Data-Driven Robust Taxi Dispatch Approaches

Fei Miao[1], Shuo Han[1], Shan Lin[2], John Stankovic[3], Qian Wang[5], Desheng Zhang[4], Tian He[4], and George J. Pappas[1]

1. ESE, University of Pennsylvania 2. ECE, Stony Brook University 3. CS, University of Virginia 4. CSE, University of Minnesota

5. China Chengxin Credit Information Co.

• Goal: Develop a system level control framework, to incorporate data information with real-time control decisions, balance vacant taxis with minimum total idle driving distance, and consider model uncertainties

Problem: Real-time GPS information provides transportation network knowledge; non-cooperative taxi service, or a greedy algorithm are not efficient

Objectives: System level optimal performance:

--Balanced supply/demand ratio

--Minimal idle cruising distance

--utilize data information with model uncertainties

Our Contributions:

First to design a receding horizon control (RHC) framework for large-scale taxi dispatch

- Both current and anticipated future costs
- Multi-objective: balance supply with minimum idle mileage

Incorporate large-scale sensing data in real-time control

- Predict demand based on realistic dataset; control feedback

Uncertain demand model: robust dispatch formulations

- Data-driven modeling of demand uncertainties
- Theorem: robust optimization form to equivalent convex form

Trace-driven analysis on realistic data sets

- RHC dispatch ***VS*** non-dispatch: the average supply demand ratio error is reduced by 45%, idle distance is reduced by 52%
- Robust ***VS*** non-robust dispatch, with demand uncertainties: the average demand supply ratio error is reduced by 31.7%, total idle driving distance is reduced by 10.13%

Robust Dispatch concave of r^k, convex of X^k

$$\min_{X^k, L^k} \max_{r^k \in \Delta} \quad J = \sum_{k=1}^{\tau} (J_D(X^k) + \beta J_E(X^k, r^k))$$

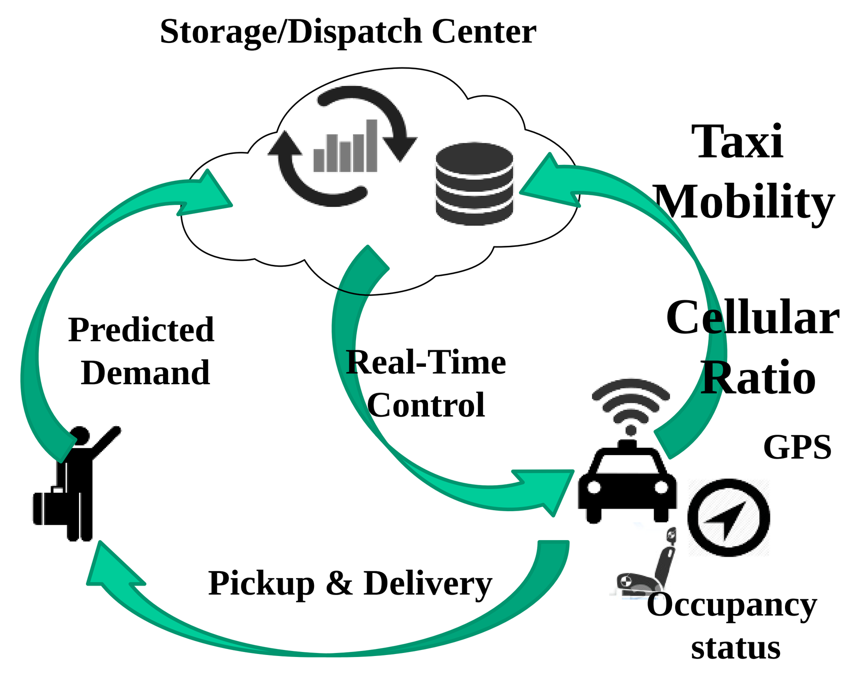
$$= \sum_{k=1}^{\tau} \sum_{i} \left(\sum_{j} X_{ij}^k W_{ij} + \frac{\beta r_i^k}{(\mathbf{1}_n^T X_{\cdot i}^k - X_{i\cdot}^k \mathbf{1}_n + L_i^k)^\alpha} \right)$$

