Using Automated Data Sources to Improve the Performance of Public Transport Systems: A Framework and Applications

Nigel H.M. Wilson
MIT
April 2017

Northwestern University
NUTC Seminar

Massachusetts Institute of Technology

Outline

- The changing environment and customer expectations
- Automated Data Collection Systems
- Framework for using Automated Data Sources (ADS)
- Analysis building blocks
  - OD Matrix Estimation
  - Inferring Train Loads
  - Measuring Service Reliability
- Future Prospects

The Changing Environment and Customer Expectations

- Many customers expect a personal relationship with service providers, e.g. Amazon
- Information technology advances raise expectations and provide new opportunities, e.g. mobile internet
- Rising incomes result in fewer captive riders
- Need to attract choice riders
- Challenges for public transport
  - Gap between customer expectations and current reality
  - Uber-type competition
  - AV in mid-term future

Automated Data Systems

- **Automatic Fare Collection Systems (AFC)**
  - increasingly based on contactless smart cards with unique ID
  - provides entry (exit) information (spatially and temporally) for individual passengers
  - traditionally not available in real-time
- **Automatic Vehicle Location Systems (AVL)**
  - bus location based on GPS
  - train tracking based on track circuit occupancy
  - available in real time
- **Automatic Passenger Counting Systems (APC)**
  - bus systems based on sensors in doors with channelized passenger movements
  - passenger boarding (alighting) counts for stops/stations with fare barriers
  - train load-weigh systems can be used to estimate number of passengers on board
  - traditionally not available in real-time

Public Transport Operators/Agencies are at a Critical Transition in Data Collection Technology

Manual	Automatic
• low capital cost	• high(er) capital cost
• high marginal cost	• low marginal cost
• small sample sizes	• large sample sizes
• "hard and soft"	• "hard"
• unreliable	• errors and biases can be estimated and corrected
• limited spatially and temporally	• ubiquitous
• not available immediately	• available in real-time or quasi real-time

ADS - Potential

- Integrated ADS database
- Models and software to support many agency decisions using database
- Monitoring and insight into normal operations, special events, unusual weather, etc.
- Large, long time-series disaggregate panel data to better understand customer experience and travel behavior

ADS - Reality

- Most ADS systems are implemented independently
- Data collection is ancillary to primary system function
  - AVL - emergency notification, stop announcements
  - AFC - fare collection and revenue protection
- Many problems to overcome:
  - not easy to integrate data
  - requires new resources and expertise

Opportunities

- ADS
  - monitoring system status
  - measuring reliability
  - understanding customer behavior
- Data + Computing
  - simulation-based predictive performance models
- Communications
  - real time information (demand)
  - operations management (supply)
- Systematic approaches for planning, operations, real-time control

Key Agency/Operator Functions

A. Off-Line Functions

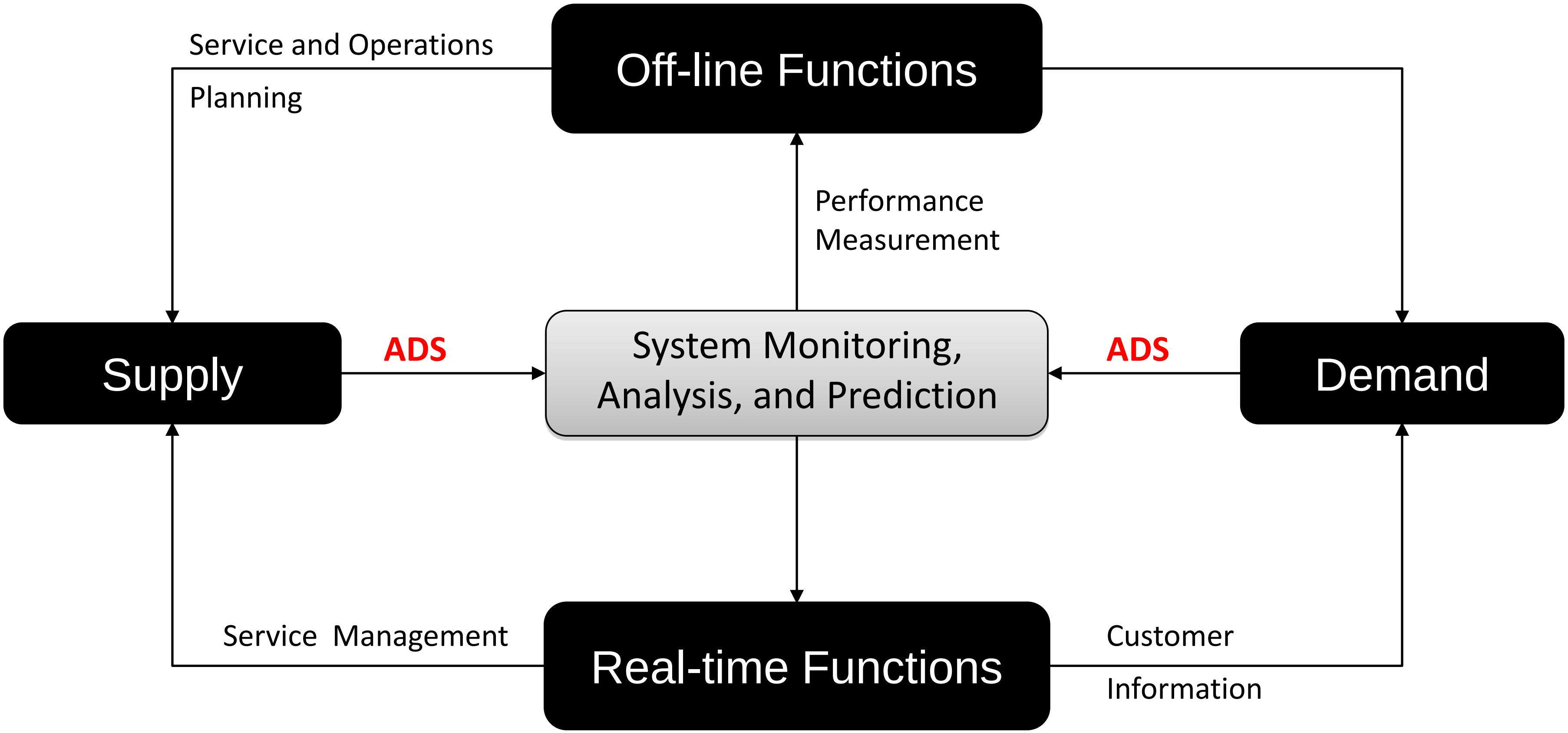
- Service and Operations Planning
  - Network and route design
  - Frequency setting and timetable development
  - Vehicle and crew scheduling
- Performance Measurement
  - Measures of operator performance against plans/contract specs
  - Measures of customer experience

