Communication-Efficient Collaborative Heterogeneous Bandits in Networks

Junghyun Lee, Laura Schmid, Se-Young Yun
KAIST

Preprint on arXiv:

KAIST AI
Kim Jaechul Graduate School

OSI Optimization and Statistical Inference LAB

informs APPLIED PROBABILITY

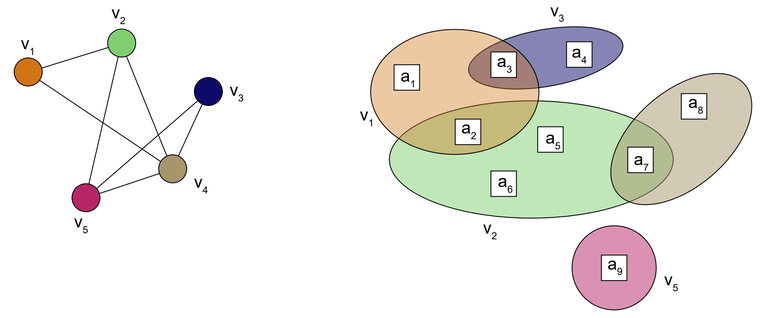
Introduction

- **Collaborative multi-agent setting for multi-armed bandits (MAB)**: large-scale decentralized decision-making (e.g., wireless networks)
- **Our setting**: Agents are **heterogeneous** in their arm sets, communicate over a **network**
- **Goal**: Minimize group regret while keeping communication complexity (CC) low
- **Our solution:** Each agent uses the UCB algorithm together with our novel **Flooding with Absorption (FWA)** protocol
- We derive rigorous regret upper bounds to compare FWA with classic Flooding, and perform extensive experiments on synthetic data

Sources of heterogeneity in multi-agent MAB

Arbitrary network topology and arm distribution

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ $\mathcal{H} = (\mathcal{K}, \mathcal{F})$



Algorithmic Solution

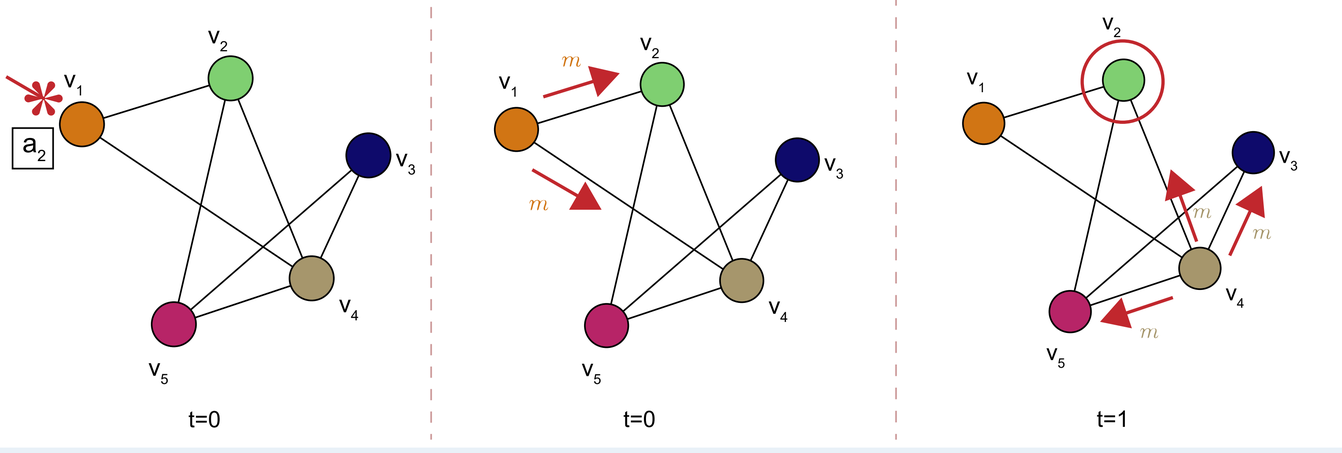
Step 1. Agent pulls one of their arms with highest UCB.

