Throughput/Energy Aware Opportunistic Transmission Control in Broadcast Networks

Sugumar Murugesan, *Member, IEEE*
Philip Schniter, *Senior Member, IEEE*
Ness B. Shroff, *Fellow, IEEE*

***Abstract*—We address throughput/energy aware opportunistic transmission control in broadcast networks, under incomplete channel state information. The channels are modeled as *i.i.d* two-state Markov chains and the transmission controller makes *transmit/idle* decisions based on past 1-bit feedback from the broadcast users. With a reward structure that captures the trade-off between sum throughput gain (*transmit*) and energy savings (*idle*), we formulate the transmission control problem as an infinite horizon, discounted reward, partially observable Markov decision process. For special cases of the system parameters, we show that the optimal control policy is either greedy or partially greedy. For the general case, we follow an indirect approach towards the control problem. We first establish thresholdability properties of the optimal control policy in a two-user broadcast. We then extrapolate these properties to the general broadcast and propose a simple threshold control policy. Extensive numerical results suggest near-optimal performance of the proposed policy. In addition, the proposed threshold policy is easy to implement with complexity being polynomial in the number of broadcast users.**

***Index Terms* – Broadcast network, transmission control, dynamic program, partially observable Markov decision process.**

I. Introduction

Broadcast networks, where a source node attempts to transmit a packet to all other nodes in the network, is an integral component of mobile ad-hoc and sensor networks [1]. In ad-hoc networks, broadcast plays a crucial role in a variety of protocols that provide basic functionality to higher layer services (e.g., [2]). In sensor networks, broadcast is used for coordinated and distributed computing (e.g., [3]). Thanks to the limited life of the mobile node batteries and a limited ability to replenish these batteries, energy aware transmission control in broadcast networks is an important design consideration. This is particularly true in sensor networks where nodes are often deployed in hard to access or hostile environments. A large volume of work (e.g., [4]-[10]) is available for energy efficient communication in wireless networks - broadcast and otherwise. The reader is directed to [11] for an excellent exposition on the topic. Much of these works, while providing valuable insights into energy efficient network design, are lacking in one of two ways: the physical channel considerations are disregarded and the problem is studied exclusively at the upper layers or, if the physical channel is indeed included in the design, the instantaneous channel state is assumed to be readily available at the controller.

In this paper, we attempt to address both these issues in broadcast networks. We consider a cross-layer throughput/energy aware transmission control problem with an explicit channel learning mechanism. In our setup, the broadcast channels are modeled by *i.i.d* two-state Gilbert Elliott Markov chains [12] with positive time correlation. The Gilbert Elliott channel model has been gaining popularity among wireless researchers (e.g., [13]-[20]) as a realistic abstraction of the fading channels. This is so, because multipath fading and shadowing in wireless channels induces time correlation (memory) in the channel that can be modeled, with reasonable accuracy, by a first-order Markov chain ([21],[22]). In each control slot, the controller makes one of the following two decisions: (1) *transmit* (broadcast) a packet to the users (2) stay *idle*. While a broadcast transmission is associated with a throughput gain (and a concurrent energy loss), an *idle* decision corresponds to energy savings (and a concurrent loss in throughput). Our reward structure reflects this trade-off – Upon *transmit* decision, the controller accrues a reward of 1 for each user that successfully decodes the broadcast packet. If an *idle* decision is made, a reward of W (reward for passivity - corresponding to energy savings) is accrued at the controller. At the end of a control slot, if a packet was broadcast in that slot, each user attempts to decode the packet and sends back bit 1 (decoding success) or bit 0 (decoding failure) to the controller, over an error-free feedback channel. The controller collects this state feedback from all the users and creates a belief value of the users' channel states in the next slot, using the Markov channel statistics. The belief values thus created are used by the controller to make *transmit/idle* decisions in future slots.

By formulating the control problem as a partially observable Markov decision process (POMDP) over an infinite horizon with discounted reward [23], we obtain the following main results: For particular ranges of the passivity reward and belief values, we show that the optimal control policy is greedy or partially greedy, depending on the case, thus significantly simplifying the control problem. For the general scenario, however, due to the dimensionality of the

This work was supported by the NSF CAREER grant 237037, NSF grants CNS-0721236, CNS-0626703, ARO W911NF-08-1-0238 and ARO W911NF-07-10376.

Murugesan and Schniter are with the Department of Electrical and Computer Engineering, and Shroff holds a joint appointment in both the Departments of Electrical and Computer Engineering, and Computer Science and Engineering at The Ohio State University, Columbus, OH 43210, USA (E-mail: {murugess, schniter, shroff}@ece.osu.edu.)

POMDP, a direct analysis of the optimal control problem appears challenging. We therefore approach the problem in two stages: (i) First, we establish thresholdability properties of the optimal control policy in a two-user broadcast (ii) We then extrapolate the thresholdability properties to the general broadcast and, based on this extrapolation, derive a threshold control policy. Extensive numerical experiments suggest that the proposed threshold policy, thanks to the optimality framework in which it is derived, has near-optimal performance. The proposed threshold policy is also computationally inexpensive to implement with polynomial complexity in the number of broadcast users. Contrast this with the complexity of the optimal POMDP solutions – for finite horizon POMDPs, the optimal solution is, in general, PSPACE-hard to compute [24], whereas infinite horizon POMDPs are, in general, *undecidable* [25].

The paper is organized as follows. The problem setup is described in Section II. In Section III-A, we characterize the optimal control policy for special cases of the system parameters. Considering a two-user broadcast, we obtain thresholdability properties of the optimal control policy in Section III-B. In Section IV, we propose the threshold control policy followed by a discussion on the numerical results in Section V. Concluding remarks are provided in Section VI.

