FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

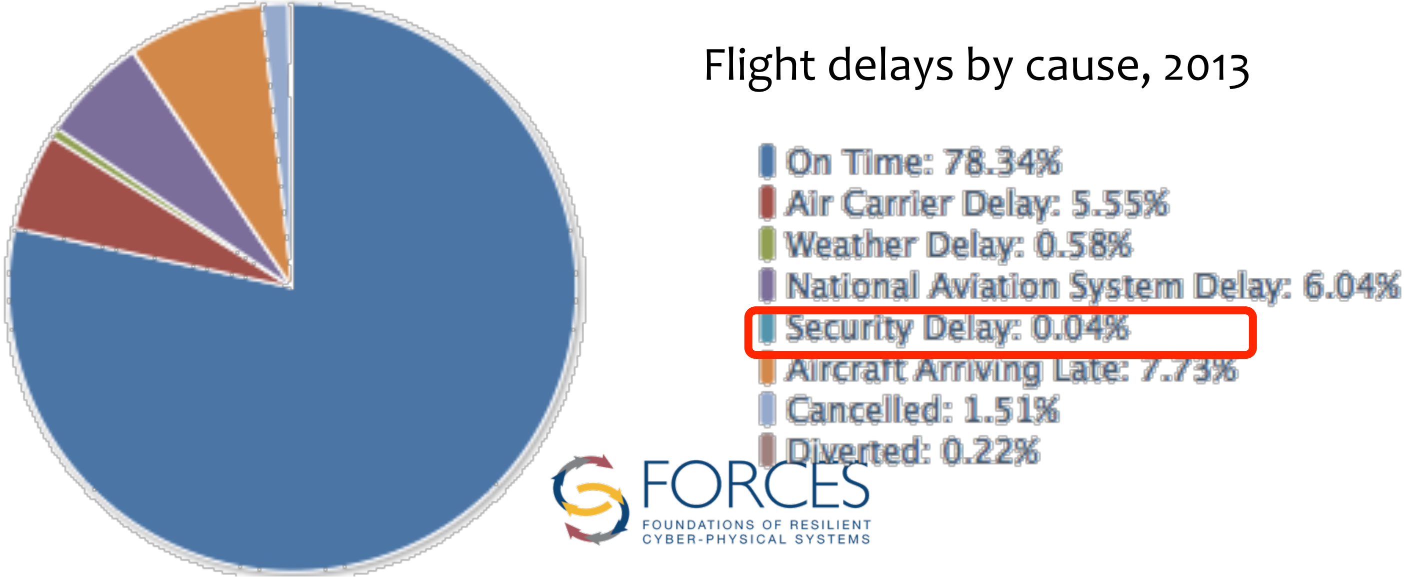
Modeling and Mitigating Disruptions in Networked, Multi-Agent CPS

Hamsa Balakrishnan
(with Karthik Gopalakrishnan, Zeb Hanley, Varun Ramanujam & Juan Jose Rebollo)
Massachusetts Institute of Technology

Berkeley UNIVERSITY OF CALIFORNIA
Massachusetts Institute of Technology
NSF
MICHIGAN
VANDERBILT UNIVERSITY

Network connectivity results in delay propagation

- Evident in the air transportation system
  - Most aircraft serve 4-6 flights a day; only 6% of flights fly only one flight/day
  - A delayed flight delays the aircraft, the crew, and the passengers
  - Large number of shared (airport and airspace) resources increase delay propagation
  - Domestic flight delays cost ~$30-40B annually



Flight delays by cause, 2013

Predicting link delays

- Current delays in the system can help predict future delays
  - Will not predict localized delays (mechanical delays, onset of unexpected events, small-scale disruptions,…)
  - But will predict delays that are "systemic" (congestion, weather disruptions,…)

- Considering temporal variables (time-of-day, day-of-week, season) as well as delay state variables (current origin, destination and link delays, current delay state, estimated type-of-day and previous day's type-of-day), can one predict link delays 2-24 hours in the future?

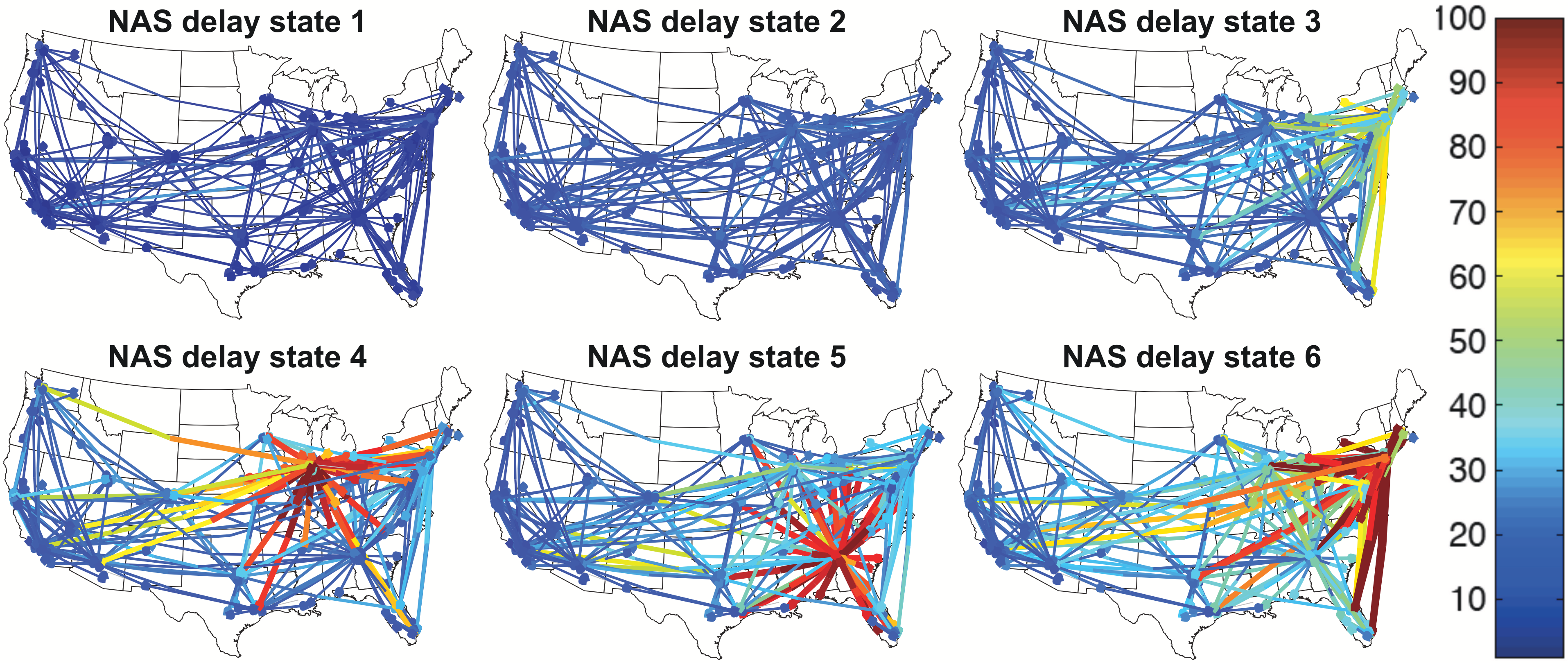
Explanatory variables

- Temporal explanatory variables:
  - Time-of-day (1 categorical variable, 24 categories)
  - Day-of-week (1 categorical variable, 7 categories)
  - Season (1 categorical variable, 3 categories)
- Network delay state explanatory variables:
  - **Local delay state variables:** Influential airports' delay states (10 continuous variables), influential OD pairs' delay states (10 continuous variables)
  - **High level delay state variables:** NAS delay state , NAS type-of-day, previous day's type-of-day
- The Kruskal-Wallis parametric ANOVA test and the multiple comparisons test used to evaluate the dependence of the future departure delay with different categories
- A Random Forest (RF) methodology was used to identify the most relevant continuous variables

