Opportunistic Spectrum Access with Multiple Users: Learning under Competition

Anima Anandkumar[1] **Nithin Michael**[2] **Ao Tang**[2]

[1]EECS, Massachusetts Institute of Technology, Cambridge, MA. USA

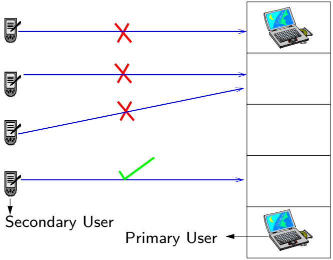
[2]ECE, Cornell University, Ithaca, NY. USA

IEEE INFOCOM 2010

Introduction: Cognitive Radio Network

Two types of users

- Primary Users
  - Priority for channel access
- Secondary or Cognitive Users
  - Opportunistic access
  - Channel sensing abilities

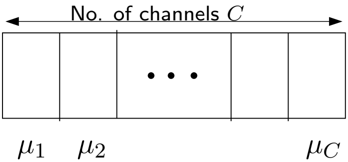


Limitations of secondary users

- Sensing constraints: Sense only part of spectrum at any time
- Lack of coordination: Collisions among secondary users
- Unknown behavior of primary users: Lost opportunities

Maximize total secondary throughput subject to above constraints

Distributed Learning and Access

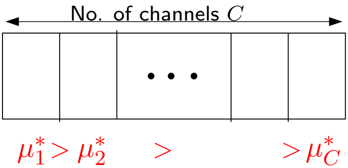


- Slotted tx. with U cognitive users and $C > U$ channels
- Channel Availability for Cognitive Users: Mean availability μ_i for channel i and $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_C]$.
- $\boldsymbol{\mu}$ unknown to secondary users: learning through sensing samples
- No explicit communication/cooperation among cognitive users

Objectives for secondary users

- Users ultimately access orthogonal channels with best availabilities $\boldsymbol{\mu}$
- Max. Total Cognitive System Throughput $\equiv$ Min. Regret

Distributed Learning and Access

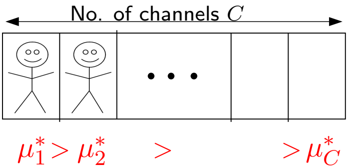


- Slotted tx. with U cognitive users and $C > U$ channels
- Channel Availability for Cognitive Users: Mean availability μ_i for channel i and $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_C]$.
- $\boldsymbol{\mu}$ unknown to secondary users: learning through sensing samples
- No explicit communication/cooperation among cognitive users

Objectives for secondary users

- Users ultimately access orthogonal channels with best availabilities $\boldsymbol{\mu}$
- Max. Total Cognitive System Throughput $\equiv$ Min. Regret

Distributed Learning and Access



- Slotted tx. with U cognitive users and $C > U$ channels
- Channel Availability for Cognitive Users: Mean availability μ_i for channel i and $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_C]$.
- $\boldsymbol{\mu}$ unknown to secondary users: learning through sensing samples
- No explicit communication/cooperation among cognitive users

Objectives for secondary users

- Users ultimately access orthogonal channels with best availabilities $\boldsymbol{\mu}$
- Max. Total Cognitive System Throughput $\equiv$ Min. Regret

Summary of Results

- Propose two distributed learning+access policies: ρ^{PRE} and ρ^{RAND}
  - ρ^{PRE}: under pre-allocated ranks among cognitive users
  - ρ^{RAND}: fully distributed and no prior information
- Provable guarantees on sum regret under two policies
  - Convergence to optimal configuration
  - Regret grows slowly in no. of access slots $R(n) \sim O(\log n)$
- Lower bound for any uniformly-good policy: also logarithmic in no. of access slots $R(n) \sim \Omega(\log n)$

