Inferring Left Behind Passengers in Congested Metro Systems from Automated Data

Yiwen Zhu, Haris Koutsopoulos, Nigel Wilson

22nd International Symposium in Transportation and Traffic Theory (ISTTT)
Chicago, Illinois, July 2017
Outline

• Background
• Methodology
• Application
  – Synthetic data
  – Actual data
• Extension to trips with route choice
• Conclusion
Background

• Issues and challenges for near capacity operations
  – Crowding
  – Queuing at elevators, gates, etc.
  – Passengers left behind, longer journey times

• Agencies are seeking approaches to better utilize current facilities
  – Better planning
  – Operational strategies
  – Customer communication (e.g. crowding information)
Key Building Block

- Model system: Passenger-to-Itinerary Assignment Model (PIAM)
Two-step Approach

Station layout
AVL
AFC

Access/Egress Time Model
Access/Egress time distributions

Left Behind Model
Passenger segmentation
Left behind probabilities by time interval

Assignment Model
Feasible itinerary set for each passenger
Probability of each itinerary
Train load, Crowding, Individual journey time components.

Aggregate
Two-step Approach

Access/Egress Time Model
- Access/Egress time distributions
- Passenger segmentation

Left Behind Model
- Left behind probabilities by time interval

Assignment Model
- Feasible itinerary set for each passenger
- Probability of each itinerary

Train load, Crowding, Individual journey time components.

Station layout
AVL
AFC

Aggregate
Disaggregate
Left Behind Model

• Estimate the distribution of the number of times passengers are left behind by time interval

• Assumptions:
  – Passenger movements
  – Known access/egress time distributions

• Sample of passengers without transfers
Left Behind Model

Origin Station Platform:

Before 1st Train Departs

Before 2nd Train Departs

Before Mth Train Departs

Boarding Train:

Passenger $i$ taps in

$P(0)$

$P(1)$

$P(2)$

$P(1)$

$P(0)$

$P(0)$

Access time distribution

$t_e = t_{out} - AT_{i1}$

$t_e = t_{out} - AT_{i2}$

$t_e = t_{out} - AT_{iM}$

Egress time distribution

Passenger $i$ taps out

Board 1st train

Board 2nd train

Board Mth train

Destination Station Platform:
Left Behind Model

Origin Station Platform:

Boarding Train:

Destination Station Platform:

\[ L(Z) = \prod_{i=1}^{N} \sum_{j=1}^{M_i} \int_{DT_{i,j}}^{DT_{i,j-1}} f_{\alpha}(t) dt \sum_{k=j}^{M_i} P_{k-j} f_{e}(JT_{i} - AT_{i,k}) \]
Left Behind Model

Origin Station Platform:

Boarding Train:

Destination Station Platform:

\[ L(Z) = \prod_{i=1}^{N} \sum_{j=1}^{M_i} \int_{DT_{i,j}}^{DT_{i,j-1}} f_a(t) dt \sum_{k=j}^{M_i} P_{k-j} f_e(JT_i - AT_{i,k}) \]
Left Behind Model

**Origin Station Platform:**

**Boarding Train:**

**Destination Station Platform:**

\[
L(Z) = \prod_{i=1}^{N} \sum_{j=1}^{M_i} \int_{DT_{i,j}} \sum_{k=j}^{M_i} f_a(t) dt \sum_{P_{k-j}} f_e(JT_i - AT_{i,k})
\]
Left Behind Model

Origin Station Platform:

Before 1st Train Departs

Before 2nd Train Departs

Before Mth Train Departs

Passenger $i$ taps in

P(0) P(1) P(0)

P(2)

P(1)

Access time distribution

Boarding Train:

Board 1st train

Board 2nd train

Board Mth train

$P(0)$

$P(1)$

$P(2)$

$t_{e} = t_{\text{out}} \cdot AT_{i,1}$

$t_{e} = t_{\text{out}} \cdot AT_{i,2}$

$t_{e} = t_{\text{out}} \cdot AT_{i,M}$

Egress time distribution

Destination Station Platform:

Passenger $i$ taps out

$L(Z) = \prod_{i=1}^{N} \sum_{j=1}^{M_{i}} \int_{DT_{i,j}}^{DT_{i,j-1}} f_{a}(t) dt \sum_{k=j}^{M_{i}} P_{k-1} f_{e}(JT_{i} - AT_{i,k})$
The distribution of the number of times passengers are left behind by time interval is estimated from journey time observations from non-transfer passengers (AFC data) using maximum likelihood or Bayesian inference.
Bayesian Inference

