Optimal Channel Probing in Communication Systems: The Two-Channel Case

Matthew Johnston and Eytan Modiano
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA
Email: {mrj, modiano}@mit.edu

***Abstract*—We consider a multi-channel communication system in which a transmitter has access to two channels, but does not know the state of either channel. We model the channel state using an ON/OFF Markovian model, and allow the transmitter to probe one of the channels at predetermined probing intervals to decide over which channel to transmit. For models in which the transmitter must transmit over the probed channel, it has been shown that a myopic policy that probes the channel most likely to be ON is optimal. In this work, we allow the transmitter to select a channel over which to transmit that is not necessarily the one it probed. We show that in the case where the two channels are i.i.d, all probing policies yield equal reward. We extend this problem to dynamically choose when to probe based on the results of previous probes, and characterize the optimal policy, as well as provide a LP in terms of state action frequencies to find the optimal policy.**

I. Introduction

Consider a communication system in which a transmitter has access to multiple channels over which to communicate. The state of each channel evolves independently of the others, and the transmitter has no knowledge of the channel states *a priori*. The transmitter probes a single channel after a predefined time interval to learn the channel state at the current time, which is either ON or OFF. Using the information obtained from the channel probes, the transmitter selects a channel in each time-slot over which to transmit, with the goal of maximizing throughput, or the number of successful transmissions.

This framework applies broadly to many opportunistic communication systems, in which there exists a tradeoff between overhead and performance. It is often impractical to learn the channel state information (CSI) of the channels before scheduling a transmission; consequently, the transmitter must make a transmission decision with only partial channel state information. The transmitter must decide *how much* information, and *which* information is needed in order to make efficient scheduling decisions.

Several works have studied channel probing policies in multichannel communication problems [1], [2], [3], [4], [5], [6], [7]. Of particular interest is the work in [8] and [9], in which the authors assume that after a channel is probed, the transmitter must transmit on that channel. They show that the optimal probing policy is a myopic policy, which probes the channel most likely to be ON.

This work was supported by NSF Grants CNS-0915988, CNS-1217048 and ARO MURI Grant W911NF-08-1-0238.

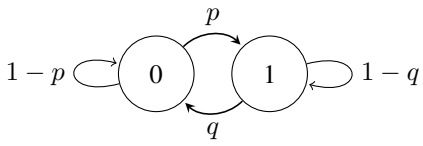

Fig. 1. Markov Chain describing the channel state evolution of each independent channel. State 0 corresponds to an OFF channel, while state 1 corresponds to an ON channel.

In this work, we consider the special case of a system with two channels. We show that when the transmitter is able to transmit over a channel other than that which was probed, the choice of which channel to probe does not affect the expected throughput. Additionally, we identify scenarios such that when the probability distribution of the channel state differs between the two channels, it is optimal to always probe one of the channels. We extend the problem of optimizing the probing epochs dynamically, by formulating a Markov Decision Process for which the optimal policy can be characterized. Lastly, we provide a linear programming formulation in terms of state action frequencies that can be used to solve for an arbitrarily good approximation of the optimal policy.

The remainder of this paper is organized as follows. We describe the model and problem formulation in detail in Section II. In Section III, we prove that all probing policies have equal throughput. In Section IV, we consider a scenario in which the two channels are statistically different. In Section V, we solve for the optimal probing intervals when a fixed cost is associated with probing. Lastly, we conclude in Section VI.

II. System Model

Consider a transmitter and a receiver that communicate using two independent channels. At every time slot, each channel is either in an OFF state or an ON state. Channels are i.i.d. with respect to each other, and evolve across time according to a discrete time Markov process described by Figure 1.

At each time slot, the transmitter chooses a channel over which to transmit. If that channel is in the ON state, then

the transmission is successful; otherwise, the transmission fails. We assume the transmitter does not receive feedback regarding previous transmissions[1]. Furthermore, at predefined epochs of T slots, the transmitter probes the receiver for the state of one of the channels at the current time, which is delivered instantaneously. The transmitter then uses the history of channel probes to make a scheduling decision. The objective is to choose a channel probing strategy to maximize the expected sum-rate throughput, equal to the number of successful transmissions over time.