II. Problem Setup

A. Channel Model and Control Structure

We consider a N user broadcast. The channel between the source node and each broadcast user is modeled by an *i.i.d* two-state Gilbert Elliott Markov chain. Assuming packetized data transmissions, each state corresponds to the degree of decodability of the packet sent through the channel. State 1 (ON) corresponds to full decodability, while state 0 (OFF) corresponds to zero decodability. Time is slotted and the channel of each user remains fixed for a slot and moves into another state in the next slot following the state transition probability of the Markov chain. The time slots of all users are synchronized. The two-state Markov channel is characterized by a 2×2 probability transition matrix

$$P = \begin{bmatrix} p & 1-p \\ r & 1-r \end{bmatrix}, \qquad (1)$$

where

$$\begin{aligned} p &:= \text{prob(channel is in ON state in the current slot} | \\ &\qquad \text{channel was in ON state in the previous slot)} \\ r &:= \text{prob(channel is in ON state in the current slot} | \\ &\qquad \text{channel was in OFF state in the previous slot).} \end{aligned}$$

The Markov channel states can be interpreted as a quantized representation of the underlying channel strength lying on a continuum. Since, in realistic scenarios, the channel strength is observed to evolve gradually over time (positive correlation), we assume $p > r$ throughout this work.

The source node is the central controller that, in each slot, makes one of the following two decisions: (1) *Transmit* (broadcast) a packet to the broadcast users (2) stay *idle*. The packets to be broadcast to the users are stored in an infinite queue at the source node (henceforth known as the controller). Upon *transmit* decision, the controller broadcasts the head of line packet to the users and drops it from the queue. Each user attempts to decode the packet and sends back bit 1 (decoding success) or bit 0 (decoding failure) to the controller, over an error-free feedback channel. By the definition of the two-state Markov channels, this 1-bit feedback from an user at the end of a slot gives the state of the channel of that user in that slot. Based on the 1-bit feedback from the users and the Markov channel statistics, the controller creates a belief value of the channel state of the users in the next slot and uses it to make future *transmit/idle* decisions. The control problem is thus a dynamic program [26], more specifically a partially observable Markov decision process [23]. We formally define the problem below.

B. Formal Problem Definition

Horizon: The number of consecutive time slots over which the control decisions are made is the horizon. Throughout this work, we focus on the infinite horizon scenario.

Action: Let $a \in \{0, 1\}$ indicate the action (control decision) taken in the current slot. Let $a = 1$ correspond to the *transmit* decision and $a = 0$ correspond to the *idle* decision.

Belief vector: Let $\pi = (\pi_1, \ldots, \pi_N) \in [0, 1]^N$ be the vector of belief values in the current slot with π_i denoting the belief value of the channel of user $i \in \{1, \ldots, N\}$. Define the operator $T(.)$ as the evolution of the belief value of a Markov channel to the next slot under the *idle* decision. Thus, if $x \in [0, 1]$ is the belief value, then $T(x) = xp + (1 - x)r$.

1-bit feedback: Upon *transmit* decision, at the end of the slot, each user i in the broadcast determines if the reception was successful and sends back a 1-bit feedback f_i (1 for success and 0 for failure). Feedback f_i is one-one mapped to the state of the channel of user i in the corresponding slot.

Expected immediate reward: In each slot, if *transmit* decision is made, the controller accrues a reward of 1 for each user that successfully decodes the broadcast packet. If an *idle* decision is made, a reward of W (reward for passivity - corresponding to energy savings) is accrued at the controller. Thus the expected immediate reward accrued by the controller as a function of belief vector and action is given by

$$R(\pi, a) = \begin{cases} \sum_i \pi_i, & \text{if } a = 1 \\ W, & \text{if } a = 0. \end{cases} \qquad (2)$$

Stationary control policy: A stationary control policy $\mathfrak{A}$ is a stationary mapping from the belief vector π to an action as follows:

$$\mathfrak{A} : \pi \rightarrow a \in \{0, 1\}.$$

Expected total discounted reward under $\mathfrak{A}$: Under a policy $\mathfrak{A}$, for initial belief vector π, the expected infinite horizon total discounted reward is given by

$$V_{\mathfrak{A}}(\pi) = R(\pi, a) + \beta E\big[V_{\mathfrak{A}}(\pi^+)\big] \qquad (3)$$

where $a = \mathfrak{A}(\pi)$ and $\beta \in [0,1)$ is the discount factor that determines the relative weight between the immediate and the future rewards. The expectation is over the belief vector in the next slot, i.e., π^+, which in turn is a function of the control decision a and (upon *transmit* decision) the 1-bit feedback from the users. We now proceed to explicitly express the total discounted reward under $\mathfrak{A}$. For notational simplicity, we first define a few quantities. Let $\Pi_0 \doteq (r, r, \ldots, r, r)$, $\Pi_1 \doteq (r, r, \ldots, r, p)$, $\Pi_2 \doteq (r, r, \ldots, p, r), \ldots, \Pi_{2^N-1} \doteq (p, p, \ldots, p, p)$. Let $P_0(\pi) \doteq (1-\pi_1)(1-\pi_2)\ldots(1-\pi_{N-1})(1-\pi_N)$, $P_1(\pi) \doteq (1-\pi_1)(1-\pi_2)\ldots(1-\pi_{N-1})(\pi_N)$, $P_2(\pi) \doteq (1-\pi_1)(1-\pi_2)\ldots(\pi_{N-1})(1-\pi_N), \ldots, P_{2^N-1}(\pi) \doteq \pi_1\pi_2\ldots\pi_{N-1}\pi_N$. The total discounted reward under $\mathfrak{A}$ is now explicitly given by

$$V_{\mathfrak{A}}(\pi) = \begin{cases} \sum_i \pi_i + \beta \sum_{j=0}^{2^N-1} P_j(\pi) V_{\mathfrak{A}}(\Pi_j), & \text{if } a = \mathfrak{A}(\pi) = 1 \\ W + \beta V_{\mathfrak{A}}(T(\pi)), & \text{if } a = \mathfrak{A}(\pi) = 0. \end{cases} \tag{4}$$