We propose order-optimal distributed learning and allocation policies

Related Work

Multi-armed Bandits

- Single cognitive user (Lai & Robbins 85)
- Multiple users with centralized allocation (Ananthram et. al 87)
  Key Result: Regret $R(n) \sim O(\log n)$ and optimal as $n \to \infty$
- Auer et. al. 02: order optimality for sample mean policies

Cognitive Medium Access & Learning

- Liu et. al. 08: Explicit communication among users
- Li 08: Q-learning, Sensing all channels simultaneously
- Liu & Zhao 10: Learning under time division access
- Gai et. al. 10: Combinatorial bandits, centralized learning

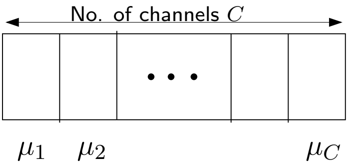
Outline

System Model

Primary and Cognitive Networks

- Slotted tx. with U cognitive users and C channels
- Primary Users: IID tx. in each slot and channel

  Channel Availability for Cognitive Users: In each slot, IID with prob. μ_i for channel i and $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_C]$.
- Perfect Sensing: Primary user always detected
- Collision Channel: tx. successful only if sole user
- Equal rate among secondary users:
  Throughput $\equiv$ total no. of successful tx.



Problem Formulation

Distributed Learning Through Sensing Samples

- No information exchange/coordination among secondary users
- All secondary users employ same policy

Throughput under perfect knowledge of $\boldsymbol{\mu}$ and coordination

$$S^*(n; \boldsymbol{\mu}, U) := n \sum_{j=1}^{U} \mu(j^*)$$

where j^* is j^{th} largest entry in $\boldsymbol{\mu}$ and n: no. of access slots

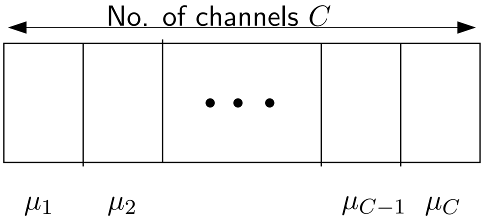
Regret under learning and distributed access policy ρ

Loss in throughput due to learning and collisions

$$R(n; \boldsymbol{\mu}, U, \rho) := S^*(n; \boldsymbol{\mu}, U) - S(n; \boldsymbol{\mu}, U, \rho)$$

Max. Throughput $\equiv$ Min. Sum Regret

Single Cognitive User: Multi-armed Bandit

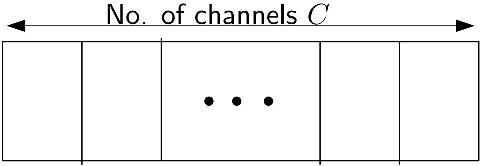


Exploration vs. Exploitation Tradeoff

- Exploration: channels with good availability are not missed
- Exploitation: obtain good throughput

Explore in the beginning and exploit in the long run

Single Cognitive User: Multi-armed Bandit



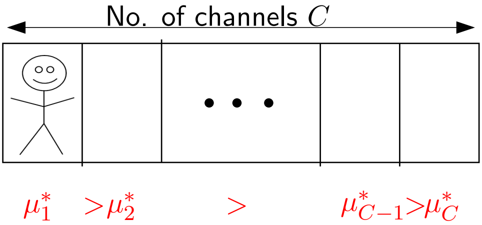
$\mu_1^* > \mu_2^* > \cdots > \mu_{C-1}^* > \mu_C^*$

Exploration vs. Exploitation Tradeoff

- Exploration: channels with good availability are not missed
- Exploitation: obtain good throughput

Explore in the beginning and exploit in the long run

Single Cognitive User: Multi-armed Bandit



Exploration vs. Exploitation Tradeoff

- Exploration: channels with good availability are not missed
- Exploitation: obtain good throughput

Explore in the beginning and exploit in the long run

Single Cognitive User: Multi-armed Bandit (Contd.)