• Prior distribution:

\[ Z = [P_{\downarrow 0}, P_{\downarrow 1}, P_{\downarrow 2}, ...]^T \sim Dir(\alpha) \]

• Posterior distribution of \( Z \) given observations \( X \) (journey times from AFC data):

\[ P_{ZX} = \frac{P(X,Z)}{P(X)} = \frac{P(X|Z)P(Z)}{P(X)} \propto PXZ \cdot P(Z) \triangleq P(Z) \]

• \( P_{ZX} \) is hard to compute \( \Rightarrow \) MCMC
Bayesian Inference

• Prior distribution:

\[ Z = [P_{\downarrow 0}, P_{\downarrow 1}, P_{\downarrow 2}, \ldots] \uparrow T \sim \text{Dir}(\alpha) \]  

Sum up to 1

• Posterior distribution of \( Z \) given observations \( X \) (journey times from AFC data):

\[ P_{Z|X} = \frac{P(X,Z)}{P(X)} = \frac{P(X|Z)P(Z)}{P(X)} \propto PXZ P(Z) \triangleq P(Z) \]

• \( P_{Z|X} \) is hard to compute \( \rightarrow \) MCMC
Bayesian Inference

• Prior distribution:

\[ Z = [P_0, P_1, P_2, \ldots]^T \sim \text{Dir}(\alpha) \quad \text{Sum up to 1} \]

• Posterior distribution of \( Z \) given observations \( X \) (journey times):

\[
P(Z|X) = \frac{P(X,Z)}{P(X)} = \frac{P(X|Z)P(Z)}{P(X)} \propto PXZ P(Z) \triangleq P(Z)
\]

• \( P(Z|X) \) is hard to compute \( \Rightarrow \) MCMC

\[
L(Z) = \prod_{i=1}^{N} \sum_{j=1}^{M_i} \int_{D_{T_{i,j}}} \int_{f_a(t)} \sum_{k=j}^{M_i} P_{k-j} f_e(JT_i - AT_{i,k})
\]
Synthetic Data

• From a major, congested subway system

• Synthetic data
  – Actual tap-in times and train arrivals and departures at stations
  – Access/Egress times based on walk speed distribution and distance based on station layout
  – Passengers board based on FCFS
  – Tap-out time generated
Sampling Process of the MH Algorithm

(a) Off-peak

(b) peak
Average Number of Times Left Behind

![Graph showing average number of times left behind over expected arrival time on the platform. The graph includes lines for actual and estimated data.](image)
Average Number of Times Left Behind

Expected Arrival Time on the Platform
ACTUAL DATA

• Left behind estimation
  • Compared with survey data
  • Stress test: occupy central
Probability of Boarding the First Train to Arrive

- Estimation of left behind
- Transfer stations

<table>
<thead>
<tr>
<th>Time of Day (For passengers leaving Station 2)</th>
<th>Probability of Boarding the First Train to Arrive</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:00</td>
<td>0.25</td>
</tr>
<tr>
<td>16:00</td>
<td>0.50</td>
</tr>
<tr>
<td>17:00</td>
<td>0.75</td>
</tr>
<tr>
<td>18:00</td>
<td>1.00</td>
</tr>
<tr>
<td>19:00</td>
<td>0.75</td>
</tr>
<tr>
<td>20:00</td>
<td>0.50</td>
</tr>
</tbody>
</table>

46.65%, Survey Data
43.89%, PIAM
EXTENSION:

TRIPS WITH ROUTE CHOICE
Three-step Approach

- **Station layout**
- **AVL**
- **AFC**

**Access/Egress Time Model**
- Access/Egress time distributions

**Left Behind Model**
- Passenger segmentation
  - Left behind probabilities by time interval

**Route Choice Model**
- Passenger segmentation
  - Route choice fractions by time interval

**Assignment Model**
- Feasible itinerary set for each passenger
  - Probability of each itinerary
  - Train load, Crowding, Individual journey time components.
Three-step Approach

- **Station layout**
- **AVL**
- **AFC**

**Access/Egress Time Model**
- Access/Egress time distributions

**Left Behind Model**
- Passenger segmentation
- Left behind probabilities by time interval

**Route Choice Model**
- Passenger segmentation
- Route choice fractions by time interval

**Assignment Model**
- Feasible itinerary set for each passenger
- Probability of each itinerary

Train load, Crowding, Individual journey time components.
Three-step Approach

Station layout
AVL
AFC

Access/Egress Time Model
Access/Egress time distributions

Left Behind Model
Passenger segmentation
Left behind probabilities by time interval

Route Choice Model
Passenger segmentation
Route choice fractions by time interval

Assignment Model
Feasible itinerary set for each passenger
Probability of each itinerary

Train load,
Crowding,
Individual journey time components.