Optimal control policy: For a given belief vector π, the *optimal* expected total discounted reward (henceforth, simply the *total discounted reward*), $V(\pi)$, is given by the Bellman equation [26]

$$V(\pi) = \max\Big\{ \sum_i \pi_i + \beta \sum_{j=0}^{2^N-1} P_j(\pi) V(\Pi_j), \; W + \beta V(T(\pi)) \Big\}. \tag{5}$$

By standard dynamic programming theory [26], a stationary control policy $\mathfrak{A}^*$ is optimal if and only if the total discounted reward under $\mathfrak{A}^*$, i.e., $V_{\mathfrak{A}^*}(\pi)$, satisfies the Bellman equation in (5), for every $\pi \in [0,1]^N$, i.e., $\mathfrak{A}^*$ is optimal if and only if

$$V_{\mathfrak{A}^*}(\pi) = \max\Big\{ \sum_i \pi_i + \beta \sum_{j=0}^{2^N-1} P_j(\pi) V_{\mathfrak{A}^*}(\Pi_j), \; W + \beta V_{\mathfrak{A}^*}(T(\pi)) \Big\}. \tag{6}$$

III. Optimal Control Policy - Partial Characterization and Thresholdability Properties

A. Partial Characterization of the Optimal Control Policy

Define $V^a(\pi)$ as the expected total discounted reward upon *transmit* (active) decision in the current slot and optimal decisions in all future slots and $V^p(\pi)$ as the expected total discounted reward upon *idle* (passive) decision in the current slot and optimal decisions in all future slots, i.e.,

$$\begin{aligned} V^a(\pi) &= \sum_i \pi_i + \beta \sum_{j=0}^{2^N-1} P_j(\pi) V(\Pi_j) \\ V^p(\pi) &= W + \beta V(T(\pi)). \end{aligned} \tag{7}$$

Let $\mathcal{A}$ and $\mathcal{P}$ be the regions in the state space, $[0,1]^N$, where it is optimal to *transmit* and *idle*, respectively. Formally,

$$\pi \in \begin{cases} \mathcal{A}, & \text{if } V^a(\pi) \geq V^p(\pi) \\ \mathcal{P}, & \text{if } V^a(\pi) < V^p(\pi). \end{cases} \tag{8}$$

We now report our result on the optimal control policy when the reward for passivity $W \notin (Nr, Np)$.

Proposition 1. *When $W \notin (Nr, Np)$, the optimal control policy is greedy, i.e.,*

$$\pi \in \begin{cases} \mathcal{A}, & \text{if } \sum_i \pi_i \geq W \\ \mathcal{P}, & \text{if } \sum_i \pi_i < W \end{cases}.$$

Proof Outline: We first reformulate the infinite horizon control problem as a limit of the finite horizon problem and then using backward induction, we show that, for any horizon length, the optimal future discounted reward after the *transmit* decision, i.e., $V^a(\pi) - \sum_i \pi_i$, equals the reward after the *idle* decision, i.e. $V^p(\pi) - W$, when $W \notin (Nr, Np)$. Details of the proof are available in [27].

Proposition 2. *For any W, the optimal control policy has the following partial structure*

$$\pi \in \mathcal{A}, \text{ if } \sum_i \pi_i \geq W$$

Proof Outline: The proof proceeds by first showing that the total discounted reward, $V(\pi)$, is component-wise convex in π. We then show that the optimal future reward after the *transmit* decision is at least as high as the optimal future reward after the *idle* decision. This establishes the proposition. Details are available in [27].

It is worth noting that the energy loss per broadcast action (*transmit*) and hence W is independent of the number of broadcast users, N. Thus as N increases, the throughput gain (*transmit*) progressively outweighs the energy savings (*idle*). It follows that the optimal policy would increasingly choose to *transmit* than *idle*, with increasing broadcast size. This intuition is supported by the result in Proposition 2.

B. Thresholdability Properties of the Optimal Control Policy in the Two-User Broadcast

For general values of the system parameters, a direct analysis of the optimal control policy appears challenging, thanks to the dimensionality of the POMDP that models the control process. Therefore, we first study the control problem in a reduced dimension - the two-user broadcast - and later use this analysis to study the general broadcast.

We first classify the two-user broadcast into two types based on the optimal control decision at steady state, as below:

- Type I: If $(\pi_{ss}, \pi_{ss}) \in \mathcal{A}$
- Type II: If $(\pi_{ss}, \pi_{ss}) \in \mathcal{P}$

Let R_I denote the region $\{(\pi_1, \pi_2); \pi_1 \in [\pi_{ss}, 1], \pi_2 \in [\pi_{ss}, 1]\}$. Let R_{II} denote the union of the regions $R_{II}^1 \doteq \{(\pi_1, \pi_2); \pi_1 \in [0, \pi_{ss}], \pi_2 \in [\pi_{ss}, 2\pi_{ss} - \pi_1]\}$ and $R_{II}^2 \doteq \{(\pi_1, \pi_2); \pi_2 \in [0, \pi_{ss}], \pi_1 \in [\pi_{ss}, 2\pi_{ss} - \pi_2]\}$. We now record our result on the thresholdability properties of the optimal control policy in the Type-I two-user broadcast.

