necessary signal processing and coordination resulting from the coexistence must be implemented in the SU system. The main goals are to minimize the mutual interference between PU and SU system and and to utilize the available (”left over”) resources as efficiently as possible. In the following subsections we discuss the assumptions for the system parameters of both systems.

A. Primary User Systems

A potential PU system is required to be using a combination of time- and frequency division multiple access (TDMA/FDMA), since only this scheme results in spectrum holes. Carrier sense multiple access with collision avoidance (CSMA/CA) is not feasible, because an allocation of the SU system will result in an interaction with the PU system. Nevertheless, there is still a great variety of potential PU systems, including e. g. DSB-AM used for aeronautical communications, FBs used for radar or mobile communication such as GSM. In this paper, we assume a scenario where multiple FBs can be used by the SU system. Further, we assume a generalized PU system where the PUs operating in each FB have different spectrum utilization characteristics. Thus, each of these FBs has a different amount of spectrum holes and therefore, available spectrum for the SU system. For simplicity reasons we assume that each of the N_{fb} FBs has the same bandwidth and that each PU system uses the same channel spacing.

B. Secondary User System

For the SU system we use OFDM, since with OFDM an efficient, flexible and sophisticated technology is available, which is shown to be a good candidate for SU systems [4].

This work was in part sponsored by NSF CAREER award no. 0541867 and the Karlsruhe House of Young Scientists (KHYS).

978-1-4244-2017-9/08/$25.00 ©2008 IEEE

One main advantage is that each subcarrier can be switched on or off individually, depending on the current allocation of the PU system. Depending on the exact specification of the SU system, the subcarrier spacing can vary, leading to a different number of subcarriers used for each channel of the PU system. The example shown in Fig. 1 uses 5 subcarriers per PU channel. Note that a smaller subcarrier spacing comes along with a longer symbol duration.

The overlay system has to periodically perform allocation measurements to determine the allocation of the primary system. In OFDM based overlay systems, this can be done efficiently without much additional costs regarding hardware by using the already existing FFT component. The result of the allocation measurement is the allocation vector (AV), indicating for each subcarrier if a primary user was detected ("1") or not ("0"). In a multi-band scenario, we have to distinguish two types of detection. On the one hand, detection has to be performed periodically in the current FB used for transmission to avoid collisions with the PU system. On the other hand, the SU system also has to observe the other FBs with respect to spectral occupancy. According to these two main tasks, the detection subsystem is split into two parts:

1) Detection of the PU System's Allocation: This part is responsible for the exact and detailed detection of the available resources in the current FB. The main characteristics are:

- Detection is performed frequently enough to avoid collisions with the PU system.
- Detection is performed for all subchannels at the same time by using an FFT.
- The detection results are used directly to determine the exact available resources

2) Detection of Spectral Resources: The task of this part of the detection subsystem is to generate a coarse overview on the average available resources in all FBs. This information is used to make more global strategic decisions, in which FB an operation would be best in near future. The main characteristics are:

- Detection is performed once in a while to obtain an estimation of the current allocation situation. In contrast to low level detection, the goal is not to avoid collisions with the PU system.
- Detection is performed using an FFT, but only one FB can be scanned at a time. When detecting in a different FB, the subsequent update interval cannot be used for transmissions in the active/current FB.
- The detection results are only used for a mid-term estimation and prediction of the allocation for making strategic decisions. They have no immediate impact on the ongoing transmission.

Since detection is a vital component for cognitive radios and dynamic spectrum access, there has recently been done a lot of research in this area. A variety of different aspects have been studied, e.g. cooperative sensing among multiple cognitive radio terminals [5], [6] or compressed sensing [7]. In [8] a Partially Observable Markov Decision Process (POMDP) is used to derive a class of decentralized cognitive MAC protocols for opportunistic spectrum access networks. Similar to our paper, also an MDP formulation is used, but a different cognitive radio scenario is assumed. For instance, in [8] the difference between the detection of the PU system's current allocation and the estimation of spectral resources is not considered. Furthermore, finding an optimum policy for the POMDP in [8] involves complex computations, whereas in this paper we use a less complex, on-line reinforcement learning algorithm for finding a trade-off between sensing and transmission. We focus on the detection of spectral resources and assume that the SU system is operating in infrastructure mode and has an access point which coordinates the detection for all SUs and decides when to use which FB.