Aggregate
Disaggregate
Extension to Trips with Route Choice

Passenger taps-in

### Route 1
- Before Train 1_1 Departs
- Before Train 1_2 Departs
- Before Train 1_3 Departs
  - Train 1_1
  - Train 1_2
  - Train 1_3
- Before Train 2_1 Departs
- Before Train 2_2 Departs
- Before Train 2_3 Departs
  - Train 2_1
  - Train 2_2
  - Train 2_3
- $t_{ie} = t_{out} - AT_{i,1,2}$
- $t_{ie} = t_{out} - AT_{i,2,2}$
- $t_{ie} = t_{out} - AT_{i,3,2}$
- Passenger taps out

### Route 2
- Before Train 1_1 Departs
- Before Train 1_2 Departs
- Before Train 1_3 Departs
  - Train 1_1
  - Train 1_2
  - Train 1_3
- Before Train 2_1 Departs
- Before Train 2_2 Departs
- Before Train 2_3 Departs
  - Train 2_1
  - Train 2_2
  - Train 2_3
- $t_{ie} = t_{out} - AT_{i,1,2}$
- $t_{ie} = t_{out} - AT_{i,2,2}$
- $t_{ie} = t_{out} - AT_{i,3,2}$
- Passenger taps out
Extension to Trips with Route Choice

**Passenger taps-in**

**Route 1**
- Before Train 1_1 Departs
- Before Train 1_2 Departs
- Before Train 1_3 Departs

**Train 1_1**
- $t_{i_e} = t_{out} - AT_{i,1,2}$

**Train 1_2**
- $t_{i_e} = t_{out} - AT_{i,2,2}$

**Train 1_3**
- $t_{i_e} = t_{out} - AT_{i,3,2}$

**Route 2**
- Before Train 2_1 Departs
- Before Train 2_2 Departs
- Before Train 2_3 Departs

**Train 2_1**
- $t_{i_e} = t_{out} - AT_{i,1,2}$

**Train 2_2**
- $t_{i_e} = t_{out} - AT_{i,2,2}$

**Train 2_3**
- $t_{i_e} = t_{out} - AT_{i,3,2}$

**Passenger taps-in**
Extension to Trips with Route Choice

Passenger taps-in

P(r₁)

Route 1

Before Train 1_1 Departs

Train 1_1

Before Train 1_2 Departs

Train 1_2

Before Train 1_3 Departs

Train 1_3

Before Train 2_1 Departs

Train 2_1

Before Train 2_2 Departs

Train 2_2

Before Train 2_3 Departs

Train 2_3

Passenger / taps out

Before Train 2_1 Departs

Train 2_1

Before Train 2_2 Departs

Train 2_2

Before Train 2_3 Departs

Train 2_3

P(r₂)

Route 2

Before Train 1_1 Departs

Train 1_1

Before Train 1_2 Departs

Train 1_2

Before Train 1_3 Departs

Train 1_3

Before Train 2_1 Departs

Train 2_1

Before Train 2_2 Departs

Train 2_2

Before Train 2_3 Departs

Train 2_3

Passenger / taps out

Before Train 2_1 Departs

Train 2_1

Before Train 2_2 Departs

Train 2_2

Before Train 2_3 Departs

Train 2_3

\( t_{ie} = t_{out} \cdot At_{i,1,2} \)

\( t_{ie} = t_{out} \cdot At_{i,2,2} \)