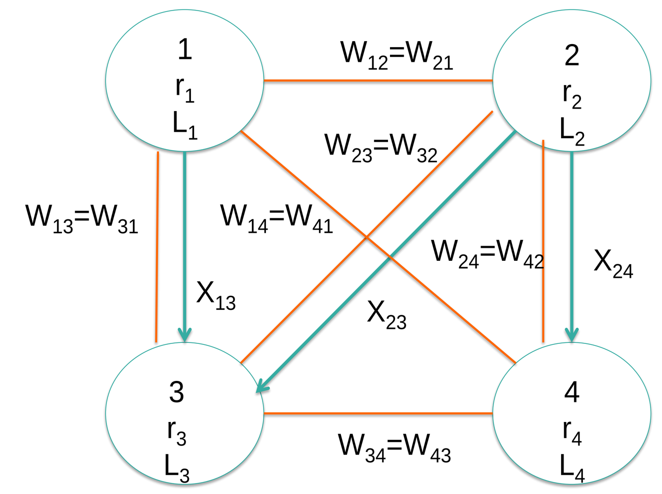
$$\text{s.t.} \quad (L^{k+1})^T = (\mathbf{1}_n^T X^k - X^k \mathbf{1}_n + (L^k)^T) P^k,$$

$$\mathbf{1}_n^T X^k - X^k \mathbf{1}_n + (L^k)^T > 0,$$

$$X_{ij}^k W_{ij} \leq m X_{ij}^k,$$

$$X_{ij}^k \geq 0, \quad i, j \in \{1, 2, \dots, n\}.$$





- **Build the uncertainty set $\mathcal{U}_\epsilon$ given ϵ from the data set ***

(1). The robust constraint is computationally tractable.

(2). The set $\mathcal{U}_\epsilon$ implies a probabilistic guarantee for $\mathbb{P}^*$ at level ϵ

$$\text{If } f(r_c, x) \leq 0, \text{for } \forall r_c \in \mathcal{U}_\epsilon, \text{ then } \mathbb{P}^*(f(r_c, x) \leqslant 0) \geq 1 - \epsilon.$$

* D.Bertsimas, V. Bupta, and N.Kallus, Data-driven robust optimization, submitted to Operations Research, 2015.

- **Theorem 1: Polytope uncertainty equivalent convex form**

$$\Delta := \{A_1 r^1 + \cdots + A_\tau r^\tau \leq b, r^k \geq 0\}$$

$$\min_{X^k, L^k, \lambda \geq 0} \quad \sum_{k=1}^{\tau} (\sum_i \sum_j X_{ij}^k W_{ij}) + b^T \lambda$$

$$\text{subject to} \quad A_k^T \lambda - \beta \begin{bmatrix} \frac{1}{(\mathbf{1}_n^T X_{\cdot 1}^k - X_{1\cdot}^k \mathbf{1}_n + L_1^k)^\alpha} \\ \vdots \\ \frac{1}{(\mathbf{1}_n^T X_{\cdot n}^k - X_{n\cdot}^k \mathbf{1}_n + L_n^k)^\alpha} \end{bmatrix} \geq 0,$$

$$\text{constraints of (2)}, \quad k = 1, \dots, \tau.$$

- **Theorem 2: SOC uncertainty set equivalent convex form**

$$\Delta = \{\hat{r}_c + y + C^T w : \exists y, w \in \mathbb{R}^{n\tau} \; s.t. \quad \|y\|_2 \leqslant \Gamma_1^B, \|w\|_2 \leqslant \sqrt{\frac{1-\epsilon}{\epsilon}}\},$$

$$C^T C = \hat{\Sigma} + \Gamma_2^B \mathbf{I}$$

$$\min_{X^k, L^k, z, t} \quad \sum_{k=1}^{\tau} (\sum_i \sum_j X_{ij}^k W_{ij}) + \beta t$$

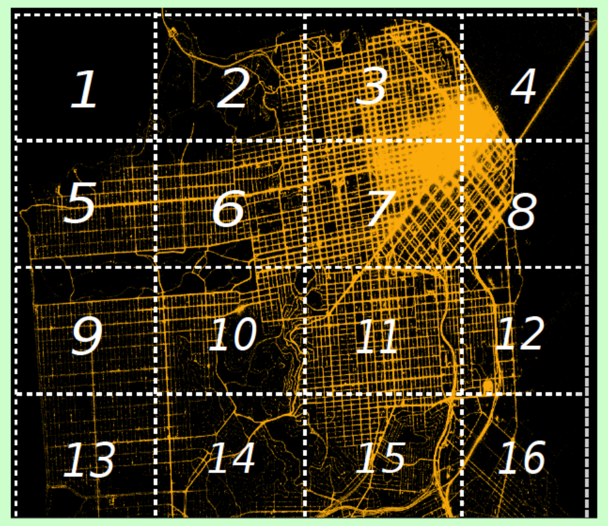
$$\text{subject to} \quad \hat{r}_c^T z + \Gamma_1^B \|z\|_2 + \sqrt{\frac{1}{\epsilon} - 1} \|C\|_F \|z\|_2 \leqslant t,$$