Proposition 3. *If the two-user broadcast is Type I, i.e., $V^a(\pi_{ss},\pi_{ss}) \geq V^p(\pi_{ss},\pi_{ss})$, then*

(1) $R_I \in \mathcal{A}$

(2) $V^a(\pi_{ss},\pi_{ss}) = V^p(\pi_{ss},\pi_{ss}) \Rightarrow R_{II} \in \mathcal{P}$

(3) $V^a(\pi_{ss},\pi_{ss}) > V^p(\pi_{ss},\pi_{ss}) \Rightarrow$ *(Thresholdability property) In the region R^1_{II}, if for $k \in [-1,0]$, $\exists$ a π_1^* and $\pi_2^* = \pi_1^* k + \pi_{ss}(1-k)$ such that $V^a(\pi_1^*,\pi_2^*) = V^p(\pi_1^*,\pi_2^*)$, then*

$$(\pi_1, \pi_2 = \pi_1 k + \pi_{ss}(1-k)) \in \begin{cases} \mathcal{A}, & \text{if } \pi_1 \in [\pi_1^*, \pi_{ss}] \\ \mathcal{P}, & \text{if } \pi_1 \in [0, \pi_1^*) \end{cases}$$

If $\nexists$ such a (π_1^*,π_2^*), then

$$(\pi_1, \pi_2 = \pi_1 k + \pi_{ss}(1-k)) \in \mathcal{A} \ \forall\ \pi_1 \in [0, \pi_{ss}].$$

Similarly, in the region R^2_{II}, if for $k \in [-1,0]$, $\exists$ a $\pi_2^* \in [0,\pi_{ss}]$ and $\pi_1^* = \pi_2^* k + \pi_{ss}(1-k)$ such that $V^a(\pi_1^*,\pi_2^*) = V^p(\pi_1^*,\pi_2^*)$, then

$$(\pi_1 = \pi_2 k + \pi_{ss}(1-k), \pi_2) \in \begin{cases} \mathcal{A}, & \text{if } \pi_2 \in [\pi_1^*, \pi_{ss}] \\ \mathcal{P}, & \text{if } \pi_2 \in [0, \pi_1^*) \end{cases}$$

If $\nexists$ such a (π_1^*,π_2^*), then

$$(\pi_1 = \pi_2 k + \pi_{ss}(1-k), \pi_2) \in \mathcal{A} \ \forall\ \pi_2 \in [0, \pi_{ss}].$$

Proof Outline: We first establish crucial structural properties of the reward functions $V(\pi), V^a(\pi), V^p(\pi)$ in the two-dimensional state space along specific axes $(\pi_1, \pi_2 = \pi_1 k + c)$, for various ranges of $k, c \in \mathbb{R}$ and show that the belief vector, upon consecutive *idle* decisions approaches the steady state in a straight line. Using these properties, with tedious algebraic manipulation, we establish the proposition. Details can be found in [27].

We now record our result on the thresholdability properties of the optimal control policy when the two-user broadcast is Type II.

Proposition 4. *If the two-user broadcast is Type II, then*

(1) $(\pi_1,\pi_2) \in \mathcal{P}$, $\forall \pi_1 + \pi_2 \leq 2\pi_{ss}$

(2) *(Thresholdability property) In the region R_I, if for $k \geq 0$, $\exists$ a π_1^* and $\pi_2^* = \pi_1^* k + \pi_{ss}(1-k)$ such that $V^a(\pi_1^*,\pi_2^*) = V^p(\pi_1^*,\pi_2^*)$, then*

$$(\pi_1, \pi_1^* k + \pi_{ss}(1-k)) \in \begin{cases} \mathcal{A}, & \text{if } \pi_1 \in [\pi_1^*, 1] \\ \mathcal{P}, & \text{if } \pi_1 \in [\pi_{ss}, \pi_1^*) \end{cases}$$

If $\nexists$ such a (π_1^*,π_2^*), then

$$(\pi_1, \pi_2 = \pi_1 k + \pi_{ss}(1-k)) \in \mathcal{P} \ \forall\ \pi_1 \in [\pi_{ss}, 1].$$

Proof Outline: The proof proceeds along the lines of the proof of Proposition 3. The reader is referred to [27] for details.

Call the set of points (π_1^*,π_2^*) identified in Proposition 3 and Proposition 4 as the threshold boundary in Type I and Type II broadcasts, respectively. Recall the definition of $V^a(\pi)$ from (7). We now characterize the threshold boundaries below.

Corollary 5. *When the broadcast is Type I: within region R_{II}, the threshold boundary is given by the upper segment of the hyperbola*

$$V^a(\pi_1,\pi_2) \quad = \quad W + \beta V^a(T(\pi_1), T(\pi_2)).$$

When the broadcast is Type II: Within region R_I, the threshold boundary is given by the upper segment of the hyperbola

$$V^a(\pi_1,\pi_2) \quad = \quad \frac{W}{1-\beta}.$$

The upper segment of the hyperbola indicates the segment of the hyperbola that lies in the first quadrant around the asymptotes.

Proof Outline: The corollary is established in two stages. We first derive the threshold equations using the thresholdability properties reported in Proposition 3 and Proposition 4 and show that it is a hyperbola. Next we study the slope of the hyperbola and the hyperbola asymptotes to show that the threshold boundary is the *upper segment* of the hyperbola for both broadcast types. The reader is referred to [27] for details.

Fig. 1 illustrates the threshold boundaries for both broadcast types.

IV. Threshold Control Policy

For the two-user broadcast, we have shown that the optimal control policy has thresholdability properties and have characterized the threshold boundaries in specific regions of the two-dimensional state space. In this section, without proof, we extrapolate the threshold boundaries to the entire state space and derive threshold control policies for the two-user broadcast and subsequently for the N-user broadcast.

The extrapolation is formally stated next: *The thresholdability property of the optimal control policy reported in Proposition 3 and Proposition 4, in regions R_{II} and R_I, respectively, extends to the entire state space* $[0,1]^2$. The extrapolated threshold boundary spanning the entire state space is given by the hyperbola equations in Corollary 5 for Type I and Type II broadcasts. Fig. 2 illustrates this extrapolated boundary for both broadcast types.

The threshold control policy for the two-user broadcast is now given below.

Step 0: (Initialization) Identify the broadcast type and evaluate the quantities $V(\Pi(0)) \ldots V(\Pi(3))$.

Step 1: With (π_1,π_2) denoting the belief vector in the current slot, solve the second order polynomial equation (polynomial in k) for the corresponding broadcast type.

$$\text{Type I}: V^a(k\pi_1,k\pi_2) - \big(W + \beta V^a(T(k\pi_1),T(k\pi_2))\big) = 0,$$

$$\text{Type II}: V^a(k\pi_1,k\pi_2) - \frac{W}{1-\beta} = 0.$$

Note that the polynomial nature of the preceding equations can be verified by examining the expression for V^a in (7). Let k_1, k_2 be the solutions to these equations. Let $k^* = \max\{k_1,k_2\}$.