- $T_{i,j}(n)$: no. of slots where user j selects channel i
- $\overline{X}_{i,j}(T_{i,j}(n))$: sample mean availability of channel i acc. to user j

Two Policies based on Sample Mean (Auer et. al. 02)

- Deterministic Policy: Select channel with highest g-statistic:

$$g_j(i;n) := \overline{X}_{i,j}(T_{i,j}(n)) + \sqrt{\frac{2\log n}{T_{i,j}(n)}}$$

- Randomized Greedy Policy: Select channel with highest $\overline{X}_{i,j}(T_{i,j}(n))$ with prob. $1-\epsilon_n$ and with prob. ϵ_n unif. select other channels, where

$$\epsilon_n := \min[\frac{\beta}{n}, 1]$$

Regret under the two policies is $O(\log n)$ for n no. of access slots

Outline

1. Introduction
2. System Model & Recap of Bandit Results
3. Proposed Algorithms & Lower Bound
4. Simulation Results
5. Conclusion

Overview of Two Proposed Algorithms

ρ^{PRE} Pre-allocation Policy: ranks are pre-assigned

If user j is assigned rank w_j, select channel with w_j^{th} highest $\overline{X}_{i,j}(T_{i,j}(n))$ with prob. $1 - \epsilon_n$ and with prob. ϵ_n unif. select other channels, where $\epsilon_n := \min[\frac{\beta}{n}, 1]$

ρ^{RAND} Random allocation Policy: no prior information

User adaptively chooses rank w_j based on feedback for successful tx.

- If collision in previous slot, draw a new w_j uniformly from 1 to U
- If no collision, retain the current w_j

Select channel with w_j^{th} highest entry:

$$g_j(i;n) := \overline{X}_{i,j}(T_{i,j}(n)) + \sqrt{\frac{2 \log n}{T_{i,j}(n)}}$$

Learning Under Pre-Allocation

If user j is assigned rank w_j, select channel with w_j^{th} highest $\overline{X}_{i,j}(T_{i,j}(n))$ with prob. $1-\epsilon_n$ and with prob. ϵ_n unif. select other channels, where

$$\epsilon_n := \min[\frac{\beta}{n}, 1]$$

Regret: user does not select channel of pre-assigned rank

$$\mathbb{E}[T_{i,j}(n)] \leq \sum_{t=1}^{n-1} \frac{\epsilon_{t+1}}{C} + \sum_{t=1}^{n-1}(1-\epsilon_{t+1})\mathbb{P}[\mathcal{E}_{i,j}(n)], \;\; i \neq w_j^*,$$

where $\mathcal{E}_{i,j}(n)$ is the error event that w_j^{th} highest entry of $\bar{X}_{i,j}(T_{i,j}(n))$ is not same as $\mu_{w_j}^*$

Regret Under Pre-allocation

Theorem (Regret Under ρ^{PRE} Policy)

No. of slots user j accesses channel $i \neq w_j^*$ other than pre-allocated channel under ρ^{PRE} satisfies

$$\mathbb{E}[T_{i,j}(n)] \leq \frac{\beta}{C} \log n + \delta, \quad \forall i = 1, \ldots, C, i \neq w_j^*,$$

when

$$\beta > \max[20, \frac{4}{\Delta_{\min}^2}],$$

where $\Delta_{\min} := \min_{i,j} |\mu_i - \mu_j|$ *is minimum separation.*

Logarithmic regret under ρ^{PRE}

Distributed Learning and Randomized Allocation ρ^{RAND}

User adaptively chooses rank w_j based on feedback for successful tx.