Key Agency/Operator Functions

B. Real-Time Functions

- Service and Operations Control and Management
  - Dealing with deviations from plans, both minor and major
  - Dealing with unexpected changes in demand
- Customer Information
  - Information on routes, trip times, vehicle arrival times, etc.
  - Increasingly dynamic

Key Functions



Analysis Building Blocks

- OD Matrix Estimation
- Inferring train loads
- Measuring Service Reliability

OD Matrix Estimation

Objective:

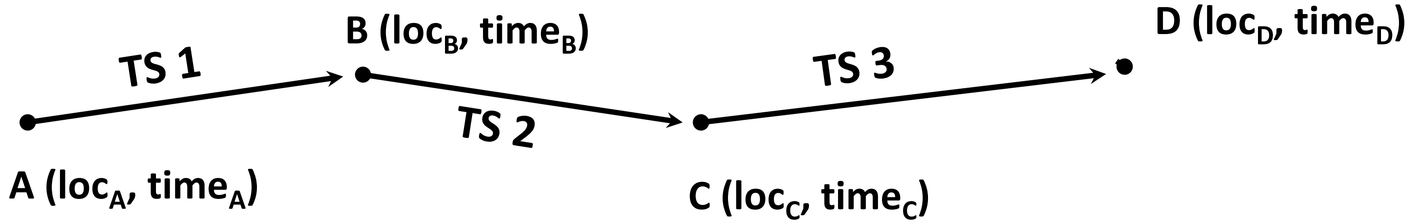
- Estimate passenger journey OD matrix at the network level using AFC and AVL data
  - Multimodal public transport passenger flows
- AFC characteristics
  - Open (entry fare control only)
  - Closed (entry+exit fare control)
  - Hybrid

Source:
"Intermodal Passenger Flows on London's Public Transport Network: Automated Inference of Full Passenger Journeys Using Fare-Transaction and Vehicle-Location Data. Jason Gordon, MST Thesis, MIT (September 2012).

Trip Chaining: Basic Idea

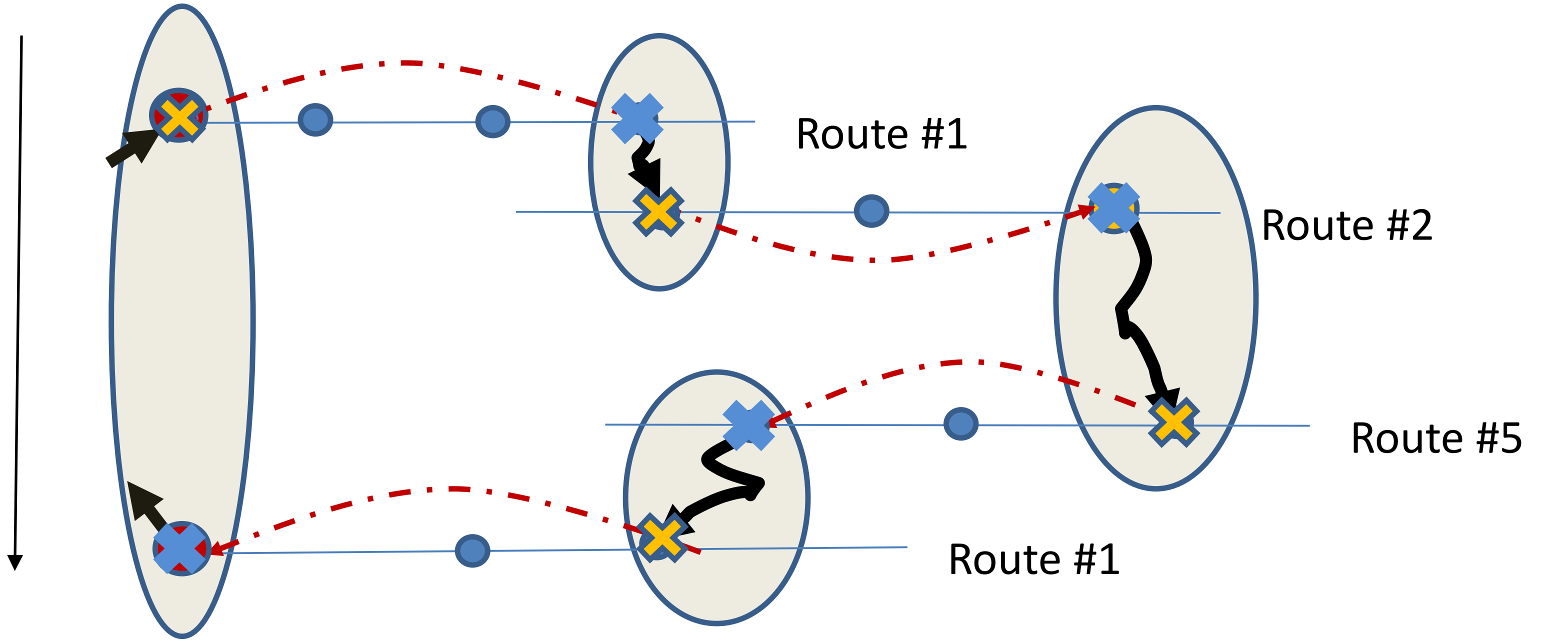
Each AFC record includes:

- AFC card ID
- transaction type
- transaction time
- transaction location: rail station or bus route and stop (either directly or based on time-matching with AVL data)

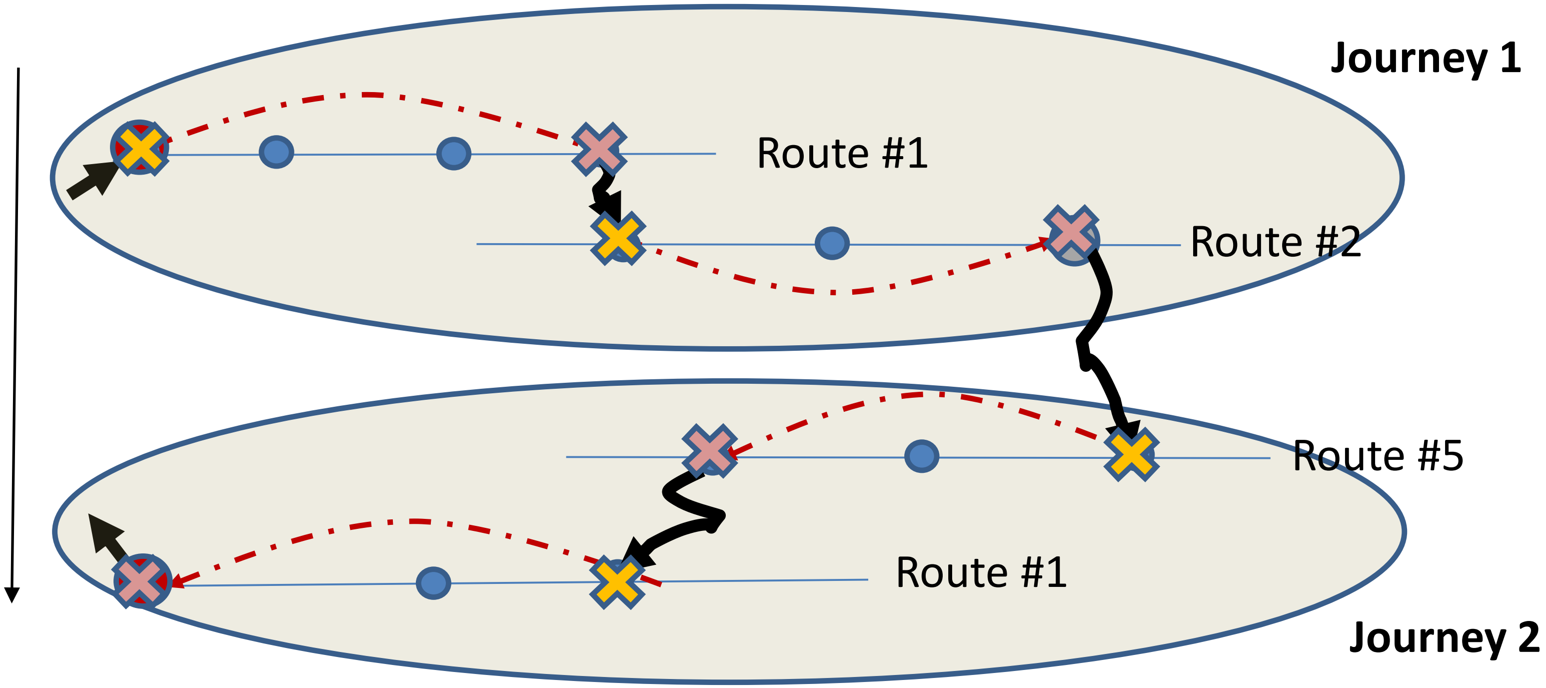


The destination of many trip segments (TS) is close to the origin of the following trip segment.

