On Optimal Control for Opportunistic Spectrum Access of Cognitive Radio Networks

Senhua Huang, Xin Liu, and Zhi Ding
University of California Davis
Davis, CA 95616, USA
Email: senhua@ece.ucdavis.edu

***Abstract*—The listen-Before-Talk (LBT) strategy has been prevalent in cognitive radio networks where secondary users opportunistically access under-utilized primary band. To minimize the amount of disruption from secondary users to primary signals, secondary users generally are required to detect the presence of the primary user reliably, and access the spectrum intelligently. The sensing time has to be long enough to achieve desirable detection performance. When the primary signals are relatively weak, long sensing time is needed which reduces the secondary transmission opportunities. In this paper, we generalize the packet-level LBT strategy by allowing the secondary user to potentially transmit multiple packets after one sensing, and study the optimal control policy to determine the conditions under which the secondary user should sense the channel.**

In spectrum overlay systems, a secondary receiver can provide feedback information (through acknowledgement) on whether a collision with the primary user is detected. However, such detection and the corresponding feedback may not always faithfully report all collisions observed by the primary user in practice. Therefore, we also consider the impact of imperfect collision detection on the access control policy. We show that the optimal spectrum access control policy has a simple threshold-based structure, where the secondary user transmits consecutive packets until the estimated probability of the primary user being idle falls below a threshold, and senses the channel otherwise. We also obtain a closed-form solution on the optimal access policy for a special case when there is no acknowledgement in the network.

I. Introduction

We consider opportunistic spectrum access by secondary users when the spectrum is allocated to a legacy user (primary user). The primary user (PU) owns the spectrum; hence, the transmission from the secondary user (SU) should lead to little interruption to the PU. In other words, the secondary spectrum access should be non-intrusive. We consider the spectrum-overlay (or spectrum weaving) scenario for cognitive radio networks, where the SU is allowed to transmit with full power when the PU is absent.

In a spectrum-overlay system, it is critical that the SU is capable to detect the presence of the PU reliably. Therefore, the spectrum access of the SU often adopts the Listen-Before-Talk (LBT) principle, according to which the SU performs sensing before transmission. While this principle is well justified, many existing works on opportunistic spectrum access assume a typical packet-level sensing model (as illustrated in Fig. 1) implicitly or explicitly, i.e., before transmitting every packet, the SU senses the PU channel for a fixed amount of time [1], [2], [3], [4]. The sensing cost in terms of time overhead is often neglected in the problem formulation for packet-level sensing.

This work is supported in part by the National Science Foundation Grant CNS-0520126 and Grant CNS-0448613.

While packet-level sensing simplifies the design of PHY/MAC layer protocol, it is not necessarily optimal, especially when the sensing may occupy a large portion of the transmission time. For example, when the PU signal strength is weak, it requires a larger number of data samples to achieve a good detection performance, which is crucial to the protection of the QoS of the PU. Alternatively, when cooperative sensing schemes are used by a group SUs, the time spent on exchanging sensing information could be in the order of packet length. In these scenarios, the overhead introduced by fixed sensing time per transmission is no longer negligible.

Additionally, perfect collision detection is often assumed in previous works. Perfect acknowledgement can accurately reflect whether the SU packet collides with the PU. However, this is not true in practical systems. In other words, it is difficult to detect the collision with the PU without error. The difficulties result from the captured effects at both the primary receiver and secondary receiver. The captured effects are often observed in multiuser wireless communication systems due to the random fluctuation on the power of desired signal and interference power from other users.

Therefore, two questions remain of strong interest to us in this work:

- What is the optimal sensing-transmitting strategy? Is it optimal to perform LBT on packet level, especially when the sensing time is large?
- What is the impact of imperfect acknowledgement, which may not faithfully indicate the presence/absence of the PU, as well as the interruption to the PU?

In this paper, we try to answer the above two questions. First, we consider the cost of sensing in terms of time, and formulate the control of the SU spectrum access using a Markov decision process. We adopt average reward per time unit as the objective function of the optimal control problem. With this generalization on the access model, we address the question of whether the packet-level LBT is optimal. It is clear that the packet-level LBT is a special case of the optimal policy. We also consider imperfect acknowledgement from the

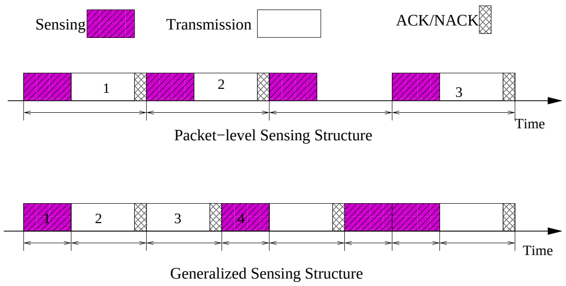

Fig. 1. Two Sensing Structures

secondary receiver, which is different from most of existing works. Therefore, the results obtained here are more general.