- If collision in previous slot, draw a new w_j uniformly from 1 to U
- If no collision, retain the current w_j

Select channel with w_j^{th} highest entry:

$$g_j(i;n) := \overline{X}_{i,j}(T_{i,j}(n)) + \sqrt{\frac{2\log n}{T_{i,j}(n)}}$$

Upper Bound on Regret

$$R(n) \leq \frac{1}{U}\sum_{k=1}^{U}\mu(k^*)\left[\sum_{j=1}^{U}\sum_{i\in U\text{-worst}}\mathbb{E}[T_{i,j}(n)+M(n)]\right]$$

- U-best: top U channels. U-worst: remaining channels
- $\sum_{i\in U\text{-worst}} T_{i,j}(n)$: Time spent in U-worst channels by user j
- $M(n)$: No. of collisions in U-best channels

Distributed Learning and Randomized Allocation ρ^{RAND}

Theorem

Under $\rho^{\textit{RAND}}$ Policy, $\mathbb{E}[\sum_{i \in U\text{-worst}} T_{i,j}(n)]$ and $\mathbb{E}[M(n)]$ are $O(\log n)$ and hence, regret is $O(\log n)$ where n is the number of access slots.

Proof for $\mathbb{E}[M(n)]$: no. of collisions in U-best channels

- Bound $\mathbb{E}[M(n)]$ under perfect knowledge of $\boldsymbol{\mu}$ as $\Pi(U)$
- Good state: all users estimate order of top-U channels correctly
- Transition from bad to good state: $\Pi(U)$ avg. no. of collisions
- Bound on no. of slots spent in bad state

Lower Bound on Regret

Uniformly good policy ρ

A policy which enables users to ultimately settle down in orthogonal best channels under any channel availabilities $\boldsymbol{\mu}$: user j spends most of time in $i \in U$-best channel

$$\mathbb{E}_{\boldsymbol{\mu}}[n - T_{i,j}(n)] = o(n^{\alpha}), \quad \forall \alpha > 0, \boldsymbol{\mu} \in (0,1)^C.$$

Satisfied by ρ^{PRE} and ρ^{RAND} policies

Theorem (Lower Bound for Uniformly Good Policy)

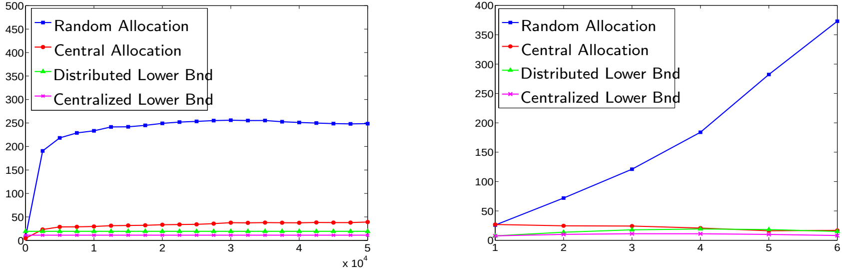
The sum regret satisfies

$$\liminf_{n\to\infty} \frac{R(n; \boldsymbol{\mu}, U, \rho)}{\log n} \geq \sum_{i \in U\text{-worst}} \sum_{j=1}^{U} \frac{\Delta(U^*, i)}{D(\mu_i, \mu_{j^*})}.$$

Order optimal regret under ρ^{PRE} and ρ^{RAND} policies

Outline

Simulation Results



Normalized regret $\frac{R(n)}{\log n}$ vs. n slots.

$U = 4$ users, $C = 9$ channels.

Normalized regret $\frac{R(n)}{\log n}$ vs. U users.

$C = 9$ channels, $n = 2500$ slots.

Probability of Availability $\boldsymbol{\mu} = [0.1, 0.2, \ldots, 0.9]$.

Outline

Conclusion

Summary

- Considered maximizing total throughput of cognitive users under unknown channel availabilities and no coordination
- Proposed two algorithms which achieve order optimality

  ρ^{PRE} policy works under pre-allocated ranks

  ρ^{RAND} policy does not require prior information

Outlook

- Imperfect sensing: logarithmic regret still achievable
- No. of cognitive users unknown to the policy: logarithmic regret still achievable
- Cognitive users with different rates and objectives