Adaptive Sensing Scheduling and Spectrum Selection in Cognitive Wireless Mesh Networks

Marco Di Felice*, Kaushik Roy Chowdhury†, Andreas Kassler‡, Luciano Bononi*
* Department of Computer Science, University of Bologna, Italy
Email: {difelice,bononi}@cs.unibo.it
† Department of Electrical and Computer Engineering, Northeastern University, Boston, USA
Email:krc@ece.neu.edu
‡ Department of Computer Science, Karlstad University, Sweden
Email: kassler@ieee.org

***Abstract*—Cognitive Radio (CR) technology constitutes a promising approach to increase the capacity of Wireless Mesh Networks (WMNs). Using this technology, Mesh Routers (MRs) and the attached Mesh Clients (MCs) are allowed to opportunistically transmit on the licensed band, but under the constraint not to interfere with the Primary Users (PUs) of the spectrum. Thus, the effective deployment of CR-WMNs require that each MR must be able to: sense the current spectrum, select an available PU-free channel and perform the spectrum handoff to a new channel in case of PU arrival on the current one. How to coordinate these actions in the optimal way which maximizes the performance of the CR-WMNs while minimizing the interference to the PUs constitutes an open research issue in CR systems. In this paper, we propose an adaptive spectrum scheduling and allocation scheme which allows a MR to identify the best schedule of (*i*) when to sense the current channel, (*ii*) when to transmit, (*iii*) when to perform a spectrum handoff. Due the large number of parameters involved, we propose Reinforcement Learning (RL) techniques to allow a MR to learn by itself the optimal balance between spectrum sensing-exploitation-exploration actions based on network feedbacks coming from the MCs. We perform extensive simulations which confirm the adaptivity and efficiency of our approach in terms of increased throughput when compared with non-learning based schemes for CR-WMNs.**

I. Introduction

The potential of Wireless Mesh Networks (WMNs) as key technology for broadband Internet wireless access has been demonstrated by their recent deployments in the urban areas of Vienna [9], Berlin [8], Chaska [5]. In most of these cases, the wireless network is organized on a three-tier architecture composed by Mesh Clients (MCs), Mesh Routers (MRs) and Internet Gateways (IGWs) [12]. The MCs access the network through the MRs which form a static wireless backbone and are responsible to forward the traffic toward the closest IGW. Compared to the wired counterpart, WMNs can provide important benefits in terms of network scalability and deployment costs. For instance, the WMN deployed in [9] can serve an area of 40km of diameter through 400 MRs which are commercial routers equipped with 802.11g technology operating on the 2.4GHz license free ISM band. However, such a benefit might turn into a drawback if we consider the increasing saturation of the ISM bands. In urban areas, WMNs must share these frequencies with several devices, including other WiFi networks, cordless phones, remote controls for appliances, computer peripherals, and so on and thus their performance might be unable to support the requests of Quality-of-Service (QoS) of the MCs.

For this reason, a large amount of research in WMNs is investigating solutions to expand the network capacity through more flexible paradigms of spectrum access and sharing. Cognitive Radio (CR) technology constitutes the most interesting approach in this research area [1]. A CR node is an highly reconfigurable wireless devices which is able to adapt its transmitting parameters based on the characteristics of the environments and on the QoS requests of the end-users/applications. Since the reconfigurability spans all layer of the protocol stack, including the physical layer, a CR node is able to dynamically decide the frequencies where to transmit by accessing all the available spectrum, even the licensed ones. Thus, one of the main advantage of using CR technology over WMNs (CR-WMNs) is the possibility to allocate each MR-MC *cluster* on a different portion of the spectrum based on the MCs bandwidth requests. However, such an allocation must not produce any interference to the licensed owners of the spectrum, also called Primary Users (PUs).

Spectrum allocation in CR-WMNs is a non-trivial task involving spectrum *sensing*, *exploration* and *exploitation* at the same time. First, each cluster must be able to select a suitable channel for its operations. This requires the exploration of the available spectrum resources by performing frequent channel switching operations. Such exploration can be achieved by the MR alone or through collaboration with the MCs. Then, each cluster must decide the availability of each channel through sensing techniques at the physical layer [1] [4]. If a channel is found free from PU activity, then it can be used for opportunistic transmission by the cluster. Otherwise, the cluster must switch to another channel, thus continuing the exploration of the licensed spectrum. Frequency and duration of sensing constitute an important tradeoff between PU protection and channel exploitation. For instance, in Figure 1 from [6] we can see that the throughput experienced by a TCP flow in a CR scenario is significantly influenced by the sensing duration interval (T_s). In the given example, the optimal throughput is achieved when $T_s = 0.2$. When the sensing interval is too short ($T_s < 0.1$), the sensing process might not be accurate,

