Throughput/Energy Aware Opportunistic Transmission Control in Broadcast Networks

Sugumar Murugesan
School of ECEE
Arizona State University

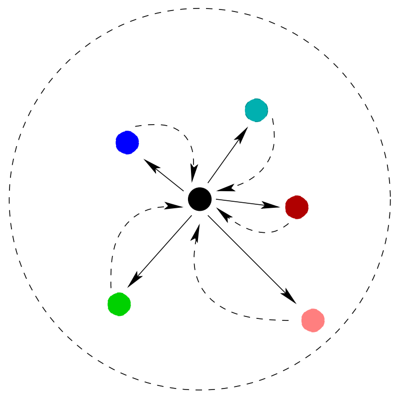
Philip Schniter, Ness B. Shroff
IPS laboratory, Dept. of ECE
The Ohio State University

Outline

- Background and Motivation
- Problem Setup
- Optimal Control Policy - Special Cases
- Thresholdability Analysis and Threshold Control Policy
- Numerical Results
- Conclusions

Background and Motivation

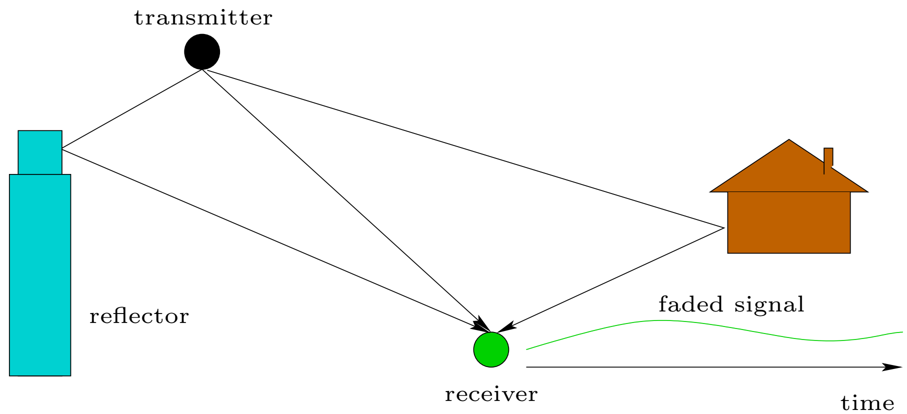
- Broadcast networks: a designated source node attempts to transmit a packet to all the users in the network



- Integral component of mobile ad-hoc and sensor networks
- Need for energy efficiency:
  - Limited battery life of mobile nodes
  - Nodes often deployed in hard to access/hostile environments (sensor networks) - battery recharge may not be an easy task

Background and Motivation

- Need a cross-layer approach to throughput/energy aware transmission control
- Opportunistic scheduling/control:
  - Shadowing, constructive/destructive interference between multiple signal paths lead to fluctuations in channel conditions - fading



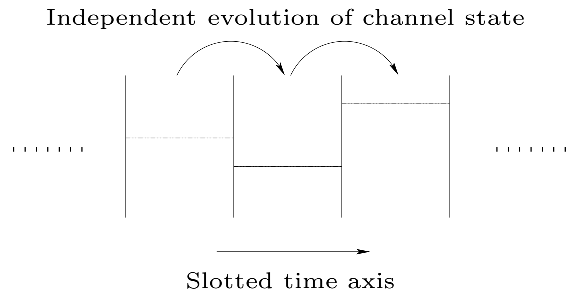
  - In a multiuser scheduling scenario, scheduler can ‘ride the peak’ channel conditions
  - In the broadcast network, source node can conserve energy by broadcasting only when the instantaneous, overall channel condition is ‘favorable’

Opportunistic Scheduling - Challenges

- Availability of channel state information (CSI) at the scheduler is pivotal to the success of opportunistic scheduling
- Significant amount of system resource is lost in learning the channel through traditional means like *pilot-aided training* - potentially offsetting the gains from opportunistic scheduling
- Problem of acquiring CSI is tightly coupled with the problem of exploiting CSI through opportunistic scheduling – need efficient joint CSI acquisition - opportunistic scheduling mechanisms that tap into the existing network resources

Solution – Exploit Channel Memory

- Take a step back - focus on the channel modeling philosophy
- Fading channels are traditionally abstracted by block fading models