A. Notation

Let $S_i(t)$ be the state of channel i at time t. The transmitter has an estimate of this state based on previous probes and the channel state distribution. In particular, the transmitter's knowledge of the state is described by the tuple $\mathbf{s} = (s_1, k_1, s_2, k_2)$, where s_1 and s_2 are the last known states of the two channels respectively, and k_1 and k_2 are the respective times since the last probe on each channel. Define the *belief* of a channel to be the probability that a channel is ON given the history of channel probes. For any channel i that was last probed k slots ago and was in state s_i, the belief x_i is given by

$$\begin{aligned} x_i(t) &= \mathbf{P}\big(\text{Channel } i \text{ is ON}|\text{probing history}\big) \\ &= \mathbf{P}\big(S_i(t) = 1 | S_i(t-k) = s_i\big) \end{aligned} \tag{1}$$

where the second equality follows from the Markov property of the channel state process. The above probability is derived from the k-step transition probabilities of the Markov chain in Figure 1:

$$\begin{aligned} p_{00}^k &= \frac{q + p(1-p-q)^k}{p+q}, p_{01}^k = \frac{p - p(1-p-q)^k}{p+q} \\ p_{10}^k &= \frac{q - q(1-p-q)^k}{p+q}, p_{11}^k = \frac{p + q(1-p-q)^k}{p+q}. \end{aligned} \tag{2}$$

Throughout this work, we assume that $p \leq \frac{1}{2}$ and $q \leq \frac{1}{2}$, corresponding to channels with "positive memory." As the CSI of a channel grows stale, the probability π that the channel is ON is given by the stationary distribution of the chain in Figure 1.

$$\lim_{k\to\infty} p_{01}^k = \lim_{k\to\infty} p_{11}^k = \pi = \frac{p}{p+q}. \tag{3}$$

B. Optimal Scheduling

Since the objective is to maximize the expected sum-rate throughput, the optimal transmission decision at each time slot is given by the maximum likelihood (ML) rule, which is to transmit over the channel that is most likely to be ON, i.e. the channel with the highest belief. The expected throughput in a time slot is therefore given by

$$\max_{i\in\{1,2\}} x_i(t). \tag{4}$$

[1]If such feedback exists in the form of higher layer acknowledgements, it arrives after a significant delay and is not useful for learning the channel state.

Following the above assumptions, the optimal scheduling decision remains the same in between channel probes.

III. Optimal Channel Probing

As described above, the transmitter chooses which channel to probe every T slots. Our first result is that for each channel probe, the expected throughput after probing channel 1 is the same as probing channel 2. In fact, we prove a more general result that when probing intervals are predetermined a priori, the total throughput is the same regardless of which channel is probed.

Theorem 1. *For a two-user system, with independent channels evolving over time according to Figure 1, if the probing instances are fixed a priori, i.e. the time between probe i and probe $i+1$ is fixed at Y_i slots, then for each channel probe, the total reward from probing channel 1 is equal to that of probing channel 2.*

Proof Outline: This proof is by reverse induction over the probing times. Consider a finite horizon problem with horizon N. At each index n, a probing decision is made, and assume there are Y_n slots between the current probe and the next probe. Assume the last probing decision is made at index $N-1$. Let the expected reward at probing index n be J_n^i if the choice is made to probe channel i. At probing time $0 \leq n < N-1$, the expected reward function is given recursively, by:

$$\begin{aligned} &J_n\big(S_1, k_1, S_2, k_2\big) \\ &\quad = \max\Big(J_n^1\big(S_1, k_1, S_2, k_2\big), J_n^2\big(S_1, k_1, S_2, k_2\big)\Big) \end{aligned} \tag{5}$$

$$\begin{aligned} J_n^1\big(S_1, k_1, S_2, k_2\big) = &\sum_{j=0}^{Y_n-1} \Big((p_{S_1,1}^{k_1}) p_{1,1}^j + (p_{S_1,0}^{k_1}) p_{S_2,1}^{k_2+j}\Big) \\ &+ (p_{S_1,1}^{k_1}) J_{n+1}(1, Y_n, S_2, k_2 + Y_n) \\ &+ (p_{S_1,0}^{k_1}) J_{n+1}(0, Y_n, S_2, k_2 + Y_n) \end{aligned} \tag{6}$$

$$\begin{aligned} J_n^2\big(S_1, k_1, S_2, k_2\big) = &\sum_{j=0}^{Y_n-1} \Big((p_{S_2,1}^{k_2}) p_{1,1}^j + (p_{S_2,0}^{k_2}) p_{S_1,1}^{k_1+j}\Big) \\ &+ (p_{S_2,1}^{k_2}) J_{n+1}(S_1, k_1 + Y_n, 1, Y_n) \\ &+ (p_{S_2,0}^{k_2}) J_{n+1}(S_1, k_1 + Y_n, 0, Y_n) \end{aligned} \tag{7}$$

and $J_N^i(\mathbf{s}) = 0$ as a base case. The proof follows by first showing that $J_{N-1}^1 = J_{N-1}^2$. Then, assuming that (6) and (7) hold for $(n+1, n+2, \ldots, N-1)$, we prove that the result holds for index n. The complete proof is omitted for brevity. Note that the case where channel probes occur every T slots is a special case of this Theorem.

