End-to-End Learning and Optimization

Pascal Van Hentenryck

H. Milton Stewart School of
Industrial and Systems Engineering

Machine Learning Center

Supply Chain and Logistics Institute

Artificial Intelligence

Georgia Tech

Machines are finally getting the best of humans at poker. DOLGACHOV/ISTOCKPH

Artificial intelligence goes deep to beat humans at poker

By **Tonya Riley** | Mar. 3, 2017 , 2:15 PM

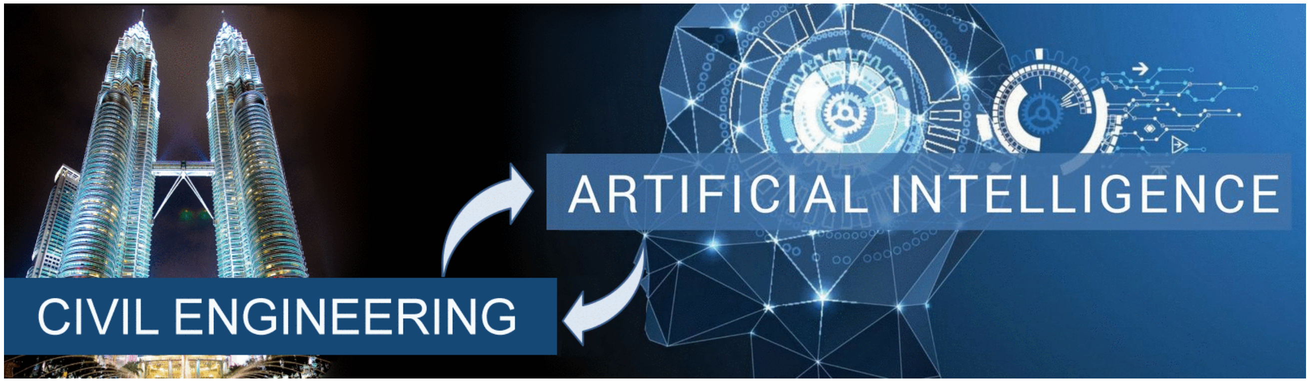
Artificial Intelligence

China Daily

China's first facial recognition payment-based shopping street opens ...

NEXT CHALLENGE

Artificial Intelligence for ISEngineering



Motivation

Georgia Tech

Learning Optimization Proxies
- ride-haring systems

Optimization over Preferences
- on-demand transit systems

Learning over Physical Systems
- security-constrained OPF

Optimization for Privacy
- releasing census data

Biases, Fairness, Privacy
- releasing census data

Motivation

Learning Optimization Proxies
- ride-haring systems

Optimization over Preferences
- on-demand transit systems

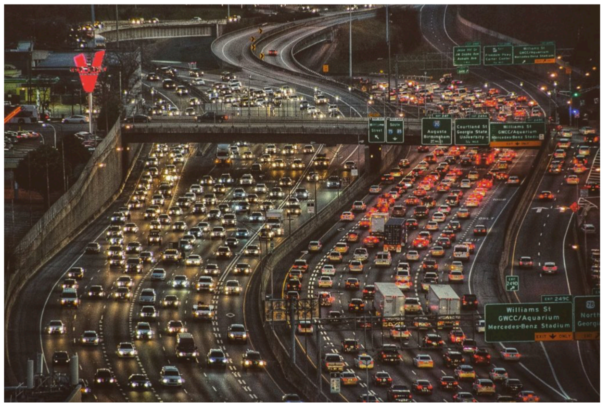
Learning over Physical Systems
- security-constrained OPF

Optimization for Privacy
- releasing census data

Biases, Fairness, Privacy
- releasing census data

Ride-Hailing Systems and Congestion

Georgia Tech



LOCAL, NEWS

Atlanta Named One of the Worst Cities For Commuting – AGAIN

By GAFollowers @GAFollowers · On May 5, 2019

RESEARCH ARTICLE | ECONOMICS

Do transportation network companies decrease or increase congestion?

Gregory D. Erhardt[1,*], Sneha Roy[1], Drew Cooper[2], Bhargava Sana[2], Mei Chen[1] and Joe Castiglione[2]

+ See all authors and affiliations

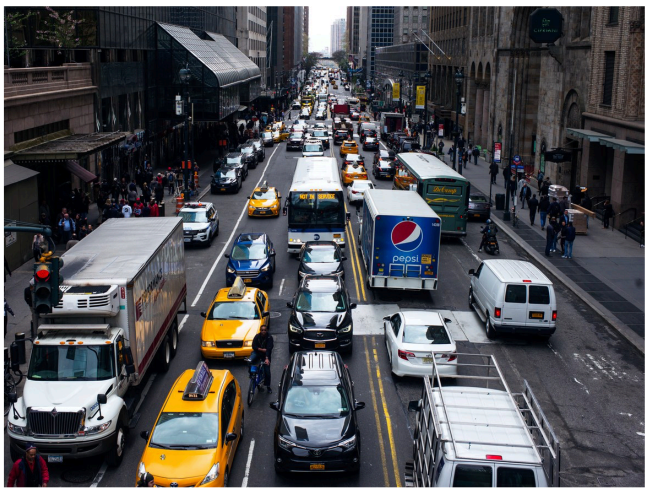
Science Advances 08 May 2019:
Vol. 5, no. 5, eaau2670
DOI: 10.1126/sciadv.aau2670

- Observations
  - Between 2010 and 2016, weekday vehicle hours of delay increased by 62% compared to 22% in a counterfactual 2016 scenario without TNCs.

Congestion

AARIAN MARSHALL TRANSPORTATION 08.08.18 08:48 PM

NEW YORK CITY GOES AFTER UBER AND LYFT



ROBERT NICKELSBERG/GETTY IMAGES

In major defeat for Uber and Lyft, New York City votes to limit ride-hailing cars

NYC becomes the first American city to restrict the explosive growth in for-hire vehicles

By Shoshana Wodinsky | Aug 8, 2018, 4:39pm EDT

SHARE

Prior Work

Georgia Tech

On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment

Check for updates

Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus

PNAS January 17, 2017 114 (3) 462-467; first published January 3, 2017 https://doi.org/10.1073/pnas.1611675114

Edited by Michael F. Goodchild, University of California, Santa Barbara, CA, and approved November 22, 2016 (received for review July 20, 2016)

This article has a correction. Please see:

Correction for Alonso-Mora et al., On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment

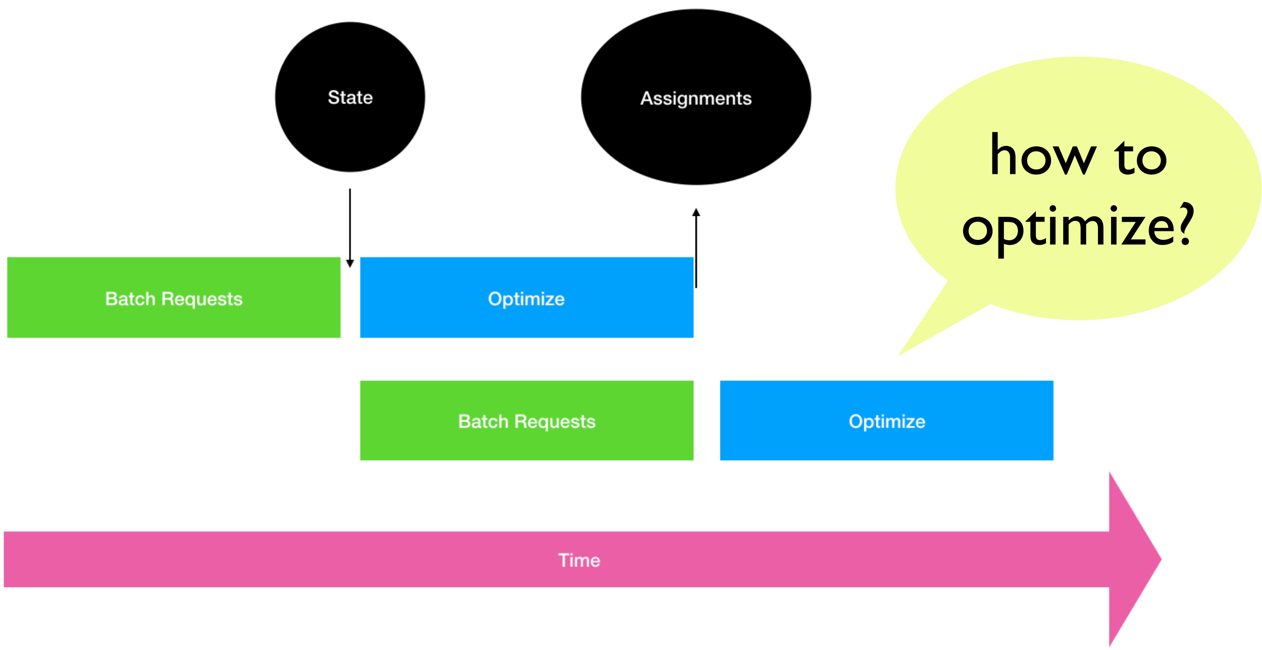
Article | Figures & SI | Info & Metrics | PDF

- **98% of riders**
- 3000 vehicles
- average wait: 3.8 min
- average deviation: 3.5 min

Large-Scale Ride-Sharing Systems

- Constraints
  - service guarantees
    - a rider will always be served
  - ride duration cannot exceed a certain threshold
    - 50% more duration or a small constant (e.g., 4 minutes)
- Objective
  - minimizing average waiting time
- Secondary objective
  - a rider cannot wait too long
- Computational challenges
  - 1/2 million requests a day
  - 500 requests every 30 seconds

Real-Time Optimization



The Master Optimization Problem

waiting times

covering the riders

one route per vehicle

$$\min \sum_{r \in R} c_r y_r + \sum_{i \in P} p_i z_i$$

subject to

$$\left(\sum_{r \in R} y_r a_i^r \right) + z_i = 1 \qquad \forall i \in P \qquad (\pi_i)$$

$$\sum_{r \in R_v} y_r = 1 \qquad \forall v \in V \qquad (\sigma_v)$$

$$z_i \in \mathbb{N} \qquad \forall i \in P$$

$$y_r \in \{0, 1\} \qquad \forall r \in R$$

The Master Optimization Problem

penalty for
lack of service

$$\min \quad \sum_{r \in R} c_r y_r + \sum_{i \in P} p_i z_i$$

subject to

$$\left(\sum_{r \in R} y_r a_i^r \right) + z_i = 1 \qquad \forall i \in P \qquad (\pi_i)$$

$$\sum_{r \in R_v} y_r = 1 \qquad \forall v \in V \qquad (\sigma_v)$$

$$z_i \in \mathbb{N} \qquad \forall i \in P$$

$$y_r \in \{0, 1\} \qquad \forall r \in R$$

riders not served
in this period

The Pricing Problem

- Significant challenge
  - minimization of waiting times
- Makes it hard to use
  - resource constrained shortest paths
- Time-expanded graphs?
  - computational issues
- Solution
  - exact anytime dedicated algorithm
  - disjoint routes
  - pruning techniques

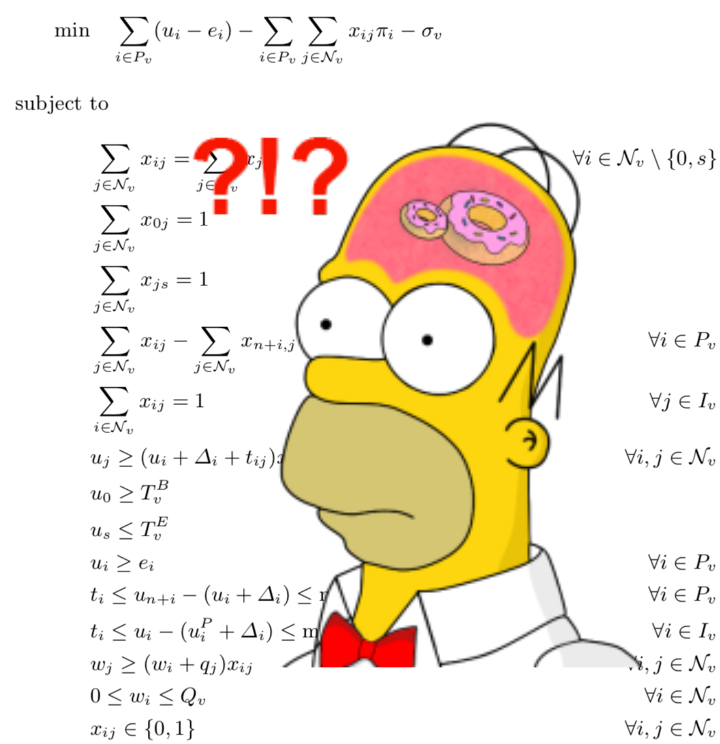
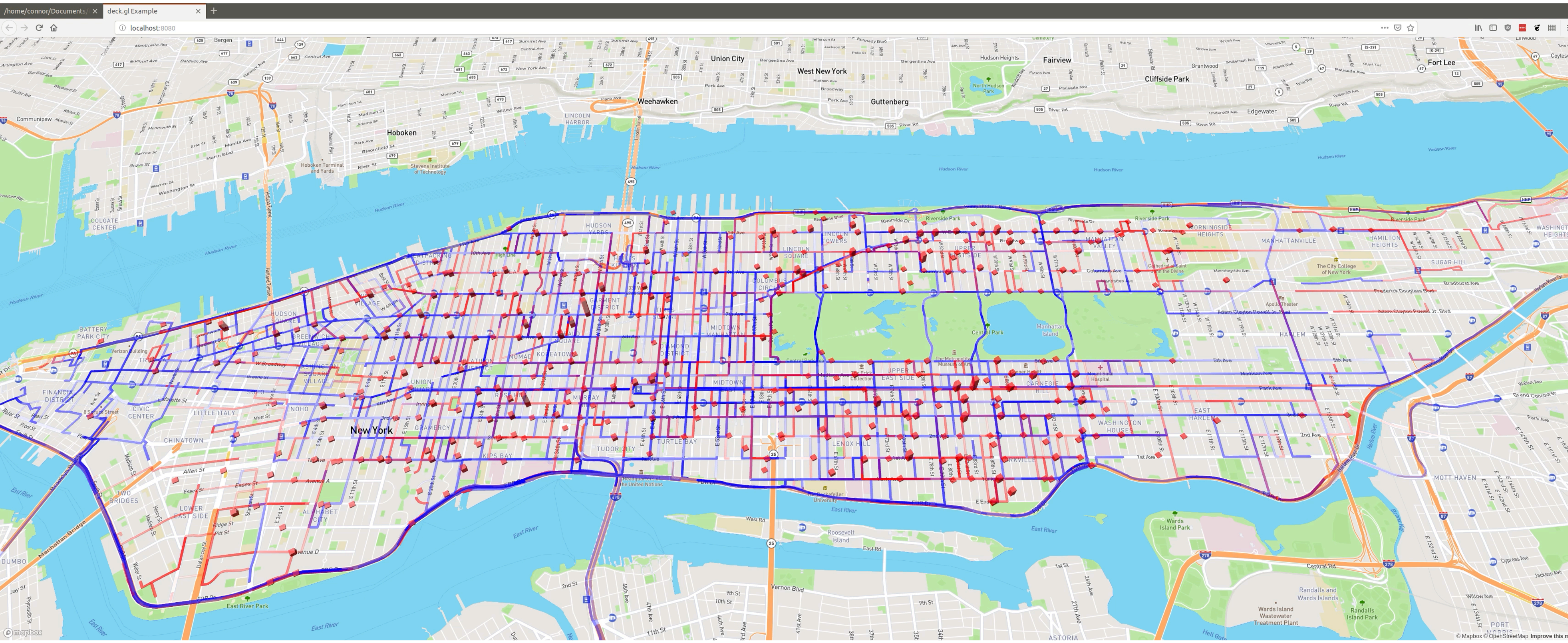


Fig. 3: The Pricing Problem Formulation for Vehicle v.

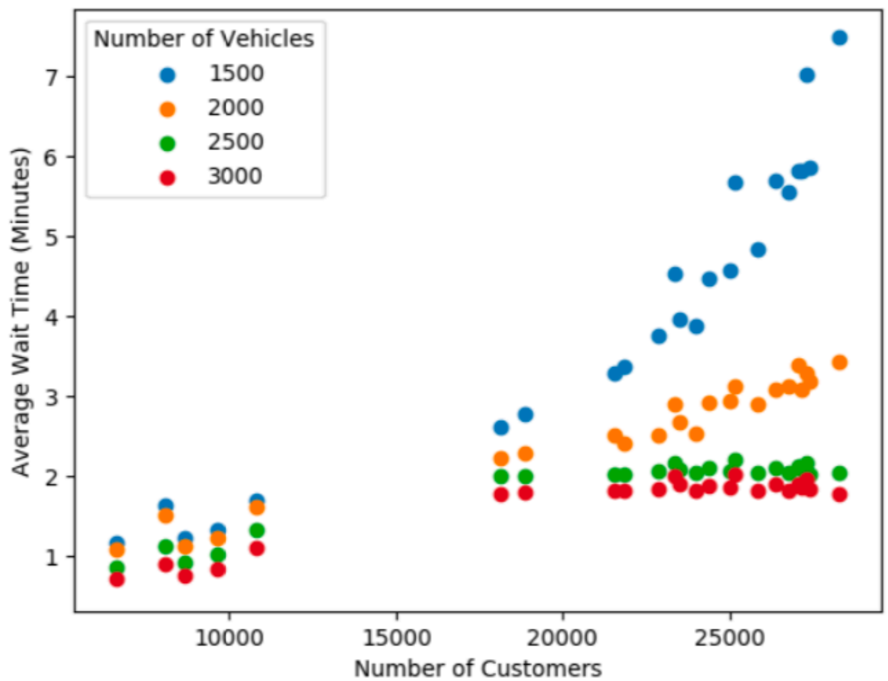
Real-Time Ride Sharing

Georgia Tech



Experimental Results on NYC

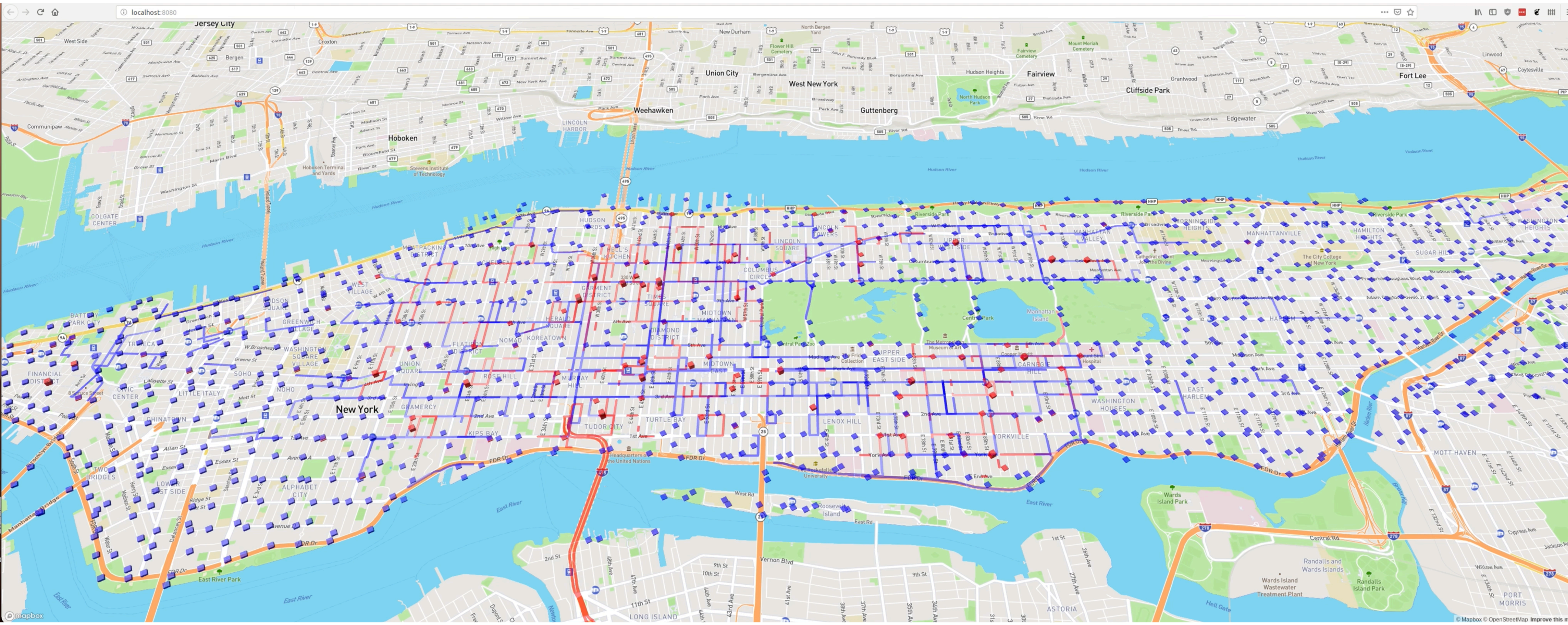
Georgia Tech



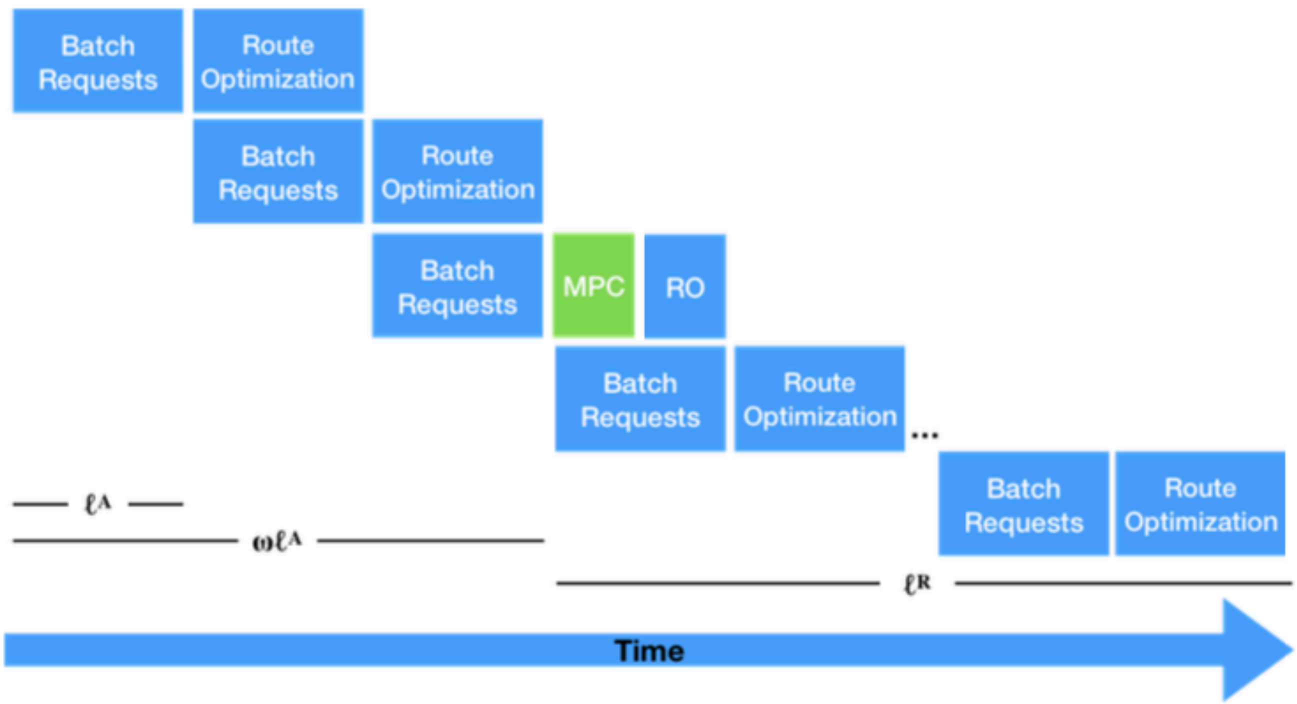
- 100% riders
- 2000 vehicles
- average wait: 3.6 min
- average deviation: 1.48 min