Rebollo and Balakrishnan, Transportation Research Part C, 2014

Characteristic NAS “delay states”

- Reflects level of delay in entire NAS
- Obtained through k-means clustering



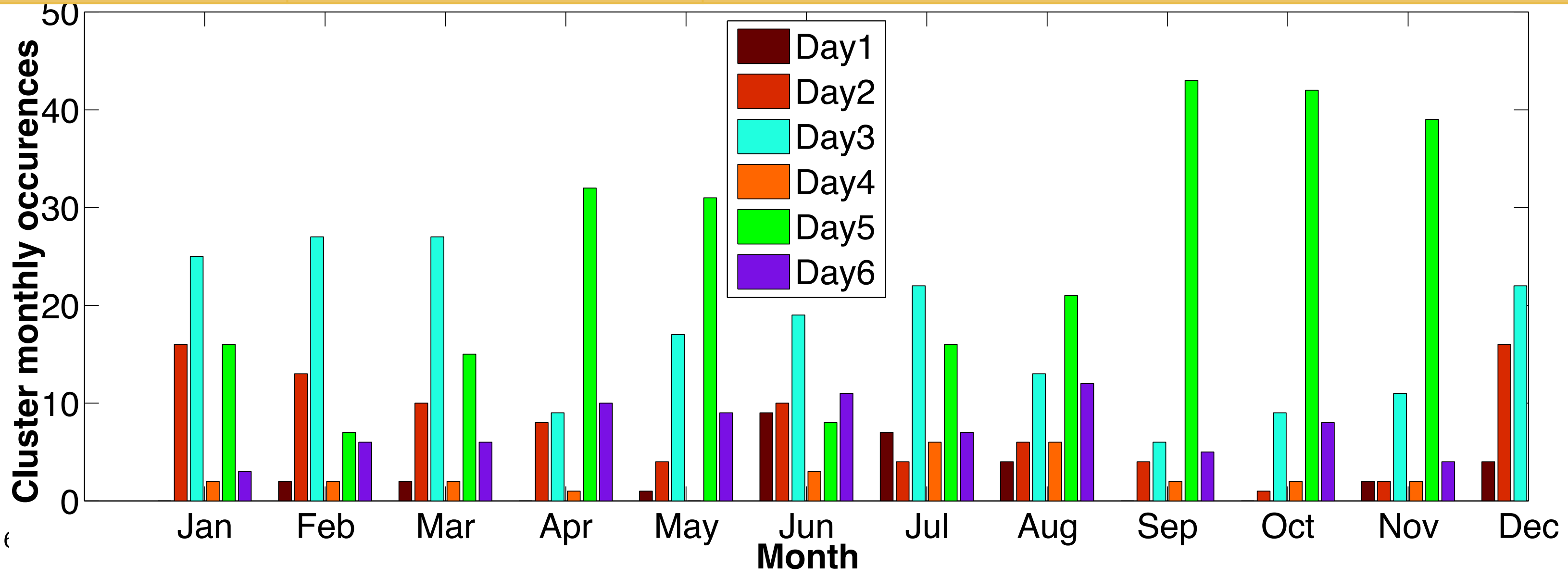
Centroids of NAS delay states. Color represents median link departure delay over 2-hr time-window

Rebollo and Balakrishnan, *Transportation Research Part C*, 2014

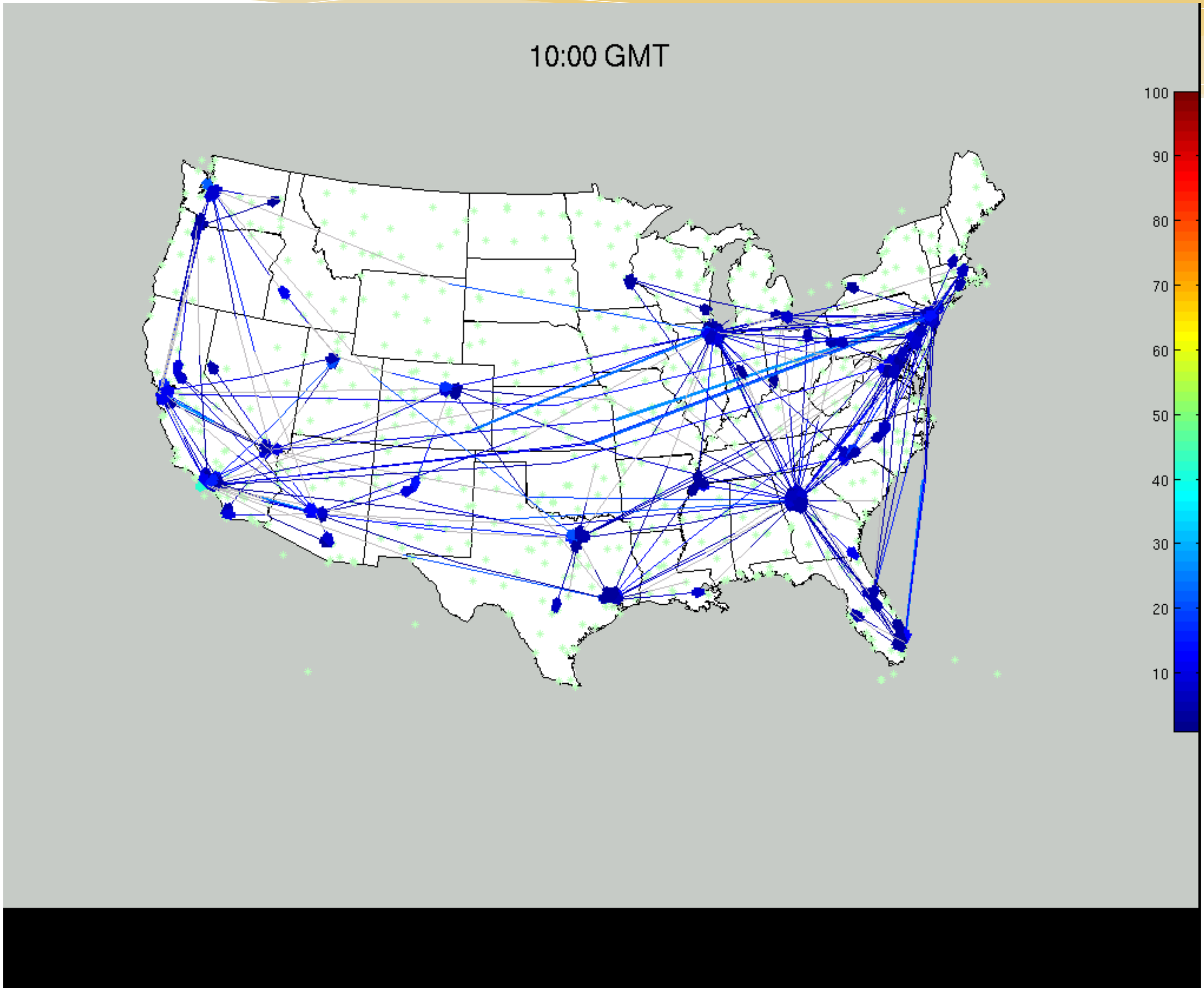
Characteristic types of NAS delay days

- Obtained through k-means clustering
- Delays: High (90 min), medium-high (60 min), medium (20 min), low (5 min)