Step 2. Agent creates and sends message containing arm index *a* and received reward to all neighbors

Step 3. Neighbors with arm *a* absorb the message, others forward it unless time-to-live (TTL) expires

Prevent routing loops: hash-based sequence number controlled flooding

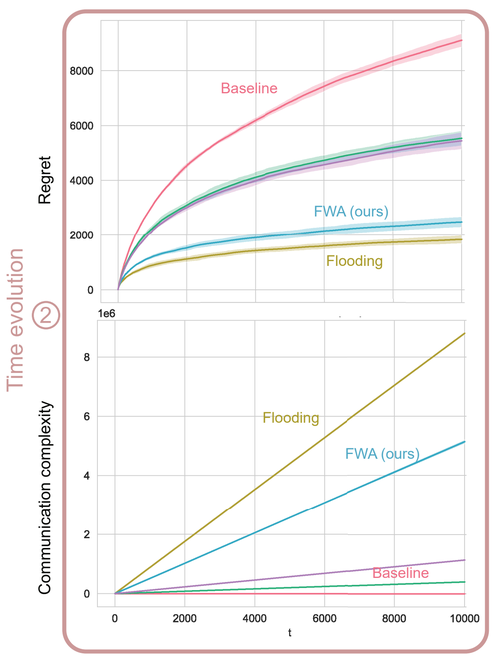
No knowledge of the network topology required!

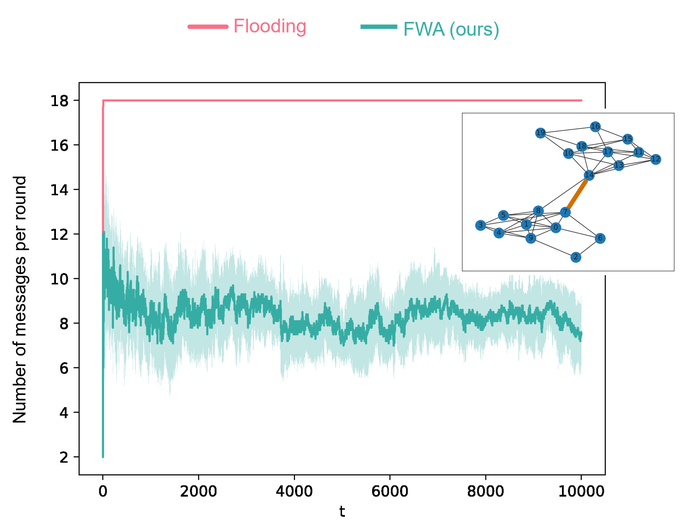


Results

Experimental evaluation:

Balancing regret vs. CC for various network topologies (here: Binary SBM)





FWA: almost optimal regret, efficient communication

Avoids link congestion: maintains small # of messages

Theoretical analysis of regret upper bounds

Flooding

$$\limsup_{T\to\infty} \frac{\mathbb{E}[R(T)]}{\log T} \le \sum_{\substack{a\in\mathcal{K}\\ \tilde{\Delta}_a>0}} \frac{8\alpha\theta([\mathcal{G}^\gamma]_{-a})}{\tilde{\Delta}_a}$$

FWA

$$\limsup_{T\to\infty} \frac{\mathbb{E}[R(T)]}{\log T} \le \sum_{\substack{a\in\mathcal{K}\\ \tilde{\Delta}_a>0}} \frac{8\alpha\theta\left([\mathcal{G}^\gamma_{(a,c)}]_{-a}\right)}{\tilde{\Delta}_a}$$

We can quantify the regret gap between Flooding and FWA

Discussion

- **Advantages of FWA**
  - Network agnostic, works in complex topologies
  - Less link congestion at minimal cost
- **Limitations of FWA**
  - Performance depends on topology & arm distribution
- **Future work**
  - Adaptive TTL → improve CC?
  - FWA in dynamic networks/nonstationary bandits

Conclusion

- Novel setting for distributed multi-armed bandits: communication on graph, differing arm sets
- To deal with large CC: new communication protocol FWA
- Derive regret upper and lower bounds for UCB algorithm with Flooding and FWA protocols
- Extensive experimental results: much improved CC at only little performance loss!