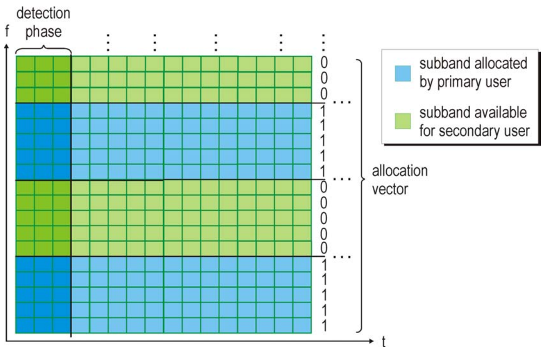
detection phase
f
t
0
0
0
1
1
1
1
1
0
0
0
0
0
1
1
1
1
1
allocation vector
subband allocated by primary user
subband available for secondary user

Fig. 1. Allocation of a primary user in the time-frequency plane and the resulting allocation vector.

III. Detection of Spectral Resources

In contrast to the detection of the current allocation, the goal of spectral resource detection is to obtain an overview on the estimated allocation situation in all FBs in the system. In case of a shortage of resources in the active FB, the SU system then can switch to another FB with currently more resources available. Since the SU system cannot operate in more than one FB at the same time, it has to tune in to all FBs periodically to perform a measurement. During this time no transmissions can be made, forcing the SU system to find an optimum trade-off between spending time for transmitting data and time for increasing the accuracy of information regarding the estimated allocation situation in the other FBs by performing detections. So basically, the SU system must decide for each stage whether it will transmit data, perform a detection or switch to another FB. This type of problem can be interpreted as a reinforcement learning problem [9], which is an approach from the area of artificial intelligence. An agent (here the SU system) learns which decisions are good and which are not so good (or even bad) in different situations by earning different rewards for the decision made at each stage.

A. Simple Problem Formulation

To begin with, we assume that switching the SU system to another FB does not result in additional costs, so the SU system can choose an arbitrary FB at each stage and immediately transmit data. When using this simple assumption for the spectral resource detection problem, it can be formulated as a learning problem similar to the n-armed bandit problem [9]. It has only a single state but several actions $a \in \mathcal{A}$ with different expected rewards $Q^*(a)$, which is called the value of a. The actually received rewards after performing action a are samples that can be used to estimate the value of a. In the context of the detection of spectral resources, each action corresponds to the cycle of choosing a FB, performing one detection phase and transmitting data. Thus, a stage corresponds to the length of an update interval. The number of bits transmitted in stage t is the immediate reward. Note that the reward directly depends on the number of currently

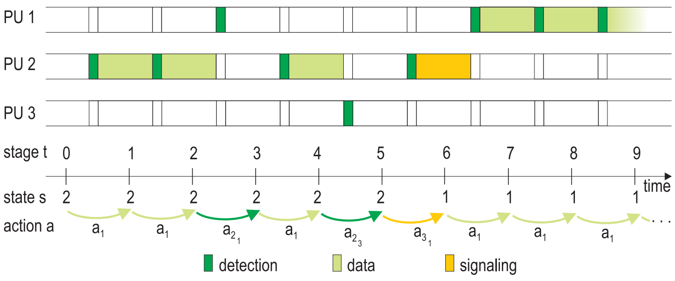

Fig. 2. Example for state transitions and actions of a SU system operating in a 3-band environment.

available subcarriers, which is determined by the detection phase at the beginning of each stage. Since the SU system does not know in advance how many resources are available in each FB, it has to find a good trade-off between exploitation and exploration, i.e. between playing safe and staying in the current FB where it knows how much data it can send, or switching to other FBs with the chance to find one with more resources, but also with the risk of switching to a band with less resources. The goal is to find the FB in which the SU system can currently transmit most data, i.e. the action with the highest value $Q^*(a)$. Since the SU system does not know the true Q-values it has to estimate them. $Q^*(a)$ is the true value of action a and it can e.g. be estimated by averaging the received rewards over time, when performing different actions. The estimated Q-values can be updated incrementally by applying the following update rule every time the action was executed:

$$Q_{k+1} = Q_k + \alpha[r_{k+1} - Q_k] \tag{1}$$

where α is a step-size parameter and r the received reward. One possibility to choose the next action based on the Q-values and achieving a trade off between exploration and exploitation is the softmax action-selection method:

$$Pr(a) = \frac{e^{Q(a)/\tau}}{\sum_b e^{Q(b)/\tau}} \tag{2}$$

giving the probability that action a is selected. τ can be used to adjust the behavior. For $\tau \to 0$ actions with the highest Q-value are preferred, whereas for large τ all actions are chosen with nearly the same probability [9]. Due to its simplicity, this scenario serves as a reference for the following scenario.

B. Problem Formulation as Markov Decision Process

In a more realistic scenario, switching the FB is not involved with zero costs, since the complete SU system with all participating stations has to be informed about the scheduled changes. This process requires additional signaling and time, and therefore reduces available resources for transmitting data. In contrast to the simple problem formulation with only one state and multiple actions, we now use a model with several states. The achieved reward in each stage depends on the action as well as the state the agent is in when performing the action. Assuming that the current state contains all necessary information regarding the history and the next action is chosen only depending on the current state, this type of problem can be modeled as a Markov decision process (MDP). A simple Markov decision process can be described by the tuple $\langle \mathcal{S}, \mathcal{A}, p, r \rangle$, where:

- $\mathcal{S}$ is the finite set of states the agent can be in.
- $\mathcal{A}$ is the finite set of all actions the agent can perform. Since in general the agent might not be able to perform every action in every state, $\mathcal{A}_s$ describes the set of possible actions in each state, with $\mathcal{A}_s \subseteq \mathcal{A}$ and $s \in \mathcal{S}$.
- $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ defines the state transition probability function $p(s'|s, a)$, giving the probability of transitioning to state $s' \in \mathcal{S}$ after performing the action $a \in \mathcal{A}$ while in state $s \in \mathcal{S}$.
- $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ defines the reward function $r(a, s)$, giving the reward the agent earns when performing action $a \in \mathcal{A}$ while in state $s \in \mathcal{S}$.

After observing the current state, the agent needs to choose an action for the next stage. This is done according to the policy $\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$, where $\pi(a, s)$ defines the probability that action a is executed when the agent is in state s. An optimum policy maximizes the cumulative expected rewards, which is usually discounted by a discount factor $\gamma \in [0, 1)$ in case of an infinite time horizon. Thus, the goal is to find an optimal policy that maximizes the expected return

$$R = E\left\{ \sum_{t=0}^{\infty} \gamma^t r_t(a_t, s_t) \right\}. \tag{3}$$

Before discussing solution strategies, we first formulate the spectral resource detection problem as an MDP. The available number of FBs N_{fb} in the scenario can be interpreted as the maximum number of states of the MDP, since the SU system (the agent) can only be in one FB at a time. Therefore, we define the set of states as $\mathcal{S} = \{1, 2, \ldots, N_{\text{fb}}\}$ and the current state is equivalent to the current FB in which the SU system is currently transmitting. The decisions the SU system has to make correspond to the actions. In each FB, the SU system can either transmit data in the current band, perform a detection phase in one of the other FBs or completely switch transmission to another FB. Hereby we assume that performing a detection phase in another FB is associated with different costs than completely switching to the other band. The set of possible actions in state s is $\mathcal{A}_s = \{a_1, a_{2_{\tilde{s}}}, a_{3_{\tilde{s}}}\}$ with $\tilde{s} \in \mathcal{S} \backslash s$ and the actions described as follows:

- a_1: perform a detection phase in the current FB s and transmit data.
- $a_{2_{\tilde{s}}}$: perform a detection phase in FB $\tilde{s}$ (Out-of-Band (OoB) detection).
- $a_{3_{\tilde{s}}}$: switch the SU system to FB $\tilde{s}$.

