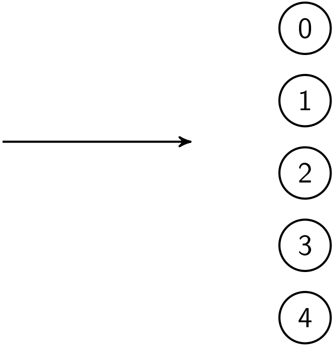
The Power of Two in Large Service-Marketplaces

Dheeraj Narasimha, **Srinivas Nomula**, Srinivas Shakkottai, Parimal Parag

Electrical Communication Engineering,
Indian Institute of Science

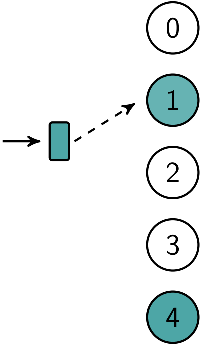
Problem setup



Objective: Maximize the revenue of all agents

- **Pricing:** How to price the services?
- **Routing:** How to route the tasks?

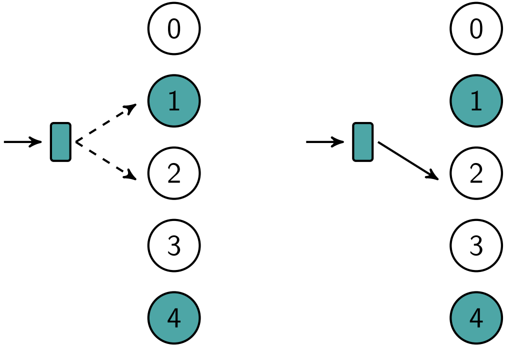
Decentralized pricing



- Each server does not cooperate with one another and picks its own price
- Task assigned only when price less than value

Power-of-two routing R_2 for decentralized pricing

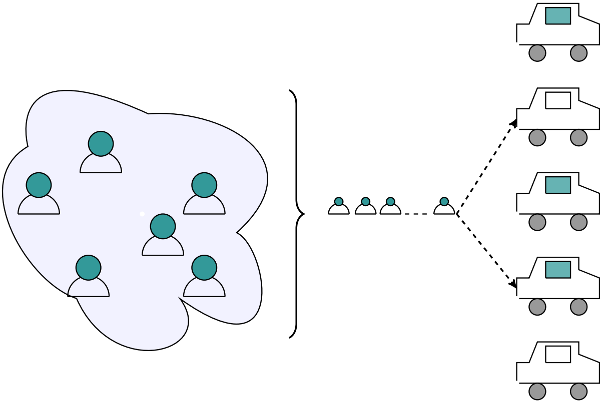
Sampling and routing



Join one of two randomly selected servers with minimum price

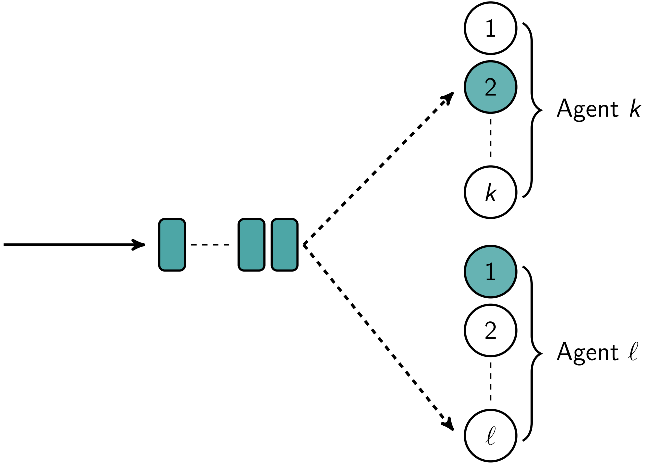
- Requires server state feedback from two servers at each arrival
- Loss when both busy servers are selected, hence no revenue

Ride sharing and on demand services



- Apps like Namma Yatri, Janata Ride in India, allows auto driver listing with independent pricing
- The two-server selection principle is similar to *sampling any two drivers for a ride* based on price and availability

Online Cloud Marketplaces



- Google cloud and AWS marketplace allow independent cloud service providers to list their services
- Multiple providers compete for customer jobs, similar to the two-server price competition model

State-of-the-art

Revenue maximizing dynamic pricing

- For a single server queue
  - Random valuation: [Naor, 1969] [1], [Borgs et al, 2011] [2]
  - Arbitrary valuation: [Ashok et al, 2023], [3]
- Multiple servers with no queues and random valuation
  - Centralized routing and pricing:[Ashok et al, 2022] [4]
  - Our work: power-of-2 routing and rational pricing

[1] P. Naor, "The regulation of queue size by levying tolls," Econometrica, vol. 37, no. 1, pp. 15–24, Jan. 1969.