- Non-negligible amount of time correlation (memory) is observed to be present in the radio frequency links that can be captured by Markov chain models
- This memory can be exploited for joint CSI acquisition - opportunistic scheduling
  - V. Krishnamurthy et al: Opportunistic file transfer over a fading channel: a POMDP search theory formulation, 2006
  - B. Krishnamachari et al: Optimality of myopic sensing in multi-channel opportunistic access, 2009

Outline

- Background and Motivation
- Problem Setup
- Optimal Control Policy - Special Cases
- Thresholdability Analysis and Threshold Control Policy
- Numerical Results
- Conclusions

Problem Setup

- Channel model:
  - Channels between the source node (controller) and the broadcast users are modeled by independent and identically distributed two-state (ON,OFF) Markov chains with parameters p (P(ON/ON)) and r (P(ON/OFF)).
  - Markov channels are positively correlated, i.e., $p > r$
- Control Model: Time is slotted. In each control slot, one of two actions is taken:
  - *transmit* (broadcast) a packet to the users - corresponds to throughput gain
  - *idle* - corresponds to energy savings
  - 1-bit feedback from each user at the end of *transmit* slots.
  - Controller maintains belief values of the channels of the users. Updates these belief values based on the end-of-slot 1-bit feedback it receives

Problem Setup (continued)

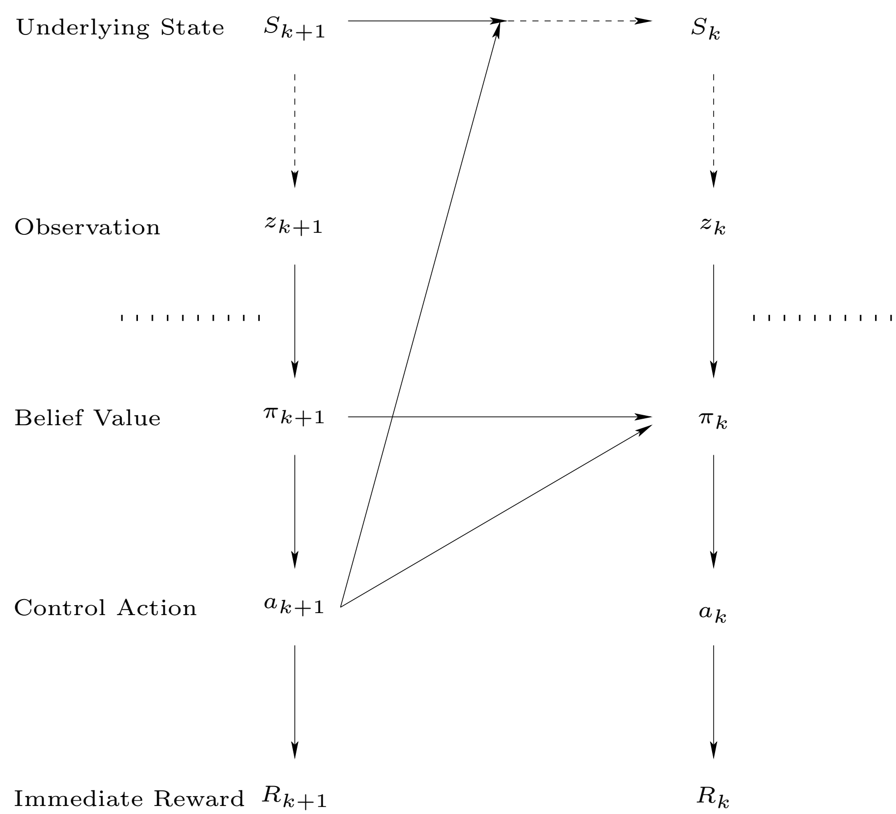
- Reward structure:
  - Upon *transmit* decision, the controller accrues a reward of 1 for each user that successfully decodes the broadcast packet
  - Upon *idle* decision, a reward of W is accrued at the controller

Control Problem Characteristics

- *idle* decision can be considered as transmission to a ‘user’ with a constant channel and throughput gain W
- *transmit* decision corresponds to transmission to a ‘cumulative user’ with a cumulative Markov channel
- Fundamental trade-off: Exploitation vs Exploration
  - Exploitation: Schedule the user with the best (perceived) channel condition at the moment - immediate gains in throughput
  - Exploration: Schedule a user and thus probe a user’s channel for better understanding of the channels, and hence better opportunistic scheduling, in the future. May need to compromise on immediate gains
- Trade-off captured by modeling the control problem as a *partially observable Markov decision process*