Using the optimal stopping theory [5] [6], we show that the optimal spectrum access policy has a simple threshold-based structure, where the posterior probability of the PU being idle is compared to the threshold. This result coincide with the heuristic that the SU should continue transmitting packets until the estimated idle probability falls below the threshold. For a special case, we derive the closed-form solution for the optimal control on the SU's spectrum access.

II. Related Works

Encouraged by the potential spectrum policy reform in FCC [7] and the progress made in DARPA XG program [8], the field of cognitive radio has thrived in recent years (see [9], [10], [11] and references therein).

Both centralized spectrum leasing/auction schemes (e.g. [12]-[13]), and distributed channel allocation and scheduling algorithms (e.g. [14], [15], [16]) have been proposed to enable the spectrum sharing among users with cognitive radio capabilities in different network scenarios. For example, in [16], authors consider the cross-layer optimization on the flow routing, scheduling, channel allocation, and power control of cognitive radios to improve the spectrum efficiency of the multihop software defined networks. However, there is a lack of consideration on the interplay between the PU behavior and SU access, which is important especially for the opportunistic spectrum access of the SUs within the spectrum overlay framework, where Listen-Before-Talk principle is emphasized.

There are also works studying the schemes where the cognitive radio actively relays the packet transmission of the PU in the hope of creating more spectrum opportunities [17], [18] for itself. However, it requires that the SU can decode the PU packets and the PU transmission adopts some retransmission mechanisms, which is difficult for PUs offering broadcast services and SUs which are not in favourable locations.

The closest related works within the framework of spectrum overlay include the joint PHY/MAC designs of channel selection, channel sensing/probing, operating points of spectrum sensor for cognitive radios [1], [2], [3], [19], [20], [21], [22], [4]. For example, in [21], adaptive queueing and Lyapunov optimization are used to design online access control and resource allocation schemes for secondary users under the collision probability constraint imposed by the PU. In [3] , authors use optimal sequential decision process theory to show that the optimal joint channel probing and transmission strategy has a threshold based-structure. However, there is no penalty on the collision with the PU. In [1][19] [20] , partially observable MDP theory (POMDP) is used to derive the structure for the optimal opportunistic spectrum access policy (including channel selection, operating point of the spectrum sensor, and the access decision) with constraint on the collision probability observed by the PU. It was shown that myopic policy is often optimal for many cases by exploiting the inner structure of the spectrum access problem. In [22], throughput limit of SU access under the collision probability constraint is found when the PU idle time follows exponential distribution. However, they assume perfect sensing which does not have a time overhead. In [4], a simple periodic sensing scheme for choosing primary channels with high idle probability is proposed for the SU to exploit the spectrum opportunities.

However, all of the aforementioned works adopt a packet level LBT structure, and/or do not consider the sensing cost or overhead. Also, they often assume that the collision detection is perfect with the acknowledgment mechanism, which is unrealistic in many practical systems. Our work differs on that we use a more flexible sensing/transmission structure based on LBT principle, and consider the impact of imperfect collision detection on the control policy of the opportunistic spectrum access.

III. System Model

We consider a system consisting of a PU, and a SU that opportunistically accesses the PU channel. The PU state is either busy or idle, with the duration of the idle and busy states following exponential distributions. The means of the busy and idle time are denoted by $1/\mu$ and $1/\lambda$, respectively. The PU transmits its traffic at will. In other words, it does not perform any sensing functionality.

The SU uses a spectrum sensor to detect the state of the PU. The access of the SU follows a slotted-structure, as shown in Fig. 2. We assume that the SU transmits a packet with a fixed length Δ. We further assume that $\lambda\Delta \ll 1$, $\mu\Delta \ll 1$, i.e., the duration of SU packet is set to be much shorter than the PU busy/idle cycles. Since shorter SU packet duration is more favorable for the SU to exploit the spectrum opportunities, this setting is reasonable. The sensing time is denoted by T_s. The PU state transition can be described as in Fig. 3. Here, we assume that the SU is aware of the transition probability of the PU states, obtained through measurement. We assume that the sensing time is long enough such that the sensing error is negligible.

After each ST's transmission, the secondary receiver may feedback an acknowledgement upon receiving the packet. Different from the traditional means of acknowledgement, the acknowledgement mechanism here serves dual purposes. First, it validates the packet transmission of the SU in the MAC layer. More importantly, it provides some information to the secondary transmitter on whether a collision with the PU may

