A POMDP Framework for Cognitive MAC Based on Primary Feedback Exploitation

Karim G. Seddik[1] and Amr A. El-Sherif[2]
[1]Electronics and Communications Engineering Department, American University in Cairo, New Cairo 11835, Egypt.
[2]Department of Electrical Engineering, Alexandria University, Alexandria 21544, Egypt.
email: kseddik@aucegypt.edu, amr.elsherif@ieee.org

***Abstract*—In this paper, a design for a cognitive MAC protocol based on Primary User (PU) feedback exploitation is proposed. A queuing approach is adopted and an infinite-state Partially Observable Markov Decision Process (POMDP) framework is proposed where the states represent the number of packets in the primary queue. The primary user quality of service (QoS) guarantee is defined through a primary queue stability constraint. Finally, we propose a greedy algorithm to simplify the design of the MAC protocol. Existing techniques and results for POMDP can then be used to develop MAC protocols.**

***Index Terms*—Cognitive Radio, POMDP, Queue Stability**

I. Introduction

Cognitive Radio technology is a communication paradigm that emerged in order to solve the spectrum scarcity problem by allowing secondary users (SUs) to exploit the under-utilized spectrum of the primary users (PUs). Coexistence of such SUs along with PUs is allowed under the condition that some minimal quality of service (QoS) level is guaranteed for PUs. Several works have considered the design of SU MAC protocols to allow for secondary access with primary QoS constraint(s).

In this paper, we consider the design of SU MAC protocol based on PU feedback exploitation. Several works have considered the use of PU feedback information to design the SU access protocol. For instance, in [1], the SU observes the automatic repeat request (ARQ) from the primary receiver. The ARQ feedback messages reflect the PU's achieved packet rate. The cognitive radio's objective is to maximize the secondary throughput under the constraint of guaranteeing a certain packet rate for the PU. Secondary power control on the basis of the primary link feedback is investigated in [2]. The objective was to maximize the SUs' utility, in a distributed manner, while maintaining a PU outage constraints. In [3], the optimal transmission policy for the SU, when the PU adopts a retransmission based error control scheme, is investigated. The policy of the SU determines how often it transmits according to the retransmission state of the packet being served by the PU.

A simple idea was introduced in a previous work [4] in which SUs refrain from accessing the channel upon hearing a NACK from the primary receiver allowing for an interference-free primary retransmission, thereby increasing secondary throughput and decreasing primary packet delay. Also, in [5], the use of PU feedback information along with soft energy sensing was used to design the SU access scheme with a PU QoS constraint defined in terms of the PU queue stability.

In this paper, we formulate the SU MAC protocol design as a partially observable Markov decision process (POMDP). Different from [4], [5], where only the last PU feedback bit is considered, we use the PU feedback history to make the SU access decisions (which allows the SU to have perfect information about the PU service process). Our formulation will results in a POMDP with infinite number of states which is, in general, very difficult to solve and this is why we resort to the design of a simple, greedy algorithm that will make the access decisions based on maximizing the instantaneous SU reward. Finally, we present some simulation results to compare our proposed greedy algorithm with the algorithm of [4]. Also, we show that our proposed greedy algorithm will guarantee the PU queue stability.

II. System Model

In our model, we consider a time-slotted network with one primary user and one secondary user (extension to multiple PUs with TDMA is straightforward). The PU has an infinite buffer for storing its incoming packets. The packet arrival process is assumed to be Bernoulli i.i.d. with an average arrival rate of λ_p packets-per-time slot. A slot duration is equal to the packet transmission time, and therefore, we assume $0 \leq \lambda_p \leq 1$ or else the queues will not be stable. Finally, we consider the case where SUs always have packets to send.

Furthermore, we adopt a collision model, for simplicity of presentation, in which whenever more than one transmission proceeds at a time, all packets involved are lost. A situation in which the interference is assumed to be too high for the receivers to have a decodable signal.