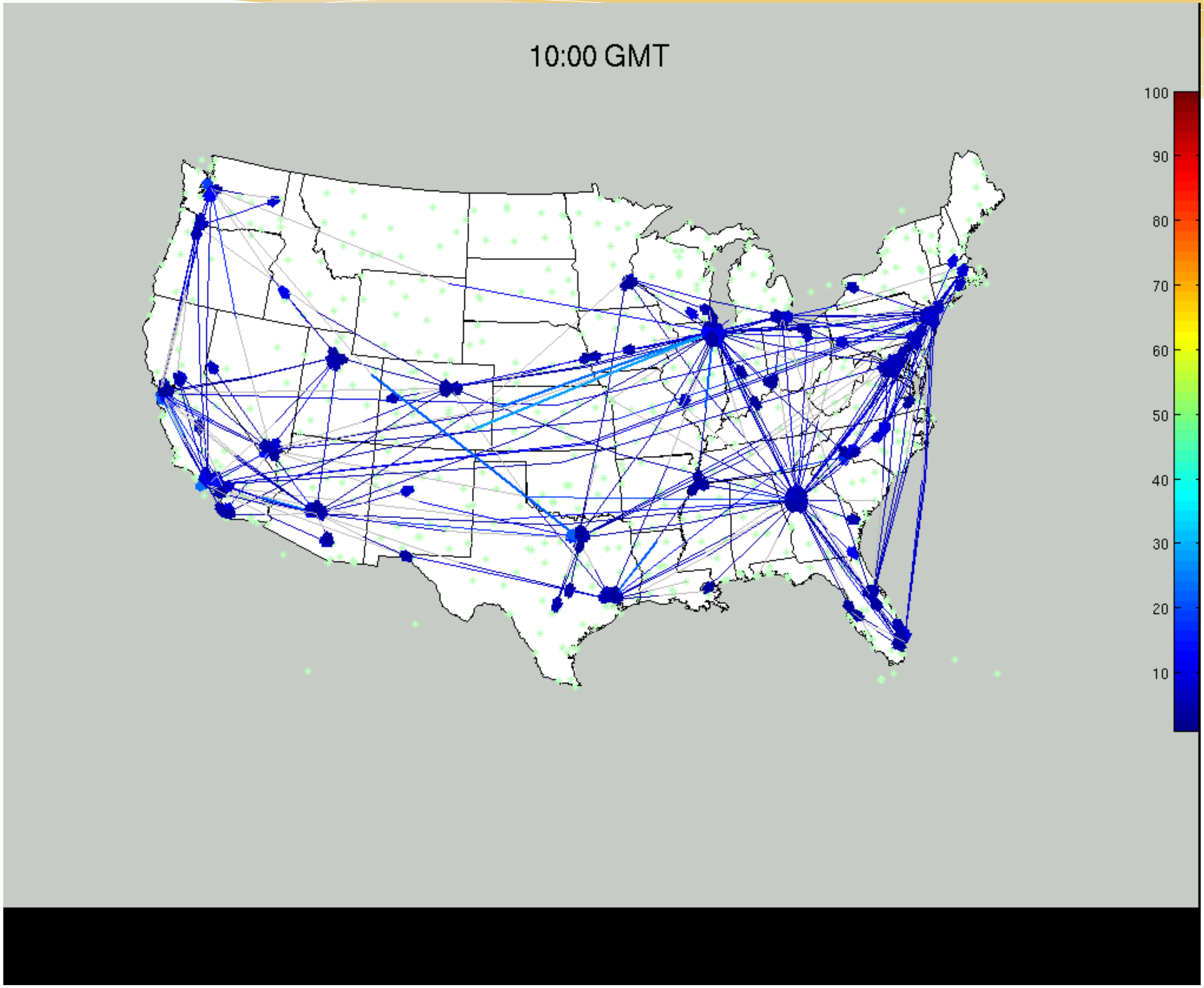
Centroids	Avg. delay (min)	Qualitative Description
Day 1	29	NYC very high, ATL, ORD high delay
Day 2	22	ORD high, NYC medium high delay
Day 3	15	NYC, ORD medium delay
Day 4	21	ATL high, NYC, ORD medium high delay
Day 5	9	Low NAS delay
Day 6	19	NYC high, ATL, ORD medium delay



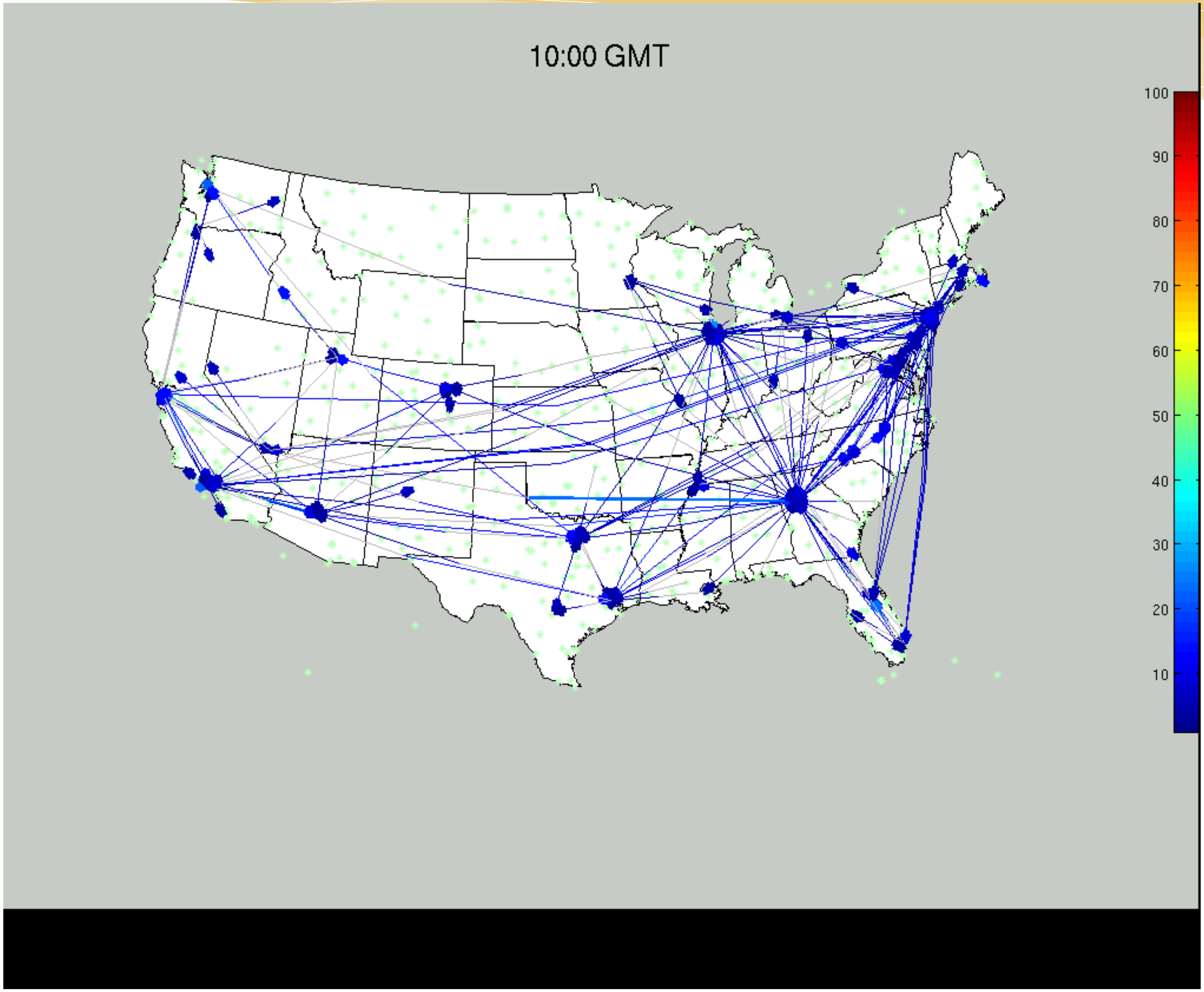
Day 1: NYC v. high, ATL, ORD high delays



