Distributed Exploration in Multi-Armed Bandits

Eshcar Hillel
Zohar Karnin
Yahoo Labs

Tomer Koren
Technion

Ronny Lempel
Oren Somekh
Yahoo Labs

Technion
Israel Institute of Technology

YAHOO! LABS
Distributed MAB setup

- **Setup**: $n$ bandit arms, stochastic rewards
- **Goal**: identify best arm (with highest reward)
- learning time $\Leftrightarrow$ no. of arm pulls
- **Question**: can we speedup by distributing to $k$ players?
  - ...without communicating too much between them
- E.g. by allowing **single** transmission per player?
**Intuition (1 transmission)**

**Hard instance**

- Even if each player explores few arms, problem might be hard
- Naïve solutions fail to provide any speedup

**But:**
- By dealing arms at random, some players get easy problems
- We can identify best arm from their outputs
Our results (for $k$ players)

Main result: by communicating only once:

- **Algorithm**: they can achieve $\sqrt{k}$ parallel speedup!
- **Tightness**: cannot do better than $\sqrt{k}$ in general

Also: by communicating only $O(\log(1/\varepsilon))$ times:

- $k$ players can find $\varepsilon$-best arm
- achieve optimal $\Omega(k)$ parallel speedup

More details at our poster: **Sun15**