Waiting Times

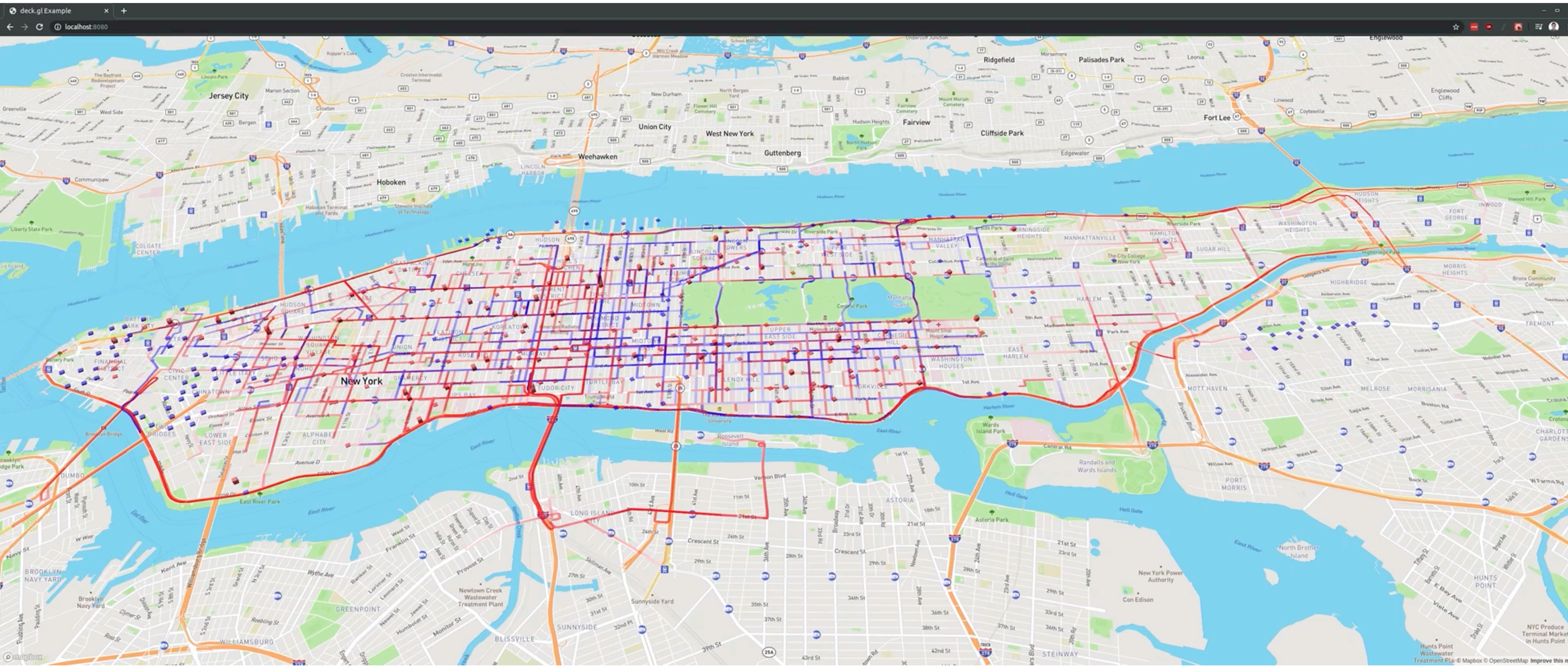
Vehicle Utilization



Optimization + ML + MPC



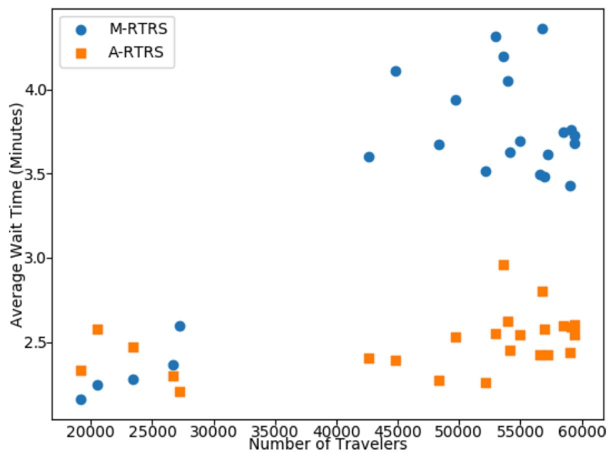
Optimization + ML + MPC



Optimization + ML + MPC

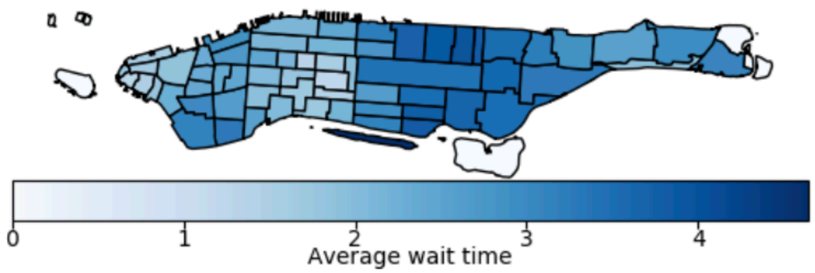
Georgia Tech

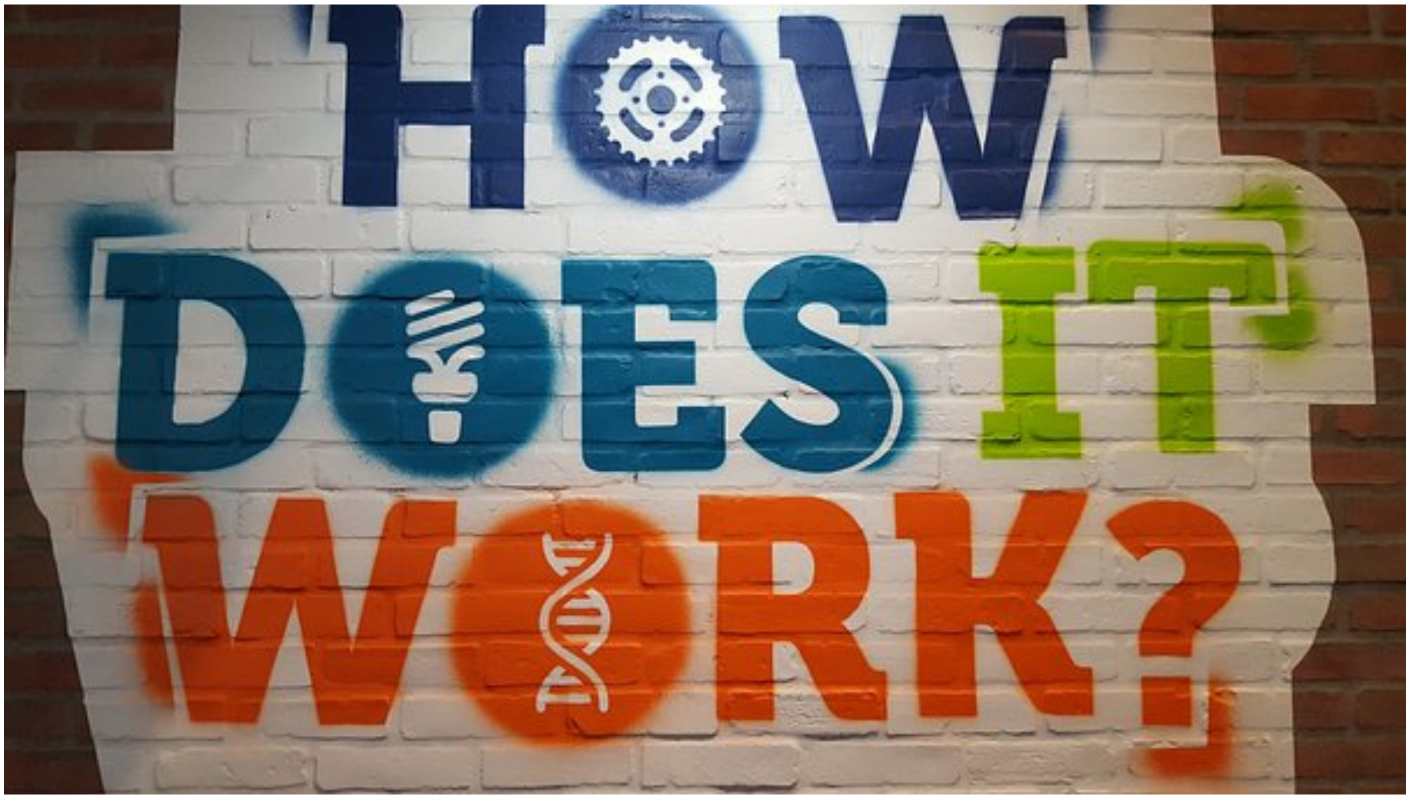
Number of Requests	M-RTRS	A-RTRS
$< 40{,}000$	2.33	2.37
$40{,}000 - 50{,}000$	3.83	2.40
$50{,}000 <$	3.78	2.56



Optimization + ML + MPC

Number of Requests	M-RTRS	A-RTRS
$< 40{,}000$	2.33	2.37
$40{,}000 - 50{,}000$	3.83	2.40
$50{,}000 <$	3.78	2.56




HOW
DOES IT
WORK?

Overview of the MPC

- Coarser Time and Space Granularity
  - zonal level instead of individual requests
  - frequency is an order of magnitude longer
- Demand forecasting
  - vector autoregression
- Model Predictive Control
  - zone balancing
  - vehicle assignment

Demand Forecasting

- Forecasting the demand from zone i to j at time t
  - sparsity may be an issue for some zones
- Two steps
  - predicting the demand in a zone z in time period t
  - approximate the zone to zone demand with historical data
- Predicting the zone demand
  - predict the weekly differenced demand
    - to handle non-stationarity
  - use the zone and its adjacent zones

$$\delta_{zt} = \phi_{zt-1}\Delta_{zt-1} + \cdots + \phi_{zt-k}\Delta_{zt-k} + \eta$$

Zone Balancing

Georgia Tech

$$\min \sum_{t=0}^{T-1} \sum_{i,j \in Z} (T-t) u_{ijt} + \sum_{t=0}^{T-1} \sum_{i,j \in Z} tt_{ij} x^r_{ijt}$$

minimize unserved requests and relocations

subject to

$$x^p_{ij0} + u_{ij0} = \lceil \bar{\lambda}_{ij0} / w_{ij} \rceil \quad (\forall i, j)$$

$$x^p_{ijt} + u_{ijt} = \lceil \bar{\lambda}_{ijt} / w_{ij} \rceil + u_{ijt-1} \quad (\forall i, j, t)$$

flow conservation for requests

$$\sum_j x^p_{ijt} + x^r_{ijt} = |A_{it-1}| + \sum_j x^p_{jit-tt_{ji}} + x^r_{jit-tt_{ji}}$$

flow conservation for vehicles

$$x^p_{ijt} \in \mathbb{Z}, x^r_{ijt} \in \mathbb{Z}, u_{ijt} \in \mathbb{Z}$$

vehicles with passengers moving from i to j at time t

empty vehicles moving from i to j at time t

scaled unserved requests from i to j at time t

Vehicle Relocation

$$
\begin{aligned}
\min \quad & \sum_{v \in A_{i0}} \sum_{j} c_{vj} y_{vj} \\
\text{subject to} \quad & \sum_{v \in A_{i0}} y_{vj} = \bar{x}^{r}_{ij0} && \forall j \in Z \\
& \sum_{j} y_{vj} \leq 1 && \forall v \in A_{i0} \\
& y_{vj} \in \{0, 1\} && \forall v \in A_{i0}, j \in Z
\end{aligned}
$$

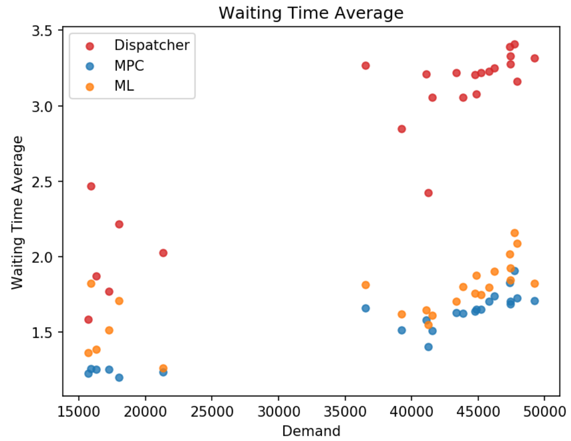
minimize relocation time

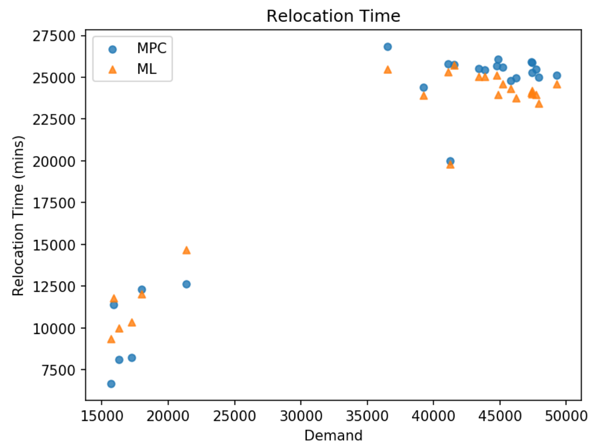
meet the relocation demand

= 1 if vehicle v relocates to zone j

Closing the Loop

- Learning the relocation model
  - overcoming the potential runtime limitations
- Optimization proxies
  - learning an optimization model
- Given
  - the current state of the system
- predict
  - the relocation decisions
- Technique
  - Kernel support vector regression





Motivation

Learning Optimization Proxies
- ride-haring systems

Optimization over Preferences
- on-demand transit systems

Learning over Physical Systems
- security-constrained OPF

Optimization for Privacy
- releasing census data

Biases, Fairness, Privacy
- releasing census data

Transforming Public Transit

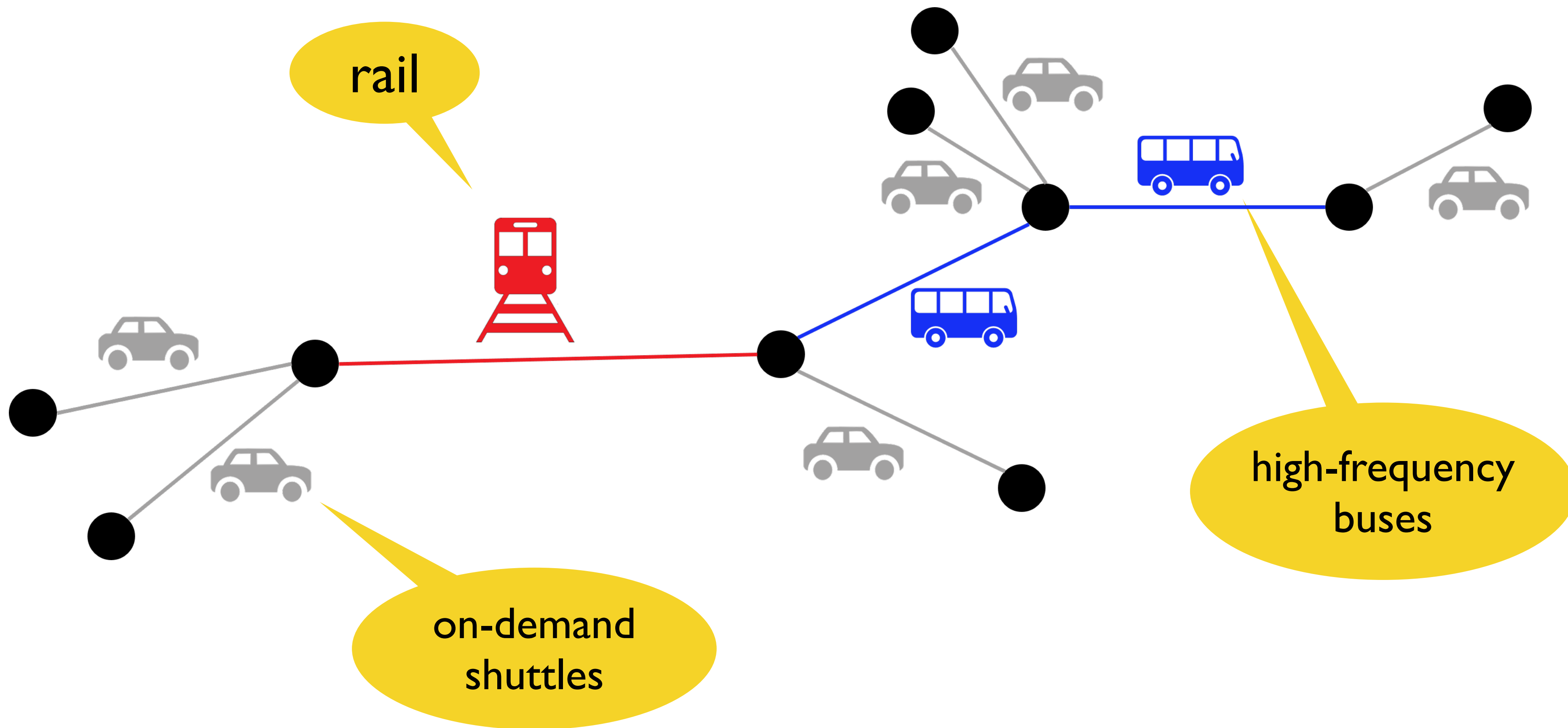
Georgia Tech

→ lyft

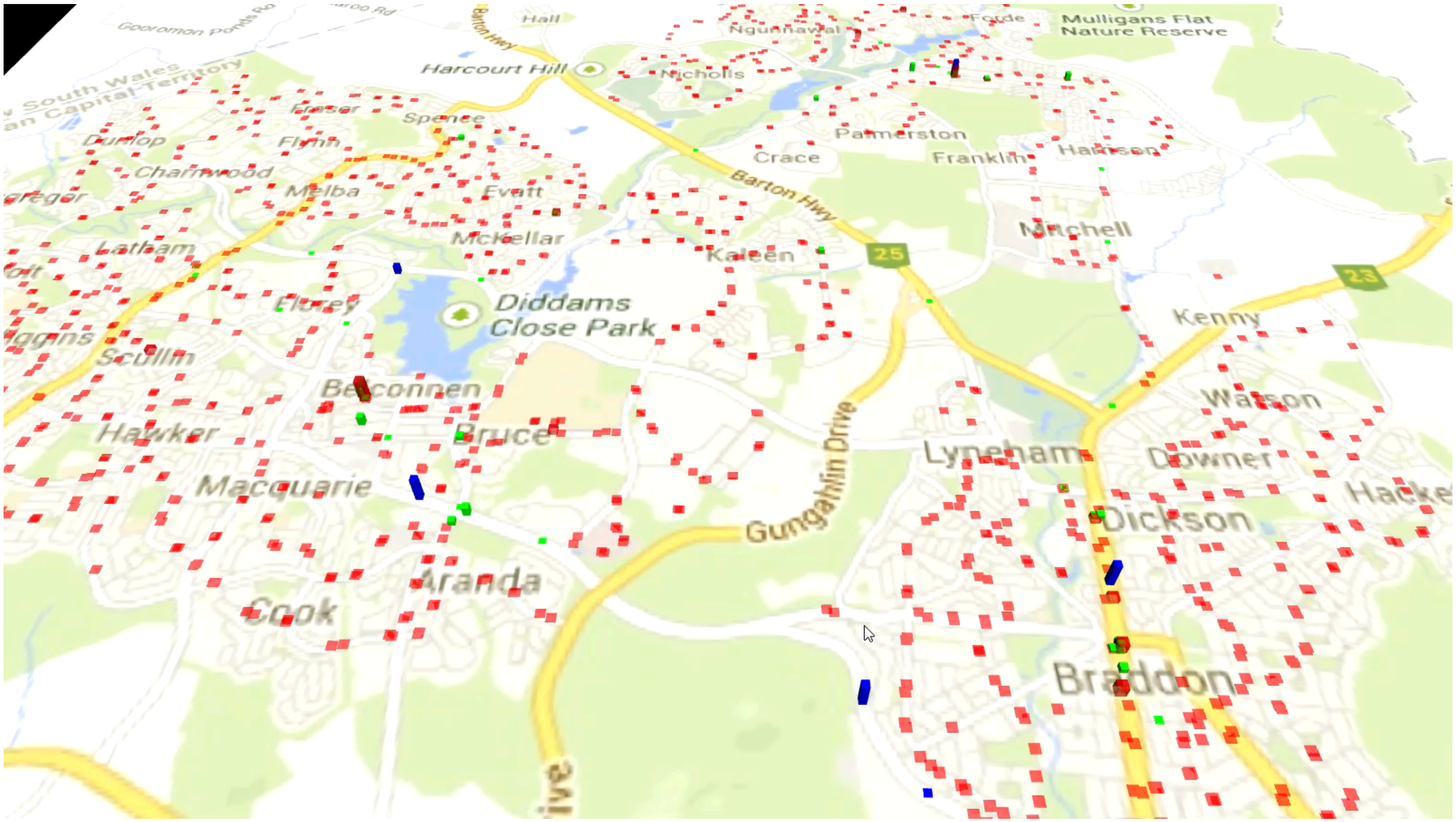
UBER

→ SOCIALLY-AWARE MOBILITY

On-Demand Multimodal Public Transit



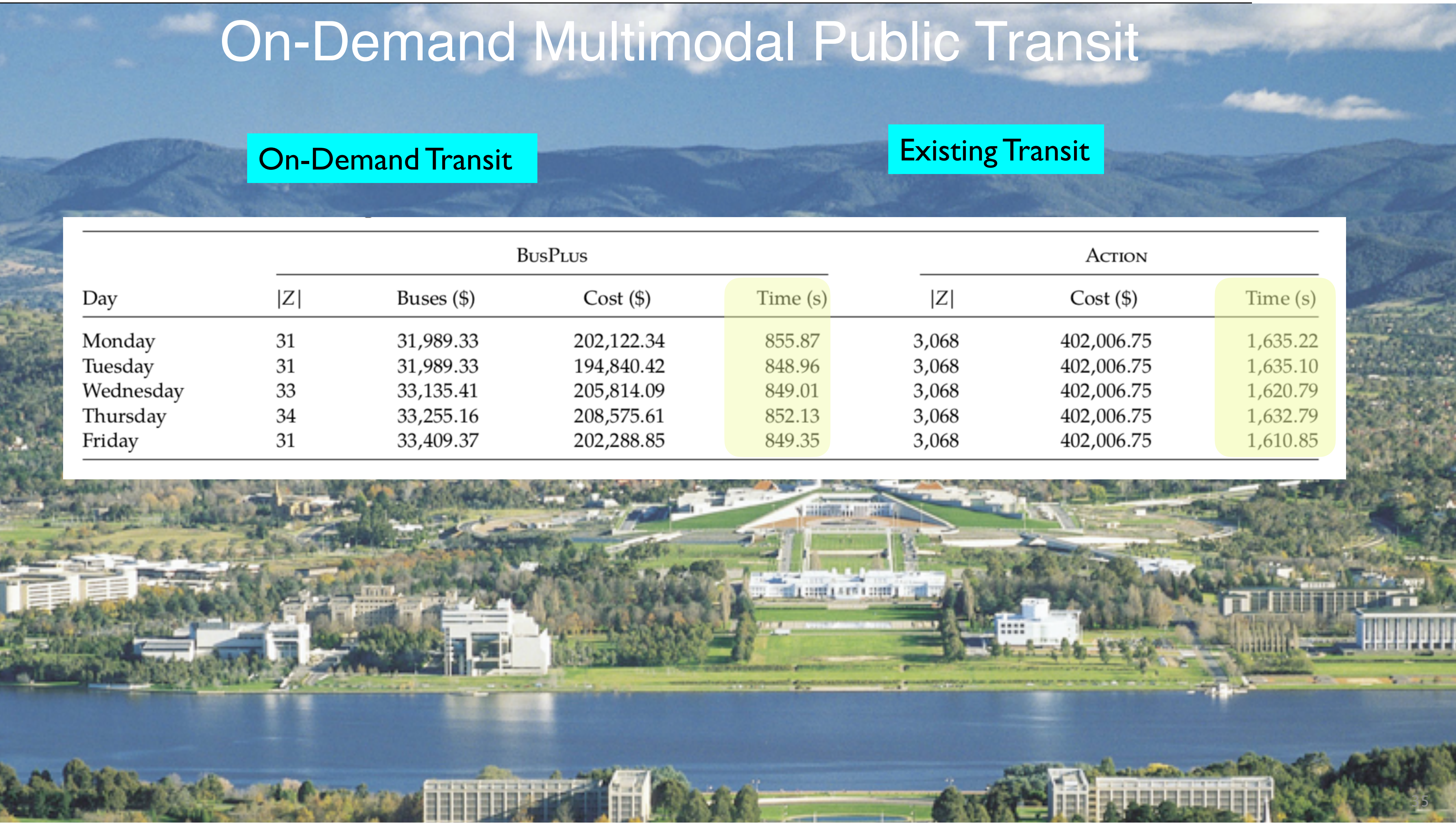
On-Demand Multimodal Transit System



On-Demand Multimodal Transit System

- On Demand and "*door to door*"
  - to address the first/last mile problem
- Multi-modal
  - to address congestion/economy of scale
  - fleets of trains, rapid transit buses, shuttles, electric bikes, scooters
- Electrified
  - no range anxiety, no emission
- Sustainable business model
  - the same price as a transit system
  - order of magnitude cheaper than Uber and Lyft
- Planned and operated holistically
  - fundamentally different from "transit + micro-transit"
- Societal impact
  - improving access to entire population segments