It is obvious that a state transition is only achieved when executing action $a_{3_{\tilde{s}}}$. Fig. 2 shows an example for a sequence of states and actions. For the immediate reward function r we use the following simple definition:

$$r(a, s) = \begin{cases} u_1(s) & \text{for } a = a_1 \\ u_2 & \text{for } a = a_{2_{\tilde{s}}} \\ u_3 & \text{for } a = a_{3_{\tilde{s}}} \end{cases} \tag{4}$$

where $u_1(s)$ is the number of radio resource goods that have been transmitted in the current stage while staying in the current FB. According to the assumed model, the PU system's allocation in each FB is a process indicating the number of available subcarriers, of which $u_1(s)$ is a sample. u_2 is the immediate reward/cost for performing a detection phase in a different FB. It is independent of the current state and setting $u_2 = 0$ often is a good choice: In the data transmission phase directly following the detection phase which was performed in a different FB, no data can be transmitted, because the

SU has no information about the allocation in the current FB, resulting in no immediate reward. On the other hand, also no additional signaling costs occur in this case. Finally, u_3 represents the costs for switching the transmission from one FB to another, including the necessary signaling effort. u_3 depends on the specific design of the applied protocol. In case that all signaling can be done within one update interval u_3 can be set to zero, but if it takes more than one update interval it has to be set to a negative value corresponding to the additional lost resources. Note that switching FBs usually becomes necessary when the current resources run low, i.e. only a small number of subcarriers is available. This means that the transmission of the signaling data is distributed over several update intervals.

C. Solution Strategy

Based on this problem formulation we now discuss a solution strategy using methods from reinforcement learning. For finding an optimal policy, many reinforcement algorithms use the concept of state-value functions $V^\pi : \mathcal{S} \to \mathbb{R}$, which map each state to a real value, describing how good it is to begin in this state and then following the policy π. For finding the optimal policy π^* and the optimal state-value function $V^*(s)$, reinforcement learning algorithms follow the idea of generalized policy iteration [9], where the policy is adapted according to the current state-value function which then evaluates the adapted policy. This iteration is repeated until the current policy and state-value function are close enough to their optimum.

Let us now summarize some of the basic properties of the described spectral resource detection problem. Although the state transition probabilities are simple and known for our problem, we do not know the exact reward of an action before executing it. This means that we don't have full knowledge regarding the system model. Additionally, we have a non-episodic scenario which continues infinitely and does not have a final state in case the average allocation in the different FBs is time variant. This recommends an on-line approach, i.e. learning is done while facing the real problem and not in a separate training phase. Furthermore, in case of a non-stationary allocation process, time has a negative effect on the precision of the state-values' estimation. The precision decreases the longer ago the last detection phase was performed in a specific FB.

Temporal-difference (TD) learning is one of the reinforcement learning approaches that is suitable for our type of problem. Especially the actor-critic methods are interesting, because they use a memory structure that explicitly represents the policy independent of the value function [9]. The actor-critic architecture is illustrated in Fig. 3. The critic uses the received reward to update the state values and to generate some information (in general the TD-error) which is then used by the actor to update the policy. In our case, the critic additionally calculates information about the reliability of the state-values.

1) Critic: The critic updates the state-values according to the following update rule:

$$V(s_t) \leftarrow V(s_t) + \beta \left[r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \right], \tag{5}$$

where β is a positive step-size parameter and γ a discount factor. The second part of Eqn. (5) represents the TD-error:

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t). \tag{6}$$

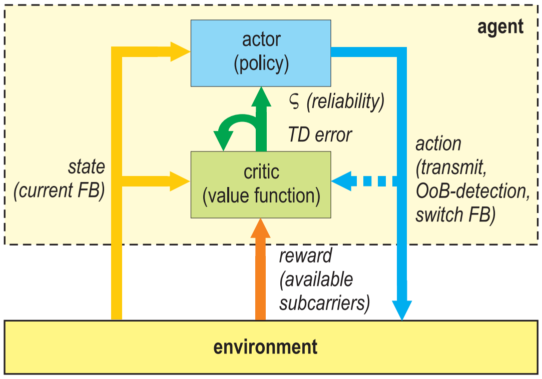

Fig. 3. Interaction of the secondary user with the environment and the applied actor-critic structure.