As a consequence of Theorem 1, it is sufficient to always probe the same channel. Assume that every T slots, a cost of c is incurred to probe a channel. Since it is optimal to always probe channel 1, the belief of channel 2 equals the steady state probability of being in the ON state π. Therefore, the average

reward less cost is given by

$$\frac{1}{T}\left(-c+\pi\sum_{i=0}^{T-1}p_{11}^{i}+(1-\pi)\pi T\right)=\frac{-c}{T}+\pi+\frac{\pi p_{10}^{T}}{T(p+q)}. \quad (8)$$

Maximizing the above equation with respect to T results in the optimal probing interval for a fixed cost c.

Intuitively, when a channel is probed, the base station receives information about the optimal decision to make until the next probe. If the probed channel is ON, it is optimal to transmit over that channel until the next probing instance. On the other hand, if the probed channel is OFF, it is optimal to transmit over the other channel. Therefore, the information gathered from probing one channel is the same as that from the other channel. This result is in contrast to the result in [9], which proves the optimal decision is to probe the channel with the highest belief. However, their model assumed that a transmission must occur on the probed channel, whereas our model allows the transmitter to choose the channel over which to transmit based on the result of the probe. Consequently, the myopic policy of [9] is not a uniquely optimal policy in this setting.

Theorem 1 holds for the case of two i.i.d. channels with fixed probing intervals. If the probing epochs are not fixed, i.e. the decision to probe depends on the results of the previous probe, then there is an advantage to probing one channel over the other. This is explored in Section V. Additionally, when the two channels differ statistically, the optimal probing decision depends on the channel statistics, as shown in Section IV.

IV. Heterogeneous Channels

Now, assume the two channels differ statistically, i.e. channel 1 evolves in time according to the Markov chain in Figure 1 with parameters p_1 and q_1, and channel 2 evolves according to a chain with parameters p_2 and q_2. Denote the k-step transition probability of channel 1 as $a_{i,j}^{k}$ and the k-step transition probability of channel 2 as $b_{i,j}^{k}$. Additionally, let π_1 and π_2 be the steady state probability of channel 1 and channel 2 respectively. Intuitively, it is optimal to probe the channel with *more memory*, as that probe yields more information.

For example, consider a channel that varies rapidly, with $p_1 = q_1 = \frac{1}{2} - \epsilon$, and a channel which rarely changes state, with $p_2 = q_2 = \epsilon$. Probing the low-memory channel provides accurate information for one or two time slots, but that information quickly becomes stale, and the transmitter must guess which channel is ON. On the other hand, probing the high-memory channel yields information that remains accurate for many time slots after the probe. This intuition is confirmed in the following result.

Theorem 2. *Assume a two-user system with channel states evolving as described above, and that the probing instances are fixed to intervals of T slots. Furthermore, assume that p_1, p_2, q_1, q_2 satisfy the following:*

$$b_{11}^{i} \geq a_{11}^{i} \qquad \forall i. \quad (9)$$

Then, the optimal probing strategy is to probe channel 2 at all probing instances.

Proof Outline: We assume a finite horizon of N probes. We can write expected reward functions J_n^1 and J_n^2 recursively similarly to the proof of Theorem 1. The proof follows by reverse induction on probing indices n to show that for all states, we have $J_n^2 \geq J_n^1$. Again, the detailed proof is omitted for brevity. To clarify the significance of this theorem, we have the following corollaries.