On-Demand Multimodal Public Transit

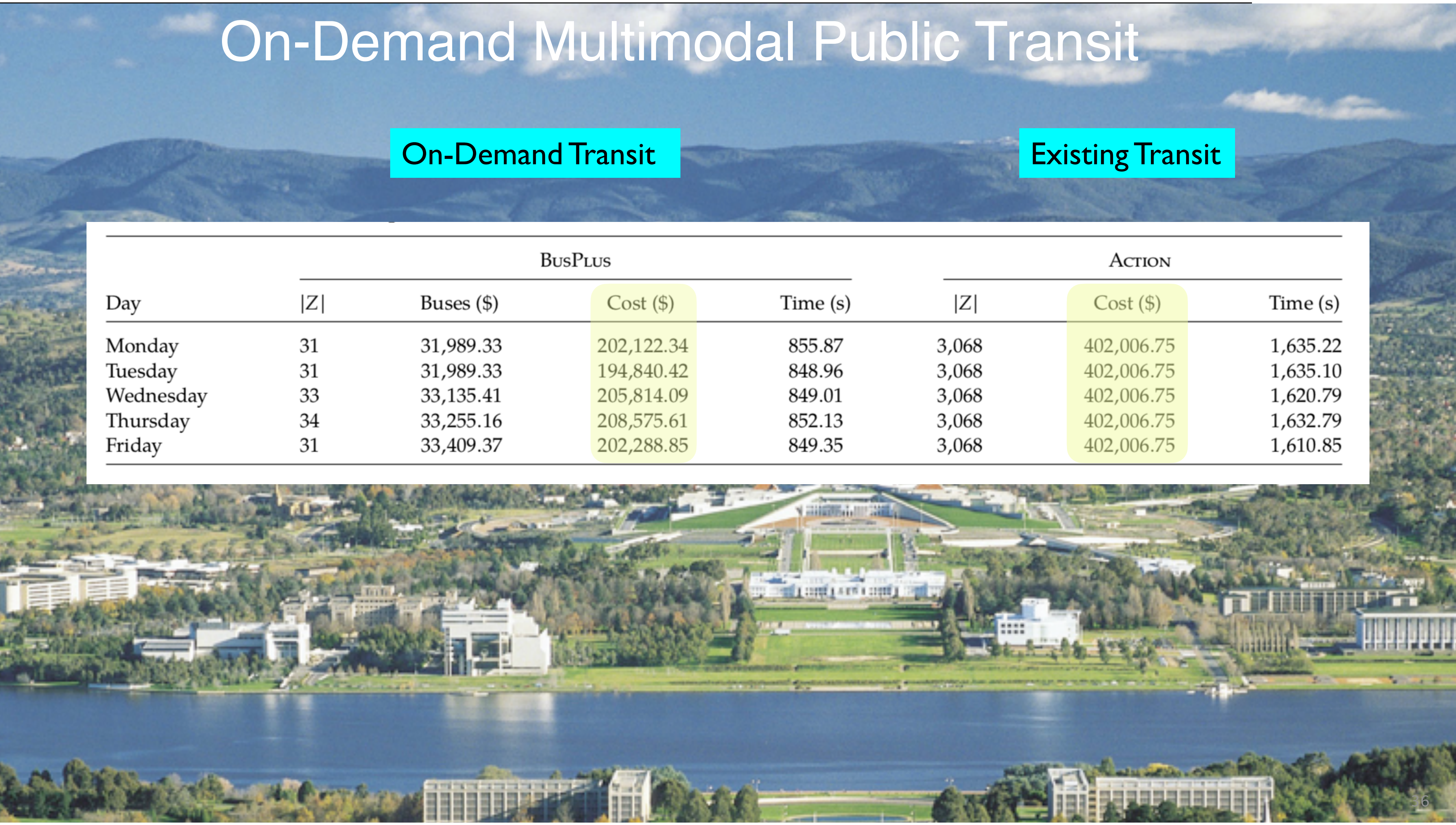


On-Demand Transit

Existing Transit

	BusPlus				Action		
Day	|Z|	Buses ($)	Cost ($)	Time (s)	|Z|	Cost ($)	Time (s)
Monday	31	31,989.33	202,122.34	855.87	3,068	402,006.75	1,635.22
Tuesday	31	31,989.33	194,840.42	848.96	3,068	402,006.75	1,635.10
Wednesday	33	33,135.41	205,814.09	849.01	3,068	402,006.75	1,620.79
Thursday	34	33,255.16	208,575.61	852.13	3,068	402,006.75	1,632.79
Friday	31	33,409.37	202,288.85	849.35	3,068	402,006.75	1,610.85

On-Demand Multimodal Public Transit

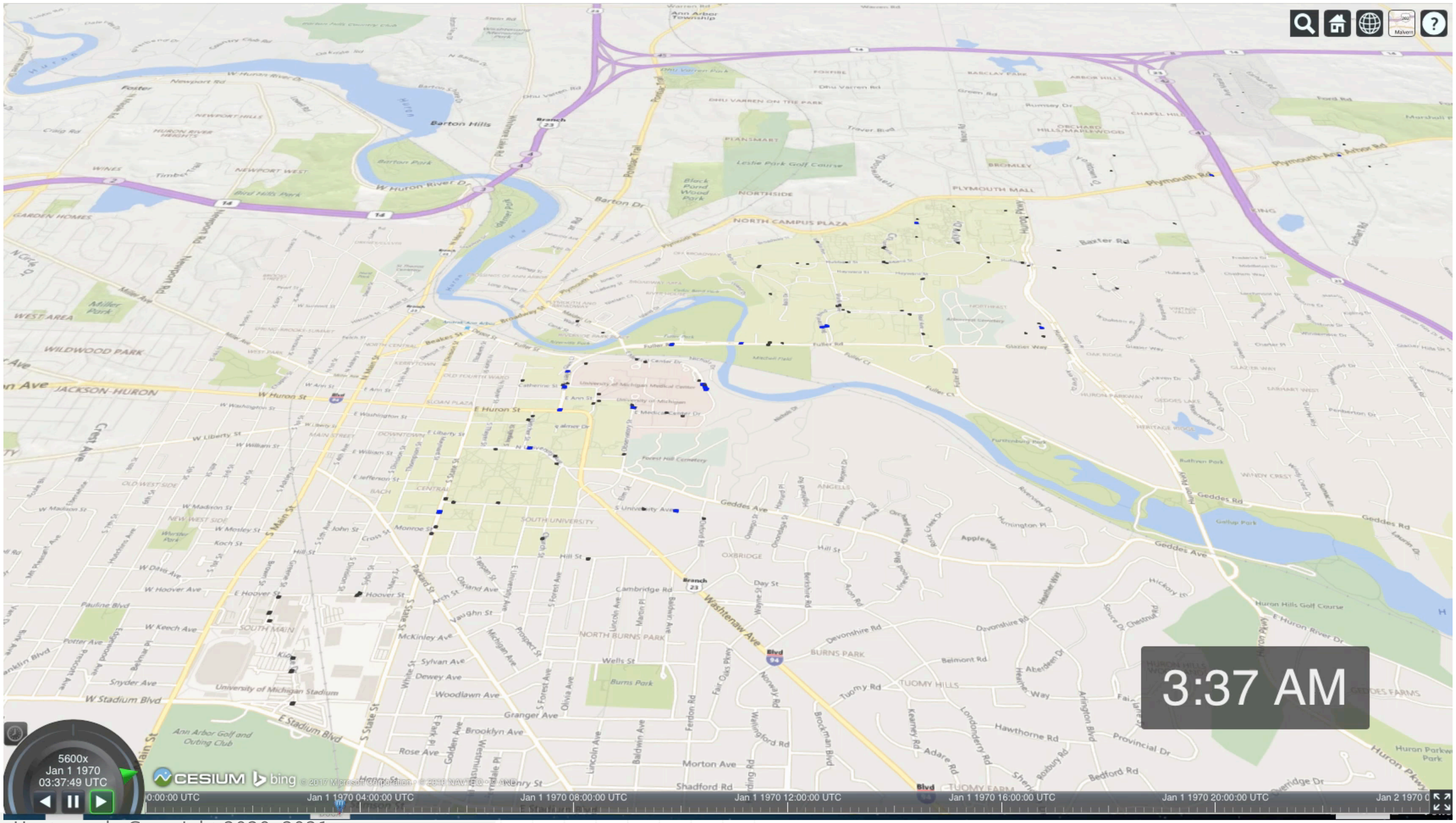


On-Demand Transit

Existing Transit

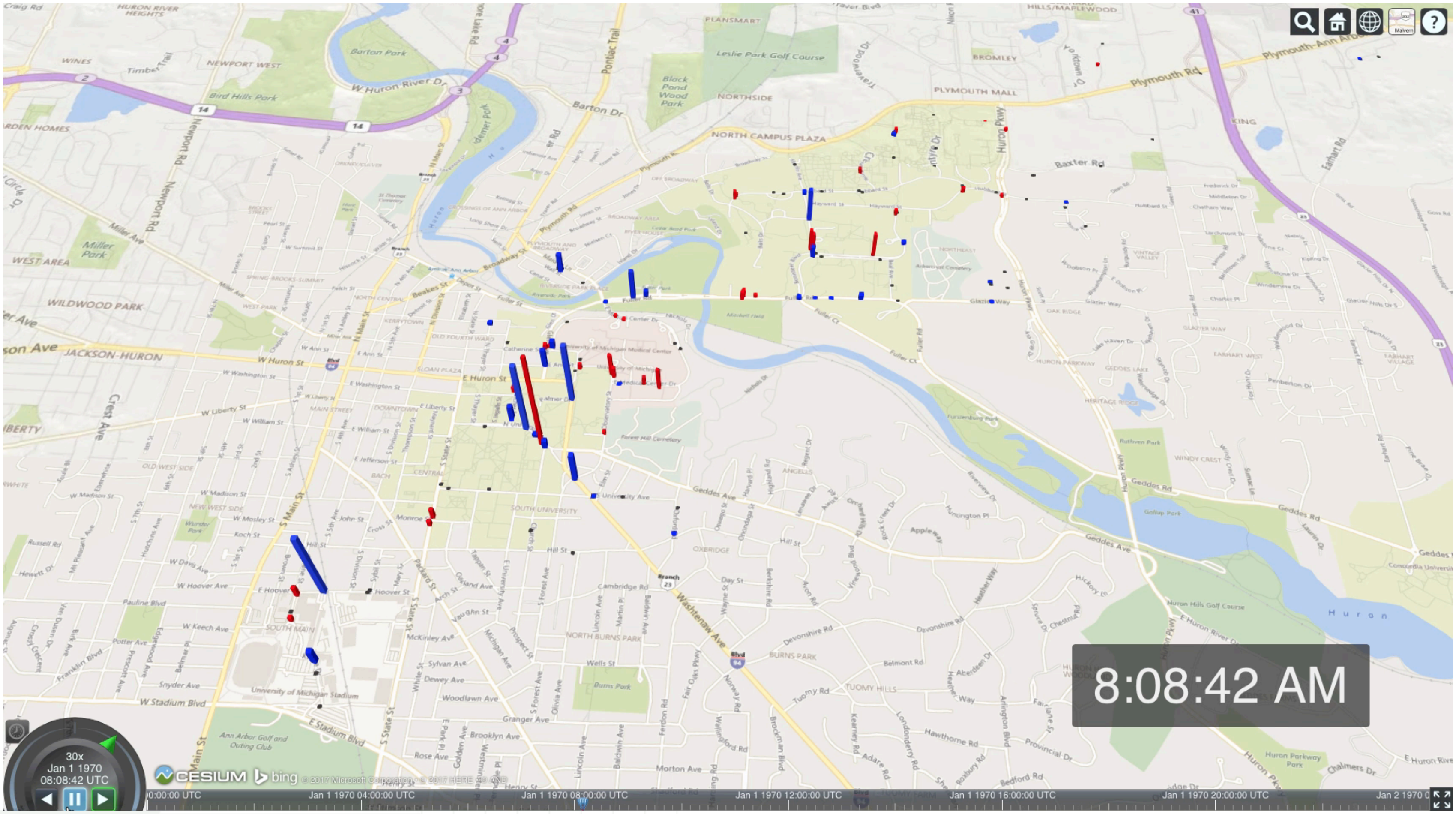
	BusPlus				Action		
Day	|Z|	Buses ($)	Cost ($)	Time (s)	|Z|	Cost ($)	Time (s)
Monday	31	31,989.33	202,122.34	855.87	3,068	402,006.75	1,635.22
Tuesday	31	31,989.33	194,840.42	848.96	3,068	402,006.75	1,635.10
Wednesday	33	33,135.41	205,814.09	849.01	3,068	402,006.75	1,620.79
Thursday	34	33,255.16	208,575.61	852.13	3,068	402,006.75	1,632.79
Friday	31	33,409.37	202,288.85	849.35	3,068	402,006.75	1,610.85