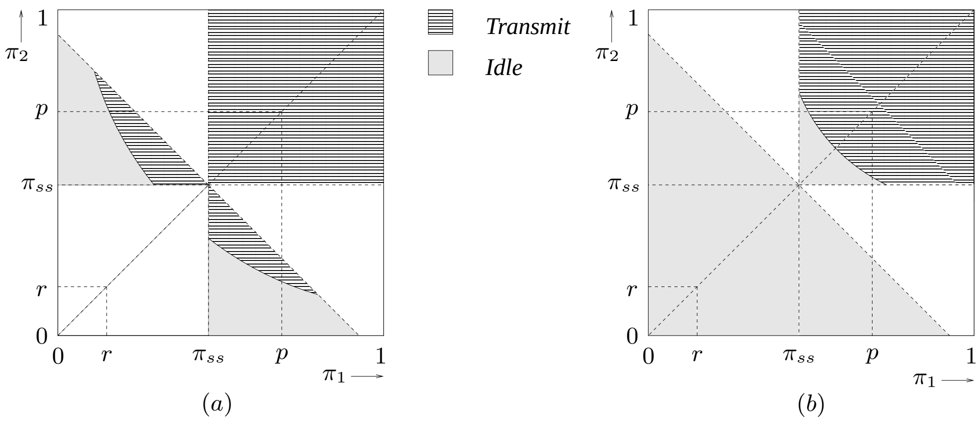

Fig. 1. Illustration of the threshold boundaries in restricted regions of the two-dimensional state space when the broadcast is (a) Type I, (b) Type II.

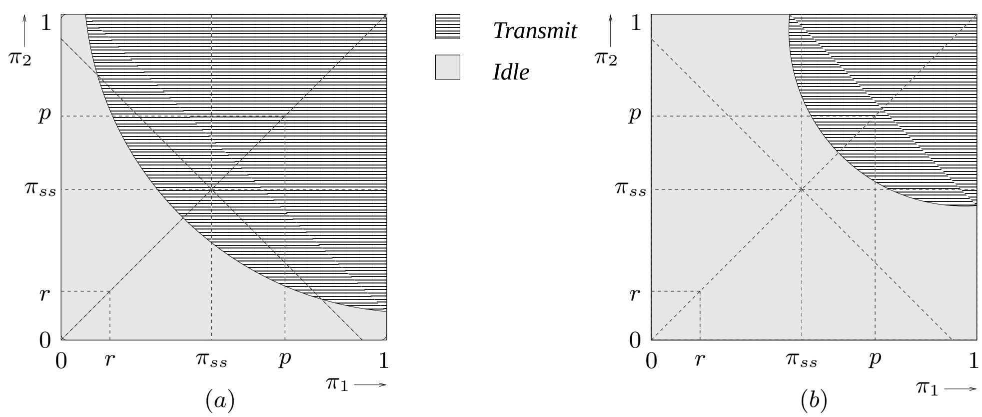

Fig. 2. Illustration of the extrapolation of the threshold boundaries to the entire two-dimensional state space when the broadcast is (a) Type I, (b) Type II.

Step 2: The threshold control policy is given by

$$\text{Transmit, if } k^* \leq 1$$
$$\text{Idle, if } k^* > 1$$

Note that Step 1 identifies the point on the extrapolated threshold boundary along the direction of the belief vector (π_1, π_2), with $k^* = \max\{k_1, k_2\}$ resulting from the fact that the threshold boundary is given by the *upper* segment of the hyperbola (Corollary 5). The second step determines the control decision based on the location of the belief vector with respect to the extrapolated threshold boundary.

We now extend the preceding threshold policy, heuristically, to the N-user broadcast. Before that, we generalize the two-user broadcast classification to the N-user case: The N-user broadcast is identified as Type I if $V^a(\pi_{ss}, \ldots, \pi_{ss}) \geq V^p(\pi_{ss}, \ldots, \pi_{ss})$ and Type II otherwise. Denote $(\pi_{ss} \ldots \pi_{ss})$ simply by π_{ss}. Using the identity $T(\pi_{ss}) = \pi_{ss}$, with elementary reasoning [27], this rule can be simplified as below:

$$\text{Type} \quad = \quad \begin{cases} I, & \text{if } V^a(\pi_{ss}) \geq \frac{W}{1-\beta} \\ II, & \text{if } V^a(\pi_{ss}) < \frac{W}{1-\beta}. \end{cases} \tag{9}$$

The threshold control policy for the N-user broadcast is now given by the following steps, when $W \in (Nr, Np)$:

Step 0: Initialization

- *Evaluate the quantities* $V(\Pi_0), \ldots, V(\Pi_{2^N-1})$*:* Note that, by the inherent symmetry in the underlying Markov channel statistics of the users, we have the following property (P): For any permutation Ψ on the belief vector x, we have $V(x) = V(\Psi(x))$. Thus the 2^N quantities, $V(\Pi_0), \ldots, V(\Pi_{2^N-1})$, can be obtained by evaluating only the following $N+1$ quantities: $\{V(\Pi_{(2^i-1)}\}, \ i \in \{0, 1, 2, \ldots, N\}$. Interpreting the infinite horizon discounted rewards as limits on the finite horizon rewards, as we did in the proof of Proposition 1, evaluate $V(\Pi_0) = \lim_{t\to\infty} V_t(\Pi_0)$, ..., $V(\Pi_{2^N-1}) = \lim_{t\to\infty} V_t(\Pi_{2^N-1})$, using an appropriate measure of convergence. The finite horizon reward $V_t(x)$ is given by the finite horizon Bellman equation [26]

$$V_t(x) \quad = \quad \max\Big\{\sum_i x_i + \beta \sum_{j=0}^{2^N-1} P_j(x) V_{t-1}(\Pi_j), \\ W + \beta V_{t-1}(T(x))\Big\}.$$