Corollary 1. *Assume the two channels satisfy $\pi_1 = \pi_2$, and that $p_1 + q_1 \geq p_2 + q_2$. Then, the optimal policy is to always probe channel 2.*

Proof: We can rewrite the k-step transition probability of the second chain from (2) as follows.

$$\begin{aligned} b_{11}^{i} &= \frac{p_2 + q_2(1-p_2-q_2)^{i}}{p_2+q_2} \\ &= \pi_2 + (1-\pi_2)(1-p_2-q_2)^{i} && (10)\\ &= \pi_1 + (1-\pi_1)(1-p_2-q_2)^{i} && (11)\\ &\geq \pi_1 + (1-\pi_1)(1-p_1-q_1)^{i} && (12)\\ &= a_{11}^{i} && (13)\end{aligned}$$

where (11) follows from the assumption that $\pi_1 = \pi_2$, and (12) follows from the assumption that $p_1 + q_1 \geq p_2 + q_2$. Therefore, $b_{11}^{i} \geq a_{11}^{i}$, and applying Theorem 2 concludes the proof. ∎

Corollary 2. *Assume the two channels satisfy $p_1 + q_1 = p_2 + q_2$, and that $\pi_1 \leq \pi_2$. Then, the optimal policy is to always probe channel 2.*

Proof: We can rewrite the k-step transition probability of the second chain from (2) as follows.

$$\begin{aligned} b_{10}^{i} &= \frac{q_2(1-(1-p_2-q_2)^{i})}{p_2+q_2} \\ &= (1-\pi_2)(1-(1-p_2-q_2)^{i}) && (14)\\ &= (1-\pi_2)(1-(1-p_1-q_1)^{i}) && (15)\\ &\leq (1-\pi_1)(1-(1-p_1-q_1)^{i}) && (16)\\ &= a_{10}^{i} && (17)\end{aligned}$$

where (15) follows from the assumption that $p_1+q_1 = p_2+q_2$, and the inequality in follows from the assumption that $\pi_1 \leq \pi_2$. Since $b_{10}^{i} \leq a_{10}^{i}$, then $b_{11}^{i} \geq a_{11}^{i}$, and Theorem 2 can be applied to complete the proof. ∎

The above two corollaries describe scenarios where asymmetries in the channel model result in the optimal policy of always probing one of the two channels. This is in contrast to Theorem 1 where the channels are homogeneous. Corollary 1 states that if the channels are equally likely to be ON in steady state, the optimal decision is to probe the channel with the smaller $p_i + q_i$. In this context, $p_i + q_i$ is the rate at which the channel approaches the steady state. Thus, probing the channel with more memory is always optimal. Corollary 2 examines a system in which the rate at which the steady state is reached

Simulation	$p_1 = q_1 = 0.1$ $p_2 = q_2 = 0.1$	$p_1 = 0.3, q_1 = 0.1$ $p_2 = 0.15, q_2 = 0.05$	$p_1 = q_1 = 0.1$ $p_2 = 0.15. q_2 = 0.05$
Probe Channel 1	0.6536	0.8240	0.7899
Probe Channel 2	0.6540	0.8652	0.8027
Probe Best Channel	0.6538	0.8450	0.8030
Probe Worst Channel	0.6538	0.8402	0.7902
Round Robin	0.6532	0.8452	0.7981

TABLE I
COMPARISON OF DIFFERENT PROBING POLICIES FOR A FIXED PROBING INTERVAL (6) AND TIME HORIZON 2,000,000.

is the same for both channels, but channel 2 is more likely to be ON in steady state than channel 1. In this case, it is optimal to probe the channel with the highest steady state probability of being ON at all probing instances.

A. Simulation Results

We simulate the evolution of a two channel system over time, and compare different fixed probing policies in terms of average throughput. We assume a time horizon of 2,000,000 probes, and assume a probe occurs every 6 slots. We consider five deterministic stationary channel probing policies: probe channel 1 always, probe channel 2 always, probe the channel with the higher belief, probe the channel with the lower belief, and alternate between the channels (round robin). The first column of Table I shows that for a system with two i.i.d. channels with parameters $p = q = 0.1$, the choice of channel probing policy does not affect the average reward earned by the system, as predicted by Theorem 1.

Additionally, we simulate a system with two statistically different channels. These results are shown in the second and third columns of Table I. The first simulation (column 2) uses two channels with the same steady state probability ($\pi = 0.75$), but with channel 1 approaching that steady state at a faster rate than channel 2. By Corollary 1, the optimal probing policy is to always probe channel 2, which is consistent with the simulation. The second simulation (column 3) uses two channels satisfying $p_1 + q_1 = p_2 + q_2 = 0.2$, and $\pi_2 > \pi_1$, as in Corollary 2. As expected, probing channel two is optimal. In this case, probing the channel with the higher belief is a good policy, since the channel with the higher steady state probability has a higher belief more often.