$$c_l(X) \leqslant z,$$

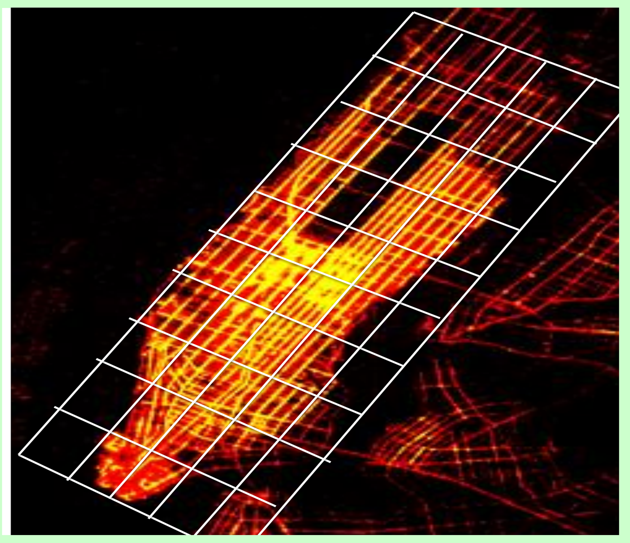
$$\text{constraints of (2)}, \quad k = 1, \dots, \tau.$$

$$c_l(X) = \begin{bmatrix} \frac{1}{(\mathbf{1}_n^T X_{\cdot 1}^1 - X_{1\cdot}^1 \mathbf{1}_n + L_1^1)^\alpha} \\ \vdots \\ \frac{1}{(\mathbf{1}_n^T X_{\cdot n}^1 - X_{n\cdot}^1 \mathbf{1}_n + L_n^1)^\alpha} \\ \vdots \\ \frac{1}{(\mathbf{1}_n^T X_{\cdot n}^\tau - X_{n\cdot}^\tau \mathbf{1}_n + L_n^\tau)^\alpha} \end{bmatrix}$$

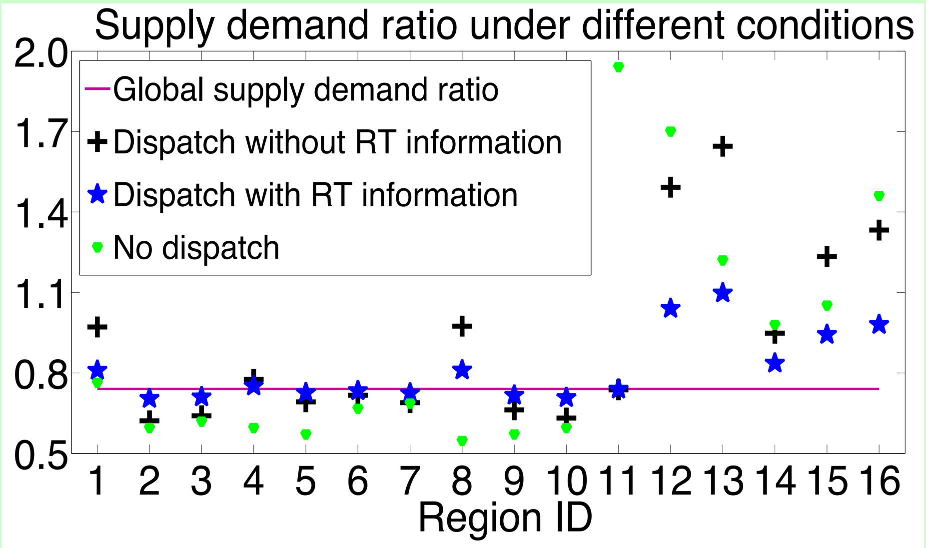
Evaluations with a SF dataset and a NYC dataset