III. The POMDP Framework

In our proposed framework, the SU will be able to receive the primary feedback bits in the form of ACKs, NACKs and No-Feedback (No-FB). Based on the received feedback sequence, the SU can form its "belief vector" about the primary queue status; the PU queue Markov chain is shown in Fig. 1.

In Fig. 1, we have two classes of states, namely, the i_F's and i_R's states where the subscript F denotes first transmission and R denotes a retransmission after a PU NACK. Our system

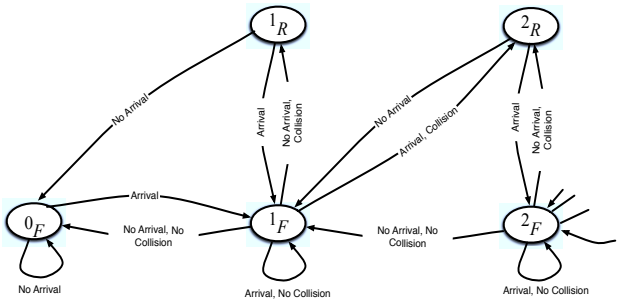

Fig. 1: The PU queue Markov chain model

model settings allow us to model the SU access decisions as a PODMP.

The POMDP is characterized by the tuple $(\mathcal{S},\ \mathcal{A},\ \mathcal{O},\ T,\ \Omega,\ R)$; the set $\mathcal{S}$ denotes the states of the PU queue's Markov chain, $S = \{\{i_F\}, \{j_R\}\}$, $i = 0, 1, \cdots$ and $j = 1, 2, \cdots$. The set $\mathcal{A}$ denotes the set of SU actions, i.e., $\mathcal{A} = \{\text{access, no access}\}$. The set $\mathcal{O}$ denotes the set of observations which is given by $\mathcal{O} = \{\text{ACK, NACK, No-FB}\}$. The set T denotes the set of transition probabilities. $T(s'|s,a)$ is the probability of the queue to move into state s' if the action a is taken while the queue is in state s. The transition probabilities are given by

$$\begin{aligned}
&T(i_R|j_F,\ \text{no access}) = 0,\ \forall i,\ j\\
&T(i_F|j_F,\ \text{access}) \quad = 0,\ \forall i,\ j \neq 0\\
&T(1_F|0_F,\ \text{access}) \quad = \lambda_p,\\
&T(1_F|0_F,\ \text{no access}) \quad = \lambda_p,\\
&T(i_R|j_F,\ \text{access}) \quad = \begin{cases} 1-\lambda_p & \text{if } i = j,\ j \neq 0\\ \lambda_p & \text{if } i = j+1,\ j \neq 0\\ 0 & \text{otherwise,}\end{cases}\\
&T(i_F|j_R,\ \text{no access}) = \begin{cases} 1-\lambda_p & \text{if } i = j-1\\ \lambda_p & \text{if } i = j\\ 0 & \text{otherwise.}\end{cases}
\end{aligned} \tag{1}$$

The set Ω denotes the set of conditional observation probabilities. $\Omega(o|s',a)$ is the probability of observing o when the state is s' after taking the action a and can be calculated as follows (details are omitted due to space limitations)[1].

$$\begin{aligned}
&\Omega(o|s' = i_F,\ \text{no access})\\
&= \begin{cases}
0 & o = \text{NACK and } \forall i_F\\
P_{\text{ACK}}(0_F) & o = \text{ACK},\ i_F = 0\\
P_{\text{No-FB}}(0_F) = 1 - P_{\text{ACK}}(0_F) & o = \text{No-FB},\ i_F = 0\\
P_{\text{ACK}}(1_F) & o = \text{ACK},\ i_F = 1\\
P_{\text{No-FB}}(1_F) = 1 - P_{\text{ACK}}(1_F) & o = \text{No-FB},\ i_F = 1\\
1 & o = \text{ACK},\ i_F \geq 2\\
0 & o = \text{No-FB},\ i_F \geq 2
\end{cases}.
\end{aligned} \tag{2}$$

[1]By abuse of notations, we set $\Pr(A|B) = 0$ if $\Pr(B) = 0$ (for example we set $\Omega(o|i_R,\ \text{no access}) = 0$ since $\Pr(i_R,\ \text{no access}) = 0$ under our collision system model assumption).