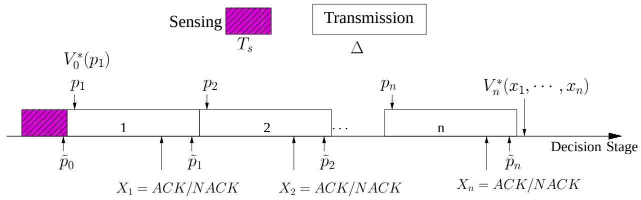

Fig. 2. Sequences of Spectrum Access

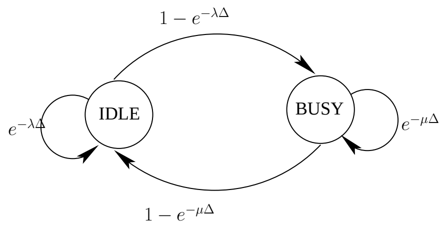

Fig. 3. Primary user's state transition

have occurred. This information assists the SU in controlling its access such that the interruption (here collision) observed by the PU is contained. Since the secondary transmitter cannot sense the channel during data transmission, if the PU returns the channel, the PU traffic will be interrupted. In an ideal scenario, an ACK signifies that the SU packet is successfully received and the PU is idle; while a NACK (or absence of ACK) means the loss of the SU packet or the return of the PU. In other words, the secondary communication pair coordinates to detect the collision with the PU perfectly.

However, under what is known as the captured effect, it is possible that the SR can decode the ST's packet even when the PT is transmitting in practical systems. Moreover, the lack of ACK could be a result of collision with the PU, or deep channel fades between the ST and SR, or interference from other users. It is also possible that the ACK itself is weak. Therefore, we view the acknowledgement as inaccurate (in the sense of collision detection), which is different from [1] [23] [24]. We define the following two probabilities:

$$\gamma_1 = Pr[\text{NACK}|\text{Collision with PU}]$$
$$\gamma_0 = Pr[\text{NACK}|\text{No Collision with PU}].$$

With perfect ACK, we have $\gamma_0 = 0$ and $\gamma_1 = 1$. Here, we assume that $\gamma_1 > \gamma_0$, i.e., the ACK/NACK feedback can offer at least some information on the collision status with the PU.

We also assume a steady channel data rate for the SU, and the SU always has packet to send. The objective of the SU is to maximize its average transmission rate by intelligently deciding whether to transmit a packet or perform sensing at each time slot. Within the framework here, the SU acquires more design freedom to intelligently access the spectrum. Note that, the listen-before-talk slot structure is a special case of the spectrum access policy considered here.

Since the PU is the legacy user of the spectrum, the SU should generate nearly imperceptible interruption to the PU. Therefore, we introduce a penalty on the SU if the SU collides with the PU traffic. This collision cost, denoted by C, is important to control the aggressiveness of the SU's access activities. For each successful transmission, the SU receives a reward $R\Delta$, where R could be the data rate of the channel. For notational convenience, we normalize $R\Delta = 1$.

IV. Problem Formulation

Under normal assumptions, the SU cannot sense the channel when it is transmitting. However, since observing (or sensing) is one of two decision outcomes, the secondary user cannot observe the state of primary activity directly and completely. In other words, the decision process at the secondary user is a partially observable decision process.

Though it is natural to formulate the spectrum access problem as a partially observable Markov decision process (POMDP) problem, we find it here easier to define the state of the system as the posterior probability that the PU is idle and the problem becomes a MDP problem with fully observable state and uncountable state space. We use p_t to denote the state at time slot t, $p_t \in [0, 1]$. It can be proved that p_t is a sufficient statistic to determine the optimal policy for the originally partially observable Markov decision process [25]. The action space of the SU is denoted by $\mathbb{A} = \{1(\text{Transmit}), 0(\text{Sense})\}$. With imperfect ACK, we define the immediate reward $r_t(p_t, a_t)$ at time slot t with state p_t and action a_t as follows:

$$r_t(p_t, 1) = p_t(1 - \gamma_0) + (1 - p_t)(1 - \gamma_1 - C)$$
$$r_t(p_t, 0) = 0.$$

The value of C can be adjusted to achieve required protection on the PU. To prevent the extreme case where the SU never stops transmitting, we assume that $\gamma_1 > \gamma_0$, and $\bar{p} - (1-\bar{p})C < 0$, where $\bar{p}$ is the limiting distribution of channel being idle. Then we have $\bar{p} = \frac{\mu}{\mu+\lambda}$, and we have $C > \frac{\mu}{\lambda} R\Delta$.

With the assumption that $\lambda\Delta \ll 1$ and $\mu\Delta \ll 1$, we have $e^{-\lambda\Delta} \sim 1$ and $e^{-\mu\Delta} \sim 1$, i.e., when the PU is idle (busy), it will remain idle (busy) with high probability. We assume that the SU should start the access activities only after sensing the channel being idle. The question is how many consecutive packets should the SU transmit before it senses the channel again. As shown in Fig. 2, at the end of decision slot t, the SU will update its estimation on the PU idle probability as $\tilde{p}_t$ based on the observation it receives after its chosen action. The observation can be either the sensing outcome or the ACK/NACK. Specifically, we have the following observation model:

$$\tilde{p}_t^{(a_t=0)}(p_t) = \begin{cases} 1, & \text{sensing “idle”} \\ 0, & \text{sensing “busy”} \end{cases}$$
$$\tilde{p}_t^{(a_t=1)}(p_t) = \begin{cases} \tilde{p}_t(\text{ACK}), & \text{ACK} \\ \tilde{p}_t(\text{NACK}), & \text{NACK}, \end{cases} \quad (1)$$

where

$$\tilde{p}_t(\text{ACK}) = \frac{p_t(1-\gamma_0)}{p_t(1-\gamma_0)+(1-p_t)(1-\gamma_1)},$$
$$\tilde{p}_t(\text{NACK}) = \frac{p_t\gamma_0}{p_t\gamma_0+(1-p_t)\gamma_1}.$$