Step 0: Initialization (continued)

- *Identify the broadcast type:*
  The broadcast type is identified using the simplified rule (9), reproduced below:

$$\text{Type} \;=\; \begin{cases} I, & \text{if } V^a(\pi_{ss}) \geq \frac{W}{1-\beta} \\ II, & \text{if } V^a(\pi_{ss}) < \frac{W}{1-\beta}. \end{cases}$$

  where $V^a(\pi_{ss})$ is evaluated using (7), simplified using property (P) as

$$V^a(\pi_{ss}) \;=\; N\pi_{ss} + \beta \sum_{j=0}^{N} NC_j (1-\pi_{ss})^{(N-j)} \times \pi_{ss}^j V(\Pi(2^j - 1)),$$

  with $V(\Pi(0)) \ldots V(\Pi(2^N - 1))$ evaluated earlier.

Step 1: Threshold control policy on belief vector π

- If $\sum_i \pi_i > W$, *transmit* (follows from Proposition 2. Skip to Step 2.
- Otherwise, solve the following N^{th} order polynomial equation for the corresponding broadcast type:

  If the broadcast is Type I:

$$V^a(k\pi) - \big(W + \beta V^a(T(k\pi))\big) = 0.$$

  If the broadcast is Type II:

$$V^a(k\pi) - \frac{W}{1-\beta} = 0,$$

  where, recall from (7),

$$V^a(x) \;=\; \sum_i x_i + \beta \sum_{j=0}^{2^N-1} P_j(x) V(\Pi(j)),$$

  with $V(\Pi(0)) \ldots V(\Pi(2^N - 1))$ evaluated in Step 0. Let $\{k_1, \ldots, k_N\}$ be the solutions to the polynomial equation. Let $k^* = \max_{i \text{ s.t. } k_i \in \mathbb{R}} \{k_i\}$.
- The threshold policy is given by

$$\text{Transmit, if } k^* \leq 1$$
$$\text{Idle, if } k^* > 1$$

Step 2: State evolution

- If *transmit* decision was made, at the end of the slot, collect the 1-bit feedback, $f_1, \ldots, f_N$, from the broadcast users and update the belief values as below.

$$\pi_i \leftarrow \begin{cases} p, & \text{if } f_i = 1 \\ r, & \text{if } f_i = 0 \end{cases}.$$

- If *idle* decision was made, update the belief vector as $\pi \leftarrow T(\pi)$.
- Repeat Step 1 in the next slot.

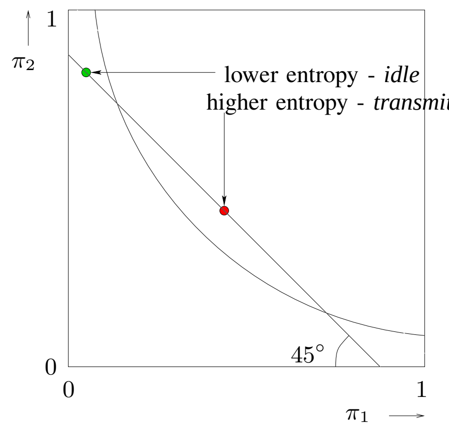

Fig. 3. Illustration of the connection between the threshold scheduling decision and the entropy of the broadcast system state.

Remark: The convexity of the threshold boundary renders optimality properties to the threshold policy in the following sense. Consider two belief vectors π^g and π^r such that $\sum \pi^g = \sum \pi^r$. Thus, if the broadcast state is in any of these two states, the immediate rewards upon *transmit* decisions are the same. Indeed, the immediate rewards upon *idle* decisions are equal to W in both states. Let the entropy of the broadcast in state π^g be lower than when the broadcast is in state π^r, i.e., $\sum_i H(\pi^g(i)) < \sum_i H(\pi^g(i))$, where $H(x)$ indicates the entropy of a channel with belief x. It is intuitive to see that if it is optimal to *transmit* at state π^g, then it is optimal to *transmit* at state π^r. This is because, with the 'exploitation' end of the trade-off equalized between π^g and π^r (since $\sum \pi^g = \sum \pi^r$), the exploration end of the tradeoff is more pronounced in π^r due to its higher entropy, essentially making it optimal to *transmit*, i.e., explore at π^r, if it is optimal to explore at π^g. Since the threshold boundary in the proposed threshold policy is convex, if the threshold decision at π^g is to *transmit*, then the threshold decision at π^r is also to *transmit*. An illustration of π^g, π^r along with the convex threshold boundary is provided in Fig. 3. Thus the threshold policy, thanks to the optimality framework in which it is derived, exhibits an implementation structure similar to that of the optimal policy.

V. Numerical Results and Discussion

We now proceed to illustrate, via numerical experiments, that the proposed policy has near-optimal performance. We first study the finite horizon performance of the proposed policy in Fig. 4. The total discounted reward corresponding to the proposed policy over a finite horizon m (denoted by $V_{\text{policy}}(m)$) is plotted alongside the optimal total discounted reward over horizon m (denoted by $V(m)$). Note that $V_{\text{policy}}(m)$ is almost indistinguishable from $V(m)$. Recall that the threshold policy was derived by extending the structural properties of the optimal control policy from the two-user broadcast to the general N-user broadcast. The superior performance of the threshold policy, indicated by Fig. 4 (and subsequent numerical results), suggests that the structural properties indeed extend to the N-user broadcast.