The UM Transit System

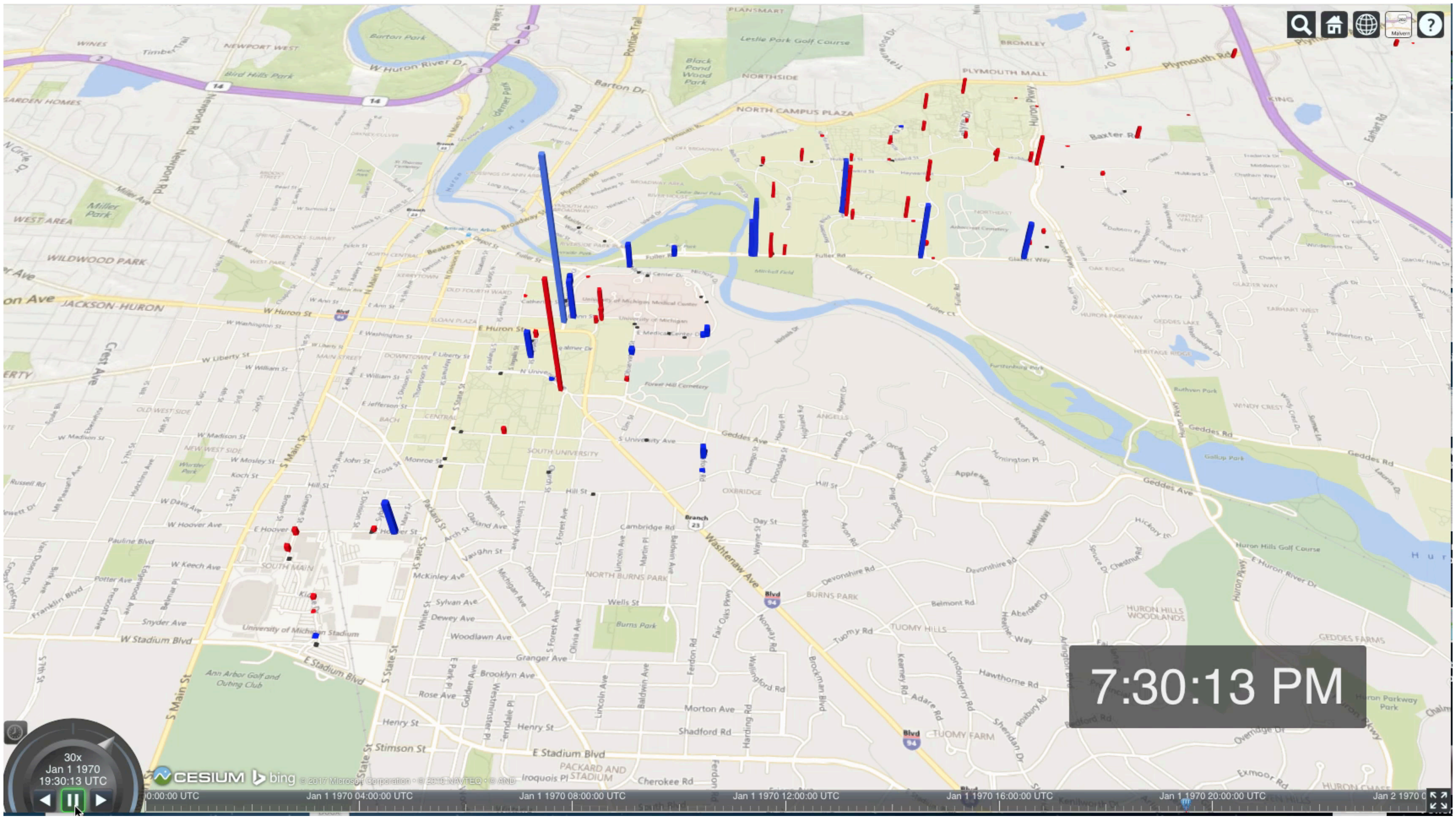


Before/After Comparisons

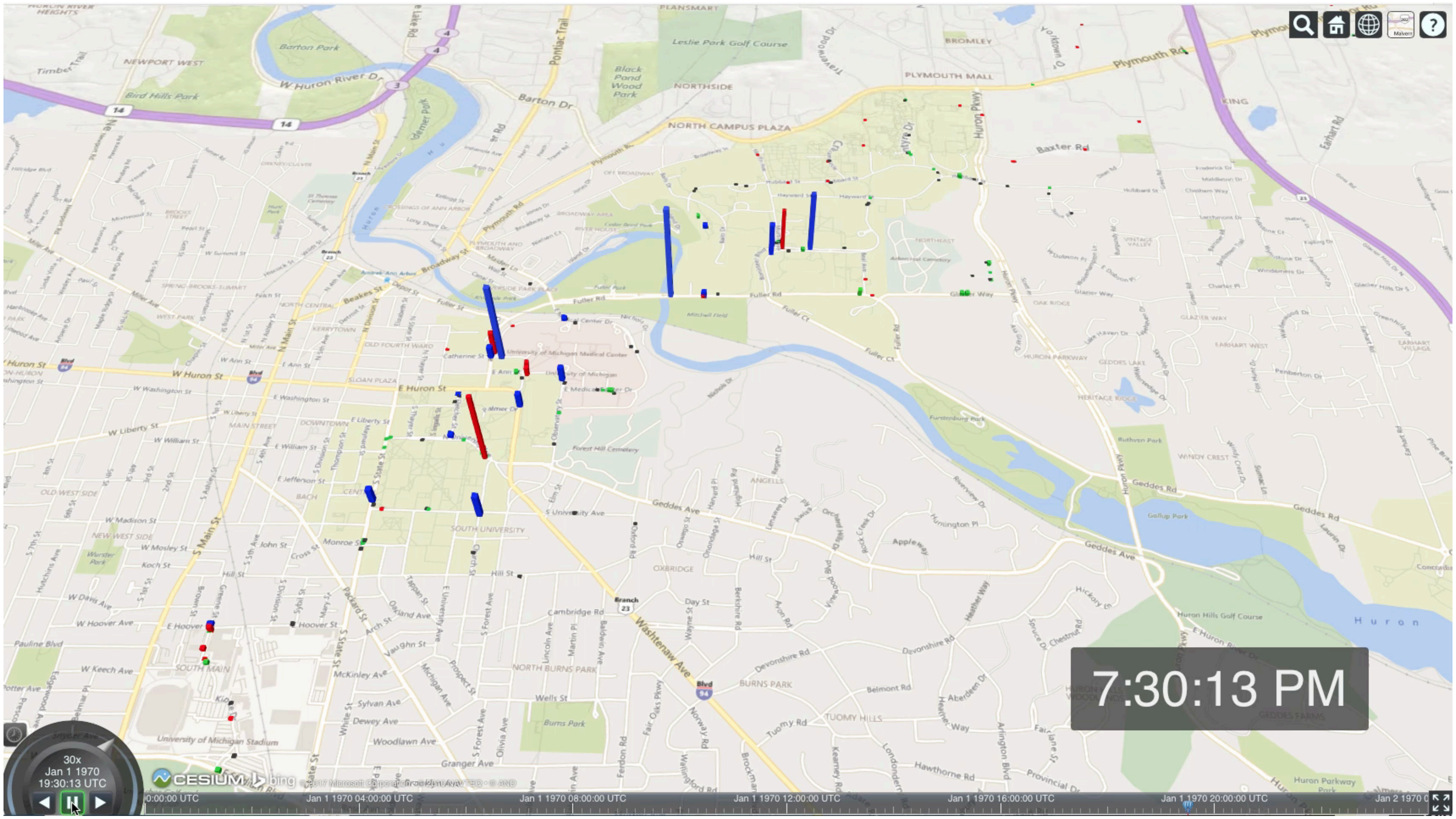
Existing Bus System



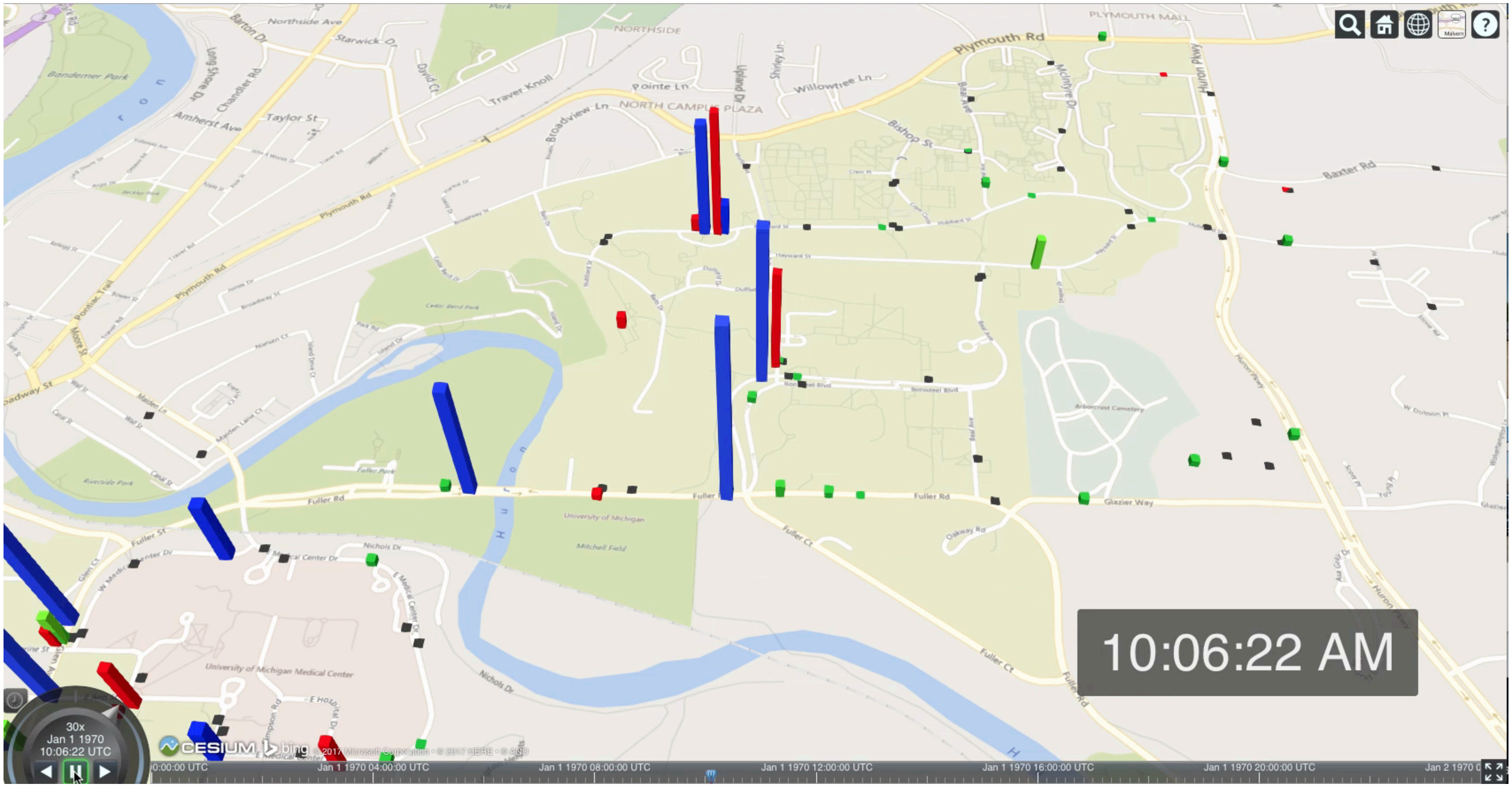
Existing Bus System



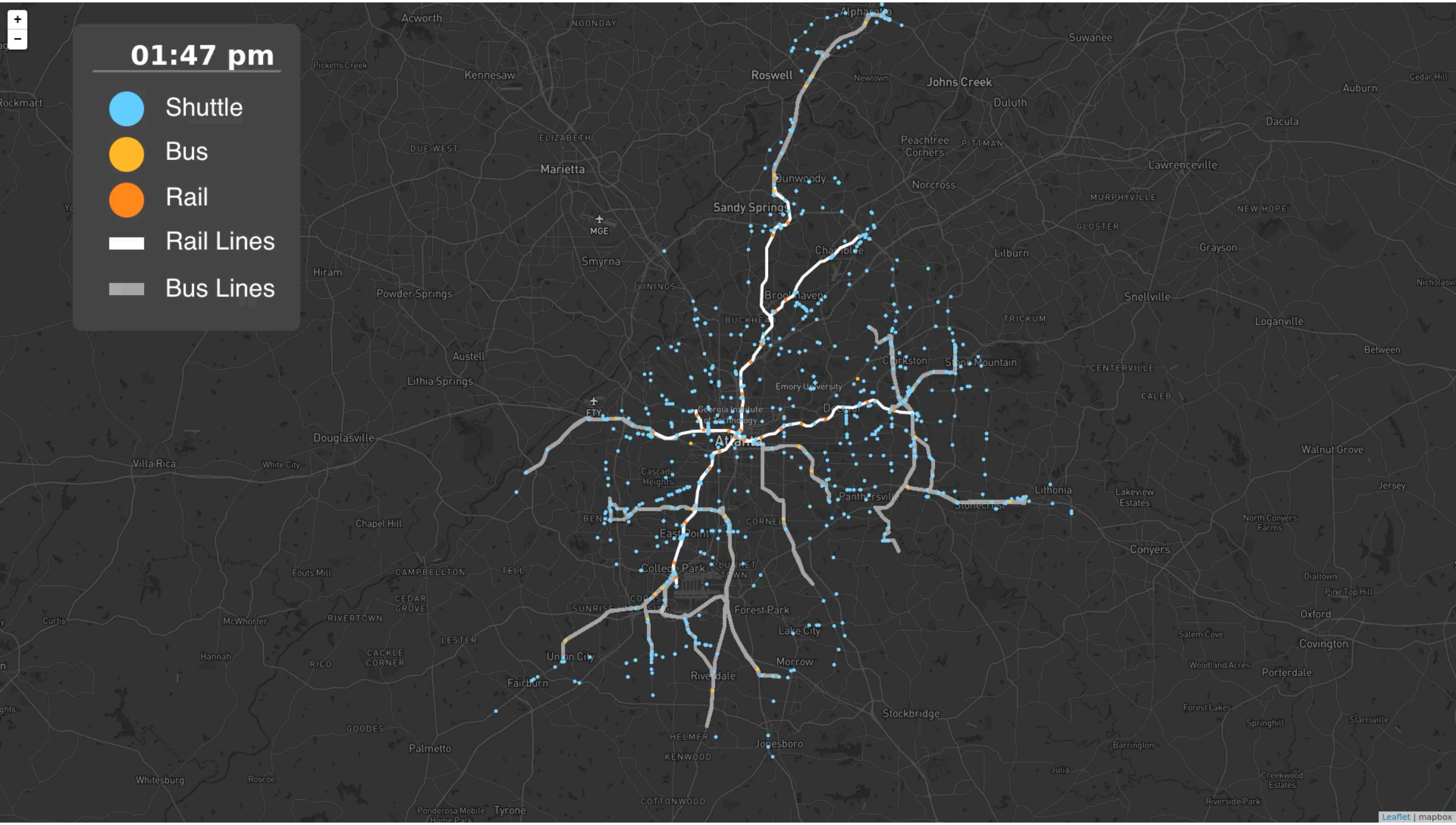
ODMTS



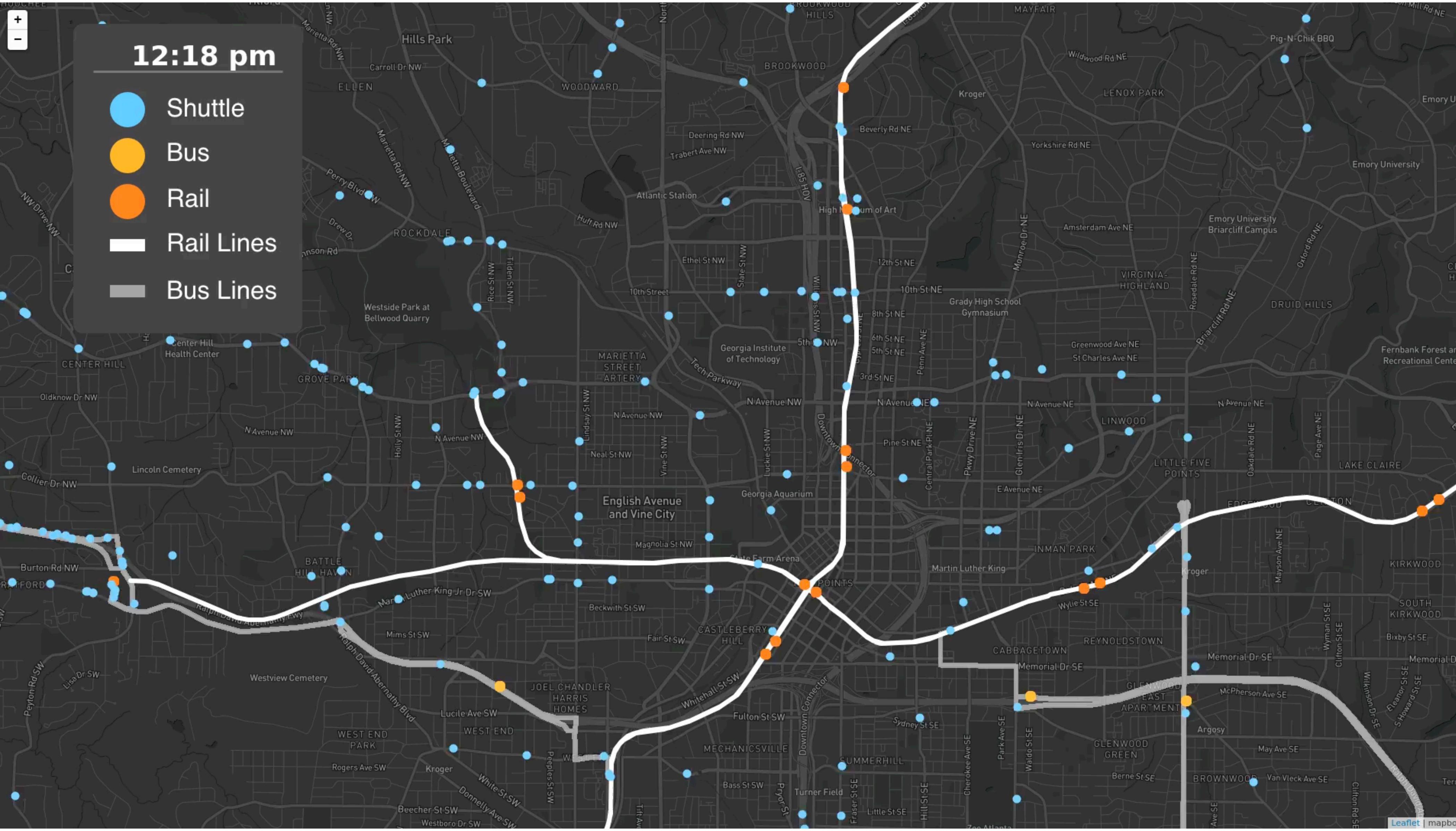
ODMTS


10:06:22 AM
30x
Jan 1 1970
10:06:22 UTC
Fuller Rd
Plymouth Rd
Glazier Way
Baxter Rd
Jan 1 1970 04:00:00 UTC
Jan 1 1970 08:00:00 UTC
Jan 1 1970 12:00:00 UTC
Jan 1 1970 16:00:00 UTC
Jan 1 1970 20:00:00 UTC

Atlanta



01:47 pm
Shuttle
Bus
Rail
Rail Lines
Bus Lines

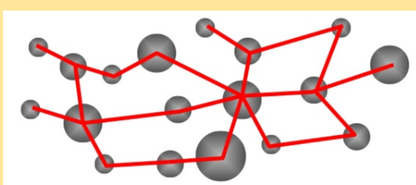


12:18 pm
Shuttle
Bus
Rail
Rail Lines
Bus Lines
Hills Park
Brookwood
Woodward
Mayfair
Lenox Park
Kroger
Emory University
Pig-N-Chik BBQ
Atlantic Station
Rockdale
Westside Park at Bellwood Quarry
Center Hill Health Center
Center Hill
Grove Park
Marietta Street Artery
Georgia Institute of Technology
Grady High School Gymnasium
Virginia-Highland
Druid Hills
Emory University Briarcliff Campus
Fernbank Forest and Recreational Center
Linwood
Little Five Points
Lake Claire
Lincoln Cemetery
English Avenue and Vine City
Georgia Aquarium
State Farm Arena
Inman Park
Martin Luther King
Kirkwood
South Kirkwood
Battle Hill Haven
Castleberry Hill
Cabbagetown
Reynoldstown
Westview Cemetery
Joel Chandler Harris Homes
West End Park
West End
Mechanicsville
Summerhill
Turner Field
Glenwood Green
Brownwood
Argosy
Leaflet | mapbox

Capturing Rider Mode Choice

- Key worry of transit agencies
  - what if adoption increases?
  - are we designing the network properly?
  - are we sizing the fleet correctly?

Leader
(Transit system designer)



Transit network design →

← User preference

Follower
(rider)

Computational Challenges

- Relax the optimization program

$$\min_x f(x, y) \text{ subject to } c(x, y) \wedge \mathcal{M}(x) = y$$

- Check for inconsistencies

$$\mathcal{M}(\bar{x}) \neq \bar{y}$$

- Add combinatorial Benders cuts

$$x = \bar{x} \Rightarrow y = \bar{y}$$

- Strengthen the cuts using domain knowledge

$$x_{|P} = \bar{x}_{|P} \Rightarrow y = \bar{y}$$

Capturing Mode Choice

- Trip inconvenience
  - function of travel time, waiting time, transfers

$$f^r(\mathbf{x^r}, \mathbf{y^r}) = \sum_{h,l \in H} (t_{hl} + S) x^r_{hl} + \sum_{i,j \in N} t_{ij} y^r_{ij}.$$

- Mode choice
  - choosing between personal vehicle or transit
  - thresholding operator

$$\mathcal{C}^r(\mathbf{x^r}, \mathbf{y^r}) \equiv \mathbb{1}(f^r(\mathbf{x^r}, \mathbf{y^r}) \leq \alpha^r \ t^r_{cur}).$$

Solution Method (4)

- Strengthening the combinatorial cuts
- Intuition:
  – if I add arcs to the network, the transit time decreases and the rider keeps adopting transit

$$\sum_{(h,l):\bar{z}_{hl}=0} z_{hl} + \sum_{(h,l):\bar{z}_{hl}=1} (1 - z_{hl}) + \delta^r \geq 1. \quad \longrightarrow \quad \sum_{(h,l):z^1_{hl}=1} (1 - z_{hl}) + \delta^r \geq 1.$$

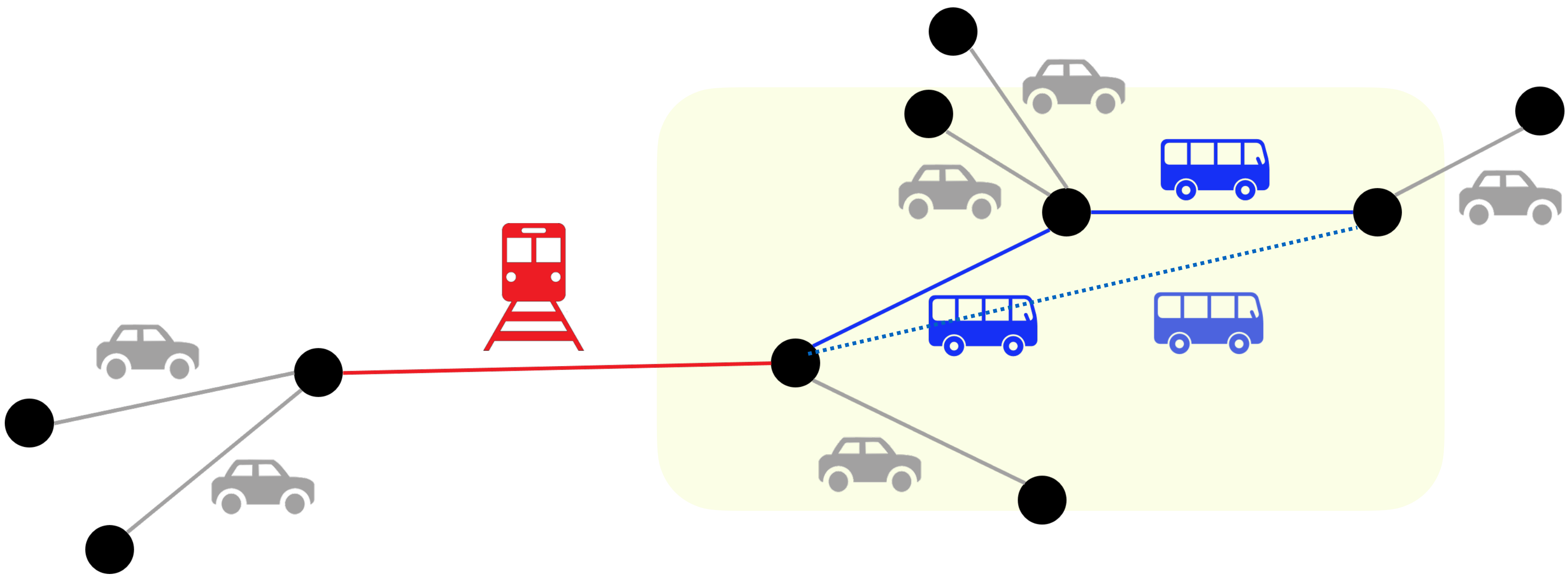
- Intuition:
  – if I remove arcs from the network, the transit time increases and the rider will keep using her vehicle

$$\sum_{(h,l):\bar{z}_{hl}=0} z_{hl} + \sum_{(h,l):\bar{z}_{hl}=1} (1 - z_{hl}) \geq \delta^r. \quad \longrightarrow \quad \sum_{(h,l):z^1_{hl}=0} z_{hl} \geq \delta^r.$$

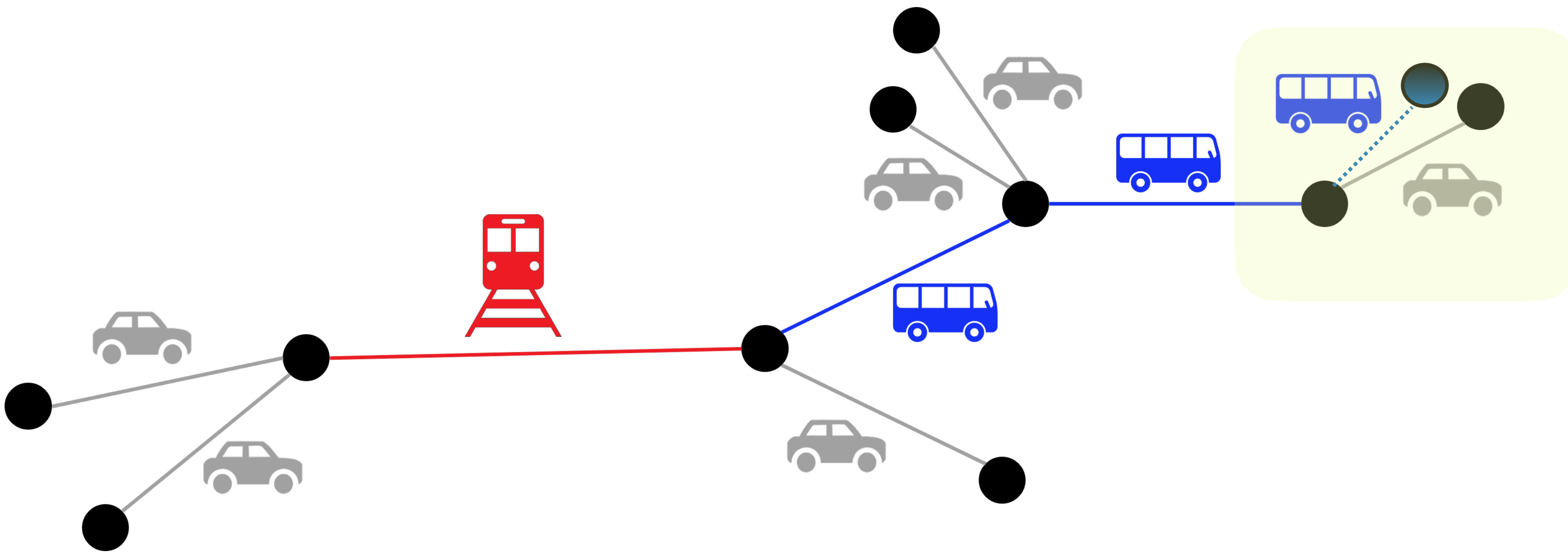
But


YOUR INTUITION
IS WRONG
YOUR STRONGEST
INTUITIONS
ARE OFTEN
WRONG

The Right Intuition

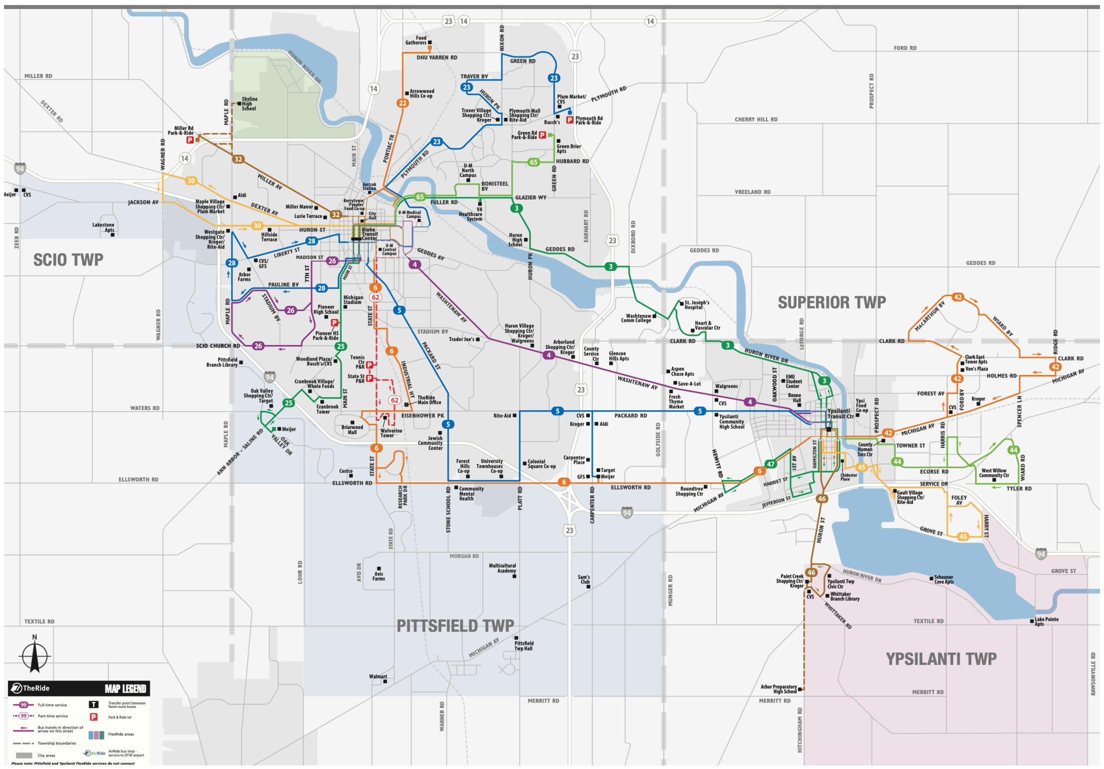


The Wrong Intuition



Experimental Settings

- Ann Arbor / Ypsilanti Transit System

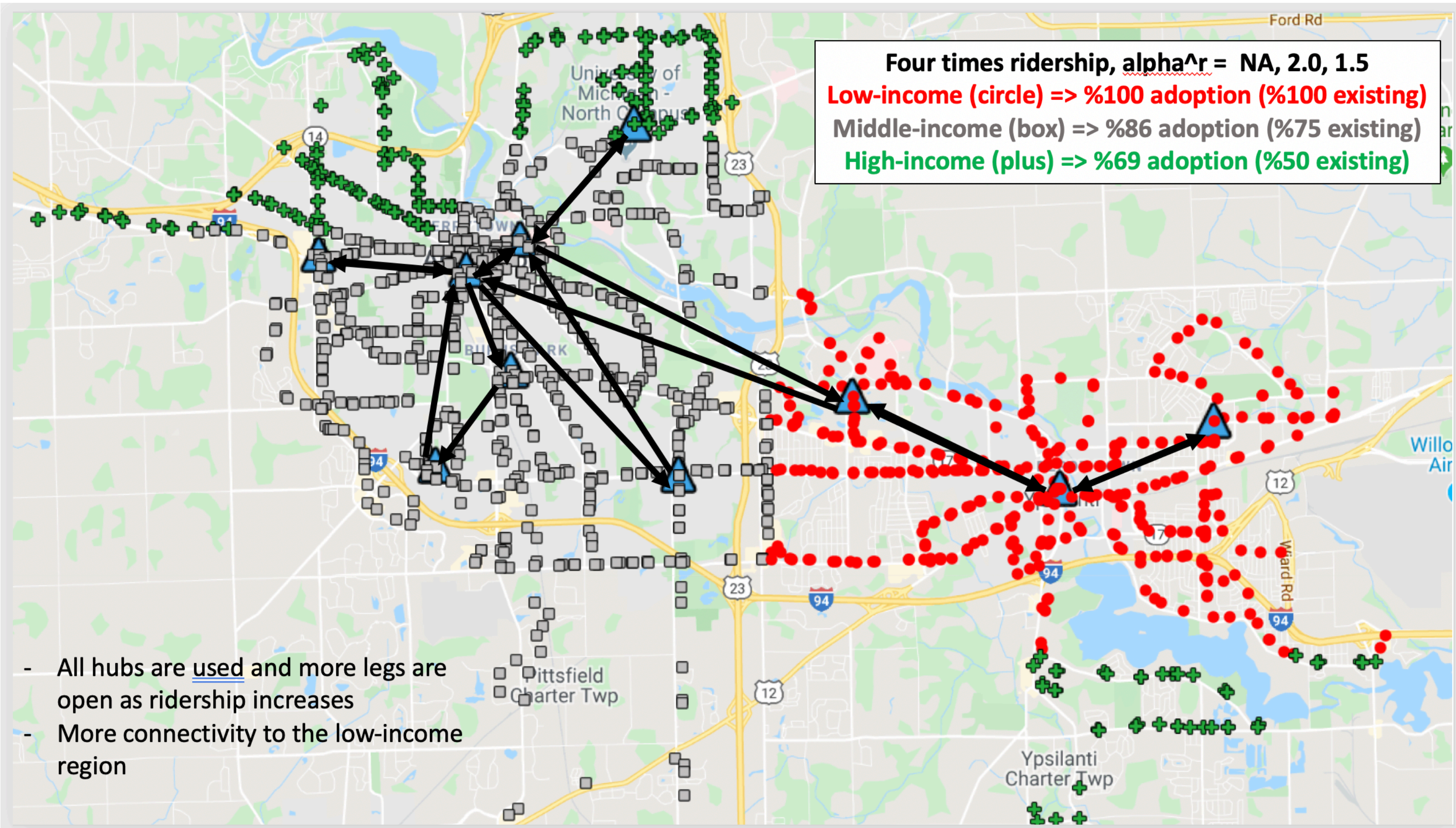


Experimental Settings

Georgia Tech

- Ridership
  - 1,754 low-income, 3,316 middle-income, and 722 high-income riders
  - 100%, 75%, and 50% are existing users
- Mode choice
  - α^r is set to 1.5 and 2.0 for high- and middle-income riders
- Transit system
  - 1,267 bus stops, 6pm-10pm period,
  - average waiting time is 7 minutes
- Prices per mile
  - $5.44 for buses and $1.61 for shuttles

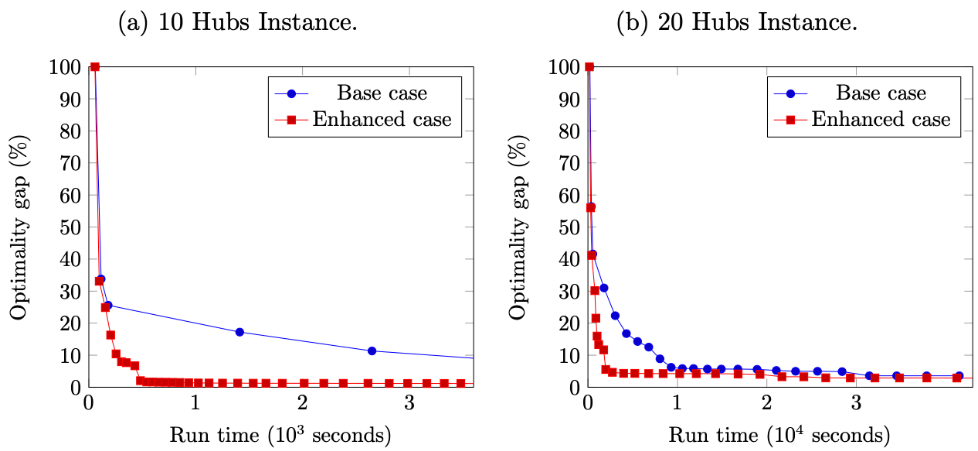
Ann Arbor / Ypsilanti Transit



Cost and Computational Performance

	Adoption			Revenue & Costs			
	MI (%)	HI (%)	# of riders	Revenue	Inv Cost	Trv Cost	NC/rider
10Hub	89	70	5792 (5402)	13505.00	2440.80	13553.31	0.46
10HubISC	85	69	5792 (5326)	13315.00	3564.59	17516.07	1.46
10HubDR	86	69	11584 (10700)	26750.00	4073.14	23847.84	0.11
10HubDRAC	86	69	11584 (10620)	26550.00	4073.14	23642.55	0.11
20HubDR	85	68	11584 (10608)	26520.00	4959.34	20285.19	-0.12

Table 6 Adoption, Cost and Revenue Comparison under Different ODMTS Settings.