V. DYNAMIC OPTIMIZATION OF PROBING EPOCHS

In this section, we extend the problem to allow the transmitter to decide whether or not to probe in each slot for a fixed cost c. Determining the optimal probing policy becomes a stochastic control problem, where at each time slot, a decision is made whether to probe channel 1, probe channel 2, or not to probe either channel.

A. Dynamic Programming Formulation

This problem can be formulated as a Markov Decision Process (MDP) or a Dynamic Programming problem (DP). At each time slot, the system state is the belief vector. Let $\tau(\cdot)$ represent the evolution of the belief of a channel over a time slot to the next when that channel is not probed. In particular,

$$\tau(x_i) = x_i(1-q) + (1-x_i)p. \tag{18}$$

We formulate a DP over a finite horizon of length N. The expected reward function at time slot n is given by

$$J_n(x_1, x_2) = \max\{J_n^0(x_1, x_2), J_n^1(x_1, x_2), J_n^2(x_1, x_2)\}, \tag{19}$$

where J_n^0 is the expected reward given that we do not probe at the current slot, and J_n^1 and J_n^2 are the expected reward functions given that we probe channel 1 and 2 respectively. When a channel probe does not occur, the reward is equal to the maximum belief. On the other hand, when a channel is probed, a reward (throughput) of 1 is earned if the probed channel is ON, and if it is OFF, a unit throughput is earned only if the remaining channel is ON. Therefore, the terminal cost at time $n = N$ is given by

$$J_N^0(x_1, x_2) = \max(x_1, x_2), \tag{20}$$

$$J_N^1(x_1, x_2) = -c + x_1 + x_2 - x_1 x_2, \tag{21}$$

$$J_N^2(x_1, x_2) = -c + x_1 + x_2 - x_1 x_2 \tag{22}$$

where c is the cost for probing a channel. For $n < N$, the reward function includes the expected future rewards as well:

$$J_n^0(x_1, x_2) = \max(x_1, x_2) + J_{n+1}\big(\tau(x_1), \tau(x_2)\big) \tag{23}$$

$$J_n^1(x_1, x_2) = -c + x_1 + x_2 - x_1 x_2 + x_1 J_{n+1}\big(1-q, \tau(x_2)\big) + (1-x_1) J_{n+1}\big(p, \tau(x_2)\big) \tag{24}$$

$$J_n^2(x_1, x_2) = -c + x_1 + x_2 - x_1 x_2 + x_2 J_{n+1}\big(\tau(x_1), 1-q\big) + (1-x_2) J_{n+1}\big(\tau(x_1), p\big) \tag{25}$$

Maximizing (19) yields the optimal probing policy at each time slot as a function of the current state. Note the state space is countably infinite, since each x_1 has a one-to-one mapping to an (S, k) pair, where S is the state at the last channel probe, and k is the time since the last probe.

Several observations can be made about the value function described in (19)-(25), as stated through the following lemmas.

Lemma 1 (Linearity). *$J_n^1(x_1, x_2)$ is linear in x_1 for fixed x_2, and similarly, $J_n^2(x_1, x_2)$ is linear in x_2 for fixed x_1.*

Lemma 2 (Commutativity).

$$J_n(x_1, x_2) = J_n(x_2, x_1) \tag{26}$$

The proofs of Lemma 1 and Lemma 2 are omitted for brevity. Let $\Phi_n(0)$, $\Phi_n(1)$, $\Phi_n(2)$ be the values of (x_1, x_2) such that it is optimal to not probe, probe channel 1, and probe channel 2 respectively at time n.

Lemma 3 (Probe Symmetry). *If $(x_1, x_2) \in \Phi_n(1)$, then $(x_2, x_1) \in \Phi_n(2)$.*

Proof: If $(x_1, x_2) \in \Phi(1)$, then $J_n^1(x_1, x_2) \geq J_n^2(x_1, x_2)$ and $J_n^1(x_1, x_2) \geq J_n^0(x_1, x_2)$. Using Lemma 2, we can then say that $J_n^2(x_2, x_1) \geq J_n^1(x_2, x_1)$ and $J_n^2(x_2, x_1) \geq J_n^0(x_2, x_1)$ which implies $(x_2, x_1) \in \Phi_n(2)$. ■

Lemma 4 (No-Probe Symmetry). *If $(x_1, x_2) \in \Phi_n(0)$, then $(x_2, x_1) \in \Phi_n(0)$.*