Destination Inference



Interchange Inference



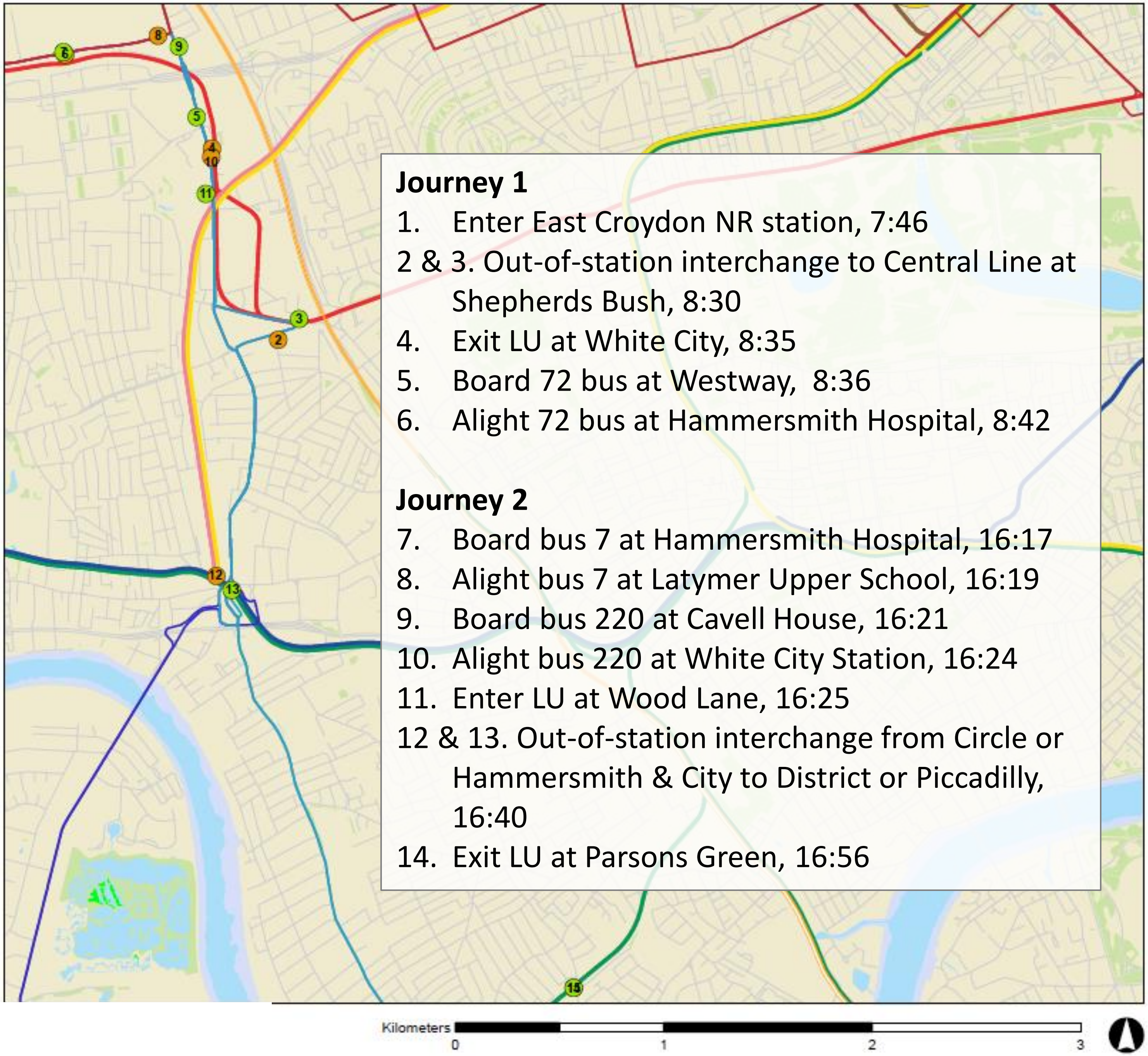
Trip-Chaining Method for OD Inference

Key Assumptions for Destination Inference to be correct:

- No intermediate private transportation mode trip segment
- Passengers will not walk a long distance
- Last trip of a day ends at the origin of the first trip of the day

Sample Daily Oyster Itinerary

Wednesday, 11 May 2011



Bus routes

- 220
- 7
- 72

Journey 1

1. Enter East Croydon NR station, 7:46

2 & 3. Out-of-station interchange to Central Line at Shepherds Bush, 8:30

4. Exit LU at White City, 8:35
5. Board 72 bus at Westway, 8:36
6. Alight 72 bus at Hammersmith Hospital, 8:42

Journey 2

7. Board bus 7 at Hammersmith Hospital, 16:17
8. Alight bus 7 at Latymer Upper School, 16:19
9. Board bus 220 at Cavell House, 16:21
10. Alight bus 220 at White City Station, 16:24
11. Enter LU at Wood Lane, 16:25

12 & 13. Out-of-station interchange from Circle or Hammersmith & City to District or Piccadilly, 16:40

14. Exit LU at Parsons Green, 16:56

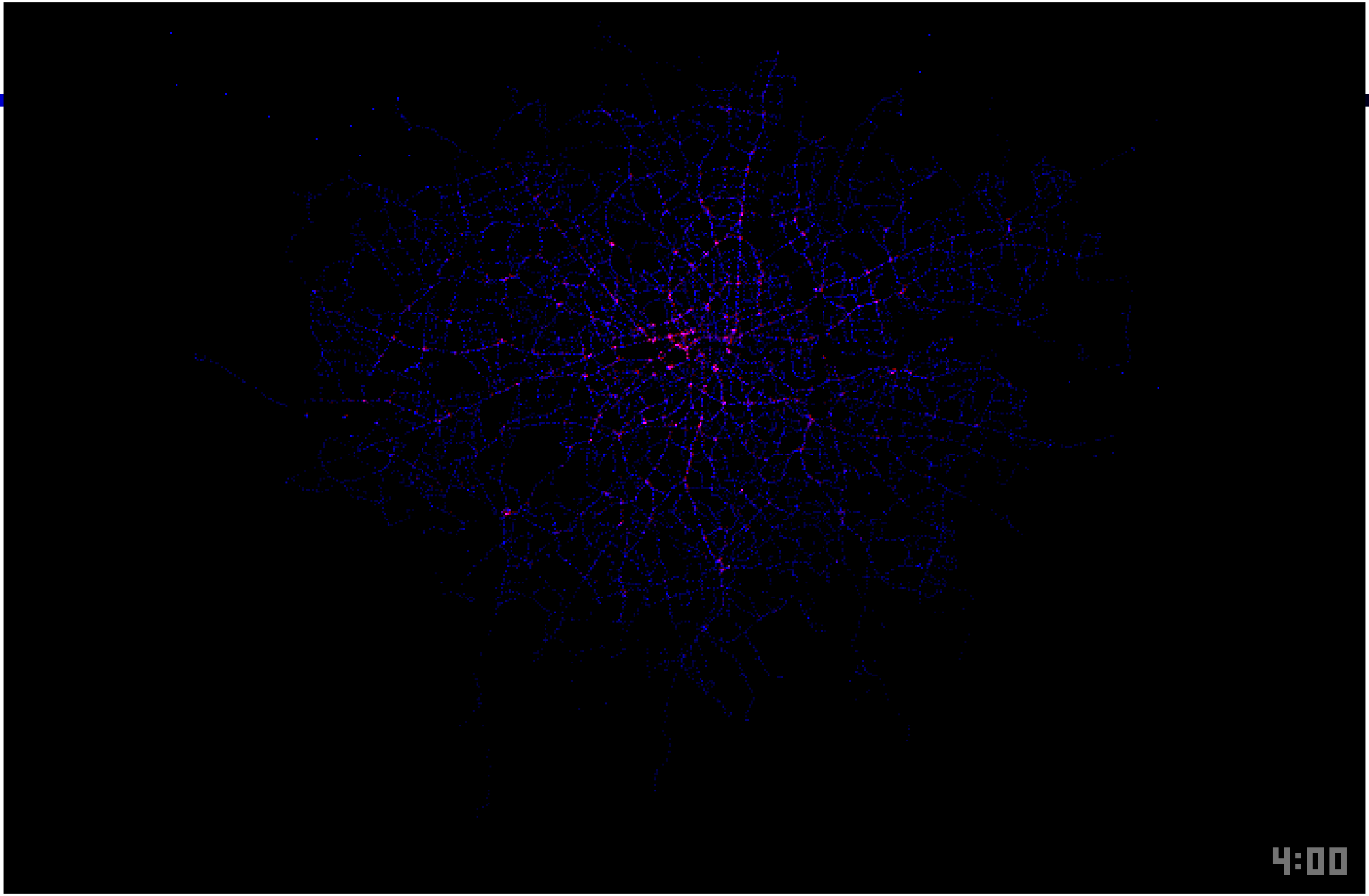
Trip-Chaining Method Steps

- Infer start and end of each trip segment for individual AFC cards
- Link trip segments into complete (one-way) journeys
- Integrate individual journeys to form seed OD matrix (by time period)
- Expand to full OD matrix using available control totals
  - station entries and/or exits for rail
  - passenger entries and/or exits by stop, trip, or period for bus