\( t_{ie} = t_{out} \cdot At_{i,3,2} \)
Extension to Trips with Route Choice

Passenger taps-in

$P(r_1)$

$P(r_2)$

**Route 1**

- Before Train 1_1 Departs
  - Train 1_1
  - $t_{i1} = t_{out} \cdot At_{i1,2}$

- Before Train 1_2 Departs
  - Train 1_2
  - $t_{i2} = t_{out} \cdot At_{i2,2}$

- Before Train 1_3 Departs
  - Train 1_3
  - $t_{i3} = t_{out} \cdot At_{i3,2}$

**Route 2**

- Before Train 2_1 Departs
  - Train 2_1
  - $t_{i1} = t_{out} \cdot At_{i1,2}$

- Before Train 2_2 Departs
  - Train 2_2
  - $t_{i2} = t_{out} \cdot At_{i2,2}$

- Before Train 2_3 Departs
  - Train 2_3
  - $t_{i3} = t_{out} \cdot At_{i3,2}$

Passenger taps out
Extension to Trips with Route Choice

\[ P(r_1) \]

\[ P(r_2) \]

**Route 1**

- **Before Train 1_1 Departs**
  - Train 1_1
  - \( t_{ie} = t_{out} \cdot AT_{i,1,2} \)

- **Before Train 1_2 Departs**
  - Train 1_2
  - \( t_{ie} = t_{out} \cdot AT_{i,2,2} \)

- **Before Train 1_3 Departs**
  - Train 1_3
  - \( t_{ie} = t_{out} \cdot AT_{i,3,2} \)

**Route 2**

- **Before Train 1_1 Departs**
  - Train 1_1
  - \( t_{ie} = t_{out} \cdot AT_{i,1,2} \)

- **Before Train 1_2 Departs**
  - Train 1_2
  - \( t_{ie} = t_{out} \cdot AT_{i,2,2} \)

- **Before Train 1_3 Departs**
  - Train 1_3
  - \( t_{ie} = t_{out} \cdot AT_{i,3,2} \)

**Passenger taps-in**
Extension to Trips with Route Choice

**Route 1**

- Before Train 1_1 Departs
  - Train 1_1
  - $t_{i\epsilon} = t_{\text{out}} - AT_{i,1,2}$

- Before Train 1_2 Departs
  - Train 1_2
  - $t_{i\epsilon} = t_{\text{out}} - AT_{i,2,2}$

- Before Train 1_3 Departs
  - Train 1_3
  - $t_{i\epsilon} = t_{\text{out}} - AT_{i,3,2}$

**Route 2**

- Before Train 2_1 Departs
  - Train 2_1

- Before Train 2_2 Departs
  - Train 2_2

- Before Train 2_3 Departs
  - Train 2_3

**P(r_1)**

- Passenger taps-in

**P(r_2)**
Extension to Trips with Route Choice

Passenger taps-in

**Route 1**

- Before Train 1_1 Departs
  - $P_0$
  - $t_{ie} = t_{out} - AT_{i,1,2}$
- Before Train 1_2 Departs
  - $P_1$
  - $t_{ie} = t_{out} - AT_{i,2,2}$
- Before Train 1_3 Departs
  - $P_2$
  - $t_{ie} = t_{out} - AT_{i,3,2}$

**Route 2**

- Before Train 1_1 Departs
  - $P_0$
  - $t_{ie} = t_{out} - AT_{i,1,1}$
- Before Train 1_2 Departs
  - $P_0$
  - $t_{ie} = t_{out} - AT_{i,2,2}$
- Before Train 1_3 Departs
  - $P_0$
  - $t_{ie} = t_{out} - AT_{i,3,2}$

Passenger taps-out
Extension to Trips with Route Choice

\[ P(t_{\downarrow} i_{\uparrow} \text{out}) = P(t_{\downarrow} i_{\uparrow} \text{out} | r\downarrow 1 )P(r\downarrow 1 ) + P(t_{\downarrow} i_{\uparrow} \text{out} | r\downarrow 2 )P(r\downarrow 2 ) \]
Route Choice Fraction Estimation

• A number of approaches have been proposed to estimate route choice shares using AFC data (e.g. Fu et al., 2014; Lee and Sohn, 2015).