The system changes to state p_{t+1} following the following rules:

$$p_{t+1} = \tilde{p}_t e^{-\lambda\Delta} + (1-\tilde{p}_t)(1-e^{-\mu\Delta}). \quad (2)$$

The access policy of the SU is denoted by $\pi = [d_0, d_1, \cdots, d_t, \cdots]$, where d_t is a function which defines a mapping from state space $\mathbb{P} = \{p_t\} = [0,1]$ to the action space $\mathbb{A}$ at time t. Assuming that the time scale of the SU's opportunistic access is much larger than the packet length, the spectrum access in a long time duration can be treated as a repeated trial of a stopping problem. The object of the SU is to maximize the average reward over the whole access period, which consists of L repeated trials, i.e.,

$$\text{maximize} \qquad \lim_{L\to\infty} \frac{(\sum_{l=1}^{L}\sum_{t=1}^{N_l} r_t)/L}{(\sum_{l=1}^{L}(T_s+N_l\Delta))/L}, \quad (3)$$

where N_l is the number of packets transmitted in the lth trial of the decision process. When it causes no confusion, we use N to denote stopping rule which decides whether to stop transmission based on the whole observation history $H_t = \{x_1, a_1, x_2, \cdots, x_t\}$, or equivalently, on the current system state p_t. For stopping rule N, define

$$Y_N = \sum_{t=1}^{N} r_t, \qquad T_N = (T_s + N\Delta), \quad (4)$$

i.e., Y_n is the accumulated reward until stage n, T_n is the total time spent to reach stage n.

Since the exponential distribution is memoryless, and the SU always "restart" the access from sensing the channel being idle, i.e., $\tilde{p}_0 = 1$, for fixed policy π, the expected sum of reward is identically and independently distributed (i.i.d.), as well as the number of packets transmitted in each run. Therefore, maximizing the average reward per unit of time is equal to maximize the rate of return $E(Y_N)/E(T_N)$ [5], i.e.,

$$\lim_{L\to\infty} \frac{(\sum_{l=1}^{L} Y_{N_l})/L}{(\sum_{l=1}^{L} T_{N_l})/L} = \frac{E(Y_N)}{E(T_N)} \quad \text{almost surely.} \quad (5)$$

Therefore, the optimal spectrum access problem can be expressed as:

$$\max_{N\in\mathcal{C}} \qquad \frac{E(Y_N)}{E(T_N)}, \quad (6)$$

where

$$\mathcal{C} = \{N : N \geq 1, E(T_N) < \infty\} \quad (7)$$

is the set of stopping rules for which $E(T_N) < \infty$. Since the collision with the PU has a penalty, we can expect that the optimal stopping rule always resides in $N \in \mathcal{C}$. The optimal average reward per time unit is then expressed as:

$$\alpha^* = \max_{N\in\mathcal{C}} \frac{E(Y_N)}{E(T_N)}. \quad (8)$$

V. Optimal Adaptive Access Control

Optimal stopping theory [5] [6] is used to characterize the structure of the optimal stopping rule for the secondary spectrum access. Define

$$S_n(p) = Y_n - \alpha T_n, \quad (9)$$

where p is the initial state, and α can be regarded as a cost per time unit to reach stage n. According to *Theorem 6.1* in [5] , if for some α, $\sup_{N\in\mathcal{C}} E(S_N) = 0$, then $\sup_{N\in\mathcal{C}} E(Y_N)/E(T_N) = \alpha$. In addition, the policy which attains $\sup_{N\in\mathcal{C}} E(S_N) = 0$ achieves the maximum rate of return, i.e., α^*. Then, we transfer the problem of maximizing rate of return to an ordinary stopping time problem with reward at stage n denoted by S_n. Define $V_0^*(p) = \sup_{N\in\mathcal{C}} E(S_N(p))$ as the maximum expected return given we start from state p. First, we show the existence of the optimal rule for the ordinary stopping time problem:

$$\max_{N\in\mathcal{C}} \qquad E(S_N). \quad (10)$$

Specifically, we have:

Proposition 1. *There exists an optimal stopping rule N^* for problem (10).*

Proof: See Appendix . ■

Relying on the optimality equation, the optimal rule has the following form:

$$N^* = \min\{n \geq 0 : S_n \geq V_n^*(x_1, \cdots, x_n)\}, \quad (11)$$

where $V_n^*(x_1, \cdots, x_n) = \sup_{N\geq n} S_N(x_1, \cdots, x_n)$ is the maximum expected reward given the observation $X_1 = x_1, \cdots, X_n = x_n$. Notice that, the optimal value obtained at decision stage n is determined by the state p_{n+1} (as illustrated in Fig. 2), which abstracts all the observed information from $(x_1, x_2, \cdots, x_n)$. This indicates a time invariance property of the optimal value function, and the expected payoffs at stage n after observing $X_1, \cdots, X_n$ is the same as it was at stage 0, except that we have an additional cost (or reward) to reach state p_{n+1}. Similar arguments can be found in *Section 4.5* of [5]. Specifically, we have the following results:

$$\begin{aligned} V_n^*(X_1, \cdots, X_n) &= V_n^*(p_{n+1}) \\ &= V_0^*(p_{n+1}) - \alpha n\Delta + \sum_{t=1}^{n} r_i, \end{aligned} \quad (12)$$

where $-\alpha n\Delta + \sum_{t=1}^{n} r_i$ denotes the total "reward" accumulated up-to stage n. Then, the rule given by the principle of optimality is reduced to the following form:

$$\begin{aligned} N^* &= \min\{n \geq 0 : S_n = V_n^*(p_{n+1})\} \\ &= \min\{n \geq 0 : S_n = V_0^*(p_{n+1}) - \alpha\Delta n - \alpha T_s\} \\ &= \min\{n \geq 0 : V_0^*(p_{n+1}) + \alpha T_s = 0\}. \end{aligned} \quad (13)$$

A. A Threshold-based Policy

For general stopping time problem with uncountable state space, it is very difficult to find the structure of the optimal stopping rule using (13). In this section, we show that for the spectrum access model considered here, the optimal policy has a threshold-based structure. With this simple structure, the calculation for the optimal policy becomes the search for the optimal threshold, which greatly simplifies the calculation of the optimal solution.

First, we introduce the following Lemma.

Lemma 1. *$V_0^*(p)$ is a convex function of p.*

Proof: See Appendix VI-B. ■

Theorem 1. The optimal stopping rule to maximize the rate of return is as belows:

$$\pi^* : a_t = \begin{cases} 1(\text{Transmit}), & \text{if } p_t \geq p^* \\ 0(\text{Sense}), & \text{o.w.} \end{cases}, \quad (14)$$

where $p^* = \max\{p : V_0^*(p) + \alpha T_s = 0\}$.

Proof: See Appendix VI-C. ■

Generally, it is difficult to obtain the closed-form expression for the optimal threshold p^* due to the complexity of obtaining the optimal value function $V_0^*(p)$. However, the optimal value of p^* can be found by Monte-Carlo simulations. The result is intuitive. When the channel is more likely to be idle ($> p^*$), the SU should continue packet transmission rather wasting spectrum opportunity on sensing. On the other hand, it worth mentioning that the typical challenge of finding optimal sequential decision is to balance between the immediate reward and all possible future payoffs. Without any structure in the optimal policy, it requires an exhaustive search over the set of all possible policies (which is practically impossible) to obtain the maximum throughput per time unit for the SU. However, with the shown well-defined structure for the optimal policy here, the complexity is significantly reduced since we only need to search over $p \in [0, 1]$.

B. Without Acknowledgement

We consider a special scenario where we can obtain a deterministic optimal control on the spectrum access for the SU. Suppose that there is no acknowledgement mechanism to facilitate the SU to detect the collision with the PU, then the state transition probability can be easily calculated by setting $\tilde{p}_t = p_t$.

Here, we assume that $\gamma_1 = 1$ and $\gamma_0 = 0$.

To simplify the description, we set $T_s = \Delta$, and assume that $\lambda = \mu$. Obviously, if we perform LBT for each packet, the maximum throughput we can obtain is reduced by half because of the sensing overhead. Since the SU access cycles always start with the sensing idle channel, the state evaluation is deterministic with random stopping time. We show that the optimal policy for this scenario is a pure policy which has a fixed stopping time, and the solution for this simplified problem has a clean structure. Specifically, we define $u(i)$ as the average reward of transmitting until time i (note that we have normalized $R\Delta = 1$), i.e.,

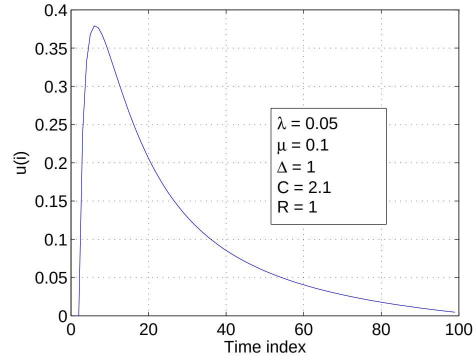

Fig. 4. Optimal Time to Stop Transmission

$$u(i) = \frac{\sum_{k=1}^{i-1}[p_k - (1-p_k)C]}{i}. \quad (15)$$

Let $i^* = \min\{i : u(i) > u(i+1)\}$, then we have the following Lemma (proof in Appendix VI-D).