The reliability of the state-values $V(s)$ is expressed by corresponding reliability values $\varsigma(s)$ with $0 \leq \varsigma \leq 1$. Small ς indicate a low reliability and large ς a high reliability of the state-values. To update the reliability values, the critic needs to know which action was performed. In each stage, only the reliability value of the FB is increased in which a detection was performed (accordingly, this happens only when executing a_1 or $a_{2_{\tilde{s}}}$). In all other cases the reliability values are decreased. Note that the received reward is not relevant for the reliability. One possible update rule for each ς is

$$\varsigma_t \leftarrow \varsigma_t + \kappa \left[d - \varsigma_t \right] \tag{7}$$

where d is a binary indicator, describing whether a detection phase was performed in the corresponding FB in the current stage ($d = 1$) or not ($d = 0$). $\kappa \in (0, 1)$ is another positive step-size parameter. Note that every ς is updated in every stage. In case of action $a_{2_{\tilde{s}}}$, additionally $V(\tilde{s})$ is updated according to the following update rule:

$$V(\tilde{s}_t) \leftarrow V(\tilde{s}_t) + \alpha \left[\tilde{r}_{t+1} - V(\tilde{s}_t) \right], \tag{8}$$

where $\tilde{r}_{t+1}$ is the detection result in the target FB and α is a step-size parameter. $\tilde{r}_{t+1}$ is not an immediate reward since no data is transmitted, but it is nevertheless used to estimate $V(\tilde{s}_t)$.

2) Actor: Based on the TD-error δ and the reliability values $\varsigma(s)$ the actor now has to update the current policy. This is done by calculating preference values p for each action, based on which the policy $\pi_t(s, a)$ can then be derived e.g. by applying the softmax action selection method:

$$\pi_t(s, a) = Pr\left\{ a_t = a | s_t = s \right\} = \frac{e^{p(s,a)}}{\sum_b e^{p(s,b)}}. \tag{9}$$

The preferences are calculated in different ways, depending on the type of action. Action a_1 (transmitting data) is updated using a common update rule:

$$p(s, a_1) \leftarrow p(s, a_1) + \beta_1 \delta_t. \tag{10}$$

The focus of actions $a_{2_{\tilde{s}}}$ (performing a detection in FB $\tilde{s}$) is exploration. Therefore, it is preferable to perform detections in FBs, where the reliability of the state-values are low. This can be achieved by the following mapping of state-values to preferences:

$$p(s, a_{2_{\tilde{s}}}) = (1 - \varsigma) \cdot V(s). \tag{11}$$

Actions $a_{3_{\tilde{s}}}$ switch the SU system to another FB. Here it is desirable to switch to a FB with a lot of resources if

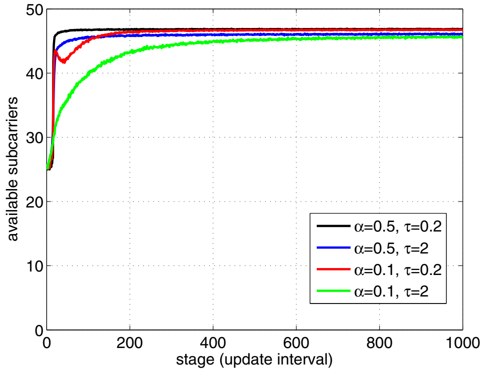

Fig. 4. Simple problem formulation: Average learning curve for different α and τ.

the information about the the resources is reliable. Unreliable information still is preferable in contrast to reliable information about low resources. The following equation gives the mapping:

$$p(s, a_{3_{\tilde{s}}}) = \varsigma \left(V(\tilde{s}) - \frac{N_{\text{fb}}}{2} \right) + \frac{N_{\text{fb}}}{2}. \qquad (12)$$

IV. Simulation Results

For all simulations we assumed $N_{\text{fb}} = 15$ different FBs, each of which having 50 channels. For the sake of simplicity the SU system uses for each PU channel one subcarrier, resulting in a maximum of 50 possibly available subcarriers at the same time. Based on the number of OFDM-symbols in each update interval and the physical layer configuration the available number of subcarriers can be directly translated to a bit rate. Each channel of the PU system is assumed to be available with the probability $P_{\text{avail}}(n)$ (with n denoting the FB), resulting in different binomial distributions with the parameters N_{fb} and $P_{\text{avail}}(n)$ for the number of available subcarriers in each FB. Note that $P_{\text{avail}}(n)$ is the same for each channel within one FB. All simulated learning curves in this paper show the average over 2000 episodes and for each episode and each FB $P_{\text{avail}}(n)$ was sampled from a uniform distribution which then stayed constant for the complete episode and therefore simulating a stationary scenario.