Day 2: ORD high, NYC medium-high delays



Day 4: ATL high; NYC, ORD med-high delays

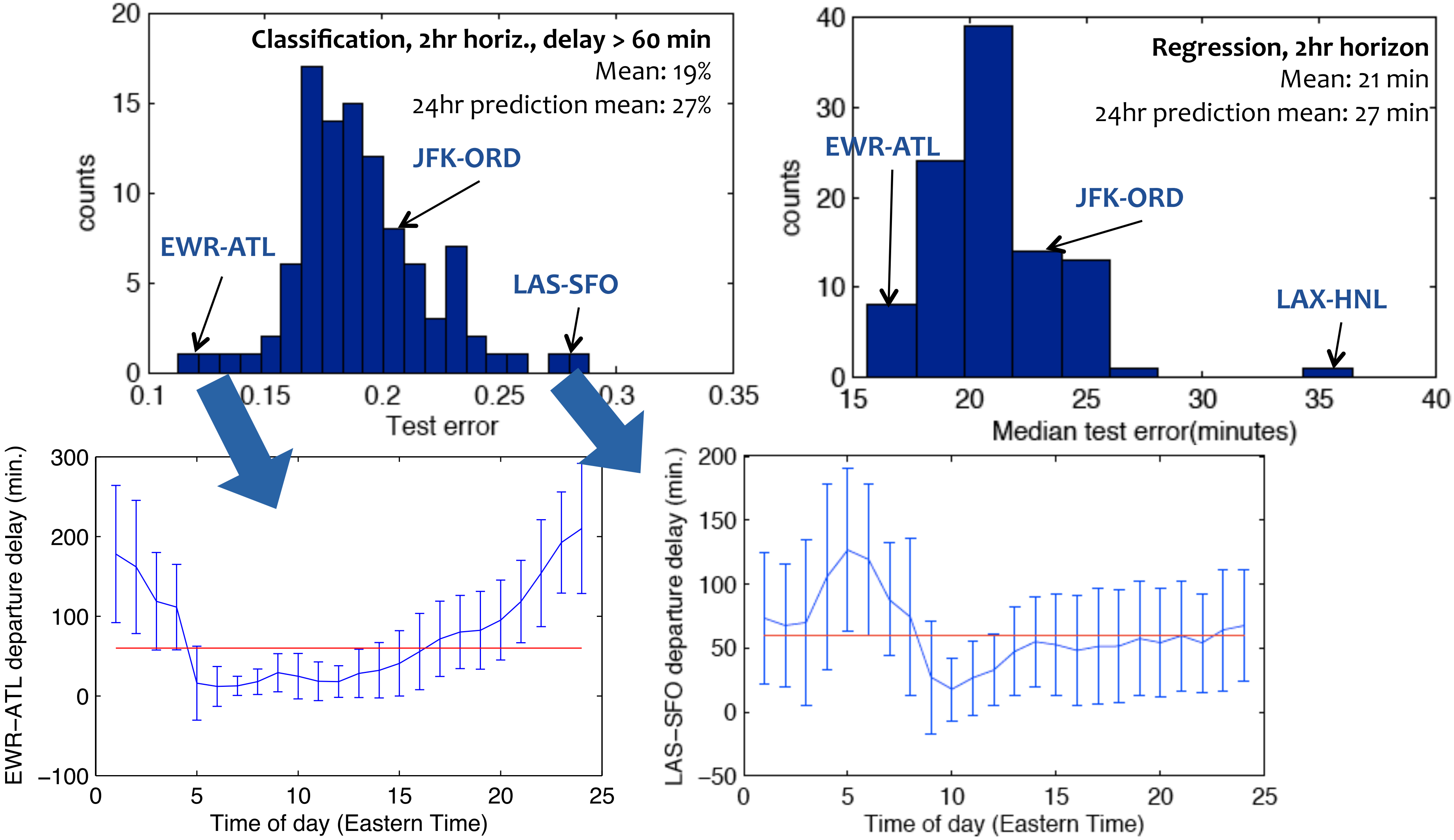


Delay prediction models

- 10 training sets (3,000 points each) and 10 test sets (1,000 points each) from 2007-2008
- Oversampled datasets to be balanced
- Classification (*is delay > or < 60 min?*) and regression algorithms based on Random Forest algorithms
- Performance of delay prediction models evaluated on 100 OD pairs with the highest average delays

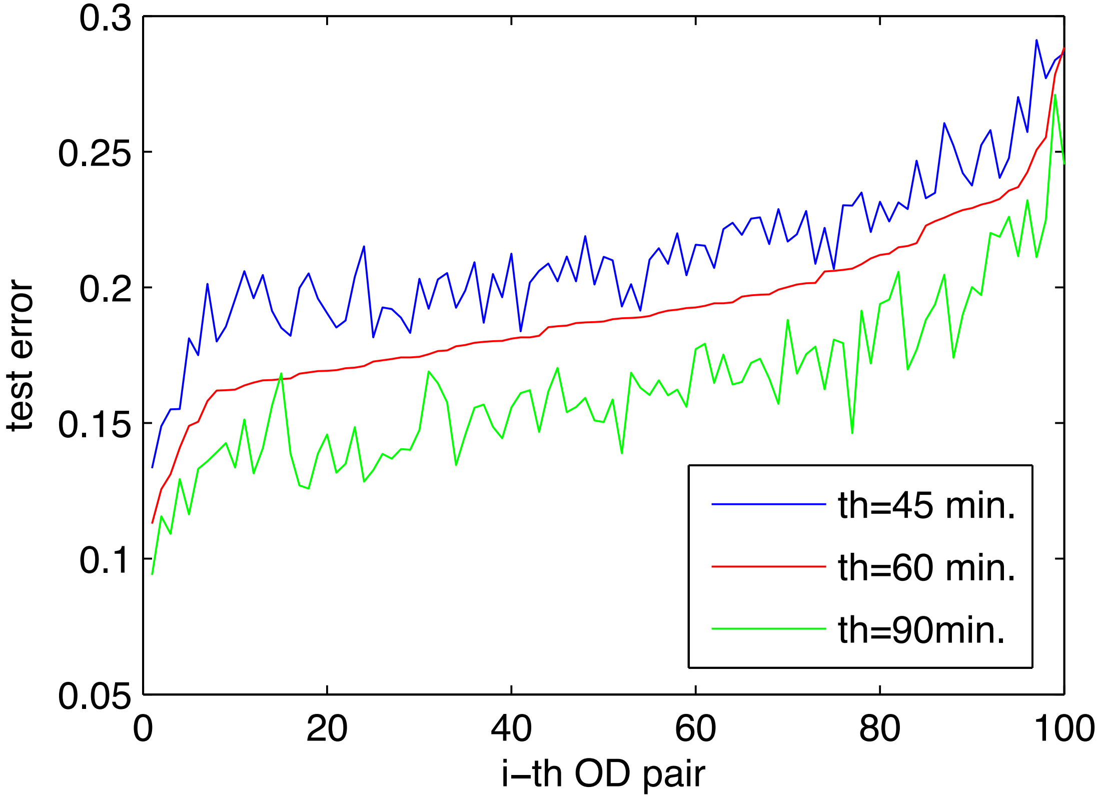
Prediction performance for the 100 most-delayed OD pairs