Control Problem as a POMDP

- *Partially Observable Markov Decision Process (POMDP)* - Controller must act based on partial observations of the underlying system state. Immediate reward accrued is a function of the underlying state and action. System evolves to the next state, probabilistically dependent on the current state and action



Control Problem as a POMDP

- Control problem is formulated as an infinite horizon, discounted reward POMDP
- Optimality (Bellman) equation:

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

  V is the optimal total discounted reward, π_i are the belief values, β is the discount factor, $P_0, \ldots, P_{2^N-1}$ denote probabilities of the 2^N cumulative states of the underlying Markov chains, $\Pi_0 \ldots \Pi_{2^N-1}$ denote belief vectors $[r, \ldots, r], \ldots, [p, \ldots, p]$, $T(x) = xp + (1-x)r$ is the belief evolution under *idle* decision

- Total reward for any stationary policy, $\mathfrak{A}$, is given by

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

  $\mathfrak{A}$ is optimal if and only if $V_{\mathfrak{A}} = V$

Outline

- Background and Motivation
- Problem Setup
- Optimal Control Policy - Special Cases
- Thresholdability Analysis and Threshold Control Policy
- Numerical Results
- Conclusions

Optimal Control Policy - Special Cases

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

$$\begin{aligned} \textit{transmit}, &\quad \text{if } \textstyle\sum_i \pi_i \geq W \\ \textit{idle}, &\quad \text{if } \textstyle\sum_i \pi_i < W \end{aligned}$$

  Sketch of proof: Established 'component-wise convexity' of optimal total discounted reward, V, over the N-dimensional state space. Using this, showed that optimal future reward after *transmit* and *idle* decisions are equal

Optimal Control Policy - Special Cases

- *Proposition*: For any W, the optimal policy has the following partial structure

$$\textit{transmit} \text{ if } \sum_i \pi_i \geq W$$

  Sketch of proof: Established that the optimal future reward after *transmit* decision is at least as high as the optimal future reward after *idle* decision

- Remarks:
  - Greedy policy is inexpensive to implement (with $N-1$ comparison operations)
  - With W being independent of the number of users, the higher the number of broadcast users, the better to *transmit*

Optimality Analysis for the General Broadcast

- Challenges:
  - POMDPs are traditionally known to be analytically intractable and computationally expensive - the 'curse of dimensionality'
  - Various 'one-size-fits-all' exact/approximate numerical solutions are available - they do not usually provide insights into the problem at hand
- Our approach:
  - Study optimal control policy in two-user broadcast
  - Strongly founded on this study, derive a near-optimal, low-complexity threshold control policy for the general broadcast

Outline

- Background and Motivation
- Problem Setup
- Optimal Control Policy - Special Cases
- Thresholdability Analysis and Threshold Control Policy
- Numerical Results
- Conclusions

Two-User Broadcast – Thresholdability Properties

- Classify the broadcast into two types: Type I, if it is optimal to *transmit* at steady state belief vector and Type II, otherwise.
- *Proposition*: The optimal controller is thresholdable in specific regions of the state space, the regions and threshold boundaries being dependent on the broadcast type

