CASPT 2018

Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data

Andre L. Carrel · Peter S.C. Lau · Rabi G. Mishalani · Raja Sengupta · Joan L. Walker

Abstract While transit agencies have increasingly adopted systems for collecting data on passengers and vehicles, the ability to derive high-resolution passenger trajectories and directly associate them with transit vehicles in a general and transferable manner remains a challenge. In this paper, a system of integrated methods is presented to reconstruct and track travelers' usage of transit at a detailed level by matching location data from smartphones to automatic transit vehicle location (AVL) data and by identifying all out-of-vehicle and in-vehicle portions of the passengers' trips. High-resolution travel times and their relationships with the timetable are then derived. Approaches are presented for processing relatively sparse smartphone location data in dense transit networks with many overlapping bus routes, distinguishing waits and transfers from non-travel related activities, and tracking underground travel in a Metro network. The derived information enables a range of analyses and applications, including the development of user-centric performance measures. Results are presented from an implementation and deployment of the system on San Francisco's Muni network. Based on 103 ground-truth passenger trips, the detection accuracy is found to be approximately 93%. A set of example applications and findings presented in this paper underscore the value of the

Corresponding Author:
Andre L. Carrel
The Ohio State University
470 Hitchcock Hall
2070 Neil Avenue
Columbus, OH 43210
Tel.: +1-614-292-2771
Fax: +1-614-292-3780
E-mail: carrel.20@osu.edu

Peter S.C. Lau, Oakland, CA
Rabi G. Mishalani, The Ohio State University
Raja Sengupta, University of California at Berkeley
Joan L. Walker, University of California at Berkeley

previously unattainable high-resolution traveler-vehicle coupled movements on a large-scale basis.

Keywords Transit Passengers · Smartphone Data · AVL Data · Location Tracking · Travel Time Components · Performance Monitoring

1 Motivation and background

Tracking the trajectories and travel times of transit passengers as they travel through a transit network can generate valuable data and insights for use in operational and planning applications. For instance, data on exact boarding, transfer, and alighting stops can be used for the generation of origin-destination (OD) matrices, and disaggregate data on individual access and egress times, wait times, transfer times, and in-vehicle travel times can support the derivation of user-based reliability and performance metrics to complement supply-side metrics that are currently in use. Nonetheless, the complexity of travel on transit makes the tracking of passengers across segments and the measurement of the individual travel time components challenging.

So far, the data that have been available for these applications were from surveys and automatic data collection systems, most notably automatic vehicle location (AVL), fare collection (AFC) and passenger count (APC) data. In fully gated systems, where the fare card needs to be tapped both upon