Near-Optimal Regret-Queue Length Tradeoff in Online Learning for Two-Sided Markets

Zixian Yang, University of Michigan, Ann Arbor
zixian@umich.edu
Joint work with Sushil Varma (Michigan) and Lei Ying (Michigan)

Two-Sided Markets

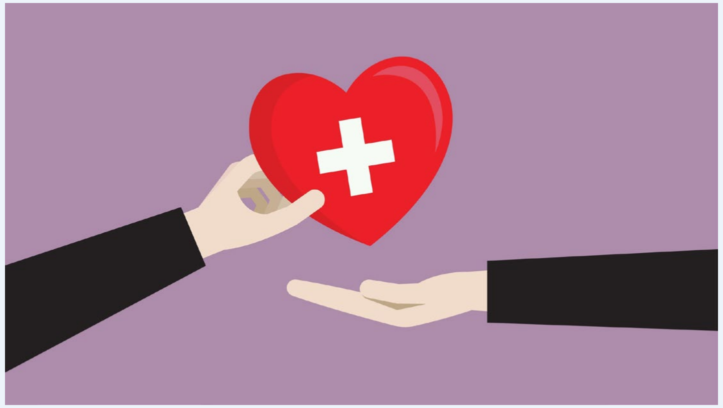
Uber

lyft

Ride-Hailing

instacart DOORDASH

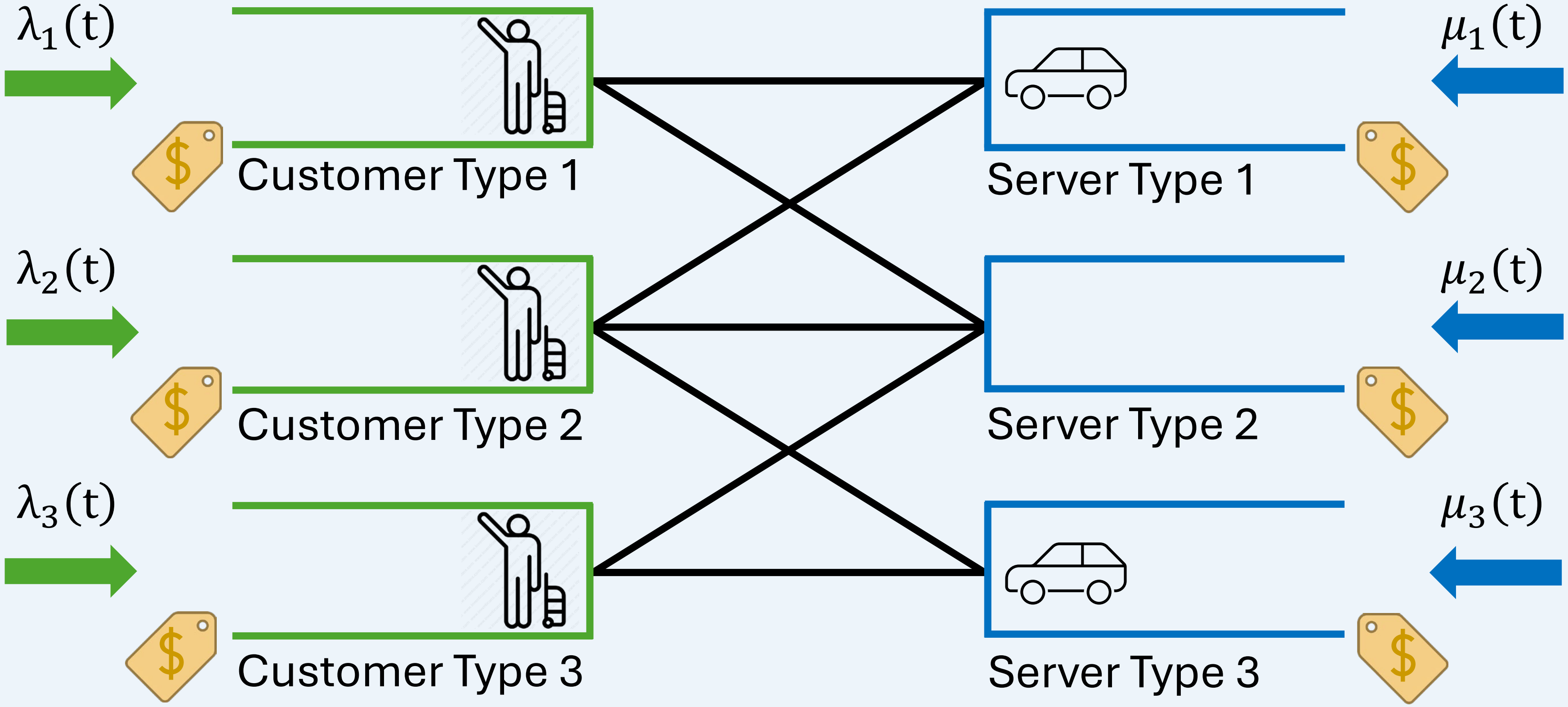
Meal/Grocery Delivery

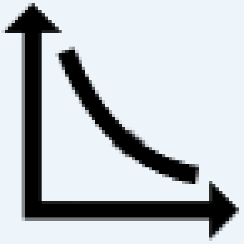


Organ Donation

Crowdsourcing

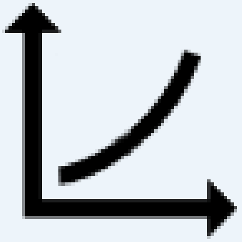
Model: Two-Sided Queues





Pricing Policy

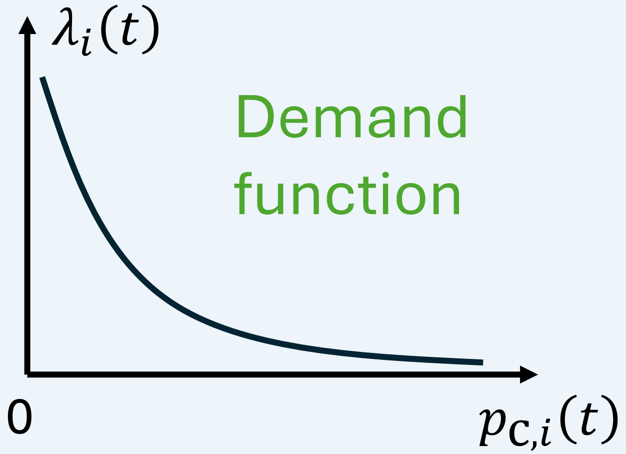
Matching Policy

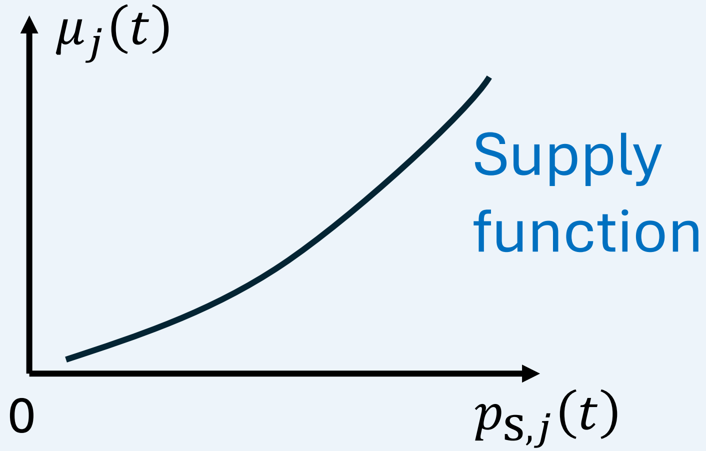


Objectives

- Maximize $E\left[\sum_{t=1}^{T} \text{Profit}(t)\right]$