978-1-4577-0638-7 /11/$26.00 ©2011 IEEE

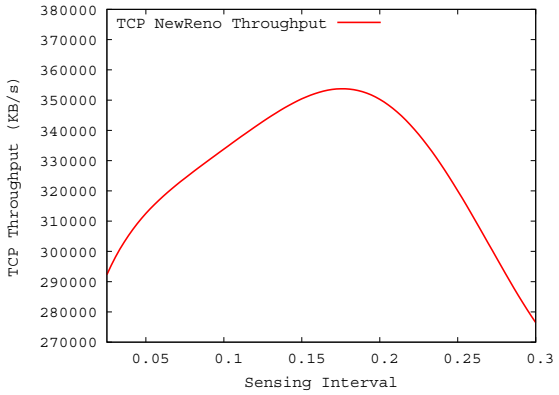


Fig. 1. TCP NewReno Throughput for different values of the sensing interval.

and thus frequent packet losses might be experienced due to PU interference on the current channel. When the sensing interval is too long ($T_s > 0.2$), the TCP source might be forced to reduce its sending rate due to the occurrence of timeout events caused by the sensing-induced delay.

In this paper, we address the issues of spectrum scheduling and allocation in CR-WMNs. Basically, we attempt to answer the following questions, on a per-cluster basis: (*i*) when and for how long to sense the spectrum, (*ii*) when to explore the licensed spectrum, (*iii*) when to transmit. Due to the large number of parameters involved and the highly dynamic nature of spectrum conditions, it is hard to formulate this problem using traditional optimization schemes [12]. For this reason, we propose an adaptive Spectrum Management (SM) scheme which allows each MR-MCs cluster to autonomously learn the best balance between spectrum sensing, switching and transmitting actions. We use Reinforcement Learning techniques [2] so that each MR will receive a network feedback after each action of sensing, transmitting or switching channel. Based on received rewards, we propose a probabilistic scheduling policy through which the MR will attempt to maximize the cluster throughput while guaranteeing the protection of PUs.

The paper is organized as follows. In Section II, we review existing proposals of SM schemes in CR-WMNs networks, by focusing on Machine Learning (ML)-based solutions. In Section III, we describe our system model. In Section IV, we formulate the problem through the RL paradigm. In Section V, we evaluate the performance of the proposed solution. Finally, Section VI contains Conclusions and Future Works.

II. Related Works

Spectrum sensing and selection constitute well investigated topics in classical CR networks. Several sensing techniques have been proposed in the literature to detect the presence of a PU signal on the current channel, e.g. energy detection [4] and cooperative sensing schemes [1]. In many cases, these techniques necessitate a periodic sensing structure, where sensing and transmission operations are performed in a periodic manner with separate observation period (e.g. T_s) and transmission period (e.g. T_x). In [10], the authors propose an optimization framework which derives the optimal value of T_s and T_x, subject to interference avoidance constraints. Also, the impact of sensing-induced delay on the upper layer of the protocol stack has been investigated in [6]. Analogously, several distributed and centralized spectrum selection schemes have been proposed for both classical multi-channel and cognitive WMNs [12]. However, few works address the mutual dependance between spectrum sensing and spectrum selection tasks. Reinforcement learning (RL) constitutes a promising approach for distributed parameter optimization in CR networks, in e.g. routing [13], spectrum sensing [3] and spectrum decision [14] tasks. Instead of addressing a single factor at a time, a RL agent can observe all the factors as a state, receive an aggregate feedback (e.g. the cost of each transmission) and optimize a general goal as a whole, e.g. throughput [13]. In [15], the authors propose a cognitive MAC protocol based on Partially Observable Markov Decision Process (POMDP). Similar to our paper, in [3] the authors use a RL approach to solve the problem of spectral resources in OFDM-based CR networks. However, the scheme proposed in [3] does not balance the trade-off between sensing and transmitting actions. Here, we extend our work proposed in [7] by considering cooperation techniques between the MR and the MCs nodes, and evaluating the ability of the RL scheme to dynamically adapt to varying PUs conditions.