Proof: If $(x_1, x_2) \in \Phi(0)$, then $J_n^0(x_1, x_2) \geq J_n^1(x_1, x_2)$

and $J_n^0(x_1,x_2) \geq J_n^2(x_1,x_2)$. Using Lemma 2, we can then say that $J_n^0(x_1,x_2) = J_n^0(x_2,x_1)$ and $J_n^1(x_1,x_2) = J_n^1(x_2,x_1)$ which implies $J_n^0(x_2,x_1) \geq J_n^1(x_2,x_1)$. Similarly, we can show $J_n^0(x_2,x_1) \geq J_n^2(x_2,x_1)$, and therefore $(x_2,x_1) \in \Phi_n(0)$. ∎

These last two lemmas show that the optimal decision regions are symmetric about the line $x_1 = x_2$.

We can use these results to prove a convexity result on the reward function, the proof of which follows by induction, but is omitted for brevity.

Theorem 3 (Convexity)**.** *For all n, $J_n(x_1,x_2)$ is convex in x_1 for fixed x_2, and is convex in x_2 for fixed x_1.*

Using the convexity of the expected reward function, we can find sufficient conditions for probing to be optimal at a certain state.

Theorem 4. *For all n, it is optimal to probe at state (x_1,x_2) if the cost to probe c satisfies*

$$c \leq \min(x_1,x_2)\big(1-\max(x_1,x_2)\big) \tag{27}$$

Proof:

$$\begin{aligned} J_n^0(x_1,x_2) &= \max(x_1,x_2) + J_{n+1}\big(\tau(x_1),\tau(x_2)\big) && (28)\\ &\leq \max(x_1,x_2) + x_1 J_{n+1}\big(p_{11},\tau(x_2)\big) \\ &\quad + (1-x_1) J_{n+1}\big(p_{01},\tau(x_2)\big) && (29)\\ &= \max(x_1,x_2) + J_n^1(x_1,x_2) \\ &\quad + c - x_1 - x_2 + x_1 x_2 && (30) \end{aligned}$$

Where (29) follows from Theorem 3. Therefore, $J_n^0(x_1,x_2) - J_n^1(x_1,x_2) \leq 0$ if

$$\begin{aligned} c - x_1 - x_2 + x_1 x_2 + \max(x_1,x_2) &\leq 0 && (31)\\ c &\leq \min(x_1,x_2)\big(1-\max(x_1,x_2)\big) && (32) \end{aligned}$$

∎

While the convexity bound yields sufficient conditions for probing optimality, necessary conditions do not follow directly from this analysis. Additionally, the convexity bound used in (29) is loose, and thus probing is often optimal even in states which do not satisfy the statement of Theorem 4.

B. State Action Frequency Formulation

The channel probing MDP can also be formulated as an infinite horizon, average cost problem. In this case, we write a linear program (LP) in terms of state action frequencies, and solve for the optimal policy. A state action frequency vector $\omega(\mathbf{s};a)$ exists for each state and potential action, and corresponds to a stationary randomized policy such that $\omega(\mathbf{s};a)$ equals the steady state probability that at a given time slot, the state is $\mathbf{s}$ and the action taken is a. Let $\mathbf{s} = (s_1,k_1,s_2,k_2)$, where s_1 and s_2 are the last known states of the two channels respectively, and k_1 and k_2 are the respective times since the last probe on each channel. We use this notation rather than the belief notation in Section V-A to emphasize the countable nature of the state space. Furthermore, the action a satisfies $a \in \{0,1,2\}$, representing the actions of not probing, probing channel 1, and probing channel 2 respectively.

While the state space is countably infinite and the resulting state action frequency LP is intractable, we can approximate the optimal solution by truncating the state space to ensure it is finite. In particular, assume that k_i takes values between 0 and $K_{\max}$, where $K_{\max}$ is a predefined constant. When $k_i = K_{\max}$, and channel i is not probed, then $k_i = K_{\max}$ at the next slot as well. Clearly, as $K_{\max}$ increases, $p_{11}^{K_{\max}} \to \pi$, and the truncated formulation approaches the countable state space formulation. Since the belief of each channel approaches steady state exponentially fast, this truncation method can be used to find a near-optimal solution to the stochastic control problem. See [1] for details.

The state action frequency formulation is presented in (33)-(44). Equation (33) is the objective, maximizing the average reward, where the reward functions are defined for each possible action in (43) and (44). Equation (34) is a normalization constraint, ensuring that the state action frequencies sum to one. Equations (39) through (42) are balance equations for the case when the action is to not probe. Note that we include constraints to deal with the truncation of the state space. Equations (35) and (37) deal with the evolution of the state when channel 1 is probed, where equations (36) and (38) deal with the case when channel 2 is probed.