Lemma 2. *For function $u(i)$ defined in (15), and state transition defined as in (2) with $\tilde{p}_t = p_t$, we have $\forall j > i^*$, $u(j+1) < u(j)$.*

This result indicates that $u(i)$ reaches its peak at certain value of i, and then monotonically decreases. The value of i^* can be easily calculated using numerically. As an illustration of this Lemma, we plot the value of $u(i)$ in Fig. 4. It follows directly that

Theorem 2. The optimal policy in the spectrum access problem without acknowledgment is to transmit i^* packets and then sense the channel, where $i^* = \min\{i : u(i) > u(i+1)\}$.

The optimal policy for the special case is simply a one-step-look-ahead policy, and has a closed-form expression. Numerical method can be used to evaluate the value of $u(i)$, and obtain the optimal solution.

When there is an acknowledgement (though imperfect) mechanism, the SU can obtain more information about the PU activities, and thus we expect the SU to achieve a higher average throughput per time unit than the case without acknowledgement.

VI. Conclusions and Future Works

We generalize the Listen-Before-Talk sensing/transmission structure at the packet level. We develop an adaptive control policy to decide whether to sense/transmit in each decision stage. We also consider the impact of inaccurate collision detection with the PU traffic on the spectrum access policy. We found the optimal spectrum access policy to have a simple threshold-based structure. The optimal access decision is to continue transmitting a packet when the posterior idle probability of the PU is higher than the threshold, and to sense the channel otherwise. With this structure, the optimal policy can be found by simply searching for the optimal threshold, with which the computation complexity is greatly reduced. In a special case when there is no acknowledgment, we derive a closed-form solution for the optimal control over the opportunistic spectrum access. This further reduces the computational complexity. One possible direction for future work is to study the problem for general idle/busy time distributions. In addition, it is also of interest to develop an adaptive MAC protocol which can adapt to the dynamic changing PU behavior.

REFERENCES

[1] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE Journal on Selected Areas in Communications (JSAC): Special Issue on Adaptive, Spectrum Agile and Cognitive Wireles Networks*, vol. 25, no. 3, pp. 589–600, 2007.

[2] L. Lai, H. E. Gamal, H. Jiang, and H. V. Poor. (2007) Cognitive medium access: Exploration, exploitation and competition. [Online]. Available: http://arxiv.org/abs/0710.1385

[3] N. B. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," in *Proc. of the 13th annual ACM international conference on Mobile computing and networking*. New York, NY, USA: ACM, 2007, pp. 27–38.

[4] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Optimal dynamic spectrum access via periodic channel sensing," in *Proc. Wireless Communications and Networking Conference (WCNC)*, 2007.

[5] T. S. Ferguson. (2006) Optimal stopping and applications. [Online]. Available: http://www.math.ucla.edu/ tom/Stopping/Contents.html

[6] S. M. Ross, *Introduction to Stochastic Dynamic Programming: Probability and Mathematical*. Orlando, FL, USA: Academic Press, Inc., 1983.

[7] "Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies, notice of proposed rule making and order," Federal Communications Commision, Report, Et docket No. 03-322, December 2003.

[8] F. W. Seelig, "A description of the August 2006 XG demonstrations at fort A.P. Hill," *Second IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN*, pp. 1–12, 2007.

[9] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[10] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access: signal processing, networking, and regulatory policy," *IEEE Signal Processing Magazine*, vol. 55, no. 5, pp. 2294–2309, 2007.

[11] M. M. Buddhikot, "Understanding dynamic spectrum access: Models, taxonomy and challenges," in *Second IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN*, 2007.

[12] M. Buddhikot, P. Kolodzy, S. Miller, K. Ryan, and J. Evans, "DIMSUMnet: New directions in wireless networking using coordinated dynamic spectrum access," in *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (IEEE WoWMoM)*, aromina/Giardini Naxos, Italy, 2005.

[13] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri, "A general framework for wireless spectrum auctions," *Second IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN*, pp. 22–33, 2007.

[14] Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda, and Y. Wu, "Allocating dynamic time-spectrum blocks in cognitive radio networks," in *Proc. of the 8th ACM international symposium on Mobile ad hoc networking and computing*. New York, NY, USA: ACM, 2007, pp. 130–139.

[15] J. Zhao, H. Zheng, and G.-H. Yang, "Distributed coordination in dynamic spectrum allocation networks," *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 259–268, 8-11 Nov. 2005.

[16] Y. T. Hou, Y. Shi, and H. D. Sherali, "Optimal spectrum sharing for multi-hop software defined radio networks," *in Proc. IEEE INFOCOM, Anchorage, AL*, pp. 1–9, 2007.

[17] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Stable throughput of cognitive radios with and without relaying capability," *IEEE Transactions on Communications*, vol. 55, no. 12, pp. 2351–2360, Dec. 2007.