- $\text{E}[\text{Profit}(t)] = \text{E}[\text{Revenue}(t)] - \text{E}[\text{Cost}(t)]$

$$\text{E}[\text{Revenue}(t)] = \sum_i \lambda_i(t) \times p_{\text{C},i}(t) \qquad \text{E}[\text{Cost}(t)] = \sum_j \mu_j(t) \times p_{\text{S},j}(t)$$





Monotone assumption

- Minimize $\text{AvgQLen}(T) = \frac{1}{T}\sum_{t=1}^{T} E[\text{total}-\text{queue}-\text{length}(t)]$

Objectives

Question:
If the demand and supply functions are **unknown**, how should the platform do pricing and matching?

Fluid Baseline and Regret

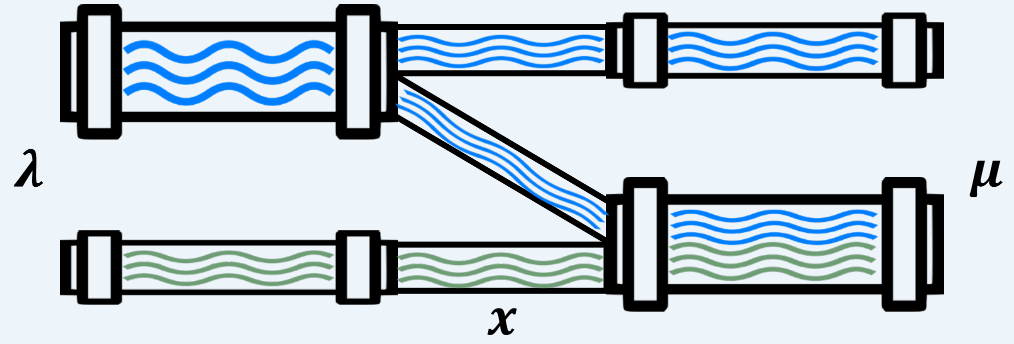
- Fluid baseline [Varma et al., 2023]:

$$\max_{\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{x}} E[\text{Profit}(\boldsymbol{\lambda},\boldsymbol{\mu})]$$