For weakly communicating finite state and action MDP's, there exists a solution to the state action frequency LP that will be a deterministic stationary policy [10]. Specifically, for all recurrent states $\mathbf{s}$ in the solution, the state action frequencies $\omega(\mathbf{s};a) > 0$ for some a, and since the optimal policy is deterministic, $\omega(\mathbf{s};a) > 0$ is satisfied for only one value of a, which is the optimal decision at that state, and $\omega(\mathbf{s};a) = 0$ for all other actions. Since transient states are only visited finitely often, they have zero state action frequencies for every action.

The solution to the state action frequency LP for sample parameters is shown in Figure 2. This graph plots the optimal decision as a function of the belief of channel 1 (x_1) and the belief of channel 2 (x_2). The system state can only reach a countable subset of the points on the x_1-x_2 plane. Under any policy, except for the policy where a channel is never probed, there is a single recurrent class of states, and only states in this class will have non-zero state action frequencies. From any recurrent state, if the optimal decision is not to probe, the system state will move to the next point $(\tau(x_1),\tau(x_2))$ on the trajectory from the current state to the point (π,π). Based on this observation, and the results in Figure 2, we can characterize the structure of the optimal probing algorithm.

For a given set of parameters, there exists a probing-region, e.g. the dotted convex region in Figure 2, and a point (π,π), denoted by the star in Figure 2. At each time slot, if the current state lies outside of the probing region, the optimal decision is to not probe, and the state moves along the trajectory to (π,π). When the state reaches the probing region, the controller probes one of the channels, and the state resets to one of the four sides of the box in Figure 2, depending on which channel is probed and the result of that probe. Then the process repeats,

Max.

$$\sum_a \sum_{s_1,s_2,k_1,k_2} \omega(s_1,k_1,s_2,k_2;a) r(s_1,k_1,s_2,k_2;a) \tag{33}$$

s.t.

$$\sum_a \sum_{s_1,s_2,k_1,k_2} \omega(s_1,k_1,s_2,k_2;a) = 1 \tag{34}$$

$$\sum_a \omega(s_1,1,s_2,k_2;a) = \sum_{k_1=1}^{K_{\max}} \sum_{s_1'} \omega(s_1',k_1,s_2,k_2-1;1) p_{s_1',s_1}^{k_1}$$
$$\forall s_1,s_2, 2 \le k_2 \le K_{\max}-1 \tag{35}$$

$$\sum_a \omega(s_1,k_1,s_2,1;a) = \sum_{k_2=1}^{K_{\max}} \sum_{s_2'} \omega(s_1,k_1-1,s_2',k_2;2) p_{s_2',s_2}^{k_2}$$
$$\forall s_1,s_2, 2 \le k_1 \le K_{\max}-1 \tag{36}$$

$$\sum_a \omega(s_1,1,s_2,K_{\max};a)$$
$$= \sum_{k_1=1}^{K_{\max}} \sum_{s_1'} p_{s_1',s_1}^{k_1} \big(\omega(s_1',k_1,s_2,K_{\max}-1;1)$$
$$+ \omega(s_1',k_1,s_2,K_{\max};1)\big) \quad \forall s_1,s_2 \tag{37}$$

$$\sum_a \omega(s_1,K_{\max},s_2,1;a)$$
$$= \sum_{k_2=1}^{K_{\max}} \sum_{s_2'} p_{s_2',s_2}^{k_2} \big(\omega(s_1,K_{\max}-1,s_2',k_2;2)$$
$$+ \omega(s_1,K_{\max},s_2',k_2;2)\big) \quad \forall s_1,s_2 \tag{38}$$

$$\sum_a \omega(s_1,k_1,s_2,k_2;a) = \omega(s_1,k_1-1,s_2,k_2-1;0)$$
$$\forall s_1,s_2, 2 \le k_1,k_2 \le K_{\max} \tag{39}$$

$$\sum_a \omega(s_1,K_{\max},s_2,k_2;a) = \omega(s_1,K_{\max}-1,s_2,k_2-1;0)$$
$$+ \omega(s_1,K_{\max},s_2,k_2-1;0) \quad \forall s_1,s_2, 2 \le k_2 \le K_{\max} \tag{40}$$