Summary Information on London Application

- Oyster fare transactions/day:
  - Rail (Underground, Overground, National Rail): 6 million (entry & exit)
  - Bus: 6 million (entry only)
- For bus:
  - Origin inference rate: 96%
  - Destination inference rate: 77%
- For full public transport network:
  - 76% of all fare transactions are included in the seed matrix
- Computationally feasible (30 mins on Intel PC for full London public transport OD Matrix for entire day, including both seed matrix and scaling)

Passenger Journeys in London Animation



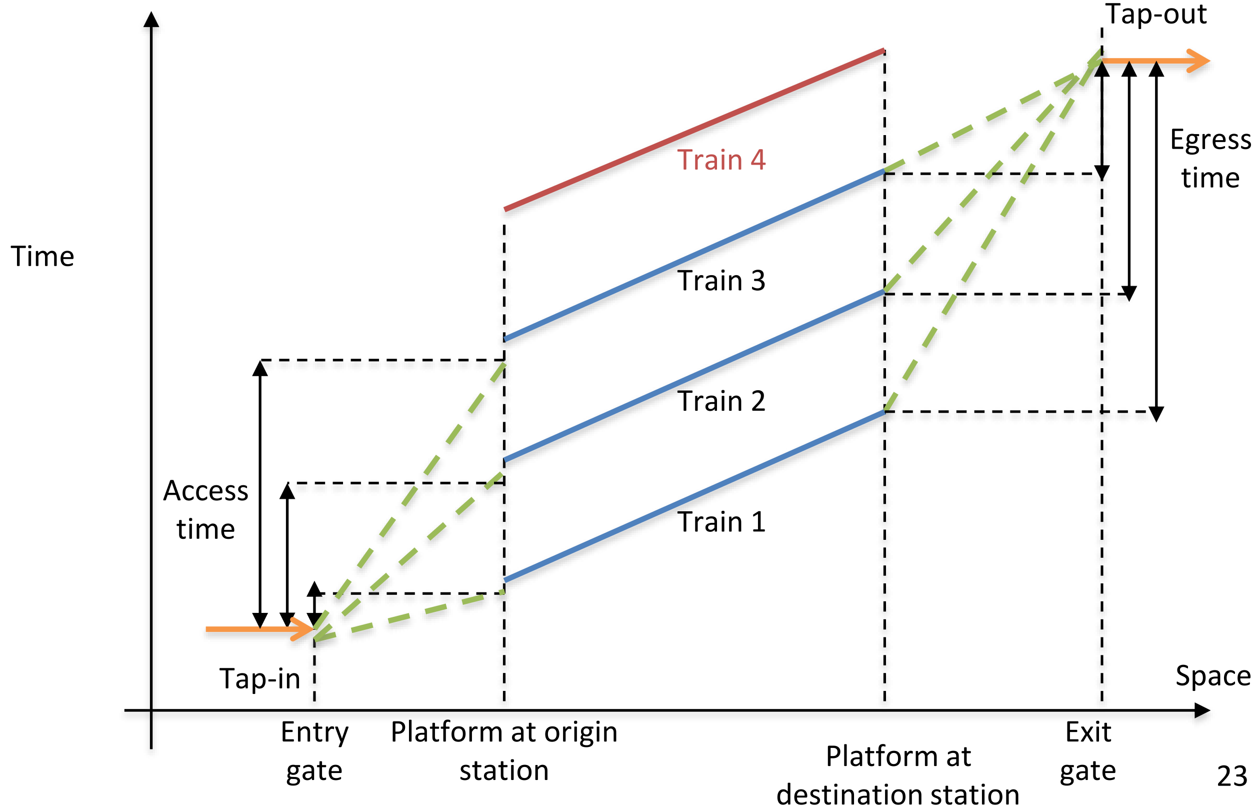
Inferring Train Loads

- Develop a methodology to "assign" passengers to trains through the use of AFC, ATR data
- The methods support:
  - Assessment of service utilization
  - Service quality metrics from the customers' point of view
    - Crowding on trains and in stations
    - Number of passengers denied boarding
    - More detailed journey time metrics

Source: "Passenger-to-Train Assignment Model Based on Automated Data."
Yiwen Zhu, Master of Science in Transportation thesis (MIT, 2014)