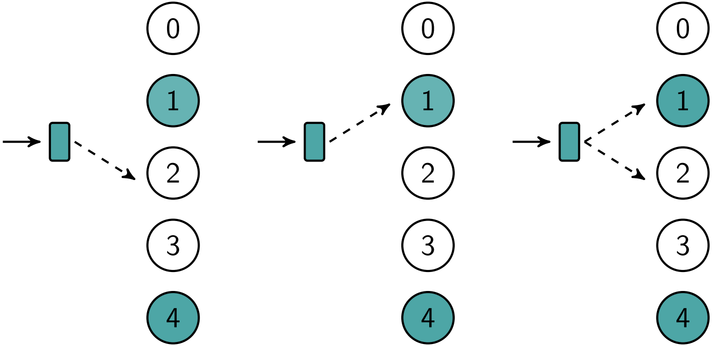
[2] C. Borgs et al, "The optimal admission threshold in observable queues with state dependent pricing," Probability in the Engineering and Informational Sciences, vol. 28, no. 1, p. 101–119, 2014.

[3] Ashok et al., "Optimal pricing in a single server system," ACM Trans. Model. Perform. Eval. Comput. Syst., vol. 8, no. 4, pp. 1–32, Dec. 2023.

[4] Ashok et al, "Optimal pricing in multi server systems," Performance Evaluation, vol. 154, p. 102282, 2022.

State-of-the-art techniques

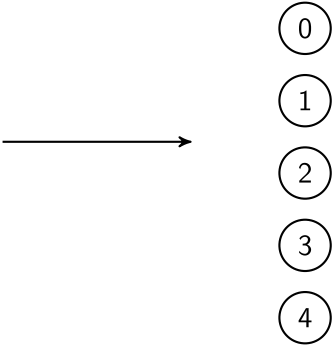
Deterministic/random routing and centralized pricing



(a) $\mathrm{D_1C}$ (b) $\mathrm{R_1C}$ (c) $\mathrm{R_2C}$

- Price rates for all the servers are decided centrally
- Random routing is agnostic to the server states

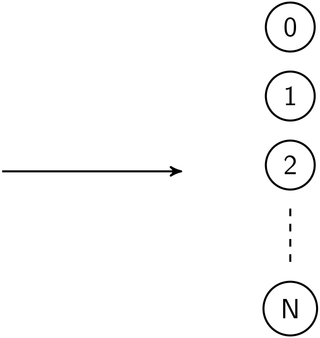
System model



N-server loss system

- Poisson arrivals of rate $N\lambda$
- Random *i.i.d.* exponential valuation and pricing
- Random *i.i.d.* unit mean exponential service times
- Server n can either be idle or busy, denoted by $X_n(t)$

Mean field game approach

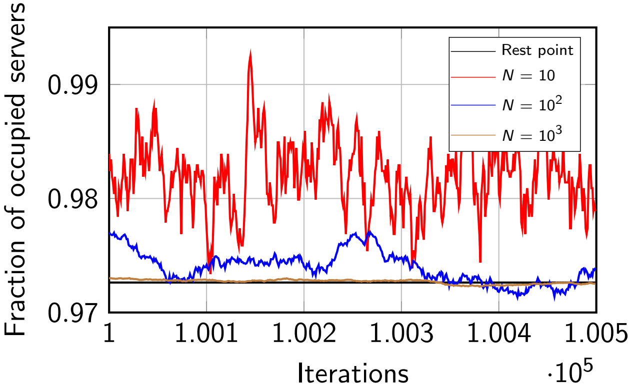


Assumption

- Choosing the price rate based on the prices of all other servers and busyness is computationally expensive
- The servers interact with the average behaviour of the other servers. Let $Z_t^N \triangleq \frac{1}{N}\sum_{n=1}^{N} X_n(t)$

Theorem 1 (Mean-field convergence)

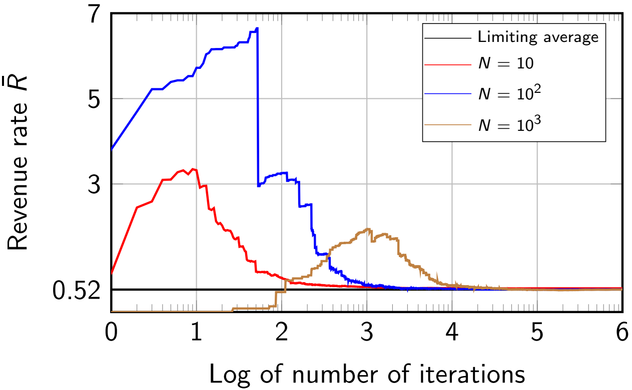
$$\lim_{N\to\infty} \lim_{t\to\infty} \mathbb{E} \left| Z_t^N - z^* \right|^2 = O\left(\frac{1}{N}\right)$$