$$\sum_a \omega(s_1,k_1,s_2,K_{\max};a) = \omega(s_1,k_1-1,s_2,K_{\max}-1;0)$$
$$+ \omega(s_1,k_1-1,s_2,K_{\max};0) \quad \forall s_1,s_2, 2 \le k_1 \le K_{\max} \tag{41}$$

$$\sum_a \omega(s_1,K_{\max},s_2,K_{\max};a) = \omega(s_1,K_{\max},s_2,K_{\max};0)$$
$$+ \omega(s_1,K_{\max}-1,s_2,K_{\max}-1;0)$$
$$+ \omega(s_1,K_{\max}-1,s_2,K_{\max};0)$$
$$+ \omega(s_1,K_{\max},s_2,K_{\max}-1;0) \quad \forall s_1,s_2 \tag{42}$$

$$r(s_1,k_1,s_2,k_2;a) = -c + p_{s_1,1}^{k_1} + p_{s_2,1}^{k_2} - p_{s_1,1}^{k_1} p_{s_2,1}^{k_2}$$
$$\forall a \in \{1,2\} \tag{43}$$

$$r(s_1,k_1,s_2,k_2;0) = \max\left(p_{s_1,1}^{k_1}, p_{s_2,1}^{k_2}\right) \tag{44}$$

and the state will follow a new trajectory to the point (π,π). If the point (π,π) lies outside of the probing region, then there exists a trajectory to (π,π) that does not intersect the probing region. Consequently, the system eventually reaches a state in which it never probes, in which case all states will be transient under the optimal policy. In summary, the optimal time between probes is given by the distance between the state immediately following a probe and the state on the border of the probing region, lying on the line between the current state and (π,π). To find the probing region, and the decisions to make at each point on the probing regions, the SAF LP in (33)-(44) must be solved.

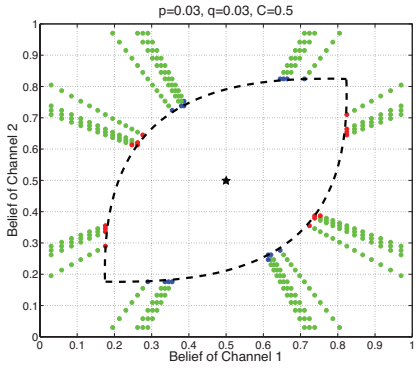

Fig. 2. Optimal decisions based on SAFs. White space corresponds to transient states under the optimal policy, and green, red, and blue dots correspond to recurrent states where the optimal action is to not probe, probe channel 1, and probe channel 2 respectively.

VI. Conclusion

In this paper, we studied the channel probing problem in a system of two channels. For fixed probing intervals, we proved that if the channels between users are i.i.d., the choice of which channel to probe is irrelevant. However, when the channels are non-identical, we characterized scenarios in which it becomes optimal to always probe one channel over the other. We then formulated the general problem of dynamic channel probe optimization, and described the form of the optimal solution as well as provided an LP to approximate the optimal solution.

References

[1] K. Jagannathan, S. Mannor, I. Menache, and E. Modiano, "A state action frequency approach to throughput maximization over uncertain wireless channels," in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011.

[2] S. Guha, K. Munagala, and S. Sarkar, "Jointly optimal transmission and probing strategies for multichannel wireless systems," in *Information Sciences and Systems, 2006 40th Annual Conference on*. IEEE, 2006.

[3] ——, "Optimizing transmission rate in wireless channels using adaptive probes," in *ACM SIGMETRICS Performance Evaluation Review*. ACM, 2006.

[4] P. Chaporkar and A. Proutiere, "Optimal joint probing and transmission strategy for maximizing throughput in wireless systems," *Selected Areas in Communications, IEEE Journal on*, 2008.

[5] N. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," in *International Conference on Mobile Computing and Networking: Proceedings of the 13 th annual ACM international conference on Mobile computing and networking*, 2007.

[6] A. Gopalan, C. Caramanis, and S. Shakkottai, "On wireless scheduling with partial channel-state information," in *Proc. Ann. Allerton Conf. Communication, Control and Computing*, 2007.

[7] K. Kar, X. Luo, and S. Sarkar, "Throughput-optimal scheduling in multi-channel access point networks under infrequent channel measurements," *Wireless Communications, IEEE Transactions on*, 2008.

[8] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *Information Theory, IEEE Transactions on*, 2009.

[9] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance," *Wireless Communications, IEEE Transactions on*, 2008.

[10] M. L. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, Inc., 1994.