- Prediction performance varies from link to link



Effect of classification threshold

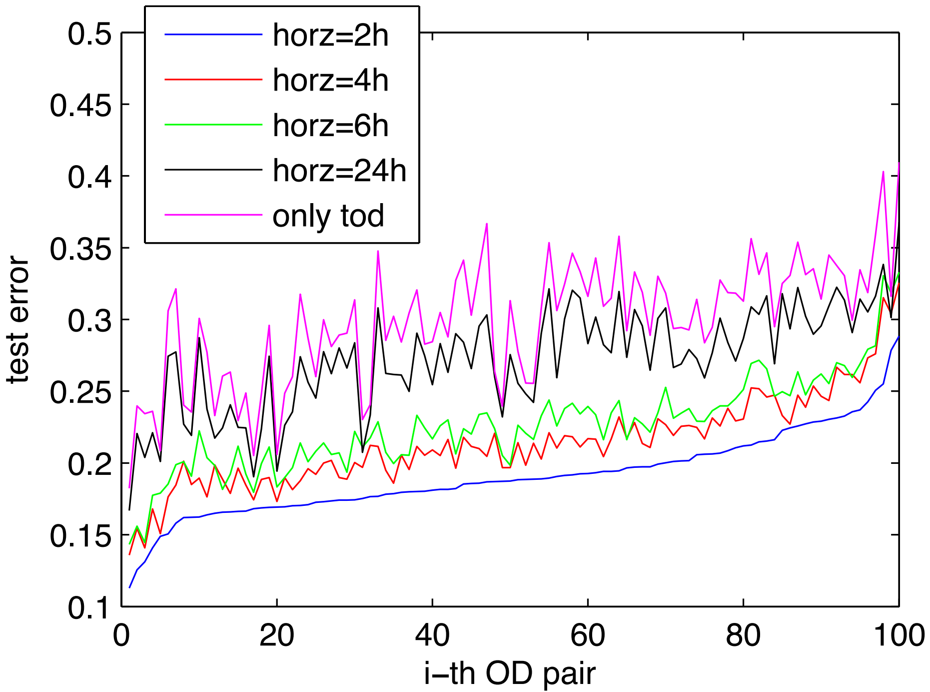
- Prediction performance improves as classification threshold increases

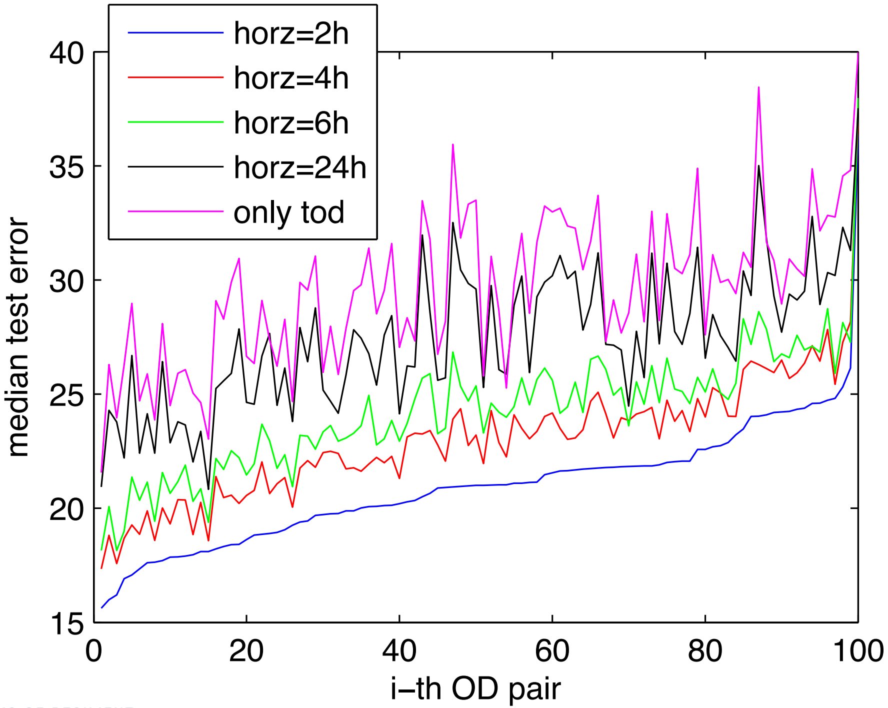


Rebollo and Balakrishnan, *Transportation Research Part C*, 2014

Effect of prediction time horizon

- Avg. classification test errors:
  - 19.1% (2h), 21.4% (4h), 22.6% (6h), and 27.2% (24h)
  - Only time-of-day: 30%
- Avg. (regression) median test error:
  - 21 min (2h), 23 min(4h), 24.3 min (6h), and 27.4 min (24h)