• Methods do not capture congestion effects.
Route Choice Fraction Estimation

- A number of approaches have been proposed to estimate route choice shares using AFC data (e.g. Fu et al., 2014; Lee and Sohn, 2015).
- Methods do not capture congestion effects.
Synthetic Data: Route Choice Fractions

- 4 OD pairs simulated
- Two routes → Station 2 or Station 4 (terminal station) as transfer station
- Passengers were assigned to different routes based on a fixed route share $P(r_1) = 0.8$ or $0.99$

(a) $P(r_1)_{true} = 0.8$

(b) $P(r_1)_{true} = 0.99$
Conclusion

- Left behind probability estimation per station and time period from AFC and AVL data
- Estimation of route choice fractions using left behind probabilities and AFC and AVL data
- Components of Passenger-to-Itinerary Model
- Possible extension
Thanks! Q & A
**Left Behind Model**

- Estimate the distribution of the number of times passengers are left behind by time interval
- Assumptions:
  - Passenger movements
  - Known access/egress time distributions
- Sample of passengers without transfers
Bayesian Inference

• Prior distribution:

\[ Z = [P_{\downarrow 0}, P_{\downarrow 1}, P_{\downarrow 2}, \ldots] \uparrow T \sim \text{Dir}(\alpha) \]  

Sum up to 1

• Posterior distribution of \( Z \) given observations \( X \) (journey times):

\[ P(Z|X) = \frac{P(X,Z)}{P(X)} = \frac{P(X|Z)P(Z)}{P(X)} \propto P(X|Z)P(Z) \triangleq P(Z) \]

• \( P(Z|X) \) is hard to compute \( \rightarrow \) MCMC
Metropolis-Hastings (MH) Algorithm

- Generate a candidate sample from a proposal distribution $Q$, as a random walk process, then accept or reject the new sample based on specific selection criteria.
  - Use Dirichlet distribution centered at the current sample as $Q$:

$$Q \uparrow^{*} \uparrow^{(\tau)} \sim \text{Dir}(\alpha^{\uparrow} (\uparrow^{(\tau)}))$$

$$\alpha^{\uparrow} (\uparrow^{(\tau)}) = \frac{\uparrow^{(\tau)}}{\min(\uparrow^{(\tau)})} \geq 1$$

- Acceptance ratio:

$$A \uparrow^{*} \uparrow^{(\tau)} = \min\{1, \frac{P(\uparrow^{*})}{P(\uparrow^{(\tau)})} \frac{Q(\uparrow^{(\tau)} | \uparrow^{*})}{Q(\uparrow^{*} | \uparrow^{(\tau)})}\}$$
Metropolis-Hastings (MH) Algorithm

- Generate a candidate sample from a proposal distribution $Q$, as a random walk process, then accept or reject the new sample based on specific selection criteria.
  - Use Dirichlet distribution centered at the current sample as $Q$:
    \[
    QZ^{\uparrow*} \sim Dir(\alpha^{\uparrow} (Z^{\uparrow}(\tau)))
    \]
    \[
    \alpha^{\uparrow} (Z^{\uparrow}(\tau)) = Z^{\uparrow}(\tau) / \min(Z^{\uparrow}(\tau)) \geq 1
    \]
  - Acceptance ratio: \[ AZ^{\uparrow*} Z^{\uparrow}(\tau) = \min\{1, P(\frac{Z^{\uparrow*}}{Z^{\uparrow}(\tau)}) \} \]
    \[
    \quad / P(Z^{\uparrow}(\tau)) \cdot Q(Z^{\uparrow}(\tau) | Z^{\uparrow*}) / Q(Z^{\uparrow*} | Z^{\uparrow}(\tau)) \}
    \]
    \[ \Rightarrow \text{Leads to convexity} \]
Dirichlet Distribution

- Flexible form for the parameters’ distribution (concave or convex depends on $\alpha$)
- Leads to sparsity
Stress Test: Occupy Central

• On 2014/10/03, demand increased by 10% at Central
Stress Test: Occupy Central

- Left behind at CEN on 2012/08/30 compared with 2014/10/03

(a) 6:00 to 6:20PM  
(b) 6:20 to 6:40PM
The probability of transferring at Station 2 decreases due to possible crowding. Some passengers prefer to transfer at Station 3, the terminal station.
Actual Data: Route Choice Fractions

- The probability of transferring at Station 2 decreases due to possible crowding. Some passengers prefer to transfer at Station 3, the terminal station.
Passenger-to-Itinerary Model (PIAM)

- Station layout
- AVL
- AFC

**Access/Egress Time Model**
- Access/Egress time distribution

**Left Behind Model**
- Passenger segmentation
- Left behind probabilities by time interval

**Assignment Model**
- Feasible itinerary set for each passenger
- Probability of each itinerary
- Train load,
  Crowding,
  Individual journey time components.
Passenger-to-Itinerary Model (PIAM)

**PIAM**

- **Station layout**
- **AVL**
- **AFC**

**Access/Egress Time Model**
- Access/Egress time distribution

**Left Behind Model**
- Passenger segmentation
- Left behind probabilities by time interval

**Assignment Model**
- Feasible itinerary set for each passenger
- Probability of each itinerary

Train load,
Crowding,
Individual journey time components.
Passenger-to-Itinerary Model (PIAM)

PIAM

- **Station layout**
- **AVL**
- **AFC**

**Access/Egress Time Model**
- Access/Egress time distribution

**Left Behind Model**
- Passenger segmentation
- Left behind probabilities by time interval

**Assignment Model**
- Feasible itinerary set for each passenger
- Probability of each itinerary

**Train load,**
- **Crowding,**
- **Individual journey time components.**

**Aggregate level**

**Disaggregate level**
• Incorporating left behind information reduces dimensionality, especially for transfer trips.
Incorporating left behind information reduces dimensionality, especially for transfer trips.

• Left behind model can also be used independently to generate performance metrics.
Opportunities

• 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 resources
Key Building Block

• Passenger-to-Itinerary Assignment Model (PIAM)
  – A platform for assigning passengers to the trains they actually boarded on a given day

• Applications:
  – Inference of a passenger’s movements at a high resolution
  – Service quality metrics, such as left behind probabilities, crowding levels at stations and on platforms
  – Enhanced customer communication
Problem Description

Probability for each passenger $i$ to board train $j$ is calculated based on access/egress time distribution.

$P_i(3|tap - out \ at \ 6:20) = 0.5$

$P_i(2|tap - out \ at \ 6:20) = 0.3$

$P_i(1|tap - out \ at \ 6:20) = 0.2$

Tap-in at 6:00

6:01~6:10

6:05~6:15

6:08~6:18

Tap-out at 6:20
Problem Description

- Large number of feasible itineraries
- Straightforward extension of the approach for non-transfer trips not adequate.
Assignment Model

- Large number of feasible itineraries

- Input
  - Access/egress/transfer time distributions
  - Left behind probabilities

\[ \text{Before } A^\text{st} \text{ Train Departs} \]
\[ \text{Before } B^\text{nd} \text{ Train Departs} \]
\[ \text{Before } M^\text{th} \text{ Train Departs} \]

\[ \text{Board } A^\text{st} \text{ train} \]
\[ \text{Board } B^\text{nd} \text{ train} \]
\[ \text{Board } M^\text{th} \text{ train} \]

\[ t_i = t_{\text{out}} - AT_{i,1} \]
\[ t_i = t_{\text{out}} - AT_{i,2} \]
\[ t_i = t_{\text{out}} - AT_{i,M} \]

\[ \text{Before } 1^\text{st} \text{ Train Departs} \]
\[ \text{Before } 2^\text{nd} \text{ Train Departs} \]
\[ \text{Before } M^\text{th} \text{ Train Departs} \]

\[ \text{Board } 1^\text{st} \text{ train} \]
\[ \text{Board } 2^\text{nd} \text{ train} \]
\[ \text{Board } M^\text{th} \text{ train} \]

\[ \text{Passenger } i \text{ taps in} \]
\[ \text{Passenger } i \text{ taps out} \]
Assignment Model

- Large number of feasible itineraries

- Input
  - Access/egress/transfer time distributions
  - Left behind probabilities
Assignment Model

- Large number of feasible itineraries

- Input
  - Access/egress/transfer time distributions
  - Left behind probabilities
Assignment Model

• Large number of feasible itineraries

• Input
  – Access/egress/transfer time distributions
  – Left behind probabilities
Previous Literature

- PIAM infers trip details on a particular day while the traditional schedule-based assignment models are planning tools that target future conditions.

- Operational applications
  - Buneman, 1984; Kusakabe et al., 2010 and Zhao et al., 2017 last train assumption based on timetable
  - Paul (2010) deterministically assigns individual passengers to train trips.
  - Horchel et al. (2017) ignores the dynamics at transfer stations.

- Our model explicitly considers the dynamics at origin, transfer, and destination stations and captures the effects of capacity constraints.

- The method is probabilistic and applicable to trips with/without transfer/route choices.
Probabilistic Framework

Station layout

AVL

AFC

Access/Egress Time Model

PIAM

Left Behind Model

Passenger Assignment Model

Crowding Model

Distribution of number of times a passenger is left behind

Crowding in stations and on platforms

Passenger movements, train loads, crowding, etc.
Assumption Validation

- The left behind probability is the same for transfer and non-transfer passengers.
Probability of Boarding the Actual Itinerary

(a) Trips 1-2

(b) Trips 2-3

(c) Trips 1-3
Benchmark to Horcher et al., 2017

Itinerary Assignment

Journey Time Components

(b) Trips 1-3
Train Load Estimation

- 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.
Train Load Estimation (Blue Line)
Probability of Using Actual Itinerary

- 4 OD pairs simulated
- Two routes → Station 2 or Station 4 (terminal station) as transfer station
- Passengers were assigned to different routes based on a fixed route share $P(r_1) = 0.8$ or 0.99

(a) $P(r_1)_\text{true} = 0.8$

(b) $P(r_1)_\text{true} = 0.99$
Three-step Approach

• Route choice estimation

\[
\max_{P(r) \forall r \in \mathcal{R}} \quad \mathcal{L} = \sum_{i=1}^{N} \log \sum_{r \in \mathcal{R}_i} \sum_{I \in \mathcal{I}_{i,r}} P(I, t^\text{out}_i | r) P(r) \\
\text{s.t.} \quad \sum_{r \in \mathcal{R}} P(r) = 1 \\
P(r) \geq 0, \forall r \in \mathcal{R}
\]

• Passenger assignment

Passenger taps-in

\[P(r_1) \quad P(r_2)\]
Synthetic Data: Route Choice Fractions

- 4 OD pairs simulated
- Two routes \( \rightarrow \) Station 2 or Station 4 (terminal station) as transfer station
- Passengers were assigned to different routes based on a fixed route share \( P(r_1) = 0.8 \) or 0.99
Actual Data: Route Choice Fractions

• The probability of transferring at Station 2 decreases due to possible crowding. Some passengers prefer to transfer at Station 3, the terminal station.
Implication: Driven by Critical Passengers

• Passengers with $P(t↓i↑out \, | \, r↓1) \ll P(t↓i↑out \, | \, r↓2)$ or $P(t↓i↑out \, | \, r↓1) \gg P(t↓i↑out \, | \, r↓2)$

|       | $P(\, | \, r↓1)$ | $P(\, | \, r↓2)$ | $P(\, | \, r↓1)P(\) + P(\, | \, r↓2)P(\)$ |
|-------|-----------------|-----------------|------------------------------------------|
| 1     | 0.99            | 0.01            | $0.99P(\) + 0.01P(\) = 0.98P(\) + 0.01$ |
| 2     | 0.99            | 0.01            | $0.99P(\) + 0.01P(\) = 0.98P(\) + 0.01$ |
| 3     | 0.99            | 0.01            | $0.99P(\) + 0.01P(\) = 0.98P(\) + 0.01$ |
| 4     | 0.5             | 0.5             | $0.5P(\) + 0.5P(\) = 0.5$ (constant)     |
| 5     | 0.5             | 0.5             | $0.5P(\) + 0.5P(\) = 0.5$ (constant)     |
| 8     | 0.01            | 0.99            | $0.01P(\) + 0.99P(\) = -0.98P(\) + 0.99$ |

- Type I
- Type II

• The estimator of $P(\) captures the share of Type I passengers among the critical passengers

• Critical passengers is a random sample of the whole population
  • Just happened to have a very reasonable egress time for one route and happened to be not likely for another
  • E.g. $egt1 = \{1s, 122s, 422s\}$, $egt2 = \{60s, 180s, 360s\}$
Log Likelihood Ratio

- Histogram of
  \[ \log \frac{P(t_i^{out} | r_1)}{P(t_i^{out} | r_2)} \]

- 0: equal likelihood
- 1: \( P(t\downarrow i\uparrow out | r\downarrow 1) \) is 10 times larger
- Many passengers has \( P(t\downarrow i\uparrow out | r\downarrow 1) \gg P(t\downarrow i\uparrow out | r\downarrow 2) \)

Figure 28: Ratio of the conditional likelihoods \( (P(r)_{true} = 0.8) \)