W	p	r	β	V	V_{policy}	%opt
3.2323	0.8147	0.7380	0.2762	4.4651	4.4651	100 %
3.8261	0.9575	0.9239	0.2946	5.4227	5.4227	100 %
1.6813	0.4218	0.3862	0.6753	5.6400	5.6399	99.9987 %
0.9212	0.2769	0.0128	0.2583	2.3176	2.3176	100 %
1.5327	0.4387	0.1674	0.6593	4.4950	4.4950	100 %
2.2140	0.6491	0.4750	0.5886	5.3344	5.3253	99.8298 %
1.9074	0.6868	0.1260	0.4211	4.0446	4.0446	100 %
1.1852	0.4868	0.2122	0.4681	3.8936	3.8935	99.9994 %
1.8291	0.6443	0.2439	0.6869	6.5860	6.5726	99.7966 %
1.7714	0.6225	0.3654	0.3246	2.7002	2.7002	100 %

TABLE I
ILLUSTRATION OF THE NEAR-OPTIMAL PERFORMANCE OF THE PROPOSED THRESHOLD POLICY. TOTAL REWARD VALUES ARE TRUNCATED TO FOUR DECIMAL PLACES. EACH ROW CORRESPONDS TO A FIXED SET OF RANDOMLY GENERATED SYSTEM PARAMETERS AND INITIAL BELIEF VALUES. NUMBER OF BROADCAST USERS = 4.

N	W	p	r	β	V_{genie}	V_{policy}	V_{nofb}	%fbgain
2	0.8139	0.4456	0.2880	0.6256	2.2523	2.2446	2.0923	95.1679 %
2	0.4801	0.2630	0.1720	0.6135	1.2585	1.2570	1.2017	97.4352 %
2	0.7949	0.6477	0.2921	0.5282	1.9907	1.9788	1.8996	86.9424 %
3	1.5819	0.5469	0.5236	0.7789	6.8312	6.7480	6.0094	89.8665 %
3	2.2272	0.8003	0.1135	0.4531	4.2901	4.2875	4.0562	98.8572 %
3	1.3724	0.5085	0.2597	0.6906	4.5968	4.5653	4.1031	93.6145 %
4	1.8299	0.5085	0.2597	0.6906	6.0284	6.0083	5.4709	96.3937 %
4	2.3315	0.7513	0.1916	0.5036	4.9462	4.9347	4.6579	96.0039 %
4	0.7165	0.4709	0.1085	0.7066	2.8552	2.8015	2.2272	91.4621 %
5	1.1834	0.6948	0.2203	0.7701	8.4060	8.4045	7.6546	99.8030 %
5	2.0542	0.4898	0.2182	0.5878	5.3549	5.3510	4.8626	99.2000 %
5	0.3981	0.1190	0.0593	0.7758	3.4376	3.3988	1.4755	98.0243 %

TABLE II
ILLUSTRATION OF THE GAIN ASSOCIATED WITH 1-BIT FEEDBACK. EACH ROW CORRESPONDS TO A FIXED SET OF RANDOMLY GENERATED SYSTEM PARAMETERS AND INITIAL BELIEF VALUES. REWARD VALUES ARE TRUNCATED TO FOUR DECIMAL PLACES.

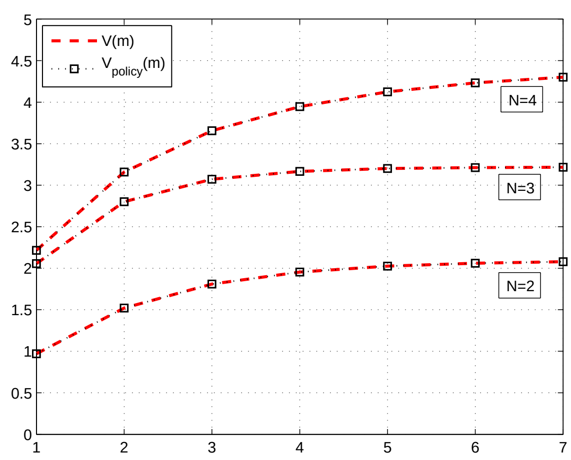

Fig. 4. $V(m)$, $V_{\text{policy}}(m)$ versus horizon m for various number of broadcast users.

In the rest of this analysis, we will focus on the infinite horizon performance of the proposed policy, compared with various system level performance limits. Note that the infinite horizon reward can be approximated by evaluating finite horizon rewards over a 'sufficiently' large horizon. From exhaustive simulations, we observed that the reward functions achieve reasonable convergence around $m = 7$ (also seen in Fig. 4). We therefore approximate the infinite horizon rewards by the rewards evaluated at $m = 7$ in the rest of this analysis. In Table I, we report the $\%$ suboptimality of the proposed policy. The quantity $\%$opt:=$\frac{V_{\text{policy}}}{V} \times 100\%$ quantifies the degree of optimality of the proposed policy. Each row in Table I corresponds to randomly generated system parameters with $N = 4$. The near optimal performance of the proposed policy is once again evident from Table I.

In Table II, we study the gains achieved by using 1-bit feedback from the users. The quantity V_{genie} corresponds to the total discounted reward under optimal scheduling in the *genie-aided* system defined as follows: at the end of each slot — independent of whether a transmit or idle decision was made in that slot — the scheduler learns about the channel states of all the users in that slot. The quantity V_{nofb} is the total discounted reward when the scheduler rejects the feedback information from the scheduled users and schedules solely based on the knowledge of the system level parameters, i.e., N, W, β and the statistics of the Markov channels, i.e., p and r. Thus with horizon $m = 7$, $V_{\text{nofb}} = \max\{W, N\pi_{ss}\}\frac{1-\beta^7}{1-\beta}$, where the steady state probability of the Markov channels, $\pi_{ss} = \frac{r}{1-(p-r)}$. The gain corresponding to the 1-bit feedback from each user, at the end of slots when *transmit* decision was made, is now quantified by the quantity %fbgain=$\frac{V_{\text{policy}}-V_{\text{nofb}}}{V_{\text{genie}}-V_{\text{nofb}}} \times 100\%$. The high value of %fbgain reported in Table II, for various randomly generated system parameters, underlines the significance of using the 1-bit feedback as well as the near-optimal performance of the proposed policy.