Rebollo and Balakrishnan, *Transportation Research Part C*, 2014

Focus on the major carriers alone

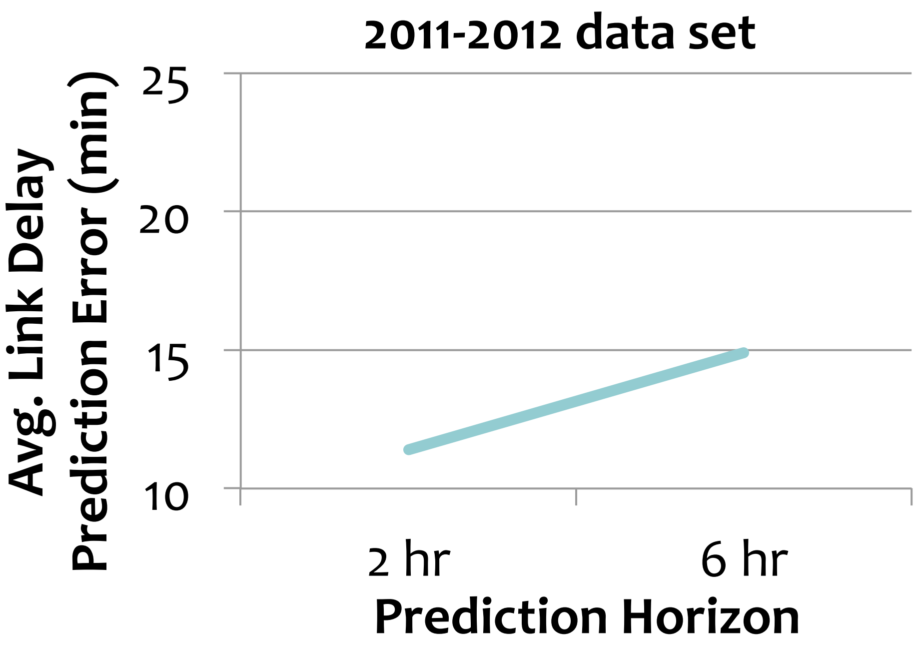
- Bureau of Transportation Statistics (BTS) data
  - Air carriers that account for at least 1% of US domestic scheduled passenger revenues
  - Used to evaluate on-time performance metrics
  - Random Forest models trained on 2011-2012 data, with 10 M+ flights
    - Tested on independent 2011-2012 data as well as 2013 data
  - Simplified network
    - OD pairs with at least 5 flights/day
    - 158 airports; account for 8 M+ flights

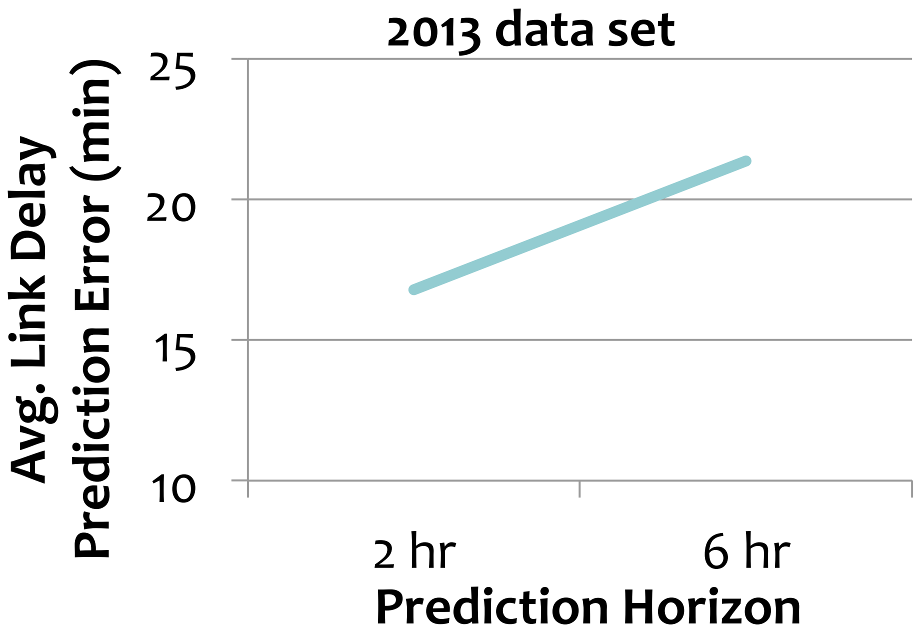
FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Hanley and Balakrishnan, working paper, 2014

Prediction performance for the 100 most-delayed routes and major carriers

- Predict delays relative to **scheduled** departure times

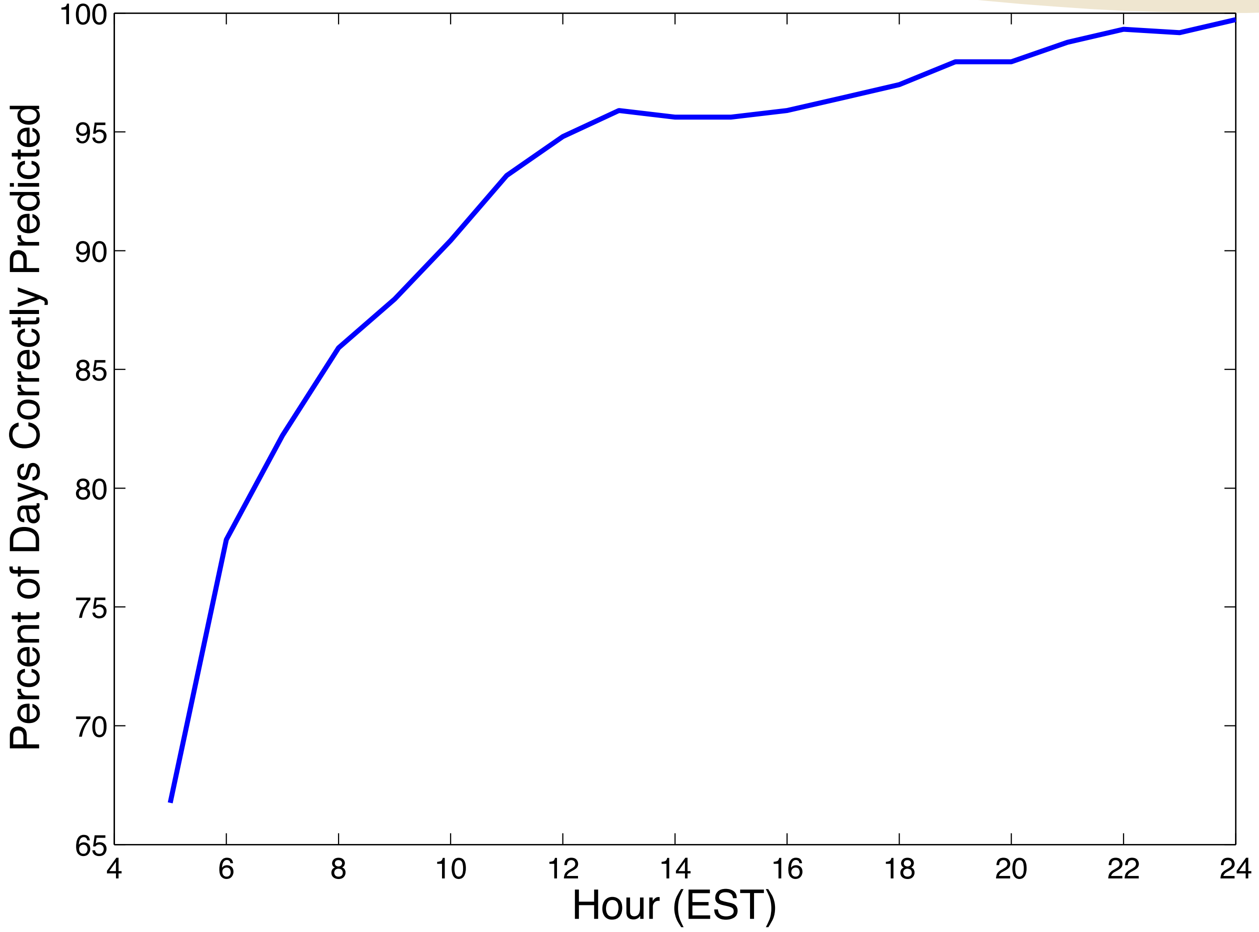




- Note: Dept. of Transportation only considers a flight delayed ***if the departure delay is more than 15 min***