[18] A. Sadek, K. Ray Liu, and A. Ephremides, "Cognitive multiple access via cooperation: Protocol design and performance analysis," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3677–3696, Oct. 2007.

[19] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053–2071, 2008.

[20] K. Liu and Q. Zhao, "Link throughput of multi-channel opportunistic access with limited sensing," *in Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[21] R. Urgaonka and M. J. neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *in Proc. IEEE INFOCOM, Phoenix, AZ, USA*, 2008.

[22] S. Huang, X. Liu, and Z. Ding, "Opportunistic spectrum access in cognitive radio networks," *in Proc. IEEE INFOCOM, Phoenix, AZ, USA*, 2008.

[23] Q. Zhao and K. Liu, "Detecting, tracking, and exploiting spectrum opportunities in unslotted primary systems," *in Proc. of IEEE Radio and Wireless Symposium (RWS)*, 2008.

[24] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Cross-layer design of opportunistic spectrum access in the presence of sensing error," *in Proc. of the 40th Conference on Information, Science, and Systems (CISS)*, 2006.

[25] R. Smallwood and E. Sondik, "The optimal control of partially observable markov processes over a finite horizon," *Operations Research*, pp. 1071–1088, 1971.

APPENDIX

A. Existence of Optimal Stopping Rule

Proof: In *Chapter 3 Theorem 1* of [5], it is shown that an optimal stopping rule exists if the following two conditions hold:

A1. $E\{\sup_n S_n\} < \infty$.
A2. $\limsup_{n\to\infty} S_n \leq S_\infty$ a.s.

First, since $C > \frac{\mu}{\lambda} R\Delta$, we have $S_\infty = -\infty$, and $\limsup_{n\to\infty} S_n \to S_\infty$. Therefore, condition A2 is satisfied.

For condition A1, suppose that there is a prophet who can foresee the idle/busy state of the PU, and $\gamma_0 = 0$, it will only stop at the end of the nth idle period and obtain reward:

$$z_n = l_1 \cdot (R-\alpha) + \sum_{k=2}^{n} l_k \cdot (R-\alpha) - (\frac{b_k}{\Delta} + \alpha)C, \quad (16)$$

where l_k and b_k are the length of the k idle/busy time with means $1/\lambda$ and $1/\mu$, respectively. If $n = 1$, $z_n = l_1(R - C)$. Obviously, $s_n \leq z_n$, and $\sup_n S_n \leq \sup_n Z_n$. Define

$$M = \sup_n \{\sum_{k=2}^{n} l_k \cdot (R-\alpha) - (\frac{b_k}{\Delta} + \alpha)C\}, \quad (17)$$

Note that $W_k = L_k \cdot (R-\alpha) - (\frac{B_k}{\Delta} + \alpha)C$ are i.i.d. with mean as:

$$\begin{aligned} E(W_k) &= \frac{R-\alpha}{\lambda} - C(\frac{1}{\mu\Delta} + \alpha) \\ &\leq \frac{R}{\lambda} - \frac{C}{\mu\Delta} \\ &< 0, \end{aligned} \quad (18)$$

where we used assumption $C > \frac{\mu}{\lambda} R\Delta$, and the fact that $\alpha^* \geq 0$. Next, we prove that $E(W^+)^2 < \infty$, where

$$W^+ = \max\{W, 0\}.$$

Let $W_k = U_k - V_k$, where $U_k = L_k \cdot (R-\alpha)$, $V_k = (\frac{B_k}{\Delta}+\alpha)C$, then we have $(W_k^+)^2 \le (U_k + V_k)^2$. Since U_k and V_k are two independent exponential random variables with finite mean, $E(W_k^+)^2 < \infty$.

Using the result in in *Chapter 4. Theorem 2* in [5], i.e., let $W, W_1, W_2, \cdots$, be i.i.d. with finite mean $E(W) < 0$, then $E(M) < \infty$ if and only if $E(W^+)^2 < \infty$. Then, we have

$$\begin{aligned} \sup_n Z_n &= \sup_n (Z_1 + \sum_{k=1}^{n-1} W_k) \\ &= \begin{cases} Z_1, & \text{if } Z_1 > Z_1 + M \\ Z_1 + M, & \text{o.w..} \end{cases} \end{aligned} \tag{19}$$

Since Z_1 is exponentially distributed with a finite mean $E(Z_1) = \frac{R-\alpha}{\lambda} < \infty$, and $E(M) < \infty$, then we know that $E(\sup_n Z_n)$ is finite. Thus, we have $E\{\sup_n S_n\} \le E\{\sup_n Z_n\} < \infty$, i.e., condition A1 is satisfied. Therefore, there exists an optimal stopping rule N^* for the ordinary stopping problem defined in Eq. (10). ■