Note that the values of $P_{\text{ACK}}(0_F)$, $P_{\text{No-FB}}(0_F)$, $P_{\text{ACK}}(1_F)$ and $P_{\text{No-FB}}(1_F)$ can be calculated based on the previous state belief vector (to be defined later) but their values will not affect our formulation as will become clear later since the SU reward under no access will always be 0.

$$\Omega(o|i_F,\ \text{access}) = \begin{cases} 0 & \forall o \text{ and } i_F \geq 2\\ 1 & o = \text{No-FB},\ i_F = 0, 1\\ 0 & o = \text{ACK or NACK},\ i_F = 0, 1\end{cases} \tag{3}$$

$$\Omega(o|i_R,\ \text{no access}) = 0 \quad \forall o \tag{4}$$

$$\Omega(o|i_R, \text{access}) = \begin{cases} 1 & o = \text{NACK}\\ 0 & \text{otherwise}\end{cases}. \tag{5}$$

The reward function, R, is defined as follows.

$$R(s,a) = \begin{cases} 1 & a = \text{access},\ s = 0_F\\ 0 & a = \text{no access},\ \forall s\\ -w & a = \text{access},\ s \neq 0_F\end{cases}. \tag{6}$$

If the SU accesses the channel while the PU queue is empty it will gain a reward of 1 successful packet transmission. If the SU decides not to access, the reward will be 0. If the SU decides to access while the PU queue is nonempty, then a collision occurs and the SU reward will be $-w$, $w \geq 0$, since in this case the PU queue will have a higher probability of being nonempty in the next time slots. The value of the design parameter w can be controlled to adjust the level of protection provided to the PU. If $w = 0$, then the optimal decision to maximize the SU reward will be always to access the channel (since there is no penalty in accessing the channel) so in this case the SU will aggressively access the channel and this can cause excessive delays and queue instability at the PU. As we increase w the SU will be less aggressive in accessing the channel and a better service will be provided to the PU.

If we start at a certain belief vector $\mathbf{b}(s_t) = [b(0_F)_t,\ b(1_F)_t,\ b(1_R)_t,\ \cdots]$, where t is the time index, then the new belief vector after taking an action a_t observing some o_{t+1} is given by

$$b(s_{t+1}) = \eta\Omega(o_{t+1}|s_{t+1}, a_t) \sum_{s_t \in \mathcal{S}} T(s_{t+1}|s_t, a_t) b(s_t), \tag{7}$$

where η is a normalization factor given by

$$\eta = \frac{1}{\sum_{s_{t+1} \in \mathcal{S}} \Omega(o_{t+1}|s_{t+1}, a_t) \sum_{s_t \in \mathcal{S}} T(s_{t+1}|s_t, a_t) b(s_t)}.$$

A. POMDP MAC Policy

In this section, we consider the MAC policy design. The policy is supposed to map the belief vector to the action space. Note that the current action affects the reward in two aspects, the current state reward and the expected reward in the next states as governed by the underlying Markov chain dynamics

(since the current action will affect the belief vector in the upcoming time instants). The MAC design can be modelled as a belief-based Markov decision process (belief MDP). The current expected reward, corresponding to a belief vector b and an action a, is given by

$$r(b,a) = \sum_{s \in \mathcal{S}} b(s) R(s,a). \tag{8}$$

The SU access policy π specifies an action $a = \pi(b)$ for any belief b. Here it is assumed that the objective is to maximize the expected total reward over an infinite horizon. The expected reward for policy π starting from belief b_0 is defined as

$$J^{\pi}(b_0) = \sum_{t=0}^{\infty} \gamma^t r(b_t, a_t) = \sum_{t=0}^{\infty} \gamma^t E\Big[R(s_t, a_t) \mid b_0, \pi\Big] \tag{9}$$

where $\gamma < 1$ is the discount factor. The optimal policy π^* is obtained by optimizing the long-term reward.

$$\pi^* = \underset{\pi}{\text{argmax}} \;\; J^{\pi}(b_0) \tag{10}$$

where b_0 is the initial belief.