Hanley and Balakrishnan, working paper, 2014

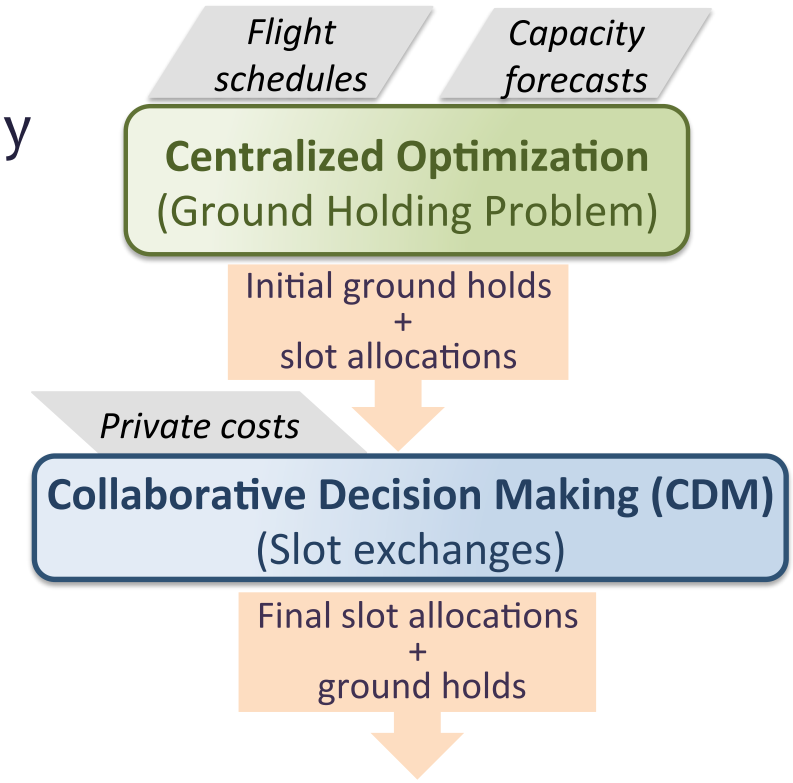
Predicting the type of delay day



Hanley and Balakrishnan, working paper, 2014

Mitigating delays: Optimization + Collaborative Decision Making

- Centralized optimization generally **assumes homogeneous delay costs**
- Airport capacity is uncertain, especially a few hours ahead of time
- Stochastic optimization formulations:
  - Static: Single-stage stochastic Integer Program (IP)
  - Dynamic: Multi-stage stochastic IP, differentiates between flights of different durations
  - *Hybrid:* Multi-stage stochastic IP, but does not differentiate between flights of different durations



Flight schedules

Capacity forecasts

Centralized Optimization
(Ground Holding Problem)

Initial ground holds
\+
slot allocations

Private costs

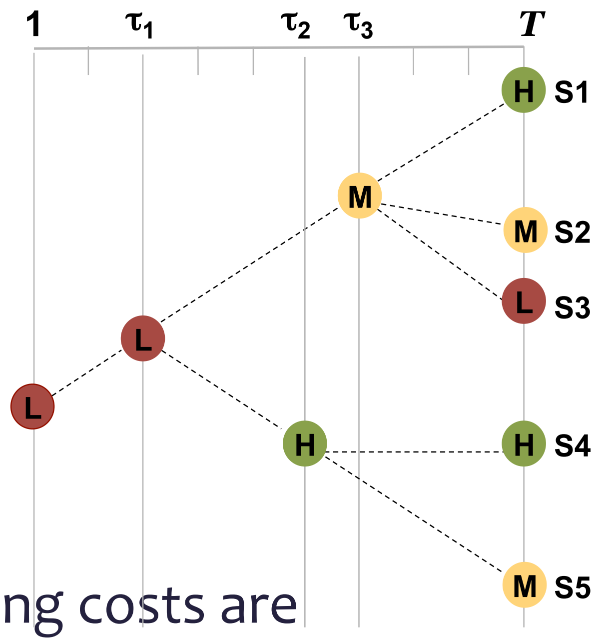
Collaborative Decision Making (CDM)
(Slot exchanges)

Final slot allocations
\+
ground holds

Static Ground Holding Problem

- Single-stage stochastic IP (Richetta & Odoni 1993)

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{n=0}^{K} C_{g,n}\Big(\sum_{t=1}^{T-n} A^{\mathrm{gq}}_{t,t+n}\Big) + \sum_{q\in Q} \pi_q\Big(C_a \sum_{t=1}^{T} A^{\mathrm{aq}}_{q,t}\Big) \\
\text{subject to} \quad & \sum_{j=t}^{t+K} A^{\mathrm{gq}}_{t,j} = A^{\mathrm{dem}}_{t},\ \forall t \in \{1,..,T\} \\
& A^{\mathrm{aq}}_{q,t} \geq \sum_{j=t-K}^{t} A^{\mathrm{gq}}_{j,t} + A^{\mathrm{aq}}_{q,t-1} - A^{\mathrm{cap}}_{q,t},\ \forall t \in \{1,..,T\}, q \in Q \\
& A^{\mathrm{gq}}_{t,j}, A^{\mathrm{aq}}_{q,t} \in \mathbb{Z}^{+},\ \forall t,j \in \{1,..,T\}, q \in Q
\end{aligned}
$$



  - LP relaxation is integer-optimal if ground-holding costs are marginally non-decreasing (Kotnyek & Richetta 2006)
  - **Investigating receding horizon implementation**

Dynamic Ground Holding Problem

- Multi-stage stochastic IP (Mukherjee & Hansen 2007)

$$\text{Minimize } \sum_{q\in Q}\pi_q\Big[\sum_{f\in F}\Big(\sum_{t=\mathrm{arr}_f}^{\mathrm{arr}_f+K} C_{g,t-\mathrm{arr}_f} X^q_{f,t}\Big) + \Big(C_a \sum_{t=1}^{T} A^{\mathrm{aq}}_{q,t}\Big)\Big]$$

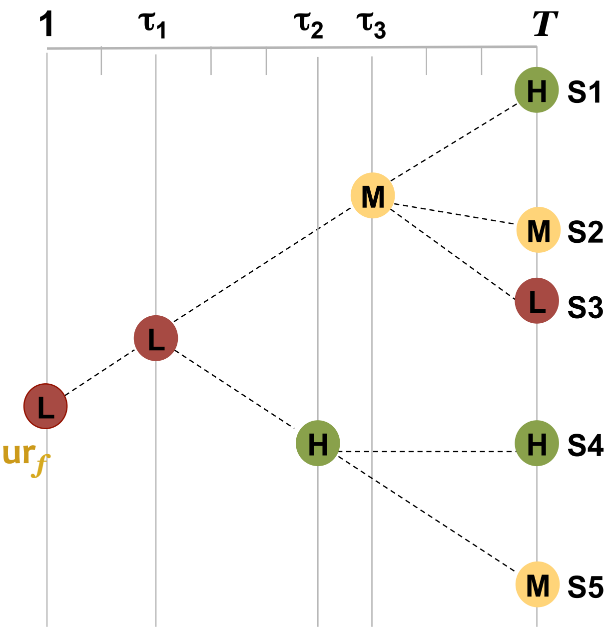
$$\text{subject to } \sum_{t=\mathrm{arr}_f}^{\mathrm{arr}_f+K} X^q_{f,t} = 1,\ \forall q\in Q, \forall f\in F$$

$$A^{\mathrm{aq}}_{q,t} \ge \sum_{f\in F} X^q_{f,t} + A^{\mathrm{aq}}_{q,t-1} - A^{\mathrm{cap}}_{q,t},\ \forall t\in\{1,..,T\}, q\in Q$$

$$X^{q_1}_{f,t} = X^{q_2}_{f,t},\ \forall q_1,q_2\in G_{t-\mathrm{dur}_f}$$

set of feasible scenarios at time $t\text{-dur}_f$

$$X^q_{f,t}\in\{0,1\},\ A^{\mathrm{aq}}_{q,t}\in\mathbb{Z}^+,\ \forall t\in\{1,..,T\}, \forall q\in Q, \forall f\in F$$