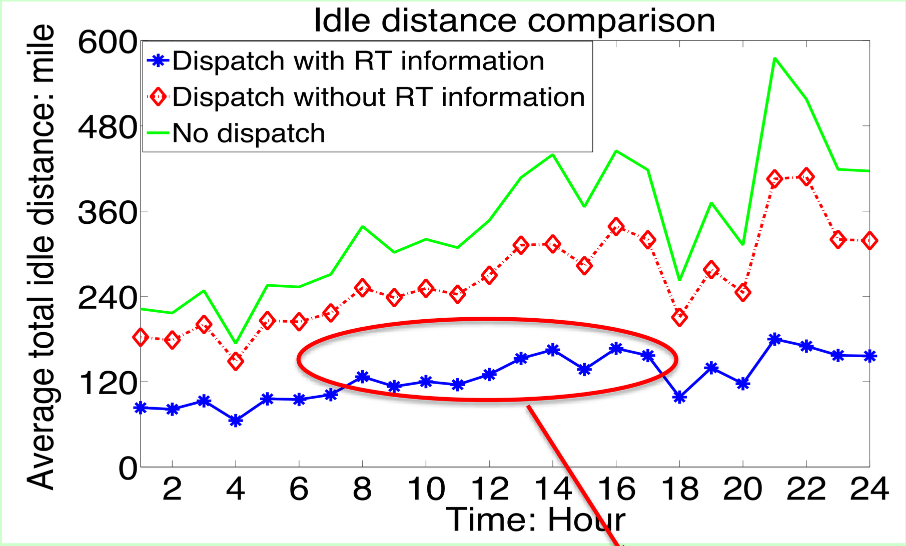
Collection Period	28 days
Record Number	1,000,000



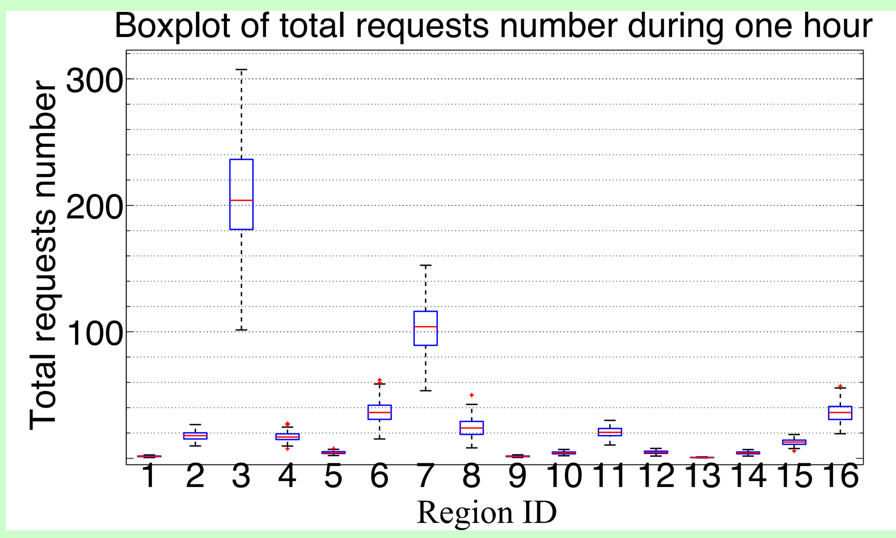
Collection Period	4 years
Data Size	100 GB