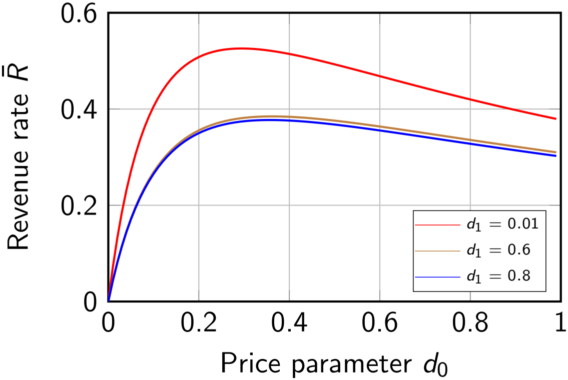
Tagged server revenue

Theorem 3 (Limiting revenue rate at tagged server 0)

$\bar{R} \triangleq \lim_{N\to\infty} \lim_{K\to\infty} \frac{1}{A_K} \sum_{k=0}^{K} P_{k,0} \xi_k^N = \frac{z^* q_1^2 + (1-z^*) q_{20}^2}{d_0(\frac{1}{2\lambda} + z^* q_1 + (1-z^*) q_{20})}$

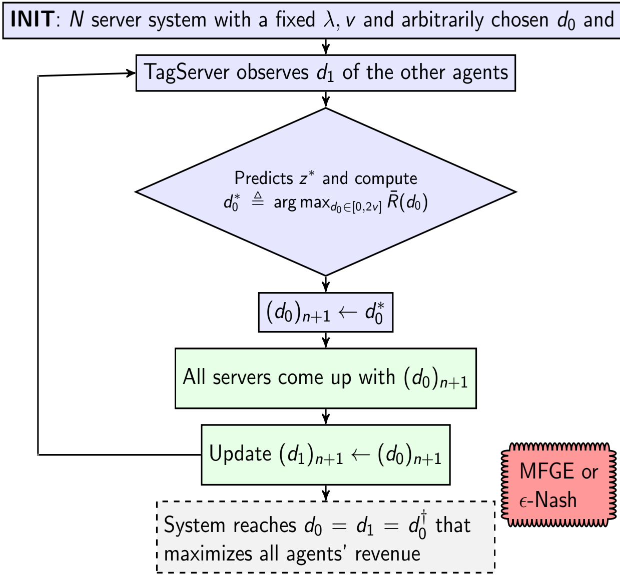


Unimodality



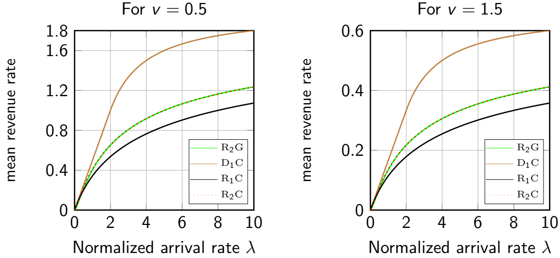
- For any chosen v and λ, there exist a unique $d_0^* \in [0, 2v]$ that maximizes $\bar{R}$

Our approach ($\mathrm{R_2G}$)

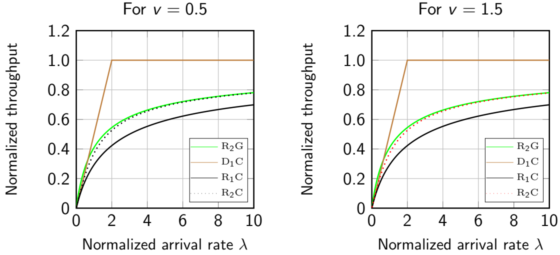


Performance comparison

Mean revenue rate



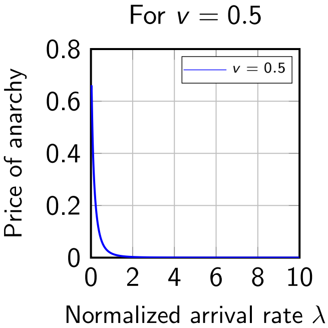
Throughput

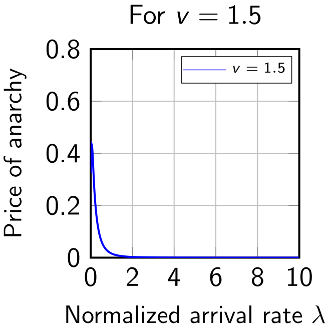


Numerical results

Price of anarchy

$$\text{PoA} \triangleq 1 - \frac{\bar{R}_{\mathrm{R_2G}}}{\bar{R}_{\mathrm{R_2C}}}$$





Conclusion

- Proved that the mean occupancy rate Z_t^N converges to the stationary point z^* and the simulated revenue rate converges to the calculated one
- Revenue rate is uni-modal over the tag price rate d_0 and proposed an algorithm which strategically chooses the price rates which maximizes the tagged server revenue rate
- Designed a decentralized algorithm that plays a game between the agents to choose the optimum price parameter which maximizes their revenue
- Calculated the revenue rates, blocking probability and the mean tag price rate for the state-of-art techniques numerically and compared the results