- In general, LP relaxation solution is not integer-optimal
- $\mathcal{O}(FT^2+T^2)$ integer decision variables

Hybrid Ground Holding Problem

- Multi-stage stochastic IP (Ramanujam & Balakrishnan CDC 2014)

$$\text{Minimize} \quad \sum_{q\in Q} \pi_q \left(\sum_{n=0}^{K} C_{g,n} \sum_{t=1}^{T-n} X^q_{t,t+n} + C_a \sum_{t=1}^{T} A^{\text{aq}}_{q,t} \right)$$

$$\text{subject to} \quad \sum_{j=t}^{t+K} X^q_{t,j} = A^{\text{dem}}_t, \ \forall t \in \{1,..,T\}, q \in Q$$

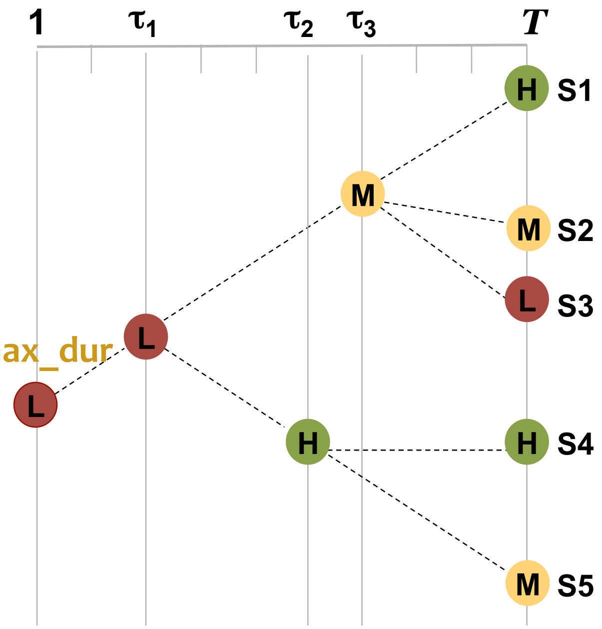
$$A^{\text{aq}}_{q,t} \geq \sum_{j=t-K}^{t} X^q_{j,t} + A^{\text{aq}}_{q,t-1} - A^{\text{cap}}_{q,t}, \ \forall t \in \{1,..,T\}, q \in Q$$

$$X^{q1}_{t,j} = X^{q2}_{t,j}, \ \forall q_1, q_2 \in G_{t-\text{max\_dur}}$$

set of feasible scenarios at time *t*-max_dur

$$X^q_{t,j} \in \mathbb{Z}^+, \ \forall t,j \in \{1,..,T\}, q \in Q$$

$$A^{\text{aq}}_{q,t} \in \mathbb{Z}^+, \ \forall t \in \{1,..,T\}, \ \forall q \in Q.$$



- In general, $\mathcal{O}(T^3)$ integer decision variables, but $\mathcal{O}(T)$ or $\mathcal{O}(T^2)$ under certain conditions

Optimization formulations of Collaborative Decision Making

Intra-airline substitution

$$
\begin{aligned}
\text{Minimize} \quad & \sum_{f_1 \in F_a} \sum_{f_2 \in F_a} C_{f_1,f_2} X_{f_1,f_2} \\
\text{subject to:} \quad & \sum_{f_1 \in F_a} X_{f_1,f_2} = 1, \; \forall f_2 \in F_a \\
& \sum_{f_2 \in F_a} X_{f_1,f_2} = 1, \; \forall f_1 \in F_a \\
& X_{f_1,f_2} \leq \mathrm{feas}_{f_1,f_2}, \; \forall f_1, f_2 \in F_a \\
& X_{f_1,f_2} \in 0,1 \, \forall f_1, f_2 \in F_a
\end{aligned}
$$

Inter-airline substitution

$$
\begin{aligned}
\text{maximize} \quad & \sum_{q \in Q} \pi_q \left[\sum_{f \in F \backslash c} \sum_{r=1}^{T} \mathcal{B}(r) Y^q_{f,r} - M d^q_c \right] \\
\text{subject to:} \quad & \\
& \sum_{t=\mathrm{ETA}_f}^{\mathrm{ETA}_f + K} X^q_{f,t} = 1, \; \forall q \in Q, \forall f \in F \\
& A^{\mathrm{aq}}_{q,t} \geq \sum_{f \in F} X^q_{f,t} + A^{\mathrm{aq}}_{q,t-1} - A^{\mathrm{cap}}_{q,t}, \; \forall t \in \{1,..,T\}, q \in Q \\
& X^{q_1}_{f,t} = X^{q_2}_{f,t}, \; \forall q_1, q_2 \in G_{\mathrm{ETA}_f - \mathrm{max\_dur}}, \\
& d^q_c = \sum_{t=1}^{T} t X^q_{c,t} - k, \; \forall q \in Q, \\
& \sum_{t=1}^{T} t X^q_{f,t} \leq \mathrm{arr}^q_f, \; \forall q \in Q, f \in F \backslash c \\
& A^{\mathrm{aq}}_{q,t} \leq A q^{\mathrm{aq,orig}}_{q,t}, \; \forall q \in Q, t \in \{1,..,T\} \\
& Y^q_{f,r} = X^q_{f, \mathrm{arr}^q_f - r + 1}, \; \forall q \in Q, f \in F \backslash c, r \in \{1,..,T\} \\
& X^q_{f,t} \in \{0,1\}, \; d^q_c \geq 0, \; \forall t \in \{1,..,T\}, \forall q \in Q, \forall f \in F
\end{aligned}
$$

Comparison of Ground Holding Problem formulations

	Static	Hybrid	Dynamic
Pre-CDM delay cost	High (Worst)	Medium	Low (Best)
Benefit from CDM	High (Best)	Medium	Low (Worst)
Equity	High (Best)	Medium	Low (Worst)
Tractability	High (Best)	Medium	Low (Worst)
Ease of implementation	High (Best)	Medium	Low (Worst)

[Ramanujam & Balakrishnan CDC 2014]

- Same under rolling horizon implementations?
- How do we handle network connectivity?

Summary

- Large number of shared resources in networked CPS increase delay propagation
  - Delay propagation implies that current delay state can be used to predict future link delays
  - Delay propagation makes it essential that resource allocation and reallocation be made as efficient as possible, especially at important nodes in the network
    - Must consider multi-stakeholder objectives and human behavior
    - Must consider heterogeneous cost functions
    - Need integration of robust control and economic incentives
  - Connectivity has implications to optimization as well