Accessibility

	Riders adopting ODMTS			Existing riders			Riders not adopting ODMTS		
Income	ODMTS	direct	AAATA	ODMTS	direct	AAATA	ODMTS	direct	AAATA
low	32.40	11.99	51.50	13.01	5.65	19.07	49.24	10.05	50.46
medium	3.71	3.17	13.69	12.06	5.03	21.53	24.71	7.30	29.31
high	4.53	4.53	14.39	10.09	5.31	21.06	20.85	8.38	30.17

Table 4 **Trip Duration Analysis under 10 Hubs Design with Doubled Ridership and Rider Choices for LILT trips.**

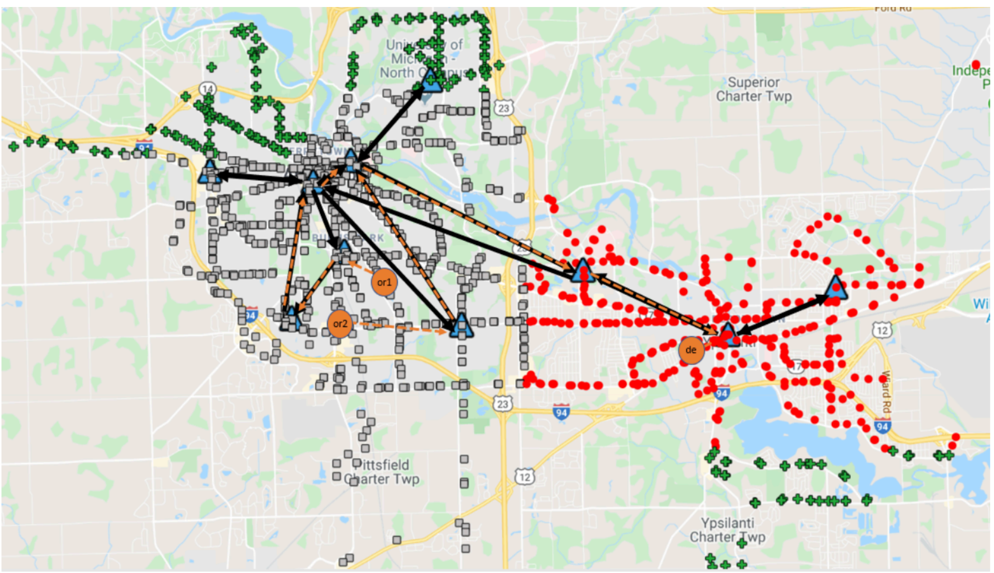


Figure 5 **Visualization of Sample LILT Trips Not Adopting ODMTS.**

Adoption

- low income: 96%
- middle income: 89%
- high income: 70%

Motivation

Learning Optimization Proxies
- ride-haring systems

Optimization over Preferences
- on-demand transit systems

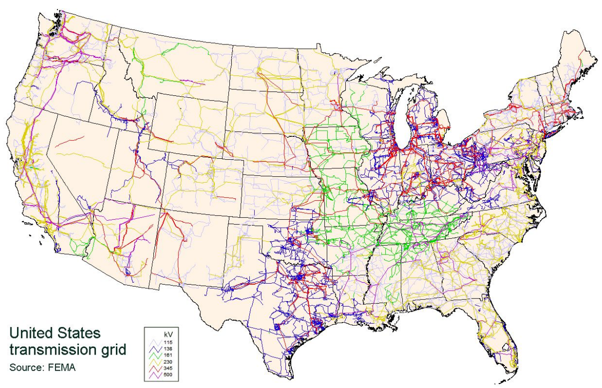
Learning over Physical Systems
- security-constrained OPF

Optimization for Privacy
- releasing census data

Biases, Fairness, Privacy
- releasing census data

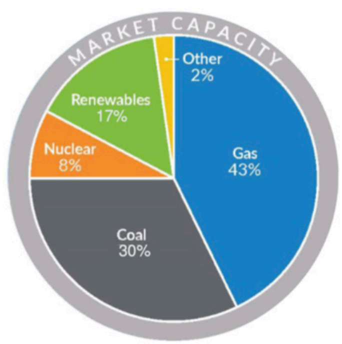
Power Systems

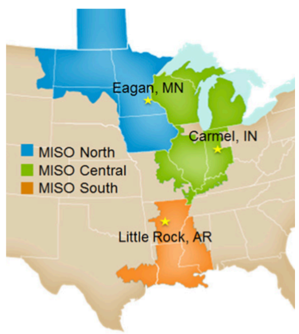
- Largest machine on earth
  - 400 billions of electricity in the US alone

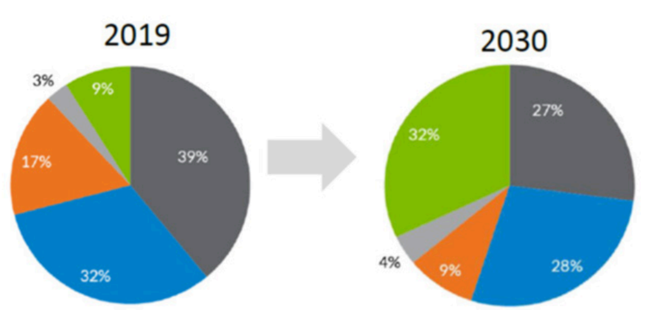


- Huge success story for optimization
  - LP, LR, MIP, Stochastic optimization, NLP, …
- Huge success story for machine learning
  - forecasting

The times, they are changing

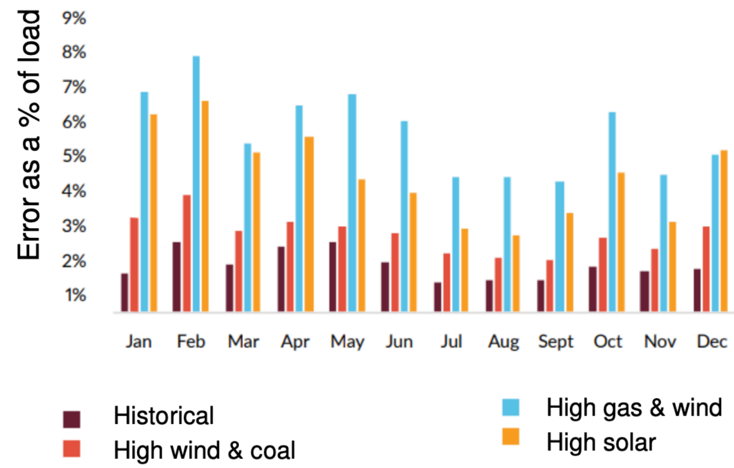






Machine Learning for Power Systems

- The new reality
  - optimization under increasing stochasticity in front and behind the meter
- Opportunity for machine learning
  - networks evolve slowly
  - repeatedly solving similar problems
- Applications
  - market-clearing algorithms
- Challenge for machine learning
  - physical and engineering constraints
  - empirical risk minimization under constraints



Optimal Power Flow

minimize: $\sum_{i \in N} \boldsymbol{c}_{2i}(\Re(S_i^g))^2 + \boldsymbol{c}_{1i}\Re(S_i^g) + \boldsymbol{c}_{0i}$

$$S_i^g - \boldsymbol{S}_i^d = \sum_{(i,j) \in E \cup E^R} S_{ij} \quad \forall i \in N$$

$$S_{ij} = \boldsymbol{Y}_{ij}^* V_i V_i^* - \boldsymbol{Y}_{ij}^* V_i V_j^* \quad (i,j) \in E \cup E^R$$

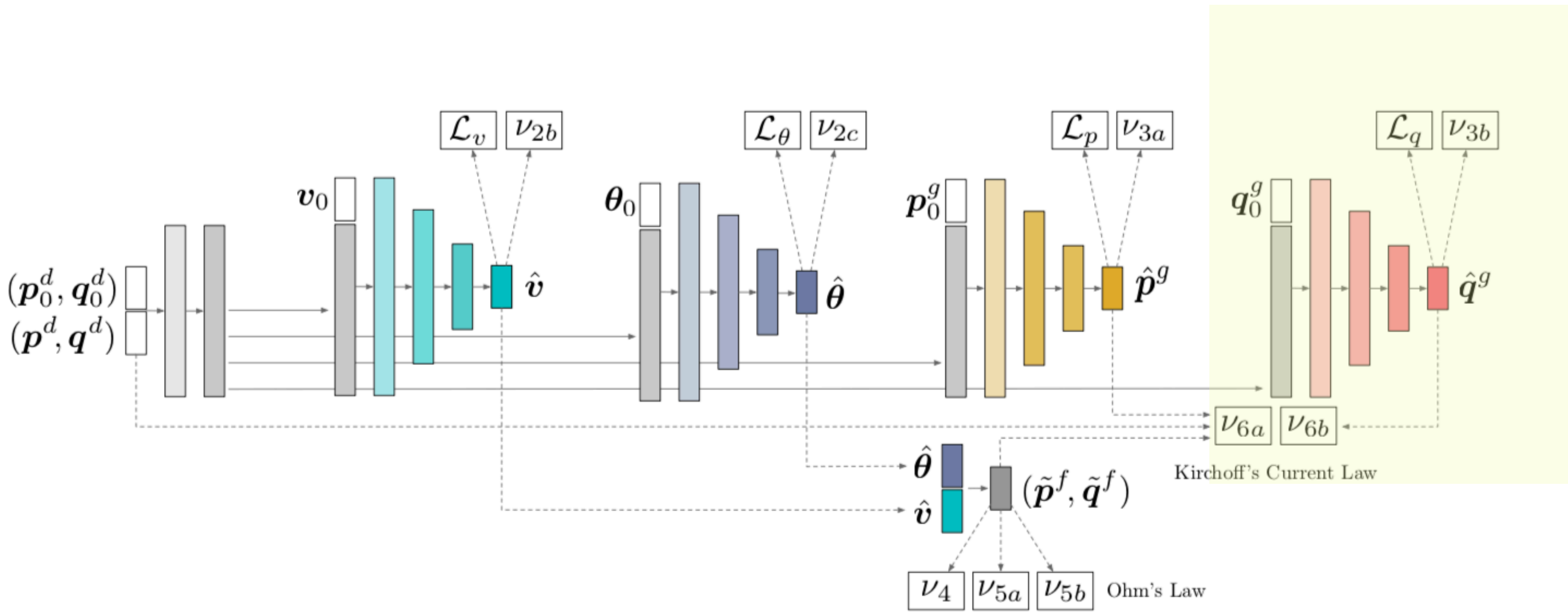
Physics

$$|S_{ij}| \leqslant \boldsymbol{s}_{ij}^{\boldsymbol{u}} \quad \forall (i,j) \in E \cup E^R$$

$$\boldsymbol{S}_i^{\boldsymbol{gl}} \leqslant S_i^g \leqslant \boldsymbol{S}_i^{\boldsymbol{gu}} \quad \forall i \in N \quad \boldsymbol{v}_i^{\boldsymbol{l}} \leqslant |V_i| \leqslant \boldsymbol{v}_i^{\boldsymbol{u}} \quad \forall i \in N$$

Engineering

Constrained Deep Learning



Constrained Deep Learning

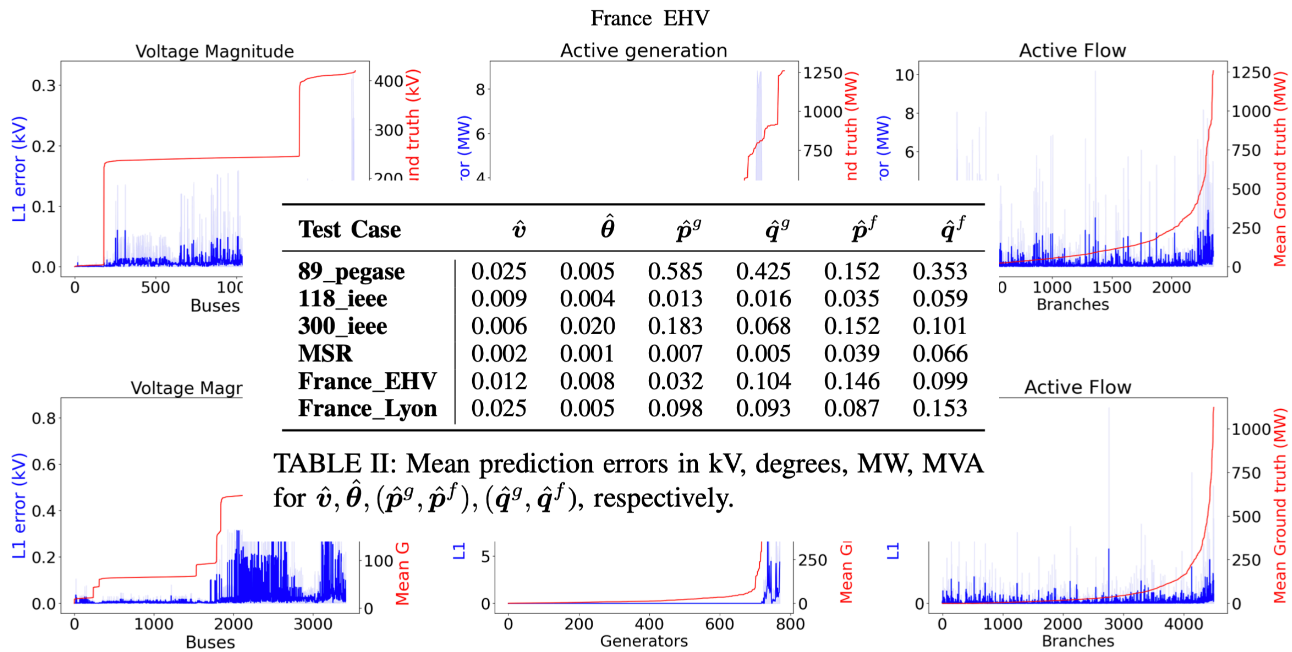
- How to find the optimal prediction?
  - Lagrangian duality
- Primal step
  - empirical loss minimization for given constraint weights
- Dual step
  - to update the constraint weights

$$\boldsymbol{w}^{k+1} = \underset{\boldsymbol{w}}{\operatorname{argmin}}\, \mathcal{L}[\boldsymbol{\lambda}^k](\boldsymbol{x}, \boldsymbol{y}, \hat{\mathcal{O}}[\boldsymbol{w}^k](\boldsymbol{x}))$$

$$\boldsymbol{\lambda}^{k+1} = \left(\lambda_c^k + \rho\, \nu_c(\boldsymbol{x}, \hat{\mathcal{O}}[\boldsymbol{w}^{k+1}](\boldsymbol{x})) \mid c \in \mathcal{C}\right)$$

- Restoring feasibility
  - loaf flow

Experimental Results



Test Case	$\hat{v}$	$\hat{\theta}$	$\hat{p}^g$	$\hat{q}^g$	$\hat{p}^f$	$\hat{q}^f$
89_pegase	0.025	0.005	0.585	0.425	0.152	0.353
118_ieee	0.009	0.004	0.013	0.016	0.035	0.059
300_ieee	0.006	0.020	0.183	0.068	0.152	0.101
MSR	0.002	0.001	0.007	0.005	0.039	0.066
France_EHV	0.012	0.008	0.032	0.104	0.146	0.099
France_Lyon	0.025	0.005	0.098	0.093	0.087	0.153

TABLE II: Mean prediction errors in kV, degrees, MW, MVA for $\hat{v}, \hat{\theta}, (\hat{p}^g, \hat{p}^f), (\hat{q}^g, \hat{q}^f)$, respectively.

Fig. 3: Prediction Errors over Buses, Generators, and Branches. The Ground Truth and Error Graphs are in Different Scales.

Experimental Results

Test Case	$\hat{v}$	$\hat{\theta}$	$\hat{p}^g$	$\hat{q}^g$	$\hat{p}^f$	$\hat{q}^f$
89_pegase	0.025	0.005	0.585	0.425	0.152	0.353
118_ieee	0.009	0.004	0.013	0.016	0.035	0.059
300_ieee	0.006	0.020	0.183	0.068	0.152	0.101
MSR	0.002	0.001	0.007	0.005	0.039	0.066
France_EHV	0.012	0.008	0.032	0.104	0.146	0.099
France_Lyon	0.025	0.005	0.098	0.093	0.087	0.153

TABLE II: Mean prediction errors in kV, degrees, MW, MVA for $\hat{v}, \hat{\theta}, (\hat{p}^g, \hat{p}^f), (\hat{q}^g, \hat{q}^f)$, respectively.

Test Case	Train time (min)	Predict time (sec)	Train Mem. (GB)	AC-OPF (sec)
89_pegase	48	0.0013	1.1	0.2
118_ieee	51	0.0013	1.1	0.2
300_ieee	54	0.0014	1.2	1.9
MSR	59	0.0016	1.5	2.2
France_EHV	142	0.0016	4.6	4.2
France_Lyon	444	0.0020	13.9	47.9

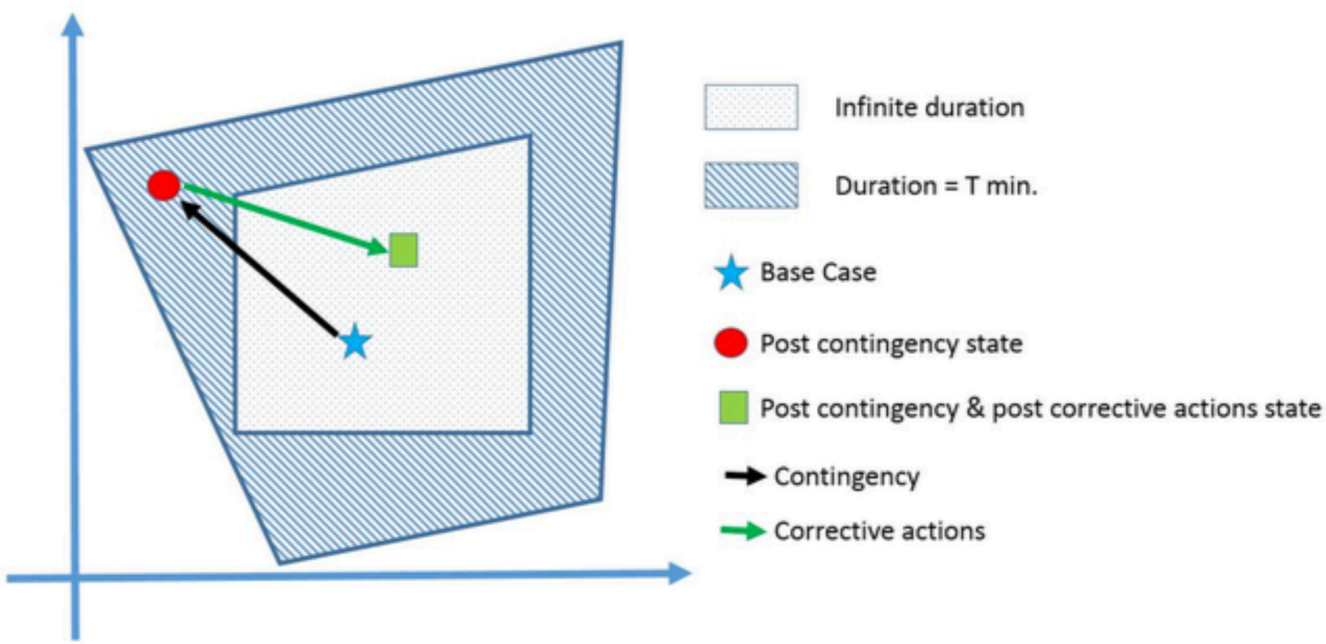
TABLE VI: Training and inference computational costs.

Test Case	(2a)	($3a$)	(3b)	(4)	($6a$)	(6b)
89_pegase	<0.01	4.52	0.45	1.39	0.25	1.01
118_ieee	<0.01	1.83	0.18	-	0.04	0.10
300_ieee	<0.01	0.43	0.20	0.21	0.16	0.17
MSR	<0.01	0.62	0.12	-	0.09	0.15
France_EHV	0.01	2.94	0.79	-	0.24	0.14
France_Lyon	<0.01	1.30	0.50	-	0.16	0.16

TABLE IV: Mean violation for violated AC-OPF constraints. The violation is expressed in kV, MW and MVA for constraints 2a, ($3a$, $6a$) and (3b, 4, 6b), respectively.

Security-Constrained OPF

- Handling contingencies
  - e.g., the loss of a generator



Security-Constrained OPF

- Assumptions
  - N-1 contingencies for generator
- Modeling framework
  - optimize a dispatch
    - pre-contingency state
  - that ensures feasibility after the contingency
    - post-contingency state
- Variables
  - $P^g_{i,0}$: pre-contingency generation at bus i
  - $P^g_{i,1}$: post-contingency generation at bus i for contingency s.