The optimal policy, denoted by π^*, yields the highest expected reward value for each belief state, compactly represented by the optimal value function V^*. This value function is the solution to the Bellman optimality equation:

$$V^*(b) = \max_{a \in \mathcal{A}} \Big[r(b,a) + \gamma \sum_{o \in \mathcal{O}} \Omega(o \mid b, a) V^*(\tau(b,a,o)) \Big], \tag{11}$$

where $\tau(\cdot,\cdot,\cdot)$ is the belief state transition function.

B. Greedy Algorithm

Solving the exact PODMP problem given in (11) is computationally demanding and this does not lead to efficient practical design of the access policy. In this section, we propose a greedy algorithm to simplify the design of the SU access policy. In the proposed algorithm, the SU access decisions will be made to maximize the instantaneous SU reward. The instantaneous expected SU reward is given by

$$r_t(a) = \begin{cases} 1 \cdot \Pr(Q_t = 0) + (-w) \cdot \Pr(Q_t \neq 0), & a = \text{access} \\ 0, & a = \text{no access} \end{cases}. \tag{12}$$

where t is the time index and Q_t is the primary queue length at t.

The SU access decision will be based on the instantaneous reward. A no-access decision will result in a reward of 0 and the reward for the access decision is $1 \cdot \Pr(Q_t = 0) + (-w) \cdot \Pr(Q_t \neq 0)$. So the access decision can be simplified to comparing $\Pr(Q_t = 0)$ to the threshold $w/(1+w)$. The access algorithm can be written as

$$\text{Greedy Algorithm: } = \begin{cases} \text{access} & \text{if } b(Q_t = 0_F) > \frac{w}{1+w} \\ \text{no access} & \text{if } b(Q_t = 0_F) \leq \frac{w}{1+w} \end{cases}, \tag{13}$$

where $b(Q_t = 0_F)$ is the belief at time instant t that the PU is empty.

Another implementation friendly aspect of the proposed greedy algorithm is that we do not need to keep the whole belief vector. At each time we need to calculate the probability of having an empty PU queue. This probability can be calculated at the secondary users by just keeping the feedback information (the number of ACKs from the last No-FB transmission). The probability of having an empty queue at the N-th time instant from the last No-FB transmission given that we have received K ACKs during this duration conditioned on having an ACK as the last feedback bit is given by

$$\begin{aligned} &\Pr(Q_N = 0_F | K \text{ ACKs, last feedback was an ACK}) \\ &= \Pr\left(\text{exactly } K \text{ arrivals in } N \text{ time slots} | \text{numbers of arrivals} \geq K\right) \\ &= \frac{\binom{N}{K} \lambda_p^K (1-\lambda_p)^{N-K}}{\sum_{n=K}^{N} \binom{N}{n} \lambda_p^n (1-\lambda_p)^{N-n}}, \end{aligned} \tag{14}$$

where, by abuse of notation, we let Q_N denote the PU queue length after N time instants from the last No-FB transmission.

Clearly, if the last feedback bit was a NACK then $\Pr(Q_t = 0_F |$ last feedback was a NACK$) = 0$ since clearly the PU queue will have at least one packet to re-transmit. If the last feedback bit was a No-FB then $\Pr(Q_t = 0_F |$ last feedback was a No-FB$) = 1 - \lambda_p$, which is the probability of no PU arrival in the last time slot.

In the case of overhearing a NACK, then our greedy algorithm will result in a no access decision at the secondary user (for the POMDP formulation, the optimum action after overhearing a NACK is also for the SU to back off, which can be easily proved).