Fig. 4 shows the simulation results for the simple problem formulation similar to the n-armed bandit problem for different α and τ. The local optimum for $\alpha = 0.1$ and $\tau = 0.2$ results from the start values for the Q-values which were set to the maximum number of possibly available channels. This encourages exploration even if the FB with the most available channels was already visited.

The simulation results of the MDP and the proposed solution strategy described in Section III-C are shown in Fig. 5. The simulations show the performance gain for increasing α (with $\beta = 0.1$, $\beta_1 = 0.1$ and $\kappa = \frac{1}{2N_{\text{fb}}}$). For large α the performance is good (average of 43 available subcarriers after 1000 stages), but not as good as in the simple problem scenario (average of 46 to 47 available subcarriers after 1000 stages), since still detection phases are performed once in a while, which result in zero rewards and therefore decrease the transmission opportunities in average. Overall, the performance of our learning algorithm is good when choosing a suitable learning factor α. It has to be chosen in advance if the SU system is being deployed in a known environment or it has to be adjusted dynamically.

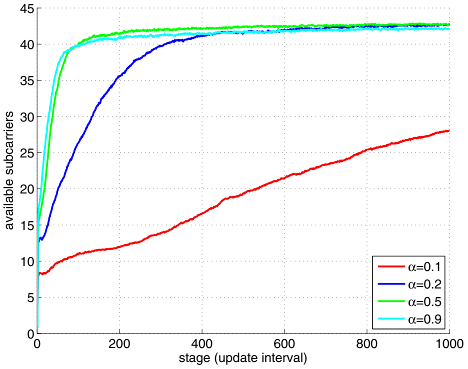

Fig. 5. Actor/critic method: Average number of available subcarriers for different α depending on the stage of the episode.

V. Conclusions and Future Work

In this paper we investigated an approach based on reinforcement learning for the detection of spectral resources in a multi-band cognitive radio scenario. Depending on the assumptions of the scenario, we developed a simple and a more complex model for describing the detection of spectral resources and proposed a solution strategy based on an actor/critic method. The simulation results show that our algorithm can quickly identify the FBs with the most available resources. The structure of the algorithm was developed in a way that enables an easy integration into the cross-layer optimization framework [10], and also is suitable for operating in an environment with a dynamically changing availability of spectral resources. These aspects are currently being investigated.

References

[1] M. McHenry, "NSF spectrum occupancy measurements," The Shared Spectrum Company, http://www.sharedspectrum.com/?section=nsf_measurements, Tech. Rep., 2005.

[2] S. Brandes, M. Schnell, U. Berthold, and F. K. Jondral, "OFDM based overlay systems - design challenges and solutions," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, 3-7 Sept. 2007, pp. 1–5.

[3] F. K. Jondral, "Cognitive radio: A communications engineering view," *IEEE Wireless Commun. Mag.*, vol. 14, no. 4, pp. 28–33, Aug. 2007.

[4] T. Weiß and F. K. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency," *IEEE Commun. Mag.*, vol. 42, pp. 8–14, 2004.

[5] U. Berthold and F. K. Jondral, "Distributed detection in OFDM based ad hoc overlay systems," in *IEEE 67th Vehicular Technology Conference, VTC Spring*, 2008.

[6] P. Papadimitratos, S. Sankaranarayanan, and A. Mishra, "A bandwidth sharing approach to improve licensed spectrum utilization," *IEEE Communications Magazine*, vol. 43, no. 12, pp. S10–S14, December 2005.

[7] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, vol. 4, 2007, pp. IV–1357–IV–1360.

[8] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Select. Areas Commun.*, vol. 25, no. 3, pp. 589–600, April 2007.

[9] R. S. Sutton and A. G. Barto, *Reinforcement learning*, [repr.] ed. MIT Press, 2004.

[10] F. Fu and M. van der Schaar, "A new theoretic framework for cross-layer optimization," in *Proc. IEEE Int. Conf. on Image Process. 2008 (ICIP 2008)*, to appear 2008.