Remark on Complexity: The proposed threshold policy involves solving a N^{th} order polynomial, the complexity of which is polynomial in N when eigenvalue based solution techniques are used. Thus the proposed policy is computationally inexpensive to implement. Contrast this with the complexity of the optimal POMDP solutions – for finite horizon POMDPs, the optimal solution is, in general, PSPACE-hard to compute [24], whereas infinite horizon POMDPs are, in general, *undecidable* [25].

VI. Conclusion

We studied throughput/energy aware opportunistic transmission control in Markov-modeled broadcast networks when channel state information is estimated using delayed 1-bit feedback from the users. Formulating the control problem as an infinite horizon, discounted reward, partially observable Markov decision process, we showed that, for specific ranges of system parameters, the optimal control policy is either greedy or partially greedy. For the general case, we proposed a threshold control policy that is derived in an optimization framework for the two-user broadcast. The proposed policy is amenable for practical implementation with complexity being polynomial in the number of broadcast users. Extensive numerical results suggest that opportunistic transmission control using only 1-bit feedback from users obtained only during the *transmit* slots is associated with significant system level gains and that almost all of these gains can be realized using the proposed threshold control policy.

References

[1] Andrew Tanenbaum, *Computer Networks,* Prentice Hall, ed. 4, 2003.

[2] I. Chlamtac and S. Kutten, "On broadcasting in radio networks - problem analysis and protocol design," *IEEE Transactions on Communications,* vol. com-33, no. 12, pp. 1240-1246, Dec. 1985.

[3] A Durresi, V. K. Paruchuri, S. S. Iyengar and R. Kannan, "Optimized broadcast protocol for sensor networks," *IEEE Transactions on Computers,* vol. 54 , no. 8, Aug. 2005.

[4] M. Agarwal, J. H. Cho, L. Gao and J. Wu, "Energy efficient broadcast in wireless ad hoc networks with hitch-hiking," *Proc. IEEE INFOCOM,* 2004.

[5] S. Singh, C. Raghavendra and J. Stepanek, "Power-aware broadcasting in mobile ad hoc networks," *Proc. IEEE INFOCOM*, Sept. 1999.

[6] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol. 9, no. 4, pp. 8-27, Aug. 2002.

[7] A. Ephremides, "Energy concerns in wireless networks," *IEEE Wireless Communications,* vol. 9, no. 4, pp. 48-59, Aug. 2002.

[8] R. Min, M. Bhardwaj, S. Cho, N. Ickes, E. Shih, A. Sinha, A. Wang, and A. Chandrakasan, "Energy-centric enabling technologies for wireless sensor networks," *IEEE Wireless Communications,* vol. 9, no. 4, pp. 28-39, Aug. 2002.

[9] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine,* vol. 43, no. 12, pp. 112-119, Dec. 2005.

[10] X Lin, N. B. Shroff and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Communications,* vol. 24, no. 8, pp. 1452-1463, Aug. 2006.

[11] G. Miao, N. Himayat, Y. Li and A. Swami, "Cross-layer optimization for energy-efficient wireless communications: a survey," *Wireless Communications and Mobile Computing,* vol. 9, no. 4, pp. 529-542, Apr. 2009.

[12] E. Gilbert, "Capacity of a burst-noise channel," *Bell Systems Technical Journal,* vol. 39, pp. 1253-1266, 1960.

[13] L. A. Johnston and V. Krishnamurthy, "Opportunistic file transfer over a fading channel: a POMDP search theory formulation with optimal threshold policies," *IEEE Transactions on Wireless Communications,* vol. 5, no. 2, Feb. 2006.

[14] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Transactions on Networking,* vol. 7, no. 4, pp. 473-489, Aug. 1999.

[15] T. Nandagopal, S. Lu, and V. Bharghavan, "A unified architecture for the design and evaluation of wireless fair queueing algorithms," *Proc. ACM Mobicom,* Aug. 1999.

[16] T. Ng, I. Stoica, and H. Zhang, "Packet fair queueing algorithms for wireless networks with location-dependent errors," *Proc IEEE INFOCOM,* (New York), vol. 3, 1998.

[17] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *Proc. ACM Workshop on Wireless and Mobile Multimedia,* (Seattle, WA), Aug. 1999.

[18] Y. Cao and V. Li, "Scheduling algorithms in broadband wireless networks," *Proc. IEEE,* vol. 89, no. 1, pp. 76-87, Jan. 2001.

[19] S. Murugesan, P. Schniter, and N. B. Shroff, "Multiuser scheduling in a Markov-modeled downlink environment," *Proc. Allerton Conf. on Communication, Control, and Computing,* (Monticello, IL), Sept. 2008.

[20] S. H. Ahmad, M. Liu, T. Javidi, Q. Zhao and B. Krishnamachari, "Optimality of myopic sensing in multi-channel opportunistic access," *IEEE Transactions on Information Theory*, vol. 55, No. 9, pp. 4040-4050, September, 2009.

[21] H. S. Wang and P. C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Vehicular Technology,* vol. 45, pp. 353-357, May 1996.

[22] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Transactions on Communications,* vol. 47, no. 11, Nov. 1999.

[23] E. J. Sondik, "The optimal control of partially observable Markov processes," *PhD Thesis*, Stanford University, 1971.

[24] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Mathematics of Operations Research,* 12, 441-450, 1987.

[25] 0. Madani, S. Hanks and A. Condon, "On the undecidability of probabilistic planning and infinite-horizon partially observable Markov decision processes," *Proceedings of the Sixteenth National Conference on Artifi cial Intelligence,* 1999.

[26] D. P. Bertsekas, *Dynamic programming and optimal control,* Athena Scientific, vol. 1, ed. 3, 2007.

[27] S. Murugesan, P. Schniter and N. B. Shroff, "Throughput/Energy Aware Opportunistic Transmission Control in Broadcast Networks," Proc. Allerton Conf. on Communication, Control, and Computing, (Monticello, IL), Sept. 2010 (www.ece.osu.edu/∼schniter).