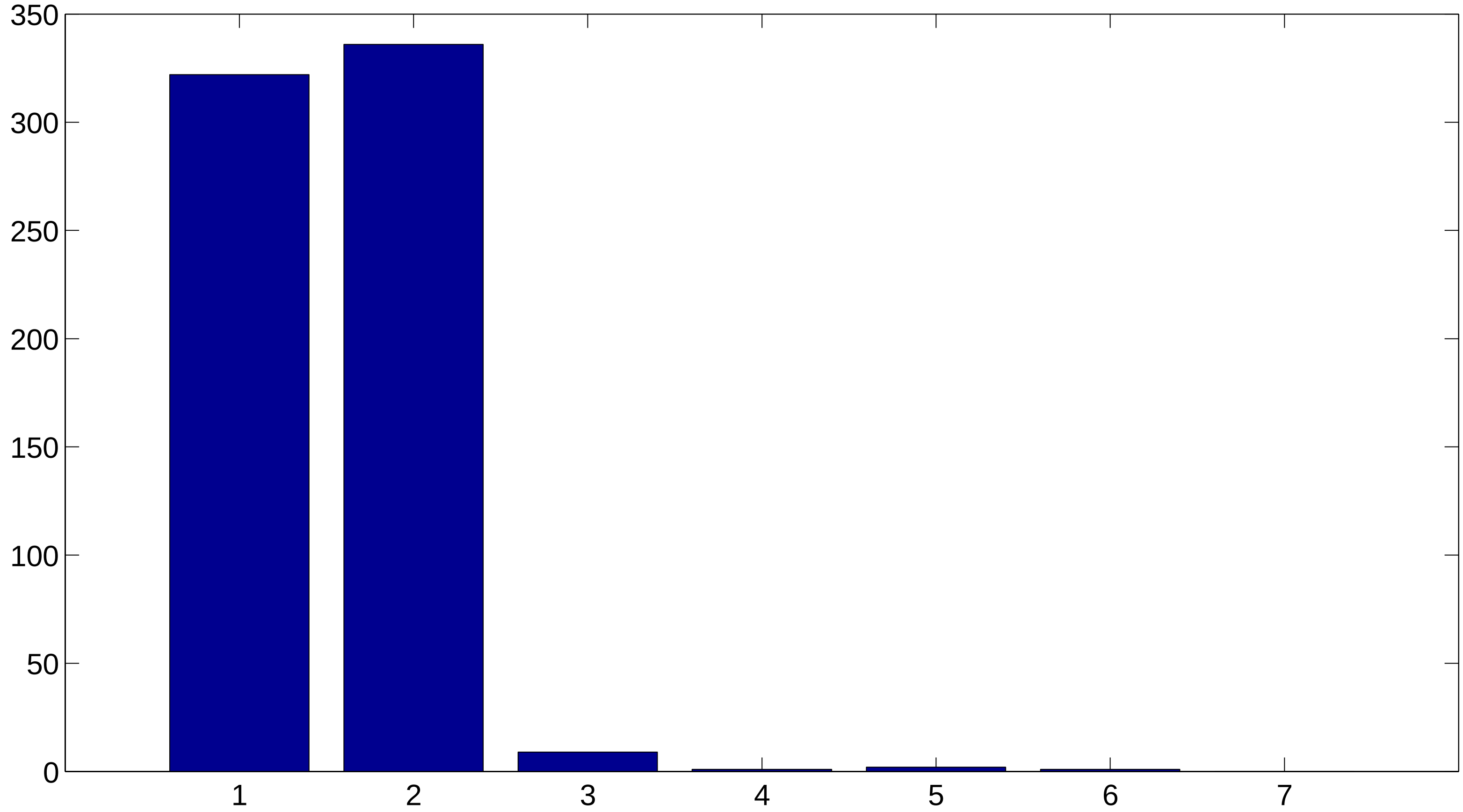
Feasible Train Itineraries

- Given: AFC & ATR data
- A train itinerary is feasible if:
  - It departs after the passenger taps in, and
  - Arrives before the passenger taps out



