A Probabilistic Framework for Passenger-to-Itinerary Assignment Using AFC and AVL Data

Yiwen Zhu\textsuperscript{a,1}, Haris N. Koutsopoulos\textsuperscript{a}, Nigel H.M. Wilson\textsuperscript{b}

\textsuperscript{a}Northeastern University, Boston, MA 02115, U.S.A
\textsuperscript{b}Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A

As transit systems in urban areas become more congested during the peak hours the level of service deteriorates, with passengers often failing to board the first train due to capacity constraints, experiencing long journey times, severe crowding, and long queues. To cope with the rapidly increasing travel demand, agencies are seeking approaches to better utilize the current facilities through better planning and operational strategies.

Using Automatic Fare Collection (AFC) and Automatic Vehicle Location (AVL) data, the paper proposes a data-driven approach to infer the impact of near capacity operations on customers and evaluate system performance from a customer's point of view, by developing a methodology to assign individual passengers to the trains they actually boarded. The problem has received little attention in the literature and prior models are based on a number of simplifying assumptions (Buneman, 1984; Kusakabe et al., 2010; Paul, 2010; Zhu, 2014).

This paper builds on prior work (Zhu, 2014; Paul, 2010) and develops a Passenger-to-Itinerary Assignment Model (PIAM) that is applicable under capacity constraints and for trips with/without transfers (but not with route choices). The tap-in and tap-out times of a passenger and the train arrival/departure times at stations are used to infer the probabilities of boarding different feasible itineraries (consisting of trains in all segments of the trip). PIAM first estimates the probabilities of being left behind at the aggregate level by passenger groups, and second, at the individual level, assigns each passenger to feasible itineraries. The contribution of this research is threefold:

1. It provides a complete inference of a passenger’s movements at a high resolution level.
2. At the aggregate level, the output information (left behind, load estimation, average waiting time, etc.) provides useful performance metrics for the operators to assess the system utilization and evaluate the impact of near capacity operations on passengers.
3. The model output can enhance customer communications with information such as expected crowding levels and congested hot-spots.

Figure 1 illustrates the overall approach. The left behind model groups passengers based on their expected arrival times at the station platform and estimates the probability of being left behind \( n \) times \( (n = 0,1,2,...) \) by station and time interval. The assignment model assigns each passenger to the feasible itineraries based on the group the passenger belongs to, the

\textsuperscript{1} Corresponding author. Email address: zhu.yiwe@husky.neu.edu (Yiwen Zhu)

Submitted to the 2\textsuperscript{nd} International Workshop TransitData2016
probability of left behind, and the walking speed distribution. The output of PIAM includes the train loads, crowding at stations, and the left behind.

Since data about the actual trains a passenger boarded is not available, in order to validate the proposed methodology, synthetic data was generated using the actual tap-in times and the train movement data for a small segment of a busy network during the peak period. It was assumed that passenger walking speeds follow a log-normal distribution with mean 1.02 m/s and standard deviation 0.20 m/s (from actual observations of passengers at various stations in the system). The movement of the passengers through the system was simulated assuming passengers board trains with available space on a first come first serve (FCFS) basis (according to their arrival times at the platform).

Figure 2: Average Number of Times Being Left Behind
Figure 2 shows the estimation of the number of times passengers are left behind at one of the most congested stations in the analyzed network compared with the synthetic data. During the peak period, the majority of the passengers were assigned to the actual itinerary with very high probability. The load on trains was also accurately estimated.

References