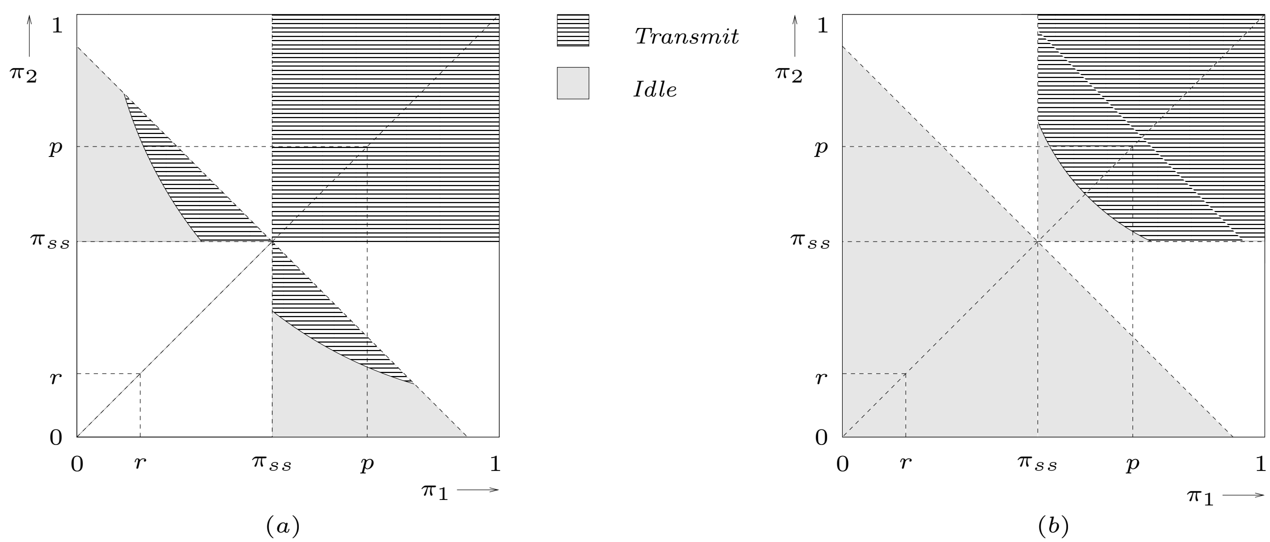


Illustration of the threshold boundaries in restricted regions of the two-dimensional state space when the broadcast is (a) Type I, (b) Type II.

Thresholdability Properties (continued)

- The threshold boundaries are convex hyperbolic segments and are explicitly characterized as
  - Type I:

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

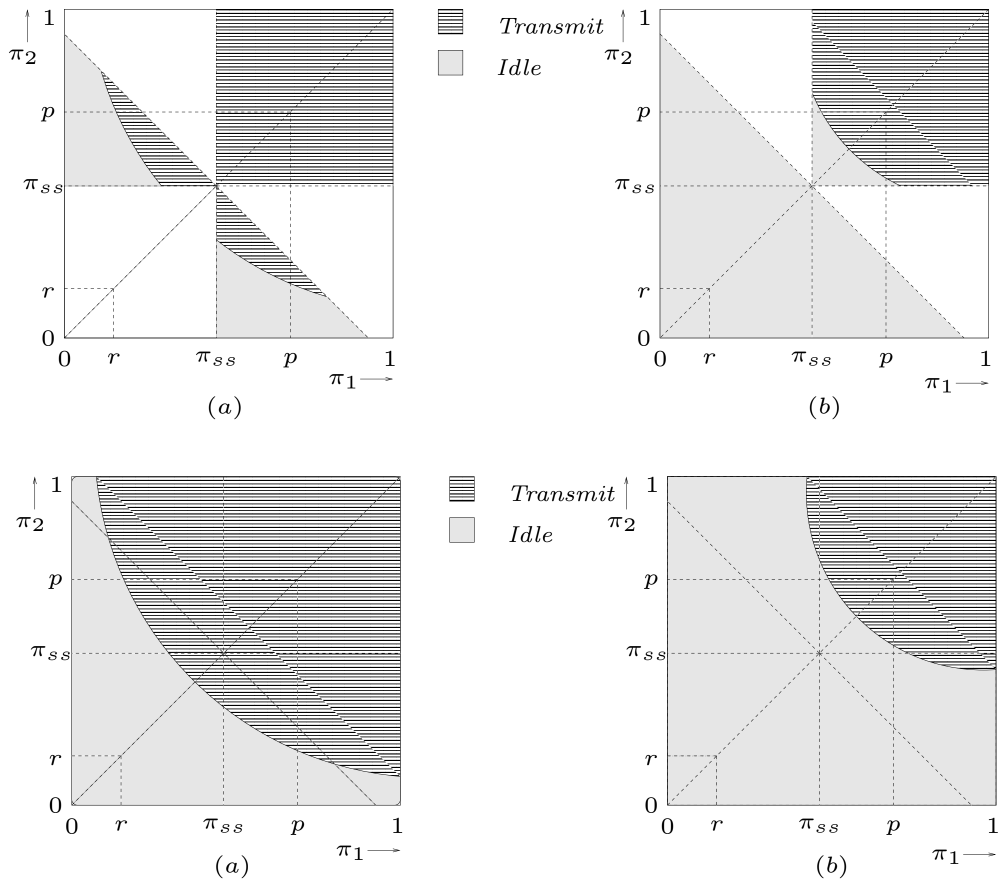
  - Type II:

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

where

$$\begin{aligned} V^a(x_1, x_2) \quad &= \quad x_1 + x_2 + \beta\big[(1-x_1)(1-x_2)V(r,r) + (1-x_1)(x_2)V(r,p) \\ &\quad +x_1(1-x_2)V(p,r) + x_1 x_2 V(p,p) \end{aligned}$$

Thresholdability Properties - Extrapolation



Extrapolation of the threshold boundaries to the entire two dimensional state space when the broadcast is (a) Type I, (b) Type II.