Security-Constrained OPF

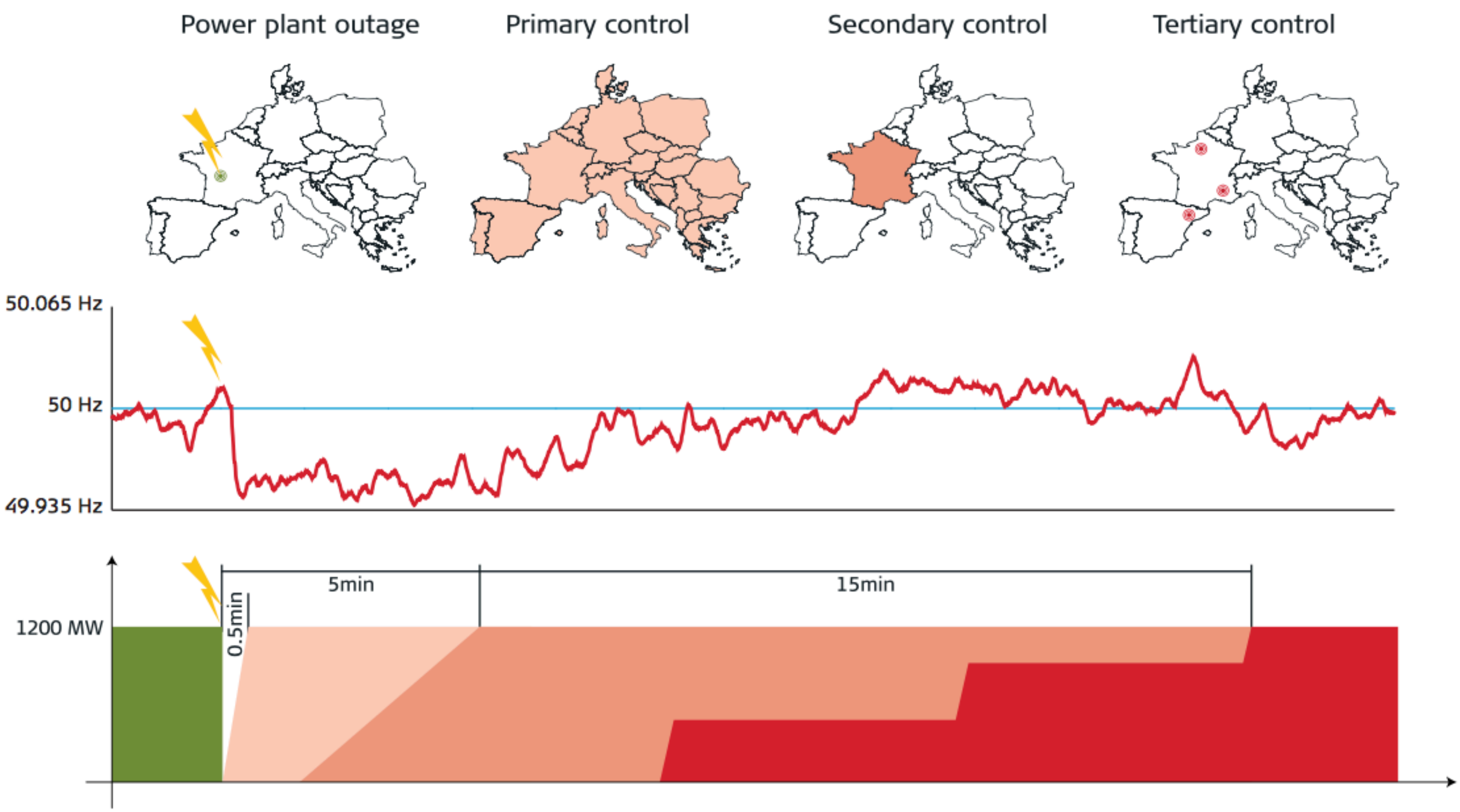
$$
\begin{aligned}
\max \quad & \sum_{i \in G} c_{2,i} \; (p^g_{i,0})^2 + c_{1,i} \; p^g_{i,0} \\
s.t. \quad & PF(\langle p^g_{i,0} \rangle, \langle q^g_{i,0} \rangle) \\
& \forall s \in G \\
& \qquad PF(\langle p^{g,s}_{i,1} \rangle, \langle q^{g,s}_{i,1} \rangle) \\
& \qquad p^{g,s}_{s,1} = 0 \\
& \qquad q^{g,s}_{s,1} = 0
\end{aligned}
$$

Security-Constrained OPF

Security-Constrained OPF

- Power systems involve control systems
  - closed loops
- We do not have full flexibility on the dispatch
  - the control systems will react
- How to capture them?
  - without modeling the dynamics

Security-Constrained OPF



Security-Constrained OPF

- Generation in post-contingency

$$p_{i,1}^{g,s} = p_{i,0}^{g} + \Delta_i^s$$

- Reserve constraints

$$\Delta_i^s \in [0, P_{i,r}^{\max}]$$

- Proportional Response

$$\Delta_i^s = N^s \; pf_i$$

participation factor in MW (from droop or AGC)

$$N^s \in [0, 1]$$

proportional response

Security-Constrained OPF

Security-Constrained OPF

Georgia Tech

- Capturing the limits

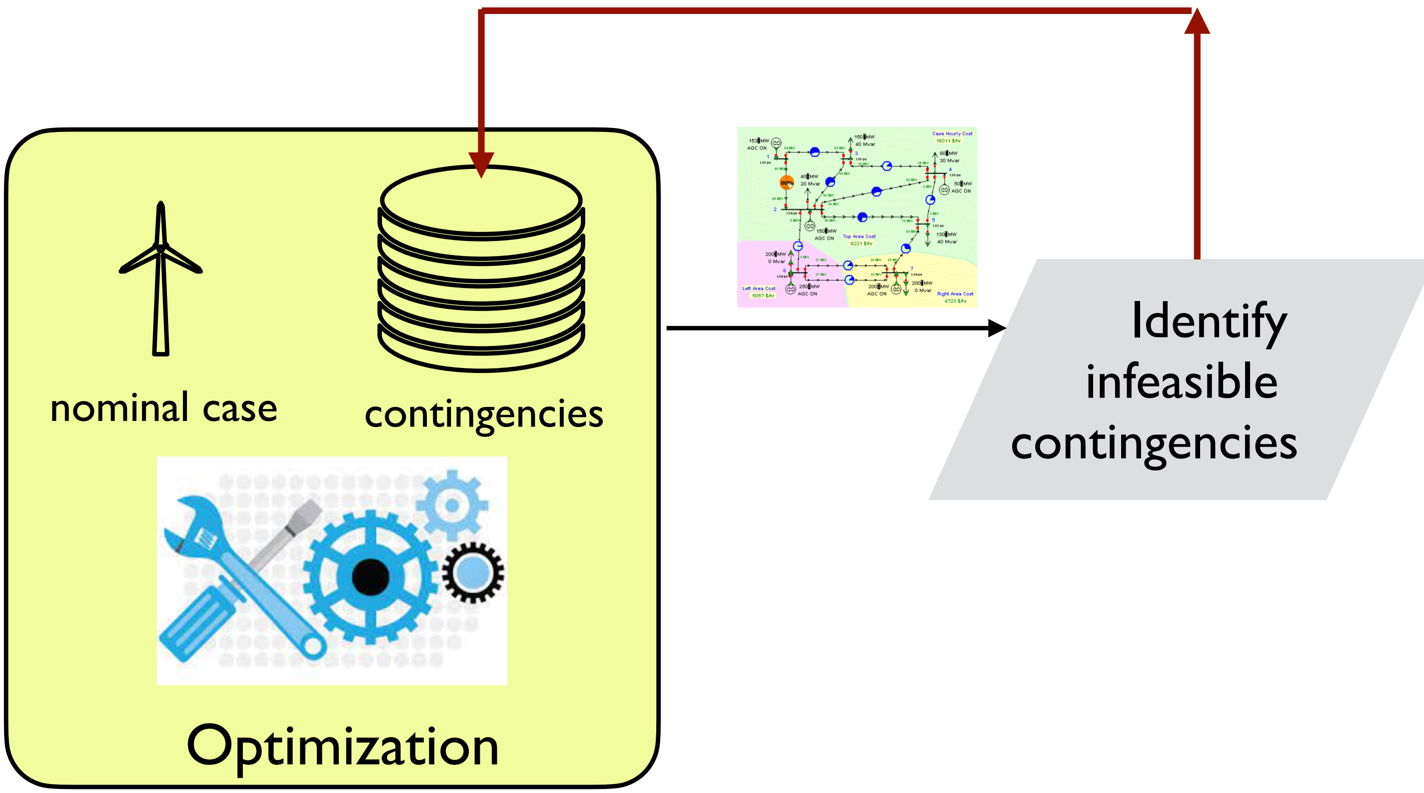
$$\Delta_i^s = \min(N^s \; pf_i, P_{i,r}^{\max})$$

- Important consequences
  - – the proportional response is only applied to generators not at the maximum generation

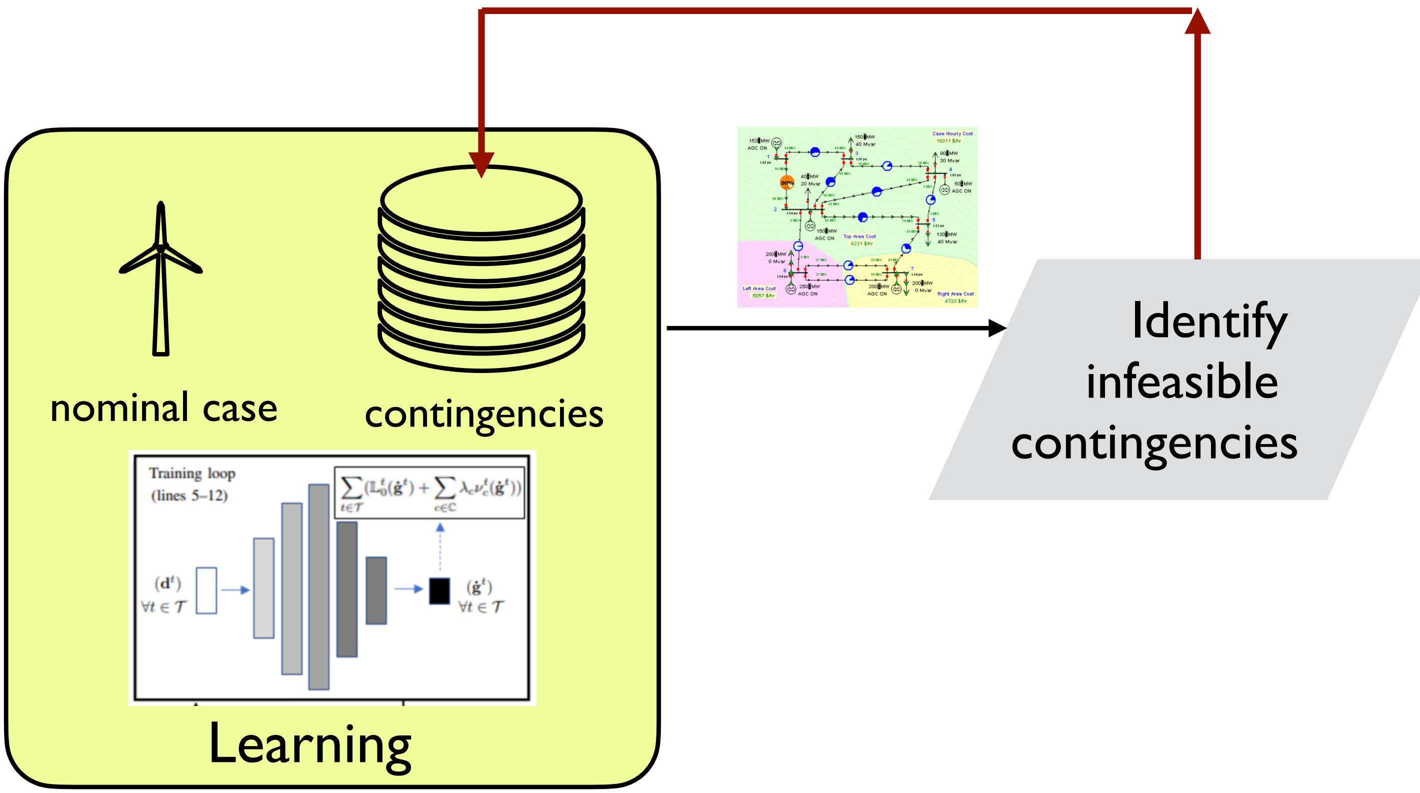
Security-Constrained OPF

$$
\begin{aligned}
\min \quad & \sum_{i \in G} c_{2,i} \, (p^g_{i,0})^2 + c_{1,i} \, p^g_{i,0} \\
s.t. \quad & PF(\langle p^g_{i,0} \rangle, \langle q^g_{i,0} \rangle) \\
& \forall s \in G \\
& \qquad PF(\langle p^{g,s}_{i,1} \rangle, \langle q^{g,s}_{i,1} \rangle) \\
& \qquad p^{g,s}_{s,1} = 0 \\
& \qquad q^{g,s}_{s,1} = 0 \\
& \qquad p^{g,s}_{i,1} = p^g_{i,0} + \Delta^s_i & (i \in G \setminus \{s\}) \\
& \qquad N^s \, pf_i \leq P^{\max}_{i,r} \Rightarrow \Delta^s_i = N^s \, pf_i \;\; \text{else} \;\; \Delta^s_i = P^{\max}_{i,r} & (i \in G \setminus \{s\}) \\
& \qquad \Delta^s_i \in [0, P^{\max}_{i,r}] & (i \in G \setminus \{s\}) \\
& \qquad N^s \in [0, 1] & (i \in G \setminus \{s\})
\end{aligned}
$$

Column and Constraint Generation



Column and Constraint Generation



Constraint and Column Generation Learning

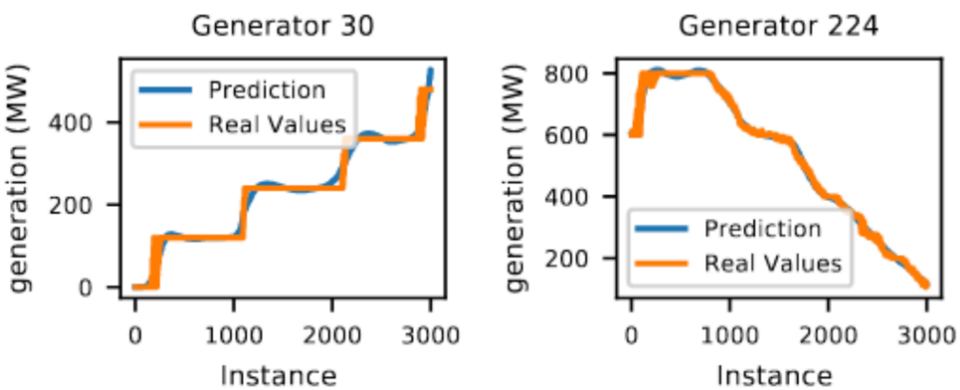


Fig. 3. Prediction of $\mathcal{M}_{ccga}$ for selected generators of the 1354-PEG System.

Constraint and Column Generation Learning

PREDICTION MEAN ABSOLUTE ERRORS (%)

System	Model	Generation Range (MW)						
		10 50	50 100	100 250	250 500	500 1000	1000 2000	2000 5000
118-IEEE	$\mathcal{M}_b$	2.3	2.8	0.7	0.3	N/A	N/A	N/A
	$\mathcal{M}_{ccga}$	2.5	3.0	0.7	0.4	N/A	N/A	N/A
1354-PEG	$\mathcal{M}_b$	2.4	1.3	1.1	0.9	0.4	0.2	0.1
	$\mathcal{M}_{ccga}$	5.0	1.8	1.2	1.0	0.4	0.3	0.2
1888-RTE	$\mathcal{M}_b$	1.3	1.2	0.7	0.4	0.3	0.1	N/A
	$\mathcal{M}_{ccga}$	1.4	1.1	0.6	0.4	0.3	0.1	N/A

Restoring Feasibility

- Load flow problem

$$\min_{\mathbf{g},[\mathbf{g}'_s]_{s\in\mathcal{S}},[\mathbf{x}_s,n_s]_{s\in\mathbb{S}}} ||\dot{\mathbf{g}} - \mathbf{g}||$$

  – find the closest generator set-points that satisfy the constraints

COST INCREASE OVER CCGA (%)

System	Model	Median	Mean	Min.	Max.	Std.
118-IEEE	CCGA-H	0.001	0.001	0.000	0.010	0.002
	FR-$\mathcal{M}_b$	0.021	0.019	-0.073	0.055	0.014
	FR-$\mathcal{M}_{ccga}$	0.027	0.030	-0.010	0.112	0.020
1354-PEG	CCGA-H	0.883	0.880	0.849	0.890	0.010
	FR-$\mathcal{M}_b$	0.020	0.021	-0.007	0.051	0.012
	FR-$\mathcal{M}_{ccga}$	0.067	0.067	0.032	0.091	0.017
1888-RTE	CCGA-H	0.349	0.340	0.323	0.357	0.013
	FR-$\mathcal{M}_b$	0.026	0.024	-0.003	0.070	0.011
	FR-$\mathcal{M}_{ccga}$	0.033	0.033	0.004	0.070	0.013

Constraint and Column Generation Learning

CPU TIME COMPARISON

System	Model	Median	Mean	Min.	Max.	Std.
118-IEEE	CCGA	0.21	0.21	0.10	3.72	0.31
	CCGA-H	0.18	0.20	0.14	3.54	0.24
	FR-$\mathcal{M}_b$	0.02	0.06	0.02	1.58	0.16
	FR-$\mathcal{M}_{ccga}$	0.03	0.07	0.02	1.37	0.17
1354-PEG	CCGA	321.75	327.21	75.59	741.80	127.10
	CCGA-H	5.26	5.32	4.40	8.65	0.49
	FR-$\mathcal{M}_b$	5.34	8.43	1.32	133.37	13.51
	FR-$\mathcal{M}_{ccga}$	1.52	2.17	0.77	8.45	1.74
1888-RTE	CCGA	5.48	7.41	3.11	30.92	7.07
	CCGA-H	3.17	3.15	2.31	13.55	0.91
	FR-$\mathcal{M}_b$	5.32	5.50	1.22	18.07	3.35
	FR-$\mathcal{M}_{ccga}$	2.12	1.95	0.91	4.54	0.92

Motivation

Learning Optimization Proxies
- ride-haring systems

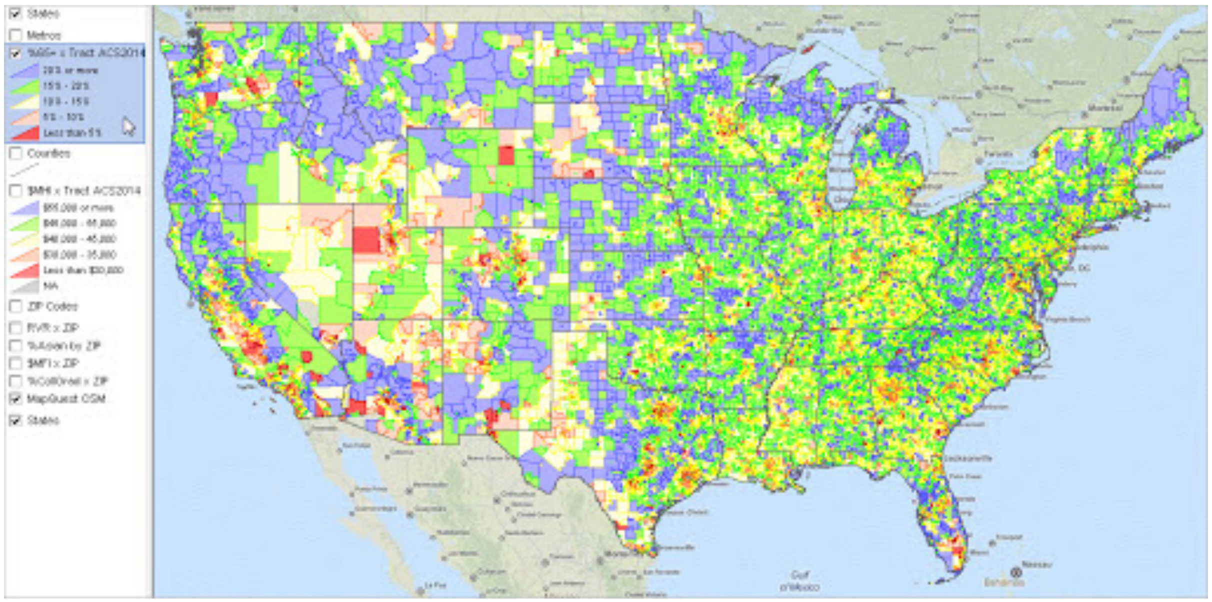
Optimization over Preferences
- on-demand transit systems

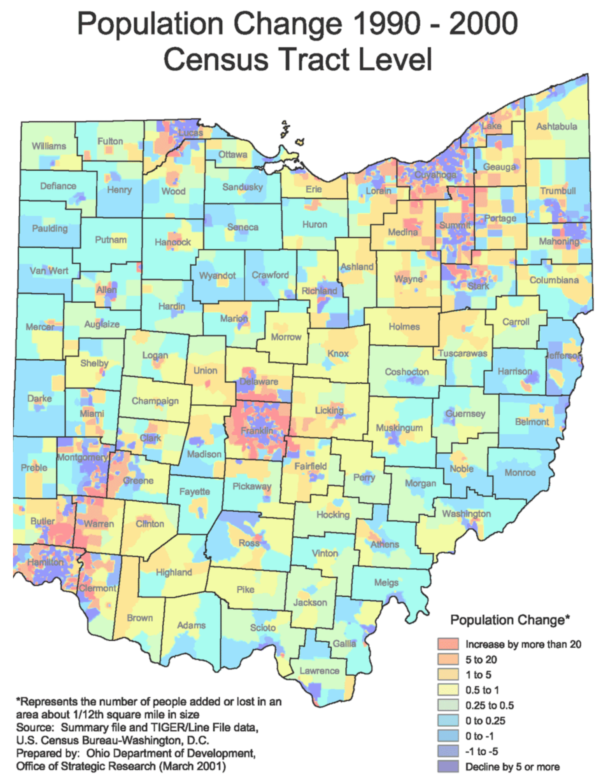
Learning over Physical Systems
- security-constrained OPF

Optimization for Privacy
- releasing census data

Biases, Fairness, Privacy
- releasing census data

Releasing Census Data

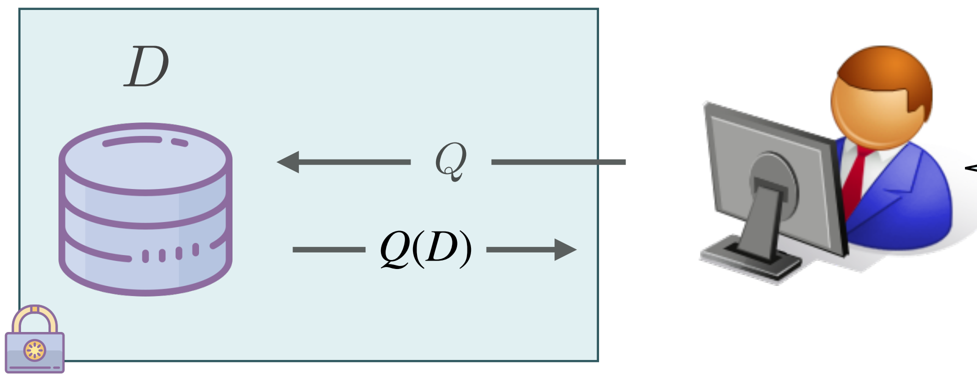




Releasing Census Data

Under Title 13, the Census Bureau cannot release any identifiable information about you, your home, or your business to law enforcement agencies. The law ensures that your private data is protected and that your answers cannot be used against you by any government agency or court.

