secondary users themselves to determine their individual best sensing window length. This decentralized allocation avoids the introduction of signaling overhead in the system.

III. POWER ALLOCATION PROBLEM FORMULATION

We consider a large circular primary cell made up of one central primary emitter and several primary receivers whose positions are unknown. The primary emitter could be a DTV broadcasting station that communicates with multiple passive receivers.

The secondary network uses the same frequency band as the primary network and consists in $L$ adjacent secondary cells. Each secondary cell is made up of one central secondary base station and multiple secondary users. For the sake of simplicity, all the secondary users (SU) are assumed to be located on the line that joins the $L$ base stations (BS) as illustrated on Figure 2. The reader is referred to [13] for more realistic assumptions regarding the geometry of the power allocation problem.

In order to protect the primary receivers from receiving harmful interference from the secondary users, a protection contour is defined around the primary emitter as a circle on which the received primary SINR must be superior to a given threshold $SINR_{th}^p$. The secondary cells are located around the protection contour. As the primary cell ray is assumed to be much larger than the secondary cells ray, the protection contour can be approximated by a line parallel to the secondary base stations line.

The secondary network is assumed to follow a Time Division Multiple Access (TDMA) scheme, so that at each time only one secondary user $SU_i$ communicates with its base station $BS_l$ in cell $l$ ($l \in \{1, \ldots, L\}$). The difference between $SU_i$ and $BS_l$ absissa is denoted $x_l$. The point on the protection contour whose distance with $SU_i$ is minimal is denoted $I_l$. We assume that each cell $l$ deploys sensors on the protection contour so that it is able to measure the primary network SINR at the point $I_l$, denoted $SINR_i^p$.

In this paper, the analysis is focused on the interference generated by the upstream transmissions of the secondary users. It is assumed that the secondary SINR at each base station $l$, denoted $SINR_i^s$, needs to be superior to a given threshold $SINR_{th}^s$ for the secondary communication to be reliable.

The power allocation problem consists in finding the optimal secondary users transmission powers $\{P_1, \ldots, P_L\}$ that minimize a cost function $f(SINR_1^s, \ldots, SINR_L^s)$ depending on the secondary SINRs, under the constraints that

$$SINR_i^s \geq SINR_{th}^s \quad \forall l \in \{1, \ldots, L\}$$

(12)

In this paper, the following cost function is considered:

$$f(SINR_1^s, \ldots, SINR_L^s) = \sum_{l=1}^{L} (SINR_i^s - SINR_{th}^s)^2$$

(13)

It is observed that the cost decreases with respect to $SINR_i^s$ until $SINR_i^s$ reaches the threshold value $SINR_{th}^s$, then the cost increases with respect to $SINR_i^s$. This should prevent secondary users from selfishly transmitting with a power higher than required, which would remove transmission opportunities for other secondary users.

The primary SINRs in Equation (12) are given by:

$$SINR_i^p = \frac{P_p}{\sigma^2 + \sum_{k=1}^{L} P_k h_{SU_k}}$$

where $P_p$ is the power that is received on the protection contour from the primary transmitter, $\sigma^2$ is the noise power and $h_{SU_k}$ is the link gain between SU and the point $I_l$ on the protection contour.

The secondary SINRs in Equation (13) are given by:

$$SINR_i^s = \frac{P_i h_{SU_k}}{\sigma^2 + \sum_{k=1, k \neq l}^{N} P_k h_{BS_l}}$$

where $h_{SU_k}$ is the link gain between SU and $BS_l$.

In this paper, we consider free space path loss. Therefore, the link gains are computed as follows:

$$h_{SU_k} = \left( \frac{4\pi f_c}{c} \sqrt{r_s^2 + (2(k-l)r_s - x_l + x_k)^2} \right)^{-2}$$

(14)

$$h_{BS_l} = \left( \frac{4\pi f_c}{c} (2(k-l)r_s + x_k) \right)^{-2}$$

(15)

where $r_s$ is the ray of the secondary cells, $f_c$ is the transmission frequency and $c$ is the speed of light in vacuum.

IV. LEARNING ALGORITHM

A. Q-Learning Algorithm

In this paper we use two multi-agent Q-learning algorithms. The first one is used to allocate the secondary user sensing times and the second one is used to allocate the secondary user transmission powers. In the sensing time allocation algorithm, each secondary user is an agent that aims to learn an optimal sensing time allocation policy for itself. In the power allocation algorithm, each secondary base station is an agent that aims to learn an optimal power allocation policy for its cell.

Q-learning implementation requires the environment to be modeled as a finite-state discrete-time stochastic system. The set of all possible states of the environment is denoted $S$. At