Threshold Control Policy for the N-User Broadcast - Outline

- Initialization: Evaluate the system level quantities $V(\Pi_0), \ldots, V(\Pi_{2^N-1})$ and identify the system type
- Threshold decision: Evaluate k^* by solving a N-dimensional polynomial in k, as below:

$$k^* = \arg\max_{k \in R} \begin{cases} V^a(k\pi) = W + \beta V^a(T(k\pi)), & \text{if Type I} \\ V^a(k\pi) = \frac{W}{1-\beta}, & \text{if Type II} \end{cases}$$

with

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

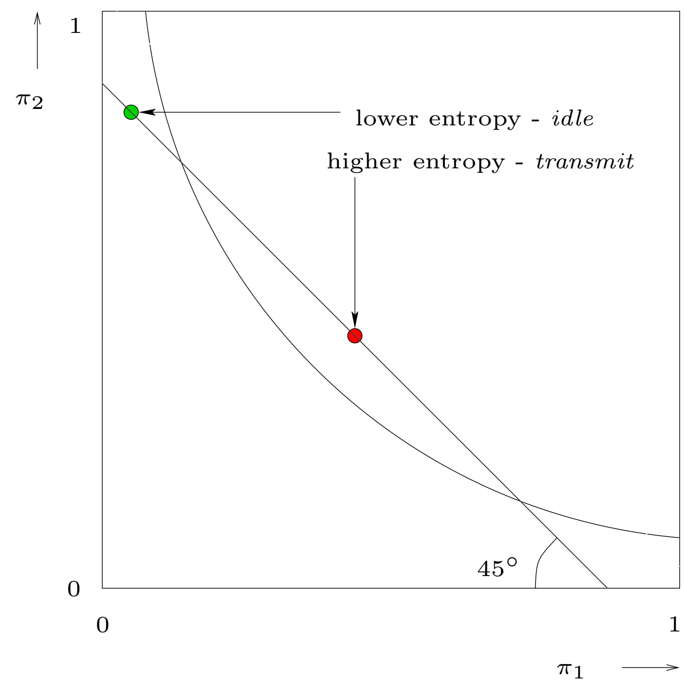
The threshold policy is given by

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

- State evolution: $\pi_i \leftarrow p$, if feedback $= 1$; $\pi_i \leftarrow r$, if feedback $= 0$; $\pi_i \leftarrow T(\pi)$, if *idle* decision was made

Threshold Control Policy - Entropy Argument

- Convexity of the threshold boundary renders optimality properties to the threshold control policy
- For the same immediate reward, more entropy ⇒ more advantageous to *transmit* and hence explore



Outline

- Background and Motivation
- Problem Setup
- Optimal Control Policy - Special Cases
- Thresholdability Analysis and Threshold Control Policy
- <u>Numerical Results</u>
- Conclusions

Numerical Results

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 %

Illustration of the near-optimal performance of the proposed threshold policy. %**opt**:$= \frac{V_{\text{policy}}}{V} \times 100\%$. Each row corresponds to a fixed set of randomly generated system parameters and initial belief values. Number of broadcast users = 4.

Numerical Results

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 %

Illustration of the gain associated with 1-bit feedback. %**fbgain:**$= \frac{V_{\text{policy}} - V_{\text{nofb}}}{V_{\text{genie}} - V_{\text{nofb}}} \times 100\%$.

Conclusions

- The 'exploitation vs exploration' trade-off vastly simplifies for special cases of broadcast parameters, with 'greedy type' policies turning out to be optimal
- For the general broadcast, the trade-off is not as simple! The threshold policy, derived in an optimality framework for the two-user broadcast, has near-optimal numerical performance. It is also computationally inexpensive to implement with complexity being polynomial in the number of broadcast users
- Significant system level gains are associated with exploiting channel memory even with minimal feedback (delayed feedback and obtained *only* during *transmit* slots). Also observed in other network settings:
  - Opportunistic scheduling using delayed ARQ in Markov-modeled downlink
  - Cooperative scheduling using ARQ in multi-cellular downlink