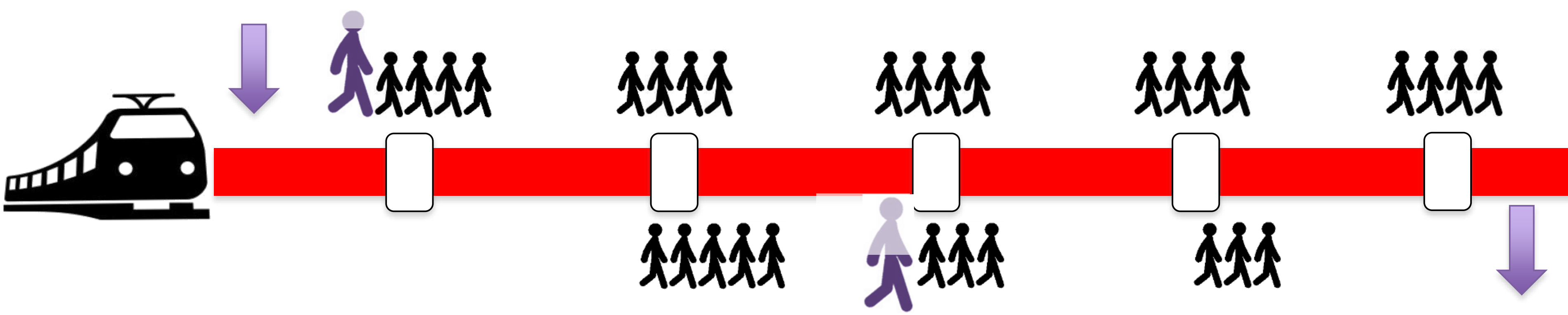
Massachusetts Institute of Technology

Feasible Train Itineraries Example



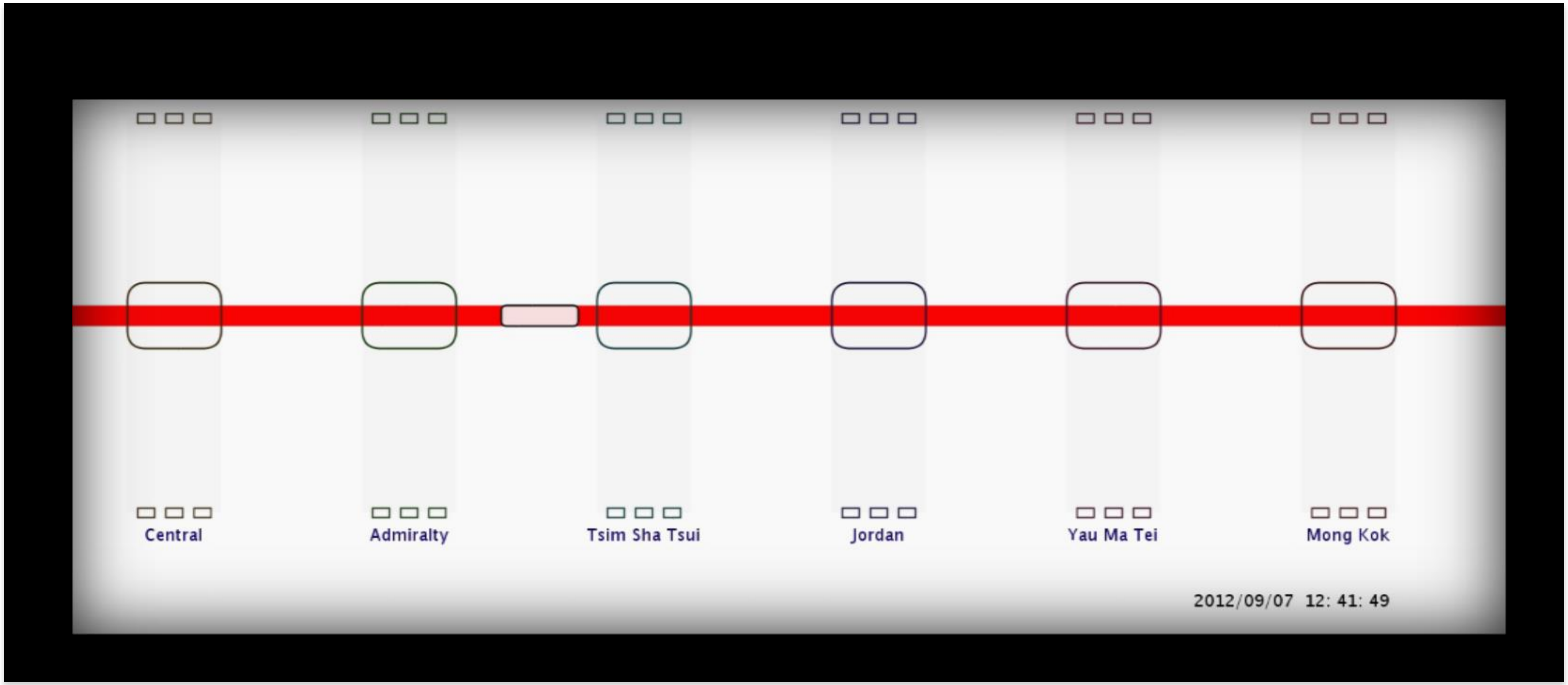
Passenger Assignment Model (PAM)

- Each station is examined in sequence starting from the terminal.
- At each station, the trainload is calculated from the corresponding probabilities of passengers whose feasible itinerary set includes this train.



Passenger Movement

- Based on the output of PAM, individual passenger movements can be presented in detail.



Recent Extensions

- Relaxing assumptions:
  - Denied boardings due to capacity constraints
  - Interchange demand

Future Work

- Advanced customer information, such as expected crowding at stations and in trains
- Real time model and application

Measuring Service Reliability

Objective:

- Define a customer-centric measure to capture effects of reliability

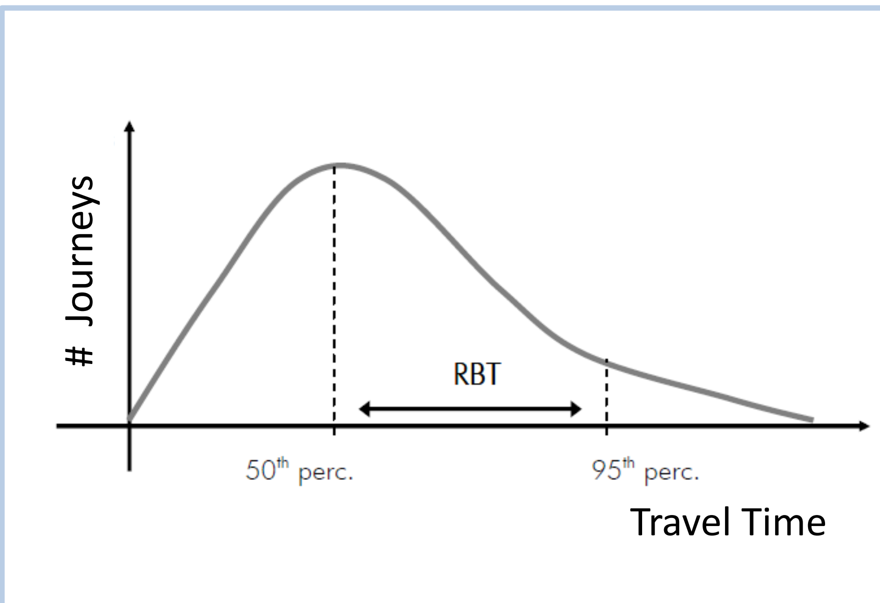
Expected benefits:

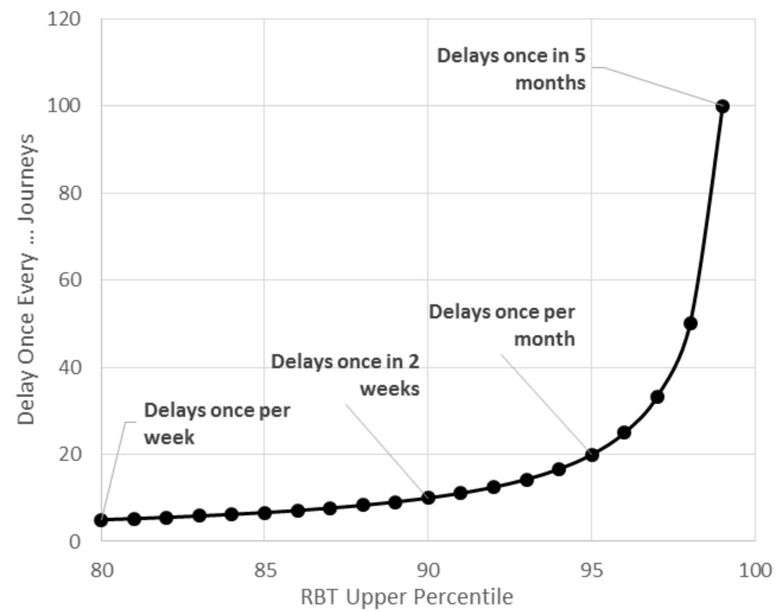
- Improve customer communication
- Capture the effects of strategies to improve service reliability

Reliability Buffer Time (RBT)

$$RBT_{OD} = JT(N^{th}) - JT(50^{th})$$

"How much additional time should I budget, beyond my typical travel time, to ensure an on time arrival N% of the time?"