Releasing Census Data

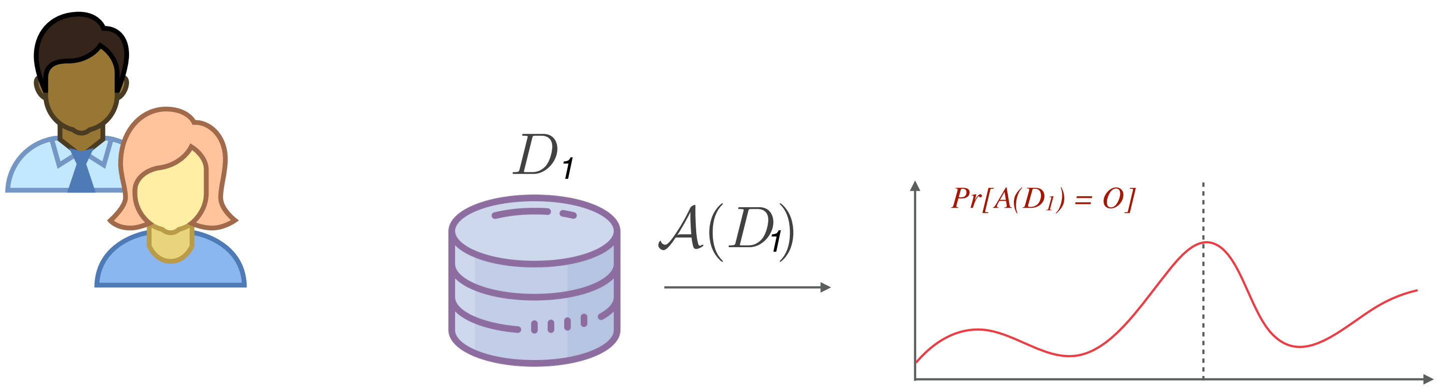


How many people live in a household that earns less than 50,000 dollars

Releasing Census Data

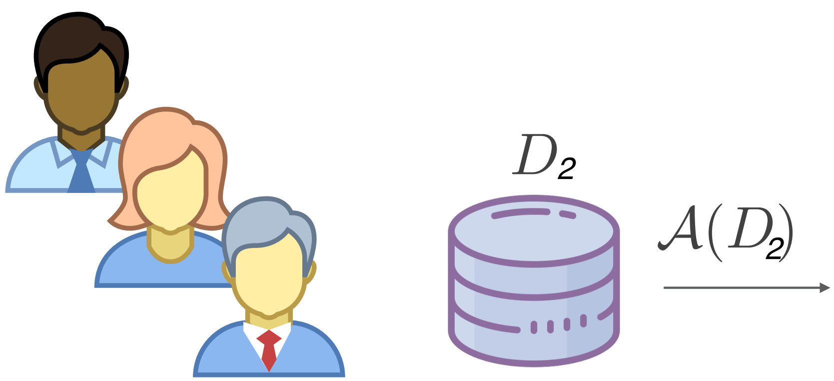
Differential Privacy

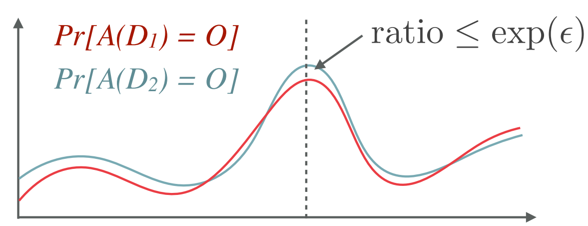
$$\frac{\Pr[\mathcal{A}(D_1) = O]}{\Pr[\mathcal{A}(D_2) = O]} \leq \exp(\epsilon)$$



Differential Privacy

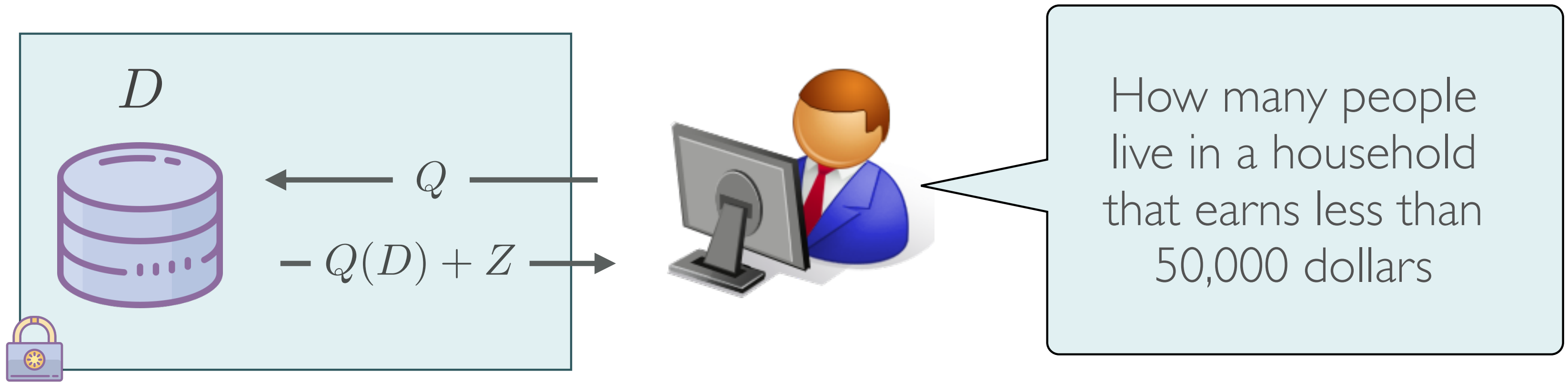
$$\frac{\Pr[\mathcal{A}(D_1) = O]}{\Pr[\mathcal{A}(D_2) = O]} \leq \exp(\epsilon)$$





The risk of a user to join the data set is bounded (by ε)

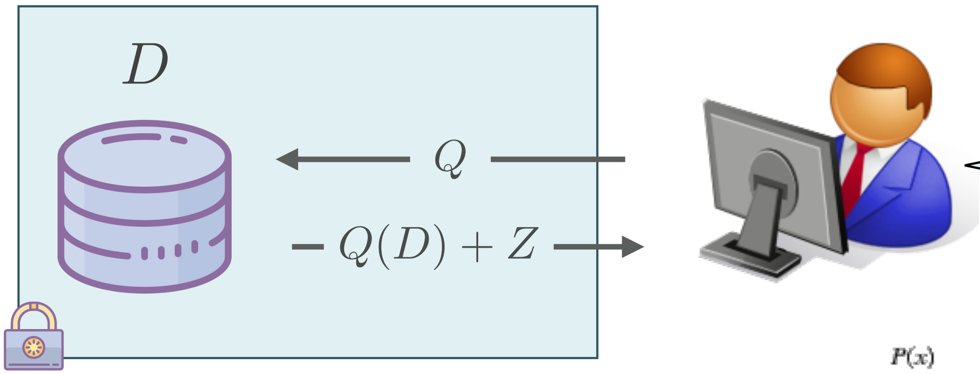
Releasing Census Data



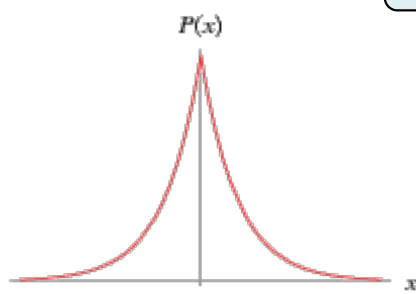
- Query sensitivities

$$\Delta_q = \max_{\mathbf{D}_1 \sim \mathbf{D}_2} \|q(\mathbf{D}_1) - q(\mathbf{D}_2)\|_1$$

Releasing Census Data



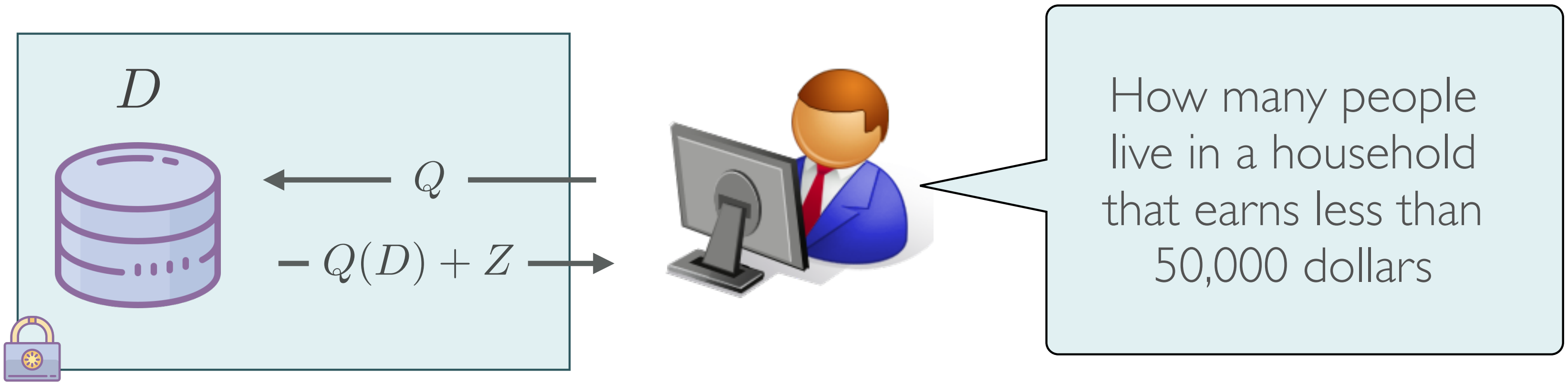
How many people live in a household that earns less than 50,000 dollars



$$Z \sim \mathrm{Lap}(\frac{\Delta_Q}{\epsilon})$$

PDF $f(x \mid \mu = 0, b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$

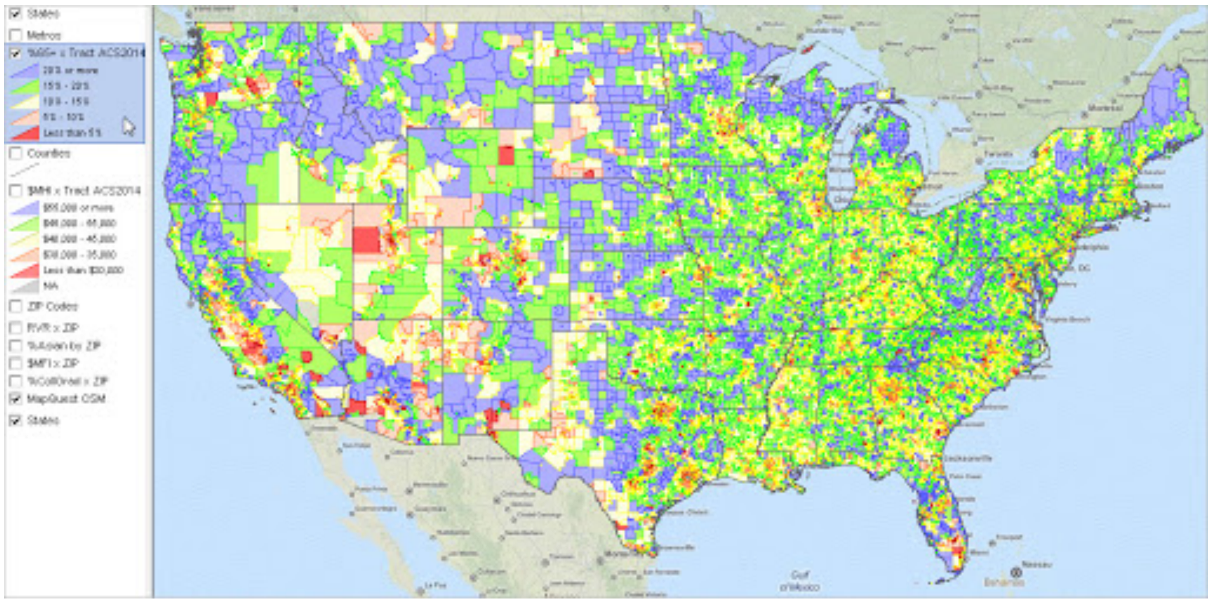
Releasing Census Data



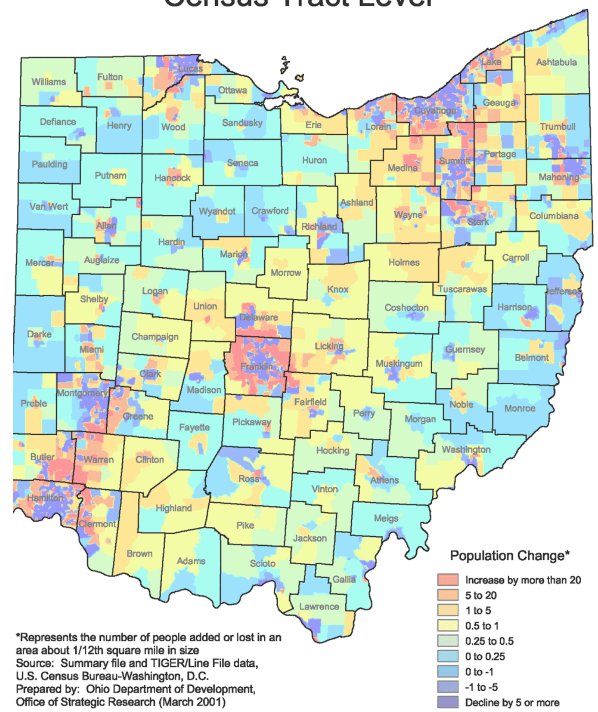
- Geometric mechanism

$$P(X=v) = \frac{1-e^{-\epsilon}}{1+e^{-\epsilon}} e^{\left(-\epsilon \frac{|v|}{\Delta q}\right)}$$

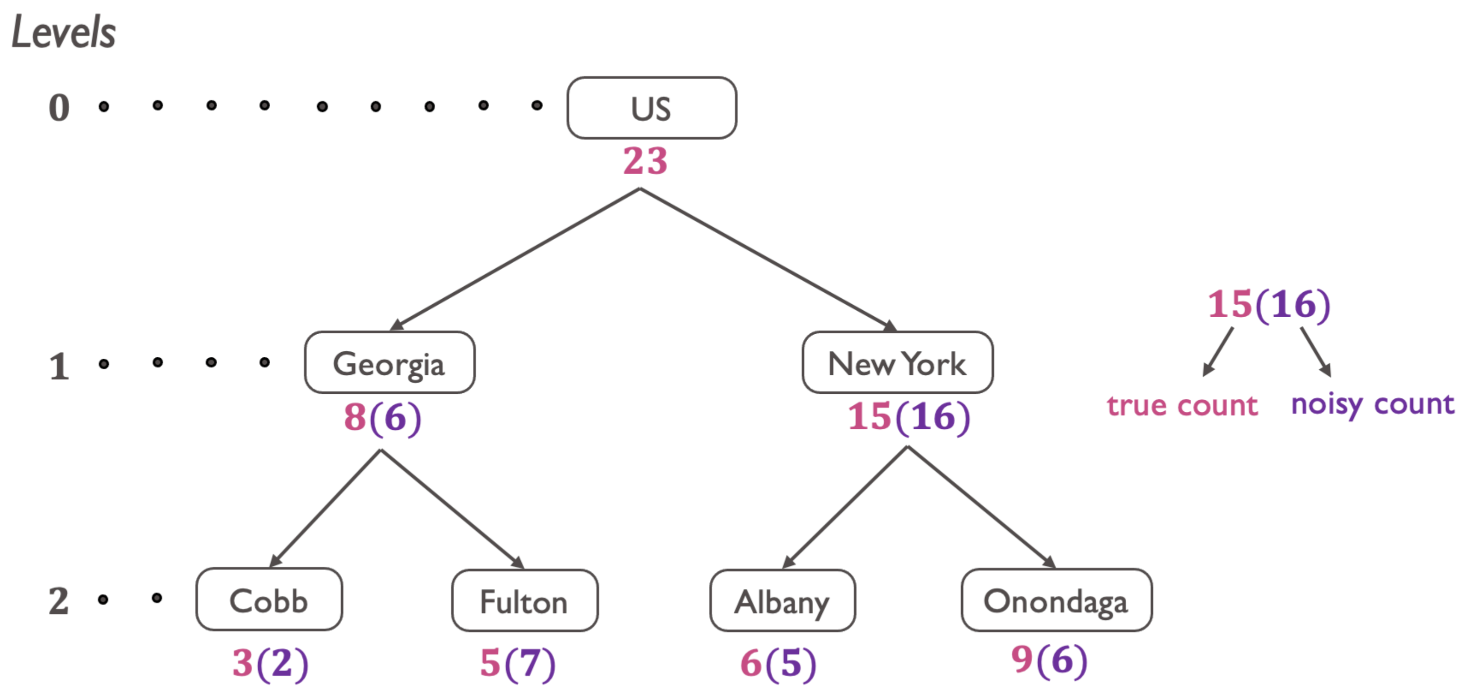
Releasing Census Data



Population Change 1990 - 2000
Census Tract Level



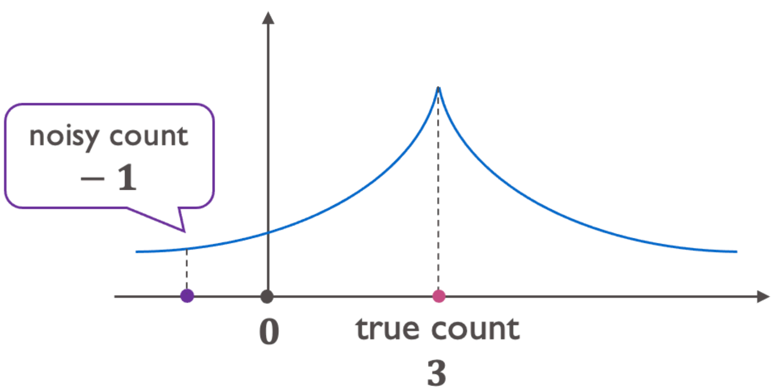
Hierarchical Release



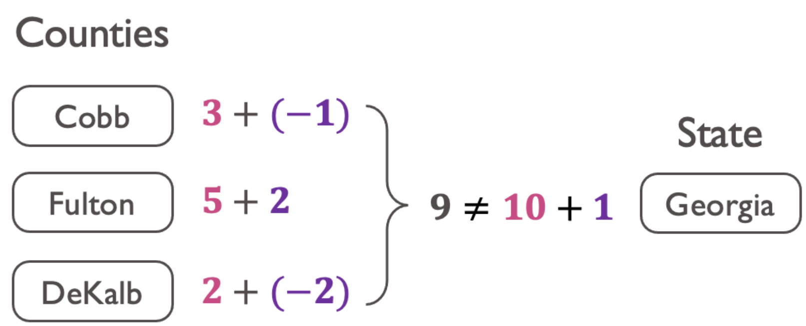
Differential Privacy in Practice

Practical Issues

Non-negativity



Consistency



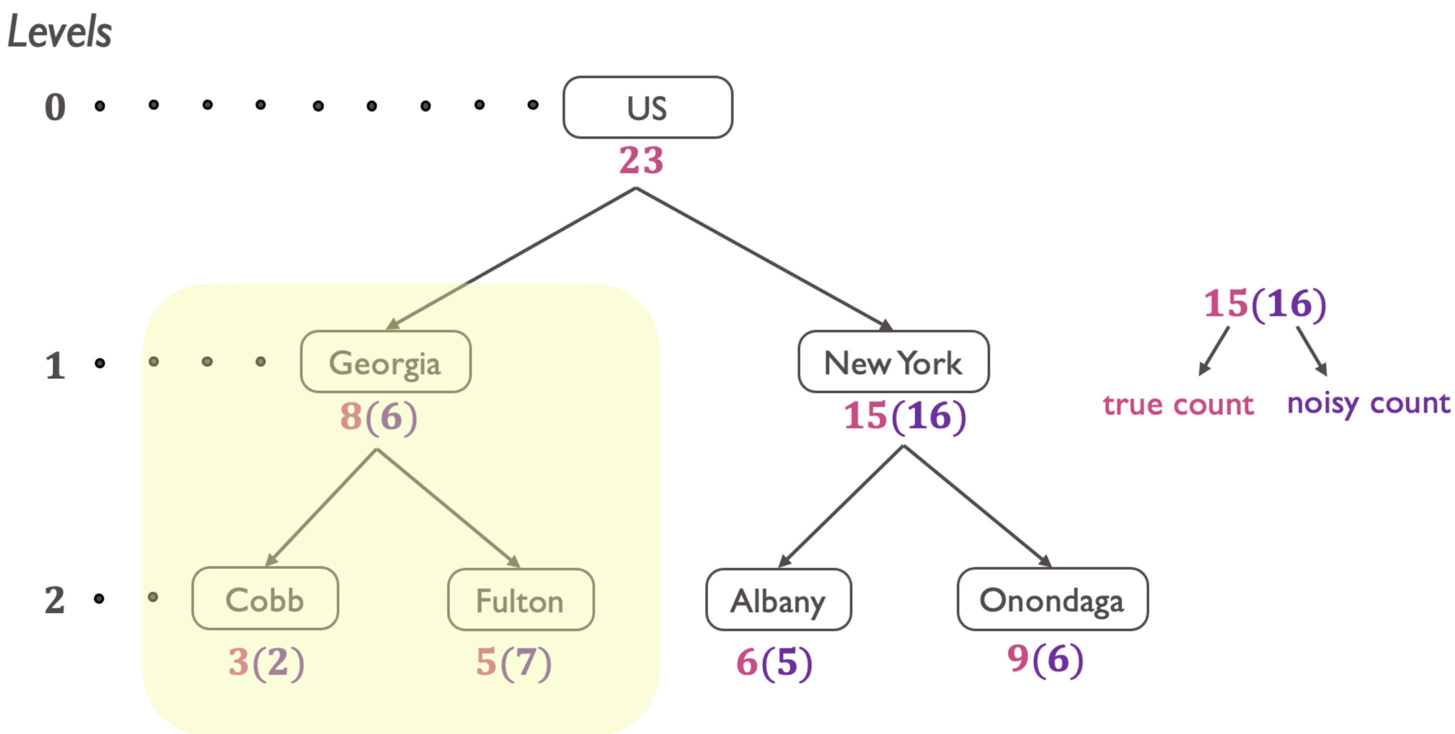
Post-Processing Immunity

Drawback Private query answers with appropriate noise added do not necessarily satisfy domain constraints.

Solution Post-processing Immunity

Private query answers can be transformed using data-independent function to meet domain constraints. Such post-processing step does not hurt privacy.

Hierarchical Release



$$x_{\text{Cobb}} + x_{\text{Fulton}} = 7, \qquad x_{\text{Cobb}} \geq 0, \qquad x_{\text{Fulton}} \geq 0.$$

Post-Processing Immunity

Model 1 The $\mathcal{M}_H$ post-processing step.

$$\begin{aligned}
\underset{\{\hat{\boldsymbol{n}}^r\}_{r\in\mathbf{R}}}{\text{Minimize}} \quad & \sum_{r\in\mathbf{R}} \|\hat{\boldsymbol{n}}^r - \tilde{\boldsymbol{n}}^r\|_2^2 \\
\text{Subject to:} \quad & \sum_{s\in[N]} \hat{n}_s^r = G && \forall r \in \mathbf{R} \\
& \sum_{c\in ch(r)} \hat{n}_s^c = \hat{n}_s^r && \forall r \in \mathbf{R}, s \in [N] \\
& \hat{n}_s^r \in D_s^r && \forall r \in \mathbf{R}, s \in [N]
\end{aligned}$$

Post-Processing Immunity

Georgia Tech

Scalability

a million dollar question

Post-Processing: Dynamic programming

- bottom-up phase
  - computes cost tables to find the optimal cost
- top-down phase
  - recover the optimal values for the optimal cost

Dynamic Programming: Bottom-Up

- Optimal cost at node r if its value is v

$$\boldsymbol{\tau}^{r}(v) = \left(v - \tilde{n}^{r}\right)^{2} +$$

$$\boldsymbol{\phi}^{r}(v) = \underset{\{x_c\}_{c \in ch(r)}}{\text{Minimize}} \sum_{c \in ch(r)} \boldsymbol{\tau}^{c}(x_c)$$

$$\text{Subject to:} \quad \sum_{c \in ch(r)} x_c = v$$

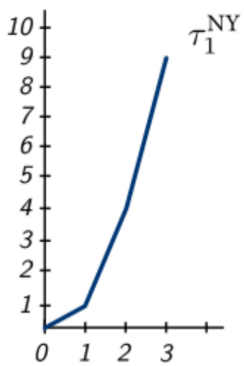
$$x_c \in D^c \qquad \forall c \in ch(r).$$

Dynamic Programming

- Optimal cost at node r if its value is v
  - the cost functions are convex piecewise linear

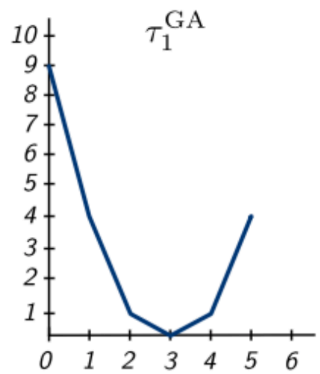
τ_1^{NY}

j	cost
0	0
1	1
2	4
3	9



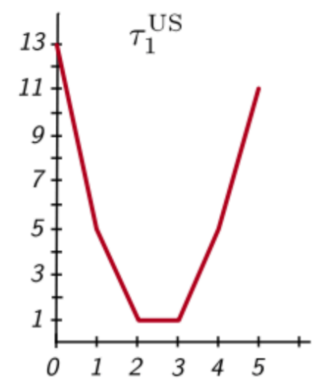
τ_1^{GA}

j	cost
0	9
1	4
2	1
3	0
4	1
5	4



τ_1^{US}

j	cost
0	$4+(0+0)$
1	$1+(0+4)$
2	$0+(0+1)$
3	$1+(0+0)$
4	$4+(1+0)$
5	$9+(1+1)$

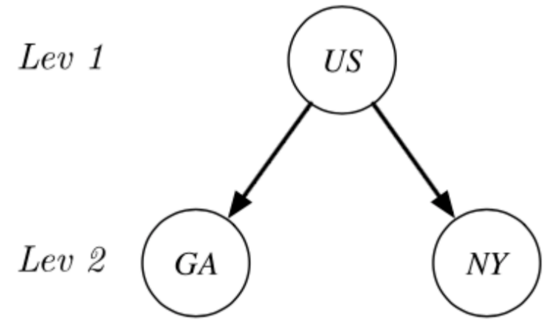


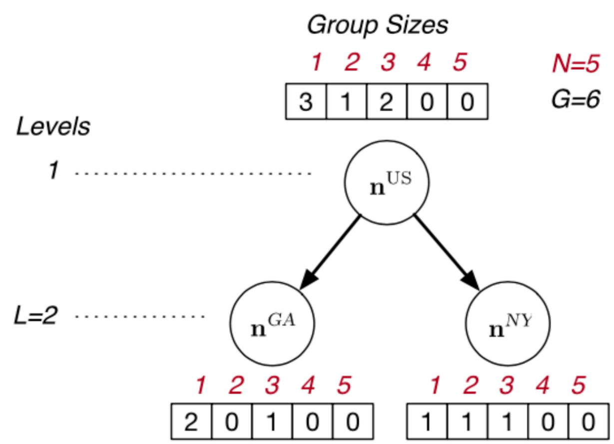
- Theorem: the cost table can be computed in O(|D| log |D|) time

The Census Bureau Release

region	u	group G_u	σ_u
GA	A	{01, 03, 04}	3
	B	{02}	1
	C	{05}	1
NY	D	{06, 08, 09}	3
	E	{07}	1
	F	{10, 11}	2

Group sizes	Lev 2		Lev 1
	GA	NY	US
1	2	1	3
2	0	1	1
3	1	1	2
4	0	0	0
$5 = \lvert\mathcal{S}\rvert = N$	0	0	0
	$\boldsymbol{n}^{GA}$	$\boldsymbol{n}^{NY}$	$\boldsymbol{n}^{US}$





Sensitivities

- Counts and sensitivities

Group	0	1	2	3
counts	4	3	2	5

- What happens if we remove one person?

Group	0	1	2	3
counts	4	4	1	5

Cumulative Counts

- Counts

Group	0	1	2	3
counts	4	3	2	5

- Cumulative counts

Group	0	1	2	3
counts	4	7	9	14

Sensitivities

Georgia Tech

- Cumulative Counts

Group	0	1	2	3
counts	4	7	9	14

- What happens if we remove one person?