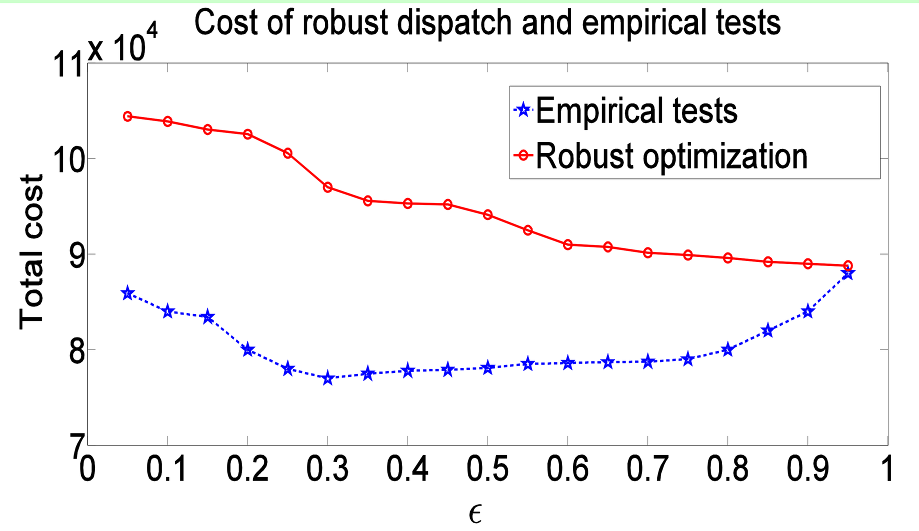


RHC dispatch *VS* non-dispatch:
Supply demand ratio error is reduced by 45%

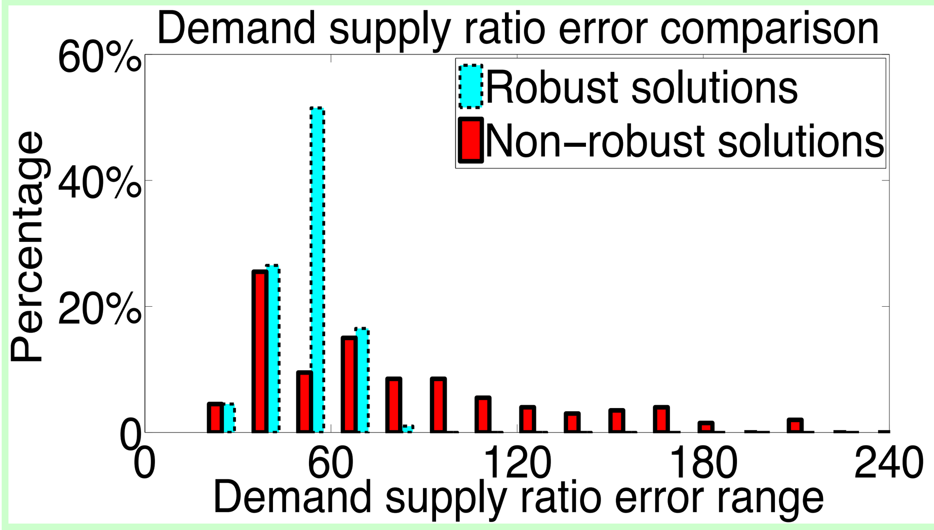


RHC dispatch *VS* non-dispatch:
Idle distance is reduced by 52%

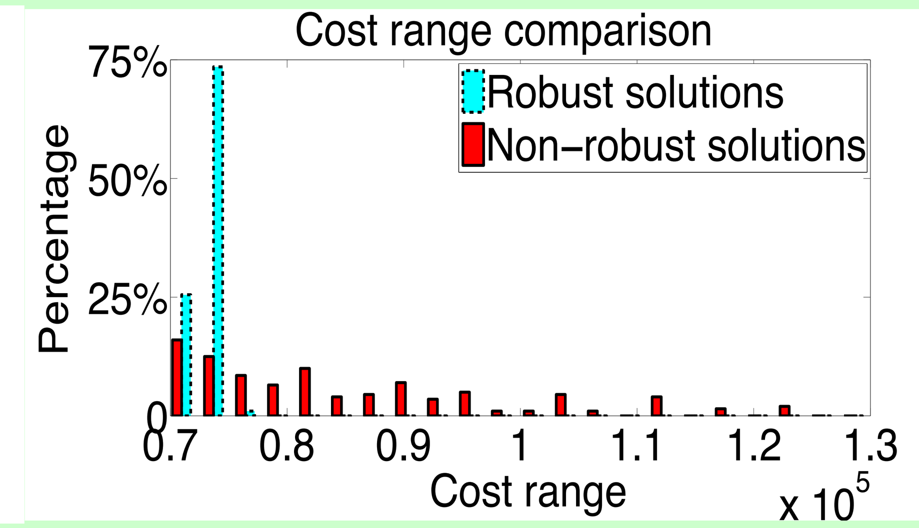




Trade-off between probabilistic guarantee level and the average cost of robust solutions



Consider demand uncertainties:
The average demand supply ratio error is reduced by 31.7% with robust dispatch solutions



Consider demand uncertainties:
The average total idle mileage is reduced by 10.13% with robust solutions

General Robotics, Automation, Sensing & Perception Lab

PENN RESEARCH IN EMBEDDED COMPUTING AND INTEGRATED SYSTEMS ENGINEERING