A concave function of $(\boldsymbol{\lambda},\boldsymbol{\mu})$

$$\text{s.t. } \lambda_i = \sum_j x_{i,j}, \quad \mu_j = \sum_i x_{i,j},$$

$$x_{i,j} \geq 0, \lambda_i, \mu_j \in [0,1], \text{for all } i,j$$



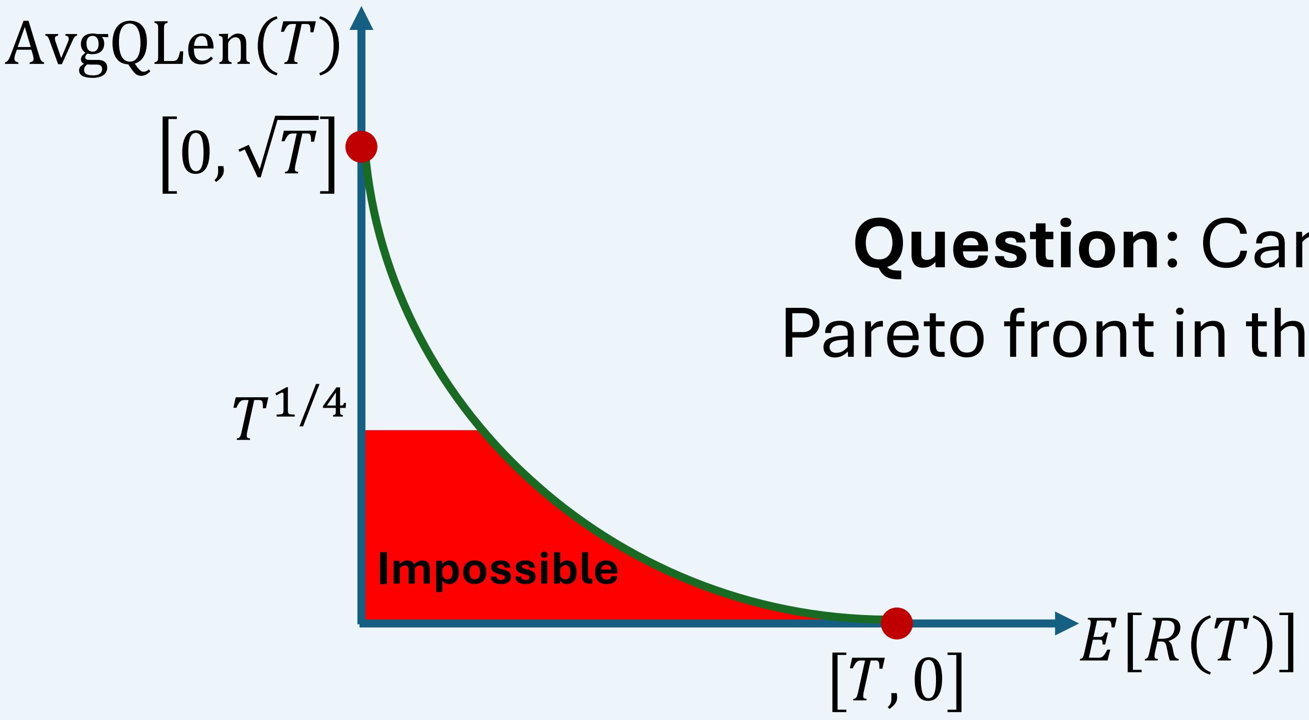
Balance equation

- Regret: $E[R(T)] = T \times \text{Profit}^* - E[\sum_{t=1}^{T} \text{Profit}(t)]$
- Minimize $E[R(T)]$

  Minimize $\text{AvgQLen}(T) = \frac{1}{T}\sum_{t=1}^{T} E[\text{total}-\text{queue}-\text{length}(t)]$

Fundamental Regret-Queue Length Tradeoff

- A lower bound result [**Y** et al., 2025]: there exists a problem instance such that, for any policy in a large class of pricing policies, if $\mathrm{AvgQLen}(T) \le T^{\gamma/2}$, $E[R(T)] = \Omega(T^{1-\gamma})$ for any $\gamma \in \left[0, \frac{1}{2}\right]$.



Question: Can we achieve the Pareto front in the learning setting?

Results [**Y** et al., 2025]

- For any $\gamma \in \left[0, \frac{1}{6}\right]$, we have $E[R(T)] = \tilde{O}(T^{1-\gamma})$, $\mathrm{AvgQLen}(T) = \tilde{O}\left(T^{\gamma/2}\right)$