Group	0	1	2	3
counts	4	8	9	14

Post-Processing Cumulative Counts

Model 2 The $\mathcal{M}_c$ post-processing step.

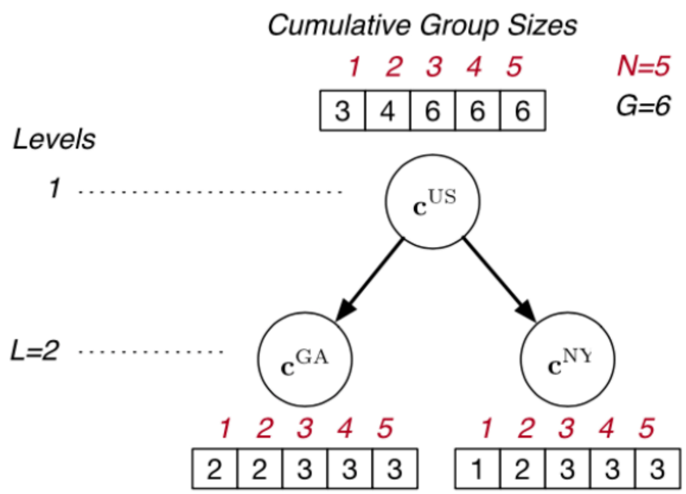
$$\underset{\{\hat{\mathbf{c}}^r\}_{r \in \mathbf{R}}}{\text{Minimize}} \sum_{r \in \mathbf{R}} \|\hat{\mathbf{c}}^r - \tilde{\mathbf{c}}^r\|_2^2$$

Subject to: $\hat{c}_N^\top = G$

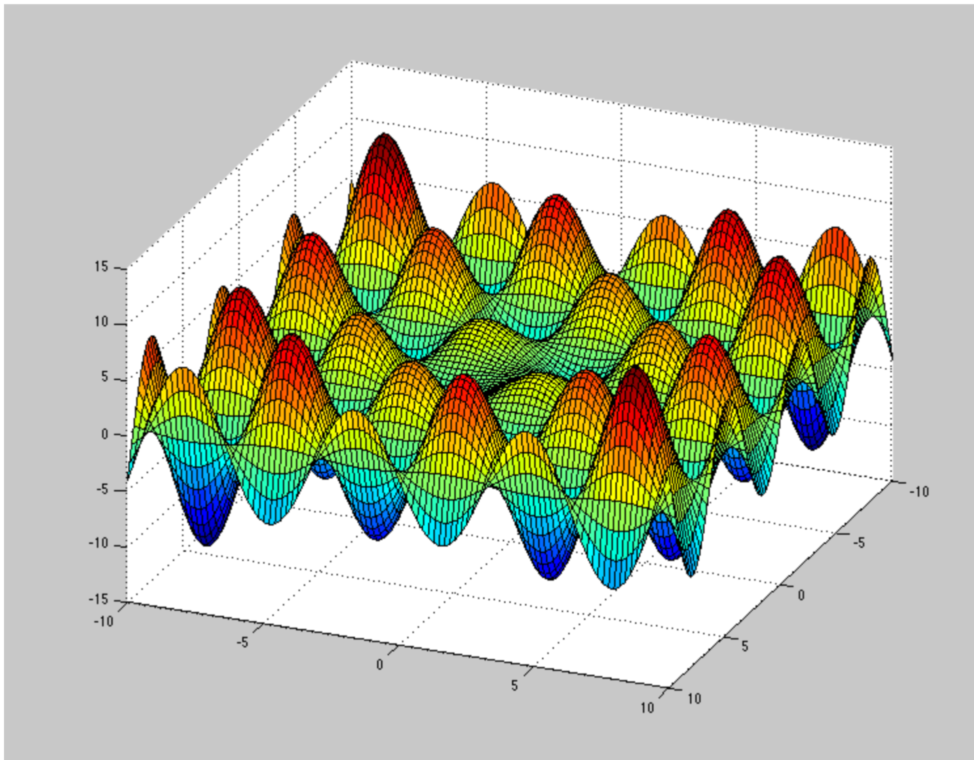
$$\hat{c}_i^r \leqslant \hat{c}_{i+1}^r \qquad \forall r \in \mathbf{R}, i \in [N-1]$$

$$\sum_{r' \in ch(r)} \hat{c}_i^{r'} = \hat{c}_i^r \qquad \forall r \in \mathbf{R}, i \in [N]$$

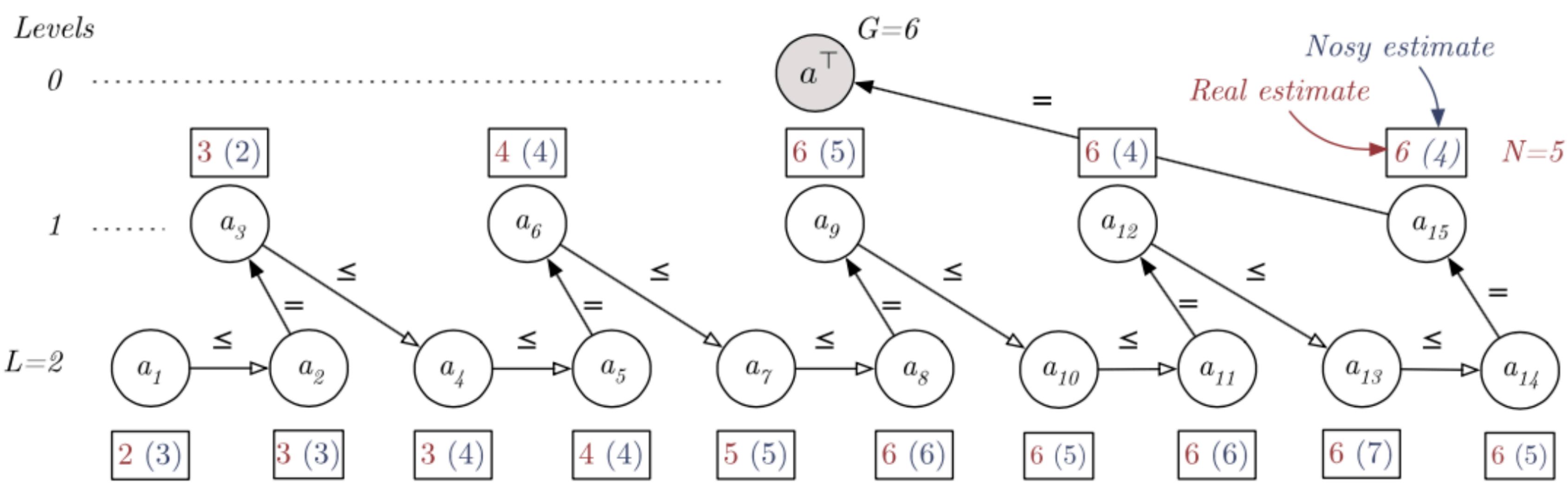
$$\hat{c}_i^r \in \{0, 1, \ldots\} \qquad \forall r \in \mathbf{R}, i \in [N]$$



Post-Processing Cumulative Counts



Flatten the Hierarchy!



Post-Processing Cumulative Counts

Model 3 The modified post-processing step.

Let $\dot{N} = |\mathbf{R}|\, N$

$$\underset{\{\hat{c}_i\}_{i\in[\dot{N}]}}{\text{Minimize}} \quad \sum_{i=1}^{\dot{N}} \|\hat{c}_i - \tilde{c}_i\|_2^2$$

$$\text{Subject to: } \hat{c}_{\dot{N}} = G$$

$$\hat{c}_i \leqslant \hat{c}_{i+1} \qquad \forall i \in \left\{ i \in [\dot{N}-1] \mid \text{lev}\,(a_i) = \text{lev}\,(a_{i+1}) \right\}$$

$$\hat{c}_i = \hat{c}_{i+1} \qquad \forall i \in \left\{ i \in [\dot{N}-1] \mid \text{lev}\,(a_i) > \text{lev}\,(a_{i+1}) \right\}$$

$$\hat{c}_i \in \{0, 1, \ldots\} \qquad \forall i \in [\dot{N}]$$

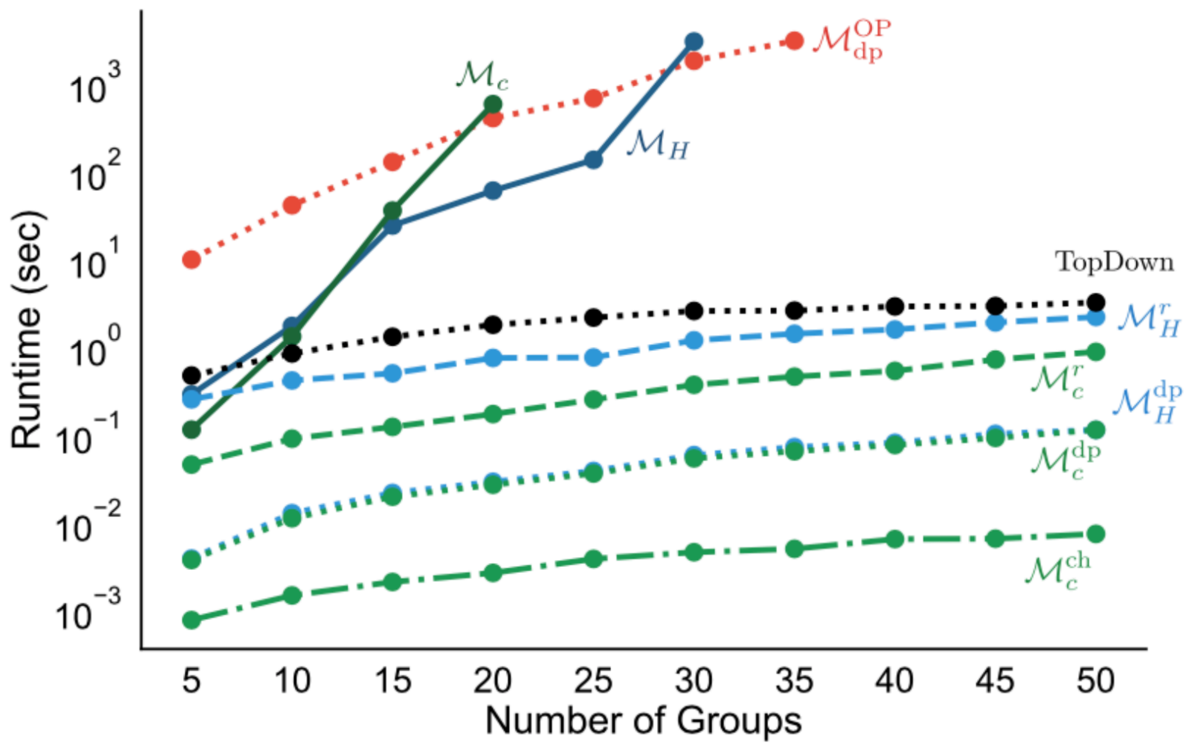
Back to Dynamic Programming

- Optimal cost at node r if its value is v

$$\boldsymbol{\tau}_{i+1}(v) = \left(v - \tilde{c}_{i+1}\right)^2 + \boldsymbol{\phi}_{i+1}(v) = \begin{cases} \boldsymbol{\tau}_i(v) & \text{if } \operatorname{lev}(a_i) > \operatorname{lev}(a_{i+1}), \\ \underset{\substack{x_i \in D_i \\ x_i \leqslant v}}{\text{Minimize}}\ \boldsymbol{\tau}_i(x_i) & \text{otherwise.} \end{cases}$$

- Theorem: the cost table can be computed in O(IDI) time
  – the overall algorithm computes a table per node

Experimental Results



Experimental Results

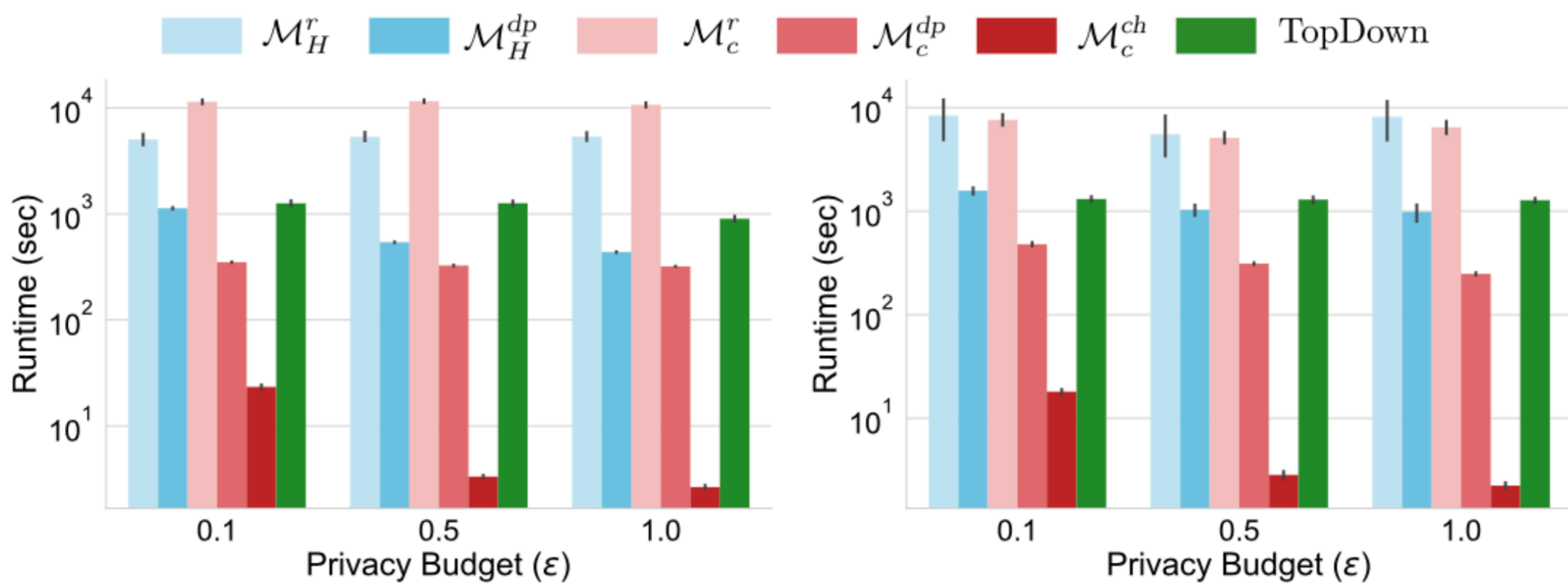


Fig. 8. The Runtime for the mechanisms: Census data (left) and Taxi data (right).

The first dataset has 117,630,445 groups, 7,592 leaves, 305,276,358 individuals, 3 levels, and N=1,000.

The second dataset has 13,282 groups, 3,973 leaves, 24,489,743 individuals, 3 levels, and N=13,282.

Experimental Results

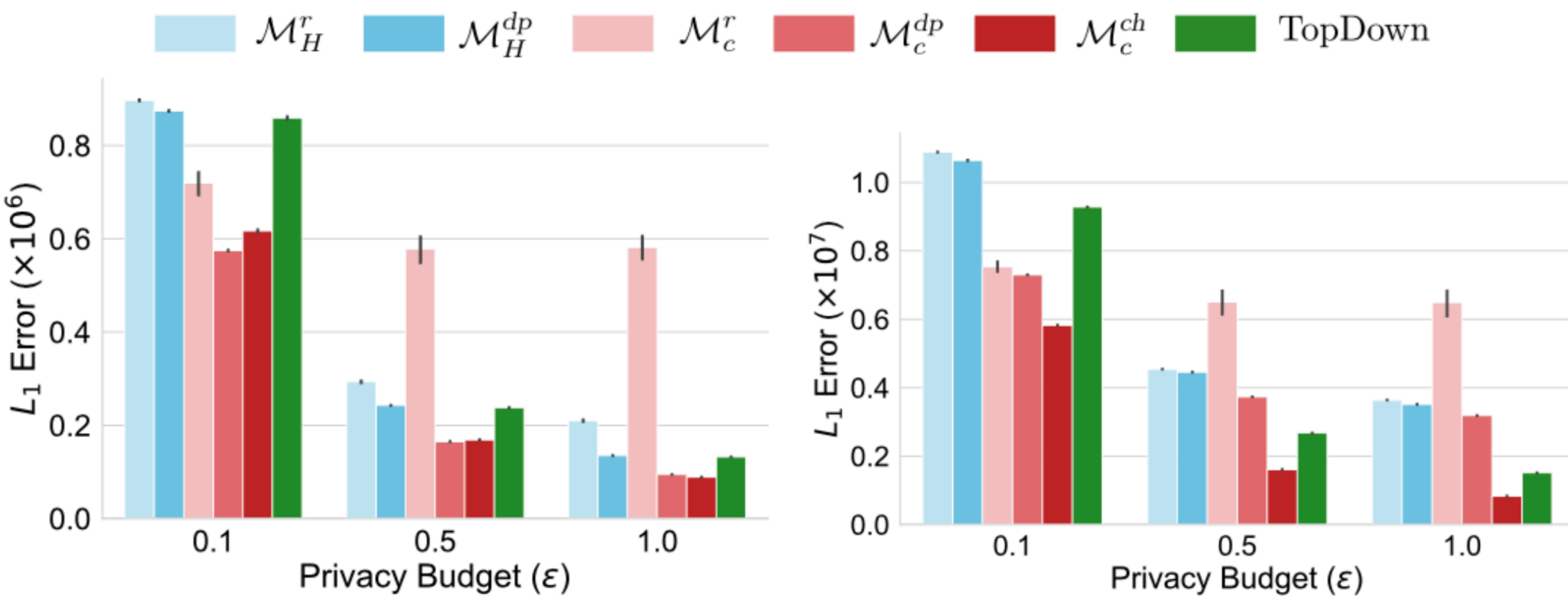


Fig. 9. The L_1 errors for the algorithms: Census data (left) and Taxi data (right).

Motivation

Learning Optimization Proxies
- ride-haring systems

Optimization over Preferences
- on-demand transit systems

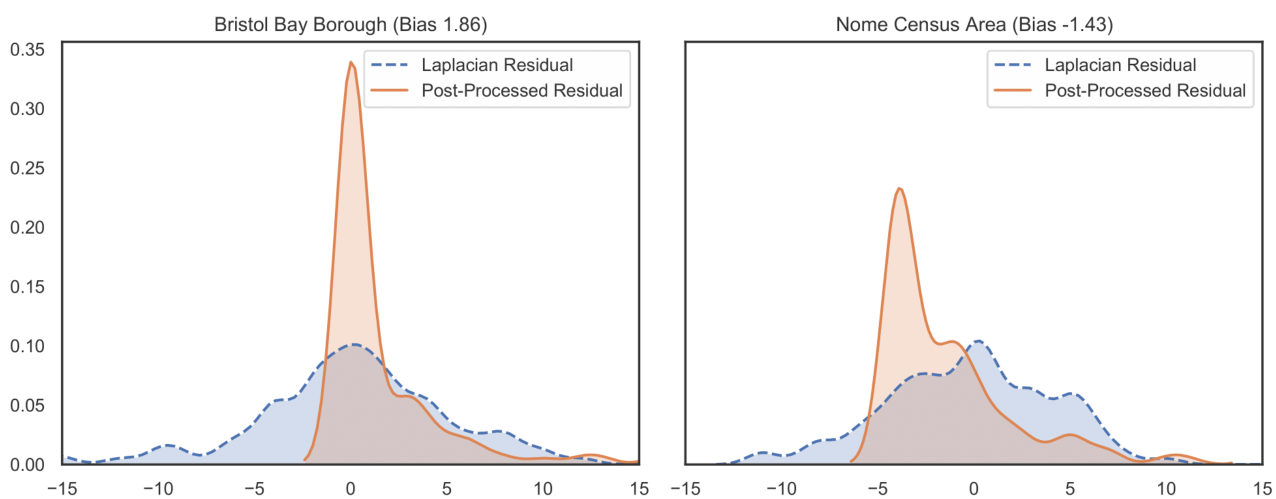
Learning over Physical Systems
- security-constrained OPF

Optimization for Privacy
- releasing census data

Biases, Fairness, Privacy
- releasing census data

Bias on Census Release

- Post-processing introduces significant bias
  - E_x˜ [p(x˜)]-x.



Studying Biases

- True (original) data denoted by

$$\boldsymbol{x} \in \mathbb{R}^n$$

- Noisy data produced by Laplace mechanism

$$\widetilde{\boldsymbol{x}} \sim \boldsymbol{x} + \mathrm{Lap}(\lambda)^n$$

- m equality constraints characterized by

$$A \in \mathbb{R}^{m \times n}, \boldsymbol{b} \in \mathbb{R}^m$$

Post-Processing

$$\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}}) \triangleq \underset{\boldsymbol{v}\in\mathbb{R}^n}{\operatorname{argmax}} \quad \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \qquad (P)$$
$$\text{s.t.} \quad A \cdot \boldsymbol{v} = \boldsymbol{b}$$

$$\text{Bias}[\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})] = \mathbb{E}_{\tilde{\boldsymbol{x}}}[\hat{\boldsymbol{x}}(\tilde{\boldsymbol{x}})] - \boldsymbol{x},$$

$$\text{Bias}[\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})] = \mathbb{E}_{\tilde{\boldsymbol{x}}}[\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}})] - \boldsymbol{x}$$

$$\hat{\boldsymbol{x}}_+(\tilde{\boldsymbol{x}}) \triangleq \underset{\boldsymbol{v}\in\mathbb{R}^n}{\operatorname{argmax}} \quad \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \qquad (P_+)$$
$$\text{s.t.} \quad A \cdot \boldsymbol{v} = \boldsymbol{b}$$
$$\boldsymbol{v} \geq \boldsymbol{0}$$

Post-Processing

- Problem P has no bias
  - all the bias comes the non-negativity constraints
- Consider problem

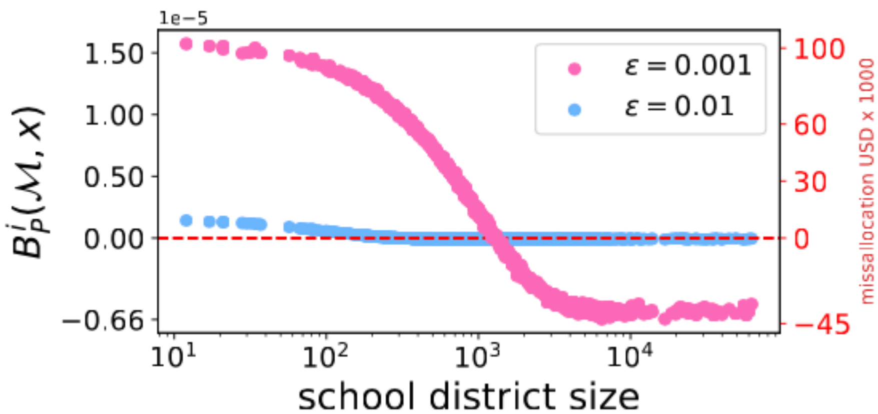
$$\hat{\boldsymbol{x}}_S(\tilde{\boldsymbol{x}}) \triangleq \underset{\boldsymbol{v} \in \mathbb{R}^n}{\operatorname{argmax}} \quad \|\boldsymbol{v} - \tilde{\boldsymbol{x}}\|_2 \qquad (P_S)$$
$$\text{s.t.} \quad \textstyle\sum_{i=1}^n v_i = b$$

Variance Given that $\tilde{x} \sim x + \mathrm{Lap}(\lambda)^n$, each entry of the post-processed solution $\hat{\boldsymbol{x}}_S(\tilde{\boldsymbol{x}})$ is of variance $2\lambda^2(1 - 1/n)$.

Asymptotics As the input dimension goes to infinity, for each $i \in [n]$, the i^{th} entry of the post-processed solution $\hat{\boldsymbol{x}}_S(\tilde{\boldsymbol{x}})$ converges to $x_i + \mathrm{Lap}(\lambda)$ in distribution.

Bias in Budget Allocation

- Bias in Title 1 funds Allocation



$$P_i^F(\boldsymbol{x}) \stackrel{\text{def}}{=} \left(\frac{x_i \cdot a_i}{\sum_{i \in [n]} x_i \cdot a_i} \right)$$

Conclusion

- Tremendous opportunities in data science
  - tight integration of ML and optimization
  - role of optimization in privacy

Summary

Learning Optimization Proxies
- machine learning models approximating optimization solutions

Optimization over machine learning models
- capturing preferences and simulations

Optimization proxies in complex optimization frameworks
- constraint and column generation security-constrained OPF

Optimization for Privacy
- post-processing differential privacy to capture domain constraints

Biases, Fairness, Privacy
- many open issues about privacy, fairness, and biases

Future Work


OPPORTUNITY
JUST AHEAD


Challenges


DECISION
MAKING UNDER
UNCERTAINTY

Thank you for listening