III. Network Model and Architecture

We consider the network architecture shown in Figure 2. We assume that N MCs are associated to each MR, thus forming a *cluster*. Each MR works also as an IGW. Both the MCs and the MR nodes are equipped with two radios: (*i*) a control radio (e.g. R_{cont}), which is tuned to a Common Control Channel (CCC), and (*ii*) a spectrum-agile radio (e.g. R_{CR}) which operates in the licensed band. The presence of a fixed control interface introduces an additional cost which can be justified by the overhead reduction for broadcast messages, which otherwise would require to send multiple copies of the same packets on all the available channels. Finding a suitable CCC is out of the scope of this work. Without loss of generality, we assume that the licensed band is divided into K channels. All the nodes of a mesh cluster operate in the same channel, but different clusters may operate on different channels to avoid inter-cluster interference problems. Moreover, each channel can be occupied by a Primary User (PU), whose activity is defined through an alternative exponential ON-OFF scheme, called as a birth-death Markov process [10]: Let α_i be the death rate (or departure rate) for a PU on channel i (e.g. PU_i), then the duration of the ON state follows an exponential distribution with mean $\frac{1}{\alpha_i}$. Similarly, let β_i be the birth rate (or arrival rate) for PU_i, then the duration of the OFF state follows an exponential distribution with mean $1/\beta_i$.

Each cluster can operate in three modes: a *sensing* mode, a *transmitting* mode and a *switching* mode. The current operation mode of the cluster is decided by the MR, that can issue the `sense`, `switch`, and `transmit` commands to the MCs over the R_{cont} interface. When the MR issues

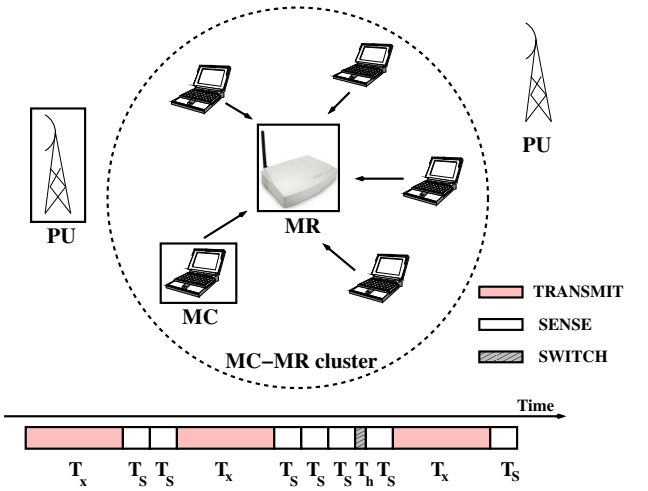


Fig. 2. The CR-WMN architecture with a MC-MR cluster. A possible schedule of commands issued by the MR is shown along the time scale.

a `sense` command, the cluster is in *sensing* mode. Then, each MC, e.g. MC_i senses the current channel, e.g. channel j, for an interval T_s. Based on the outcome of the energy-detector scheme, MC_i decides between the hypothesis H_0 (i.e. no PU detected on channel j) or H_1 (i.e. PU is active on channel j) and communicates its decision d_i to the MR on R_{cont}. When the MR issues a `switch` command, the cluster enters the *switching* mode for a time interval T_h. The `switch` command issued on the R_{cont} contains the channel information which the cluster will use at the end of the switching delay, and forces the MR and MCs to switch to the new channel in a synchronous fashion. Note that no data transfer occurs in the *sensing* and *switching* mode. Finally, the `transmit` command poses the cluster in the *transmitting* mode for a T_x interval, and allows the MCs and MR to engage in bi-directional communication over the R_{CR} interface.