Calculating the RBT

- Calculated for **each hour** of the day over some **period of time**
- RBT measures the **total** variation from all portions of a journey:

Access Time + Wait Time + In Vehicle Time + Egress Time

Variants of the RBT

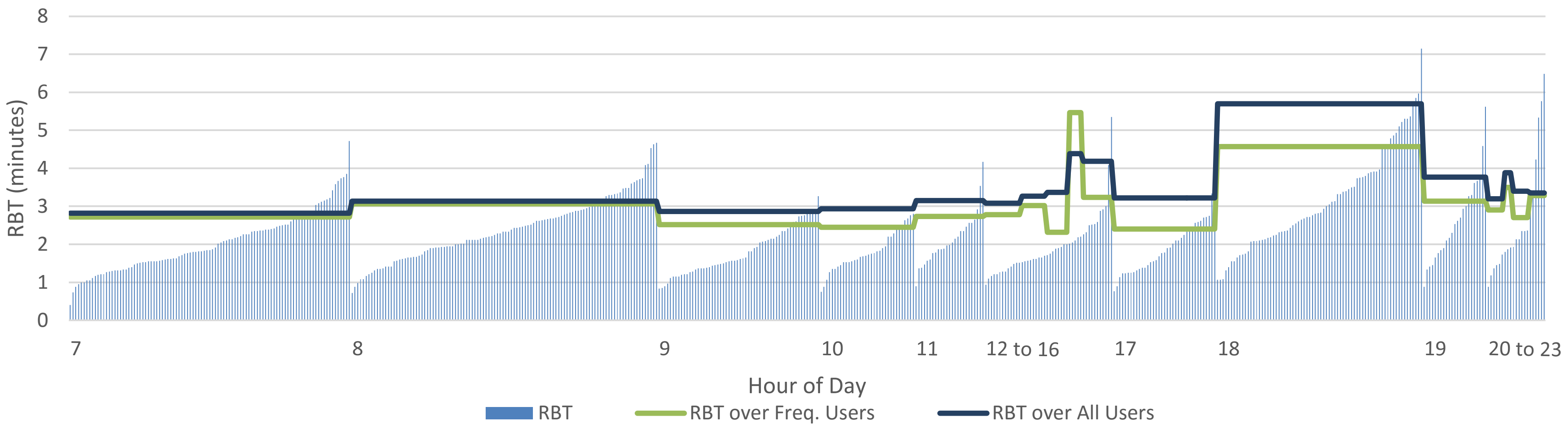
Journey Component Variants

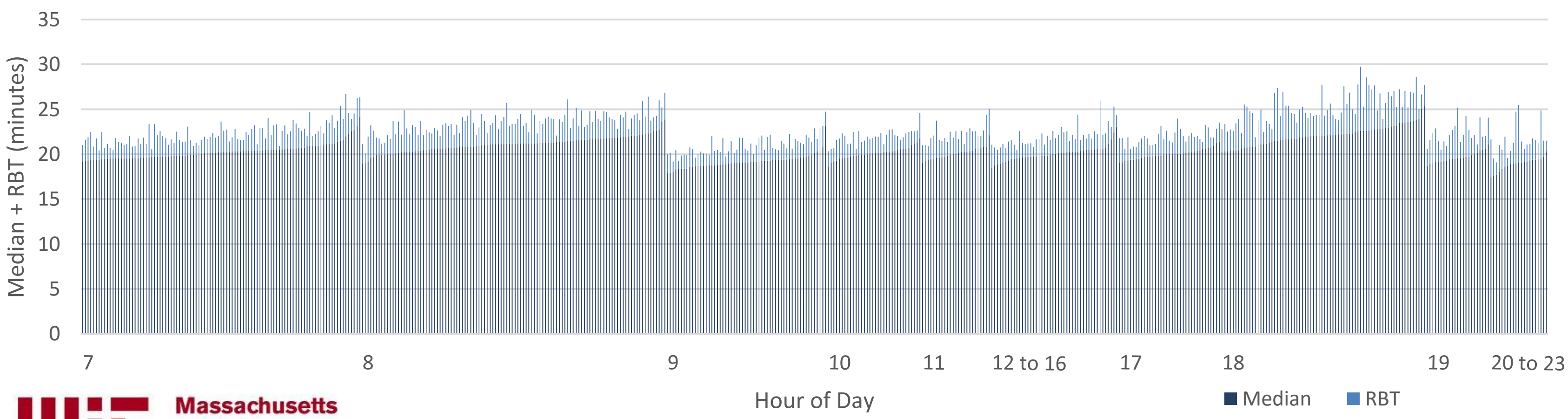
- Full Journey Time
- Platform to Platform Time
- In-vehicle Time

Passenger Variants

- All Customers
- Groups of Passengers
- Individual Passengers

Individual vs Group RBTs Example: Single OD Pair, 2 Months Data





Summary

- Complete Journey OD Estimation practical with ADS
  - foundation for many analyses related to customer experience
- Realistic to assess service reliability for individuals and journeys
  - most critical aspect of customer experience
- Targeted on-line surveys an efficient alternative to other survey methods
- Customer classification is critical in understanding the customer experience
- Developing predictive models is a critical research need

Future Prospects

- Panel data combined with full journey OD estimation and journey time provides the basis for extensive customer experience and behavior analysis including:
  - understanding impacts of changes in service and price
  - understanding customer attraction, retention, and attrition
  - informing "information push" customer information strategies
  - documenting the impacts of marketing and promotional strategies
- Strategies in light of Uber-type service and AV technology