B. Convexity of $V_0^*(p)$

Proof: For $0 \le \beta \le 1$, $0 \le p_1, p_2 \le 1$, we want to show that $V_0^*(\beta p_1 + (1-\beta)p_2) \le \beta V_0^*(p_1) + (1-\beta)V_0^*(p_2)$. This is accomplished by following the same argument as in *Chapter 3 Lemma 3.1* in [6]. Suppose that the initial state p is determined by flipping a biased-coin with probability β that a head appears. If a head appears, we choose initial state as $p = p_1$, else, $p = p_2$. If we know the outcome of the coin flipping, the most reward we can get is $\beta V_0^*(p_1) + (1-\beta)V_0^*(p_2)$. On the other hand, the most reward we can get if we do not know the outcome of the coin flipping is $V_0^*(\beta p_1 + (1-\beta)p_2)$. Since we can always do no worse with more information, we have the result that $V_0^*(p)$ is convex in p. ■

C. Optimality of Threshold-Based Policy

Proof: By the principle of optimality, the optimal policy has the following form:

$$\begin{aligned} N^* &= \min\{n \ge 0 : S_n = V_n^*(x_1, \cdots, x_n)\} \\ &= \min\{n \ge 0 : S_n = V_0^*(p_{n+1}) - \alpha\Delta n + \sum_{t=1}^{n} r_t\} \\ &= \min\{n \ge 0 : V_0^*(p_{n+1}) + \alpha T_s = 0\} \end{aligned} \tag{20}$$

based on the definition $S_n = \sum_{t=1}^n r_t - \alpha\Delta n - \alpha T_s$. Note that $V_0^*(0) = 0 - \alpha T_s$, and

$$\begin{aligned} V_0^*(1) &= \max\{S_1, V_1^*(x_1)\} \\ &\ge S_1 \\ &= (1-\gamma_0)R\Delta - \alpha\Delta - \alpha T_s. \end{aligned} \tag{21}$$

As long as $(1-\gamma_0)R \ge \alpha$, we have $V_0^*(1) \ge V_0^*(0)$. Note that the maximum return of rate α^* is upper bounded by $(1-\gamma_0)R$, which is achieved when there is a genie telling the SU the exact state of the PU in the future. Furthermore, since $V_0^*(p)$ is convex in p, $V_0^*(p)$ is a function having the shape as in Fig. 5. Define $p^* = \max\{p : V_0^*(p) + \alpha T_s = 0\}$, then for all $p \le p^*$, we have $V_0^*(p) = -\alpha^* T_s$, and the policy which transmit zero packet after sensing is optimal. This completes the proof. ■

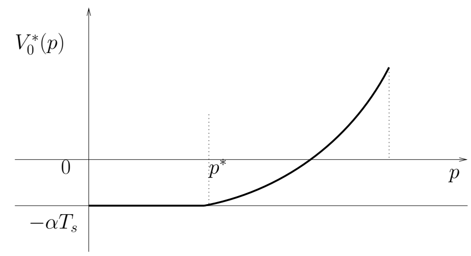

Fig. 5. Illustration of Function $V_0^*(p)$

D. Property of $u(i)$

Proof: Since the SU starts from sensing the channel being idle, i.e., $p_0 = 1$, we can first show that with assumption $\lambda = \mu$, for $p_{i-1} \ge 1/2$, we have $p_i \ge 1/2$. According to (2) and $\tilde{p}_t = p_t$, we have:

$$\begin{aligned} p_i &= p_{i-1}e^{-\lambda\Delta} + (1-p_{i-1})(1-e^{-\lambda\Delta}) \\ &= p_{i-1}(2e^{-\lambda\Delta} - 1) + 1 - e^{-\lambda\Delta} \\ &\ge 1/2. \end{aligned} \tag{22}$$

By induction, we can prove that $\forall j > i^*$, $u(j+1) < u(j)$. By the definition of i^*, we have $u(i^*) \ge u(i^*+1)$, which is the same as having

$$C + \sum_{k=1}^{i^*-1} p_k(1+C) \ge i^* p_{i^*}(1+C). \tag{23}$$

For $j > i^*$, we want to show that $u(j+1) \le u(j)$, which is equivalent to showing that: $C + \sum_{k=1}^{j-1} p_k(1+C) \ge j \cdot p_j(1+C)$. When $j = i^*+1$, we have

$$\begin{aligned} \text{RHS} &= (i^*+1)[p_{i^*}(2e^{-\lambda\Delta} - 2) + 1 - e^{-\lambda\Delta})](1+C) \\ \text{LHS} &= \sum_{k=1}^{i^*-1} p_k(1+C) + C + p_{i^*}(1+C) + C \\ &\ge i^* p_{i^*}(1+C) + p_{i^*}(1+C) + C \qquad \text{(by 23)} \end{aligned}$$

Then, using (23) and the condition $p_{i^*} \ge 1/2$, we have

$$\begin{aligned} \text{LHS} - \text{RHS} &\ge (i^*+1)(1+C)(2p_{i^*} - 1)(1 - e^{-\lambda\Delta}) \\ &\ge 0 \end{aligned}$$

Following the same procedure, we can prove that $\forall j > i^*$, $u(j+1) \le u(j)$. ■