IV. Learning-based Spectrum Management

Reinforcement learning (RL) [2] is a biologically-inspired learning paradigm in which an agent interacts with its environment over a potentially infinite sequence of discrete time steps, say *t*= 1,2,3,... Formally, a RL model is described by the Markov Decision Process (MDP), consisting in a triple: $\{S, A, R\}$. Here, $S = \{s_1, s_2, ..., s_N\}$ is the set of states, A is the set of actions permissible in each state, and $R : S \times A \longrightarrow \Re$ is a function which maps each state-action pair to a numeric reward. At each step t, the agent finds itself in a state, e.g. s_t chooses a permissible action a_t, and observes the ensuing reward $r_t(s_t, a_t)$. The action to be performed at each state is decided by the policy function $\pi : S \longrightarrow A$. The goal of the agent is to find the optimal policy π^* which chooses the best actions a^* in each state in order to maximize the long-term expected reward. Several algorithms have been proposed in the literature to determine the optimal policy π^*, e.g. Q-learning, Sarsa, Actor-Critic and so on. Here, we use the Q-learning algorithm with eligibility trace [2]. In Q-learning, each agent stores some additional information for each state/action, i.e.: (*i*) a value function $V : S \longrightarrow \Re$ which measures the opportunity for the agent to be in state $s, \forall s \in S$, (*ii*) a state-action function $Q : S \times A \longrightarrow \Re$ which measures the opportunity for the agent to be in state s and to perform action a, $\forall s \in S$ and $\forall a \in A$ and, (*iii*) an eligibility trace $E : S \longrightarrow \Re$ which is a metric reflecting the temporal validity of a state value $V(s)$, $\forall s \in S$.

In the following, we instantiate the RL model to the sensing scheduling and spectrum allocation problem for CR-WMNs.

Agent, States, Actions. In our problem, an agent is a MR. The set of states S is the set of channels (i.e. $S = K$). Thus, s_1 denotes channel 1, s_2 channel 2 and so on. On each state s, the MR can choose between the actions of sensing, transmitting or switching, i.e. $A = \{a_{se}, a_{sw}, a_{tx}\}$. When the MR performs a a_{sw} action from channel s to channel $\widetilde{s}$, this produces a state change from s to $\widetilde{s}$ in the MDP. No state changes occur in case of a_{tx} and a_{se} actions.

Eligibility Trace. Each time an agent performs an action a_t in state s_t, it updates the eligibility trace $E(s)$ of each state s as follows:

$$E(s) = \begin{cases} 1 & \text{if } s_t = s \text{ and } a_t = a_{se} \\ \phi \cdot E(s) & \text{otherwise} \end{cases} \quad (1)$$

where $0 < \phi < 1$ is a discount factor. Through Equation 1, we define the eligibility of a channel s as the temporal freshness of the sensing information. Each time we sense channel s, we reset its eligibility trace to the maximum value (i.e. 1). Otherwise, we decrease the $E(s)$ value of a ϕ factor if a different action is performed.

Sensing Reward. Each time a MR performs an a_{se} action, it issues a `sense` command on the CCC. Then, each MC_i senses the current channel s for a duration T_s and performs its binary decision (stored in d_i) between the hypothesis H_0 and H_1. Since different MCs can experience different propagation characteristics of the PU signal based on their own locations, the d_i values might not be in agreement for all the MC_i. For this reason, the MR computes the local reward of the a_{se} action as the average of the d_i values over the N MCs:

$$r(s, a_{se}) = \frac{\sum_{i=0}^{N} d_i}{N} \quad (2)$$

Basically, $0 \leq r(s, a_{se}) \leq 1$ estimates the probability that PU is active on the current channel s based on the majority of the decisions taken from the MCs. After computing $r(s, a_{se})$, the MR updates the Q-value as follows:

$$Q_{t+1}(s, a_{se}) = Q_t^i(s, a_{se}) + \gamma \cdot \overline{E(s)} \cdot \Big(r(s, a_{se}) - Q_t(s, a_{se}) \Big) \quad (3)$$

where $\overline{E(s)} = 1 - E(s)$ and $0 \leq \gamma \leq 1$ is a parameter which governs the convergency of learning.

Transmitting Reward. Each time a MR performs an a_{tx} action, it issues a `transmit` command on the CCC. We define the reward of the a_{tx} action as the average Packet Delivery Ratio

(PDR) experienced by a MC of the cluster, i.e:

$$r(s, a_{tx}) = \frac{\sum_{i=0}^{N} ACK_r^i}{\sum_{i=0}^{N} DATA_s^i} \quad (4)$$

where $DATA_s^i$ is the cumulative number of MAC frames sent by MC_i to the MR during the last T_x period and ACK_r^i is the number of MAC acknowledgment received by MC_i from the MR. It is easy to see that $r(s, a_{tx})$ estimates the reliability of channel s considering the aggregated impact of: PU-interference, interference from other clusters using the same channel, channel errors. After each T_x interval, the MR collects statistics from the MCs on the R_{CR} interfaces, computes the $r(s, a_{tx})$ value through Equation 4 and updates the Q-value as follows:

$$Q_{t+1}(s, a_{tx}) = Q_t^i(s, a_{tx}) + \gamma \cdot E(s) \cdot \Big(r(s, a_{tx}) - Q_t(s, a_{tx})\Big) \quad (5)$$

State Value. Based on the $Q_t(s, a_{tx})$ and $Q_t(s, a_{se})$ values, we define the state value of channel s (i.e. $V(s)$) as the empirical probability that a MC will successfully transmit a packet to the MR without causing any interference to the PU, i.e.:

$$V(s) = \Big(1 - Q_t(s, a_{se})\Big) \cdot Q_t(s, a_{tx}) \quad (6)$$

Switching Reward. We do not associate any specific reward to the a_{sw} action, i.e. $r(s, a_{tx}) = 0$. Instead, we estimate the Q-value of the a_{sw} action as the potential gain in terms of state value which can be obtained if the cluster switches from the current channel s to the channel $\widetilde{s}$ with the highest state value, i.e.:

$$Q_{t+1}(s, a_{sw}) = min\Big\{V(\widetilde{s}) - V(s), \theta_{t+1}\Big\} \quad (7)$$

Here, $\widetilde{s} = argmax_{h \in S} V(h)$ and θ_{t+1} is the probability to perform random channel exploration. In our approach, we perform intensive channel exploration during the network setup and we progressively reduce it over time, but without suppressing it at all. Thus, we initialize θ to θ_{MAX}, and we discount it of a δ factor after each action till a minimum value θ_{MIN} is reached, i.e.:

$$\theta_{t+1} = max\Big\{\delta \cdot \theta_t, \theta_{MIN}\Big\} \quad (8)$$

The optimal values of θ_{MAX}, θ_{MIN} and δ can not be determined a-priori, since they are scenario-dependant. Moreover, we force each cluster to stay tuned to the current channel for at least T_M time intervals to avoid the risk of excessive channel exploration without adequate channel exploitation.

Policy. Based on the Q-values, each MR updates its probabilistic policy π at the end of each period. The probability to select action a in state s at time t (i.e. $\pi_t(s, a)$) is computed by using the *soft-max* action selection method:

$$\pi_t(s, a) = \frac{e^{Q_t(s,a)}}{e^{Q_t(s,a_{tx})} + e^{Q_t(s,a_{se})} + e^{Q_t(s,a_{sw})}} \quad (9)$$

The complete learning scheme is shown by Algorithm 1.

Algorithm 1 Learning-based algorithm

```
for each time step t do
  decide next action a through the policy π
  if a == a_se then
    update Q(s, a_se) through Equations 2 and 3
  end if
  if a == a_tx then
    update Q(s, a_tx) through Equations 4 and 5
  end if
  if a == a_sw then
    decide next spectrum s̃ which maximizes V(s)
    perform state transition to s̃
    return
  end if
  update eligibility trace E(s) through Equation 1
  update Q(s, a_sw) through Equations 6,7,8
  update the policy π through Equation 9
end for
```

V. Performance Evaluation

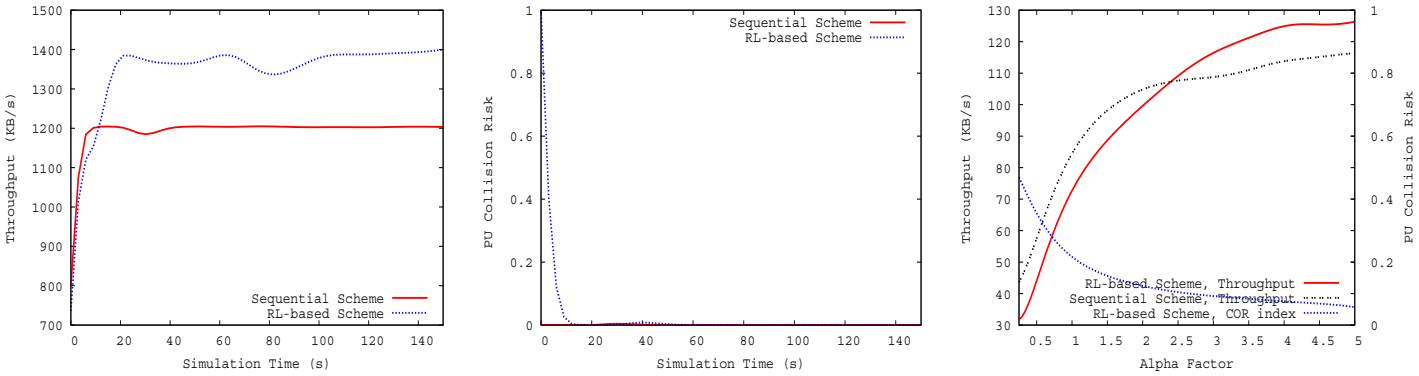
In this section, we investigate the performance of the RL-based Spectrum Management (SM) scheme described in Section IV. We use the NS-2 tool with the extension for the modeling and simulation of CR networks [6]. We consider three different analysis:

- *Single-cluster analysis*. We evaluate the accuracy and speed of convergence of the learning process in a scenario composed by a single MCs-MR cluster (Section V-A).
- *Sensing balancing analysis*. We analyze the ability of our SM scheme to balance the sensing and transmitting activities in an adaptive way under different PU activity patterns (Section V-B).
- *Multi-cluster analysis*. We investigate the ability of our SM scheme to provide effective channel allocation in a multi-cluster scenario. (Section V-C).

Unless stated otherwise, we use this configuration: T_x=1s, T_s=0.1s, T_h=0.001s, T_M=1s, θ_{MAX}=0.8, θ_{MIN}=0.2, ϕ=0.05, δ=0.05. The network topology and the traffic characteristics are described separately in the Sections below.

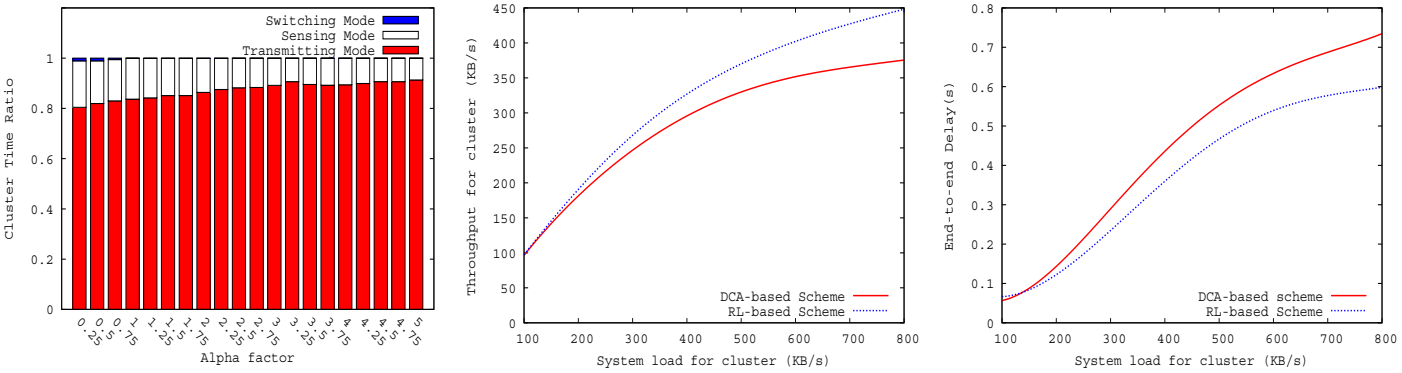
A. Single-cluster analysis

In Figures 3(a), and 3(b), we consider a CR network composed by a cluster of 10 MCs and 1 MR. Each MC delivers UDP packets to the MR with a Constant Bit Rate (CBR) of 100 KB/s and a packet size of 1000 bytes. We assume that the licensed band is divided into 5 different channels (i.e. $K = 5$). Each channel i can be occupied by a PU_i, whose activity is described by $< \alpha_i, \beta_i >$ parameters. In our configuration, we have: $\alpha = \{0.1, 0.1, 0.1, 0.1, 1\}$ and $\beta = \{1, 1, 1, 1, 0.1\}$. Here, we do not model the interference caused by adjacent clusters operating in the same channel. Instead, we consider only the interference caused by active PUs which transmit on the same channel used by the MCs-MR cluster. Under these assumptions, we evaluate the ability of our RL-based



(a) (b) (c)

Fig. 3. The network throughput and the COR index for the single-cluster scenario are shown in Figures 3(a) and 3(b), respectively. Figure 3(c) shows the same metrics as a function of the α factor which regulates the average ON time of each PU.



(a) (b) (c)

Fig. 4. Figure 4(a) shows the percentage of time spent in each cluster mode, as a function of the α parameter. The network throughput and the delay for the multi-cluster scenario are shown in Figures 4(b) and 4(c), respectively.

SM scheme to: (*i*) identify the channel which provides the highest transmission opportunity over time, and (*ii*) maximize the cluster throughput while limiting the impact to the PUs. We compare the performance of two SM schemes:

- *Sequential* SM scheme[1]: each cluster sequentially performs sensing (for a T_s interval) and transmitting activities (for a T_x interval). Once a PU signal is detected on the current channel, the cluster switches to a new channel in a round-robin order, till a PU-free channel is identified.
- *RL-based* SM scheme[2]: each cluster decides for sensing, transmitting or switching actions based on Algorithm 1.

Moreover, we consider two evaluation metrics: (*i*) the *network throughput*, which is defined as the amount of bytes per second successfully received by the MR from the MCs, and (*ii*) the *collision risk* (COR) index which is defined as the probability that a MC transmission will interfere with an active PU on the current channel. Basically, the network throughput measures the performance of the CR-WMN, while the COR index estimates the interference caused by the CR-WMN to the PUs. Figure 3(a) shows the network throughput over time (on the x-axis). Both the RL-based and the Sequential schemes experience an initial phase of *exploration* which is required to determine the best channel, i.e. the channel which provides the highest transmission opportunity for the CR-WMNs (channel 5 in our configuration). However, the RL-based scheme guarantees a better *exploitation* of the channel, since it progressively reduces the amount of sensing performed by the cluster under regime of low PU activity, while increasing the transmission opportunities for the MCs. As a result, the RL-based scheme provides higher throughput than the Sequential scheme. Figure 3(b) shows the collision risk (COR) index over simulation time (on the x-axis). Figure 3(b), shows that after the initial exploration phase the throughput enhancement provided by the RL-based scheme does not come at the expense of additional interference caused to the PUs.

[1]shortened to "Sequential Scheme" in Figures 3(a) and 3(b).
[2]shortened to "RL-based Scheme" in the Figures 3(a) and 3(b).

B. Sensing-balancing analysis

In this analysis, we evaluate the performance of the two SM schemes under different PU activity patterns. To this aim, we consider the single-cluster scenario defined above, with the same network and traffic parameters. However, we vary the α and β parameters which govern the ON/OFF states of the PU activity. Based on the α-β PU model described in

Section III, we recall here that the average ON time is given by the value of $\frac{1}{\alpha}$. Thus, increasing the α value is equivalent in reducing the average PU ON time on each channel, i.e. in increasing the spectrum opportunity for the CR-WMN cluster. In Figure 3(c), we show the aggregated network throughput (on the $y1$-axis) and the COR index (on the $y2$-axis) as a function of the α factor (on the x-axis). The β factor is fixed and equal to 1 for all the channels. Again, we compare the performance of the Sequential SM scheme and the RL-based SM scheme. Figure 3(c) confirms that the throughput of the CR-WMNs increases with the α factor, i.e. when we reduce the PU activity on each channel. For $\alpha < 2$, the Sequential scheme slightly enhances the RL-based scheme, due to the fact that it deterministically interrupts any transmission attempt till a PU-free channel is identified. Conversely, the RL-based scheme provides a consistent performance improvement under moderate PU activity (i.e. $\alpha > 2$), due to the balancing of channel exploration and exploitation provided by the learning algorithm. Moreover, such a balancing does not introduce any additional interference to the PUs, as shown by the COR index in Figure 3(c). In Figure 4(a) we show the ratio of time spent from the cluster in transmitting, switching and sensing modes for the RL-based scheme, as a function of the α factor (on the *x*-axis). Figure 4(a) shows how the RL-based scheme can dynamically adjust the exploration-exploitation tradeoff based on network feedbacks, and without assuming any a-priori knowledge of the PU activity. Under high PU activity conditions (i.e. α=0.25), the cluster is involved in channel sensing 20% of the time, in order to maximally protect the PU transmissions. However, such a percentage is reduced to less than 9% under moderate PU activity conditions (i.e. α=5), while the transmission time is increased up 90% in order to maximize the CR-WMN performance.

C. Multi-cluster analysis

In this scenario, the network is composed of 16 MRs placed on a square of 4×4 node grid. There are 4 MCs associated with each MR, thus the total number of the nodes of the CR-WMN is 80. We assume K=5, with the following PU activity pattern: $\alpha = \{0.01, 1.0, 1.0, 0.01, 1.0\}$ and $\beta = \{1.0, 0.01, 1.0, 1.0, 0.01\}$. Conversely to the single-cluster analysis, here we evaluate the ability of our SM scheme to perform distributed channel allocation under PU-interference and interference caused by other CR-WMN nodes. As term of comparison, we consider the dynamic channel assignment scheme (DCA) [11]. In DCA, each cluster chooses the less-interfered channel in its neighborhood, considering the number of mesh node interferers as selection metric. Moreover, the DCA scheme implements the Sequential SM schedule described in Section V-A. Figure 4(b) shows the average throughput for cluster, as a function of the system load produced in each cluster (on the x-axis). The DCA scheme avoids the interference caused by other CR-WMN nodes, but it experiences suboptimal spectrum selection because it does not take into account the PU interference. Thus, it might incur in frequent spectrum handoff operations caused by the PU detection on the current channel. Instead, the RL-based scheme takes into account both the cluster performance and the amount of PU interference in each channel, through the channel value metric (Equation 6). As a result, the RL-based scheme can provide higher throughput (Figure 4(b)) and lower delay (Figure 4(c)) than the DCA scheme under moderate and high traffic loads produced by the MCs.

VI. Conclusion and Future Works

In this paper, we have proposed the application of CR technology and RL techniques to increase the capacity of WMNs. We have described an adaptive sensing scheduling and spectrum allocation scheme through which a MR can learn an efficient balancing of channel sensing and transmitting actions. The simulation results have revealed that this approach can provide significant end-to-end performance gain for the WMNs, while guaranteeing the protection of the PU activity on each channel. We plan to investigate the relationship between cooperative and individual learning as future work, and to implement the proposed solutions on a CR testbed.

Acknowledgments

This research is supported by grant YR2009-7003 from Stiftelsen för internationalisering av högre utbildning och forskning (STINT).

References

[1] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty. NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey. *Computer Networks Journal*, 50(1), pp. 2127- 2159, 2006.

[2] A.G. Barto and R. Sutton. In *Reinforcement Learning: An Introduction*, MIT Press, Cambridge 1998.

[3] U. Berhold, F. Fu, M. Van Der Schaar and F. K. Jondral. Detection of Spectral Resources in Cognitive Radios Using Reinforcement Learning. In *Proc. of IEEE Dyspan*, pp. 1-5, 2008.

[4] D. Cabric, S. M. Mishra, and R. W. Brodersen. Implementation issues in spectrum sensing for cognitive radios in *Proc. of IEEE ACSSC*, pp. 772-776, 2004.

[5] Chaska wireless solutions. http://www.chaska.net/.

[6] M. Di Felice, K. R. Chowdhury and L. Bononi. Modeling and performance evaluation of transmission control protocol over cognitive radio ad hoc networks. In *Proc. of ACM MSWIM*, pp. 4-12, 2009.

[7] M. Di Felice, K. R. Chowdhury, W. Meleis and L. Bononi. To Sense or To Transmit: A Learning-based Spectrum Management Scheme for Cognitive Radio Mesh Networks. In *Proc. of IEEE WIMESH*, pp. 1-6, 2010.

[8] Freifunk wireless community. http://www.freifunk.net.

[9] FunkFeuer wireless community. http://www.funkfeuer.at.

[10] W. Y. Lee and I. Akyildiz. Optimal Spetrum Sensing Framework for Cognitive Radio Networks. in *IEEE Transactions on Wireless Communication*, 7(10), pp. 3845-3857, 2008.

[11] K. Pradeep and N. H. Vaidya. Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks. *ACM Mobile Computing and Communications Review*, 10(15):31–43, 2006.

[12] H. Skalli, S. Ghosh, S.K. Das, L. Lenzini and M. Conti. Channel assignment strategies for multiradio wireless mesh networks issues and solutions. in *IEEE Communications Magazine*, 45(11), pp. 86-95, 2007.

[13] B. Wahab, Y. Yang, Z. Fan and M. Sooriyabandara Reinforcement Learning Based Spectrum-aware Routing in Multi-hop Cognitive Radio Networks. In *Proc. of IEEE CROWNCOM*, 2009.

[14] K.-L. A. Yau, P. Komisarczuk and P. D. Teal. A Context-aware and Intelligent Dynamic Channel Selection Scheme for Cognitive Radio Networks. In *Proc. of IEEE CROWNCOM*, 2009.

[15] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework. in *IEEE JSAC*, 25(3), pp. 589-600, 2007.