Note that the model presented here can be easily extended to incorporate any spectrum sensing approach, e.g., energy detection. The *soft information* from any spectrum sensing can be used to modify the current belief vector and our proposed greedy algorithm can still be applied.

C. PU Queue Stability for the Greedy Algorithm

Stability can be loosely defined as keeping a quantity of interest bounded, in our case, the queue size. For a more general and rigorous definition of stability, see [6] and [7]. If the arrival and service processes of a queuing system are strictly stationary, one can apply Loynes' theorem to check for stability [8]. This theorem states that if the average arrival rate is less than the average service rate of a queuing system whose arrival and service processes are strictly stationary, then the queue is stable, otherwise it is unstable. In our setup, the stability of the PU queue can be guaranteed if

$$\lim_{t \to \infty} \Pr(Q_t = 0_F) > 0,$$

i.e., there is always a non-zero probability of having an empty PU queue.

In our proposed feedback-based framework, the SU has "perfect knowledge" of the service process of the PU queue

(through overhearing the PU feedback bits). It is straightforward to prove that any $w > 0$ will guarantee that the proposed greedy algorithm will result in a stable PU queue. Since for any $w > 0$ the probability of an empty PU queue will never reach 0 since in this case the SU will decide not to access the channel (as the reward from accessing the channel will be $-w$ so the no access will result in a higher SU reward of 0). So the greedy algorithm will always guarantee that the probability of having an empty PU queue is bounded away from 0 as long as $w > 0$.

IV. Numerical Results

In [4], a MAC protocol was designed where the SU employs random access based on the last overheard feedback bit under a PU queue stability constraint. In this paper, we have considered the use of the feedback information from all the previous slots to make access decisions.

In Figures 2 and 3, we compare the performance of our proposed greedy algorithm and the algorithm that was proposed in [4], denoted by FB in the figures, in terms of the SU throughput and the PU delay. In the results shown, it is assumed that the channels are perfect and the only source of error at the receivers is the collisions between the primary and secondary transmissions.

It is noted that the case of $w = 0$ corresponds to the case where the SU always accesses the channel. As w is increased, the SU becomes less aggressive in accessing the channel. Therefore, for small values of the PU arrival rate λ_p the SU is losing some of the available transmission opportunities and its throughput is lower than that in the case of $w = 0$ and the FB algorithm, as shown in Fig. 2. On the other hand, for larger values of λ_p, the greedy algorithm outperforms the FB algorithm, since it always has a better estimate of the PU activity, hence avoiding collisions and better utilizing the free time slots. Increasing w in this case limits the throughput region of the SU. For instance, for $w = 0.5$ the SU is able to access the free time slots for up to a PU $\lambda_p = 0.65$. When w is increased to 2.0, the SU refrains from any transmission attempt for λ_p beyond 0.35.

From the delay point of view as shown in Fig. 3, for $w = 0$ the delay is increasing rapidly as the SU is always trying to access the channel. For increasing values of w, the delay is slowly decreasing with λ_p. Since as λ_p increases, more feedback information will be available to the SU, which will have better estimate of the PU state leading to less collisions. After a given λ_p (this threshold is decreasing in w), the PU's delay drops, indicating that the SU has decided not to transmit at all, which is in conformance with the observation about the SU's throughput. Note that for the algorithm that was proposed in [4], the optimum access probability on overhearing an ACK or No-FB was 1 for $\lambda_p \leq \frac{1}{3}$, and this is why the curves for the algorithm that was proposed in [4] and our greedy algorithm with $w = 0$ coincide for $\lambda_p \leq \frac{1}{3}$.

Note that the value of w controls the SU throughput and the PU delay and it is clear that for every range of PU arrival rates we can optimize the value of w for some cost function

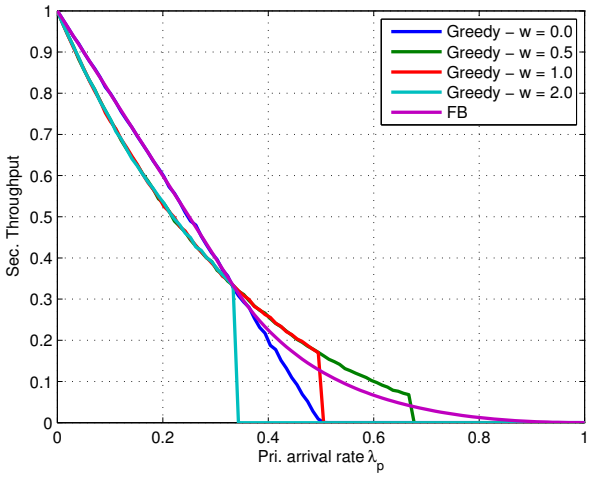

Fig. 2: The SU throughput

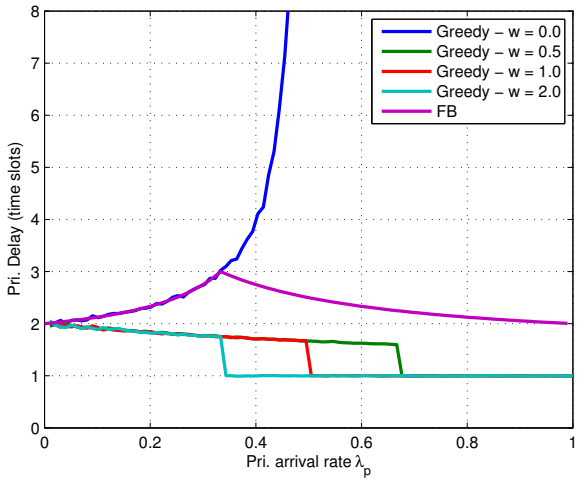

Fig. 3: The PU delay

(like maximizing the SU throughput under some PU delay constraint), which will be a point for future consideration.

V. Conclusions

In this paper, we have considered the design of a SU MAC protocol with PU feedback exploitation. The MAC protocol design was formulated in a partially observable Markov decision process (POMDP) framework with infinite number of states. Due to the inherent complexity of solving POMDP with infinite number of states, we have proposed a greedy algorithm that maximizes the SU instantaneous reward, in which the SU access decisions will depend on the overheard PU feedback bits (which allows the SU to have perfect knowledge of the PU service process). The proposed greedy algorithm will guarantee the PU stability and it can be designed to provide a PU QoS constraint in terms of average PU delay.

REFERENCES

[1] K. Eswaran, M. Gastpar, and K. Ramchandran, "Bits through arqs: Spectrum sharing with a primary packet system," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.

[2] S. Huang, X. Liu, and Z. Ding, "Distributed power control for cognitive user access based on primary link control feedback," in *IEEE International Conference on Computer Communications (INFOCOM)*, San Diego, CA, March 2010.

[3] M. Levorato, U. Mitra, and M. Zorzi, "Cognitive interference management in retransmission-based wireless networks," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3023–3046, 2012.

[4] K.G. Seddik, A.K. Sultan, A.A. El-Sherif, and A.M. Arafa, "A feedback-based access scheme for cognitive radio systems," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, San Francisco, CA, June 2011.

[5] A.M. Arafa, K.G. Seddik, A.K. Sultan, T. ElBatt, and A.A. El-Sherif, "A feedback- soft sensing-based access scheme for cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 7, pp. 3226–3237, 2013.

[6] R.R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple-access system," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 918–930, Sep. 1988.

[7] W. Szpankowski, "Stability conditions for some distributed systems: buffered random access systems," *Advances in Applied Probability*, vol. 26, no. 2, pp. 498–515, June 1994.

[8] RM Loynes, "The stability of a queue with non-independent inter-arrival and service times," *Cambridge Philosophical Society*, vol. 58, no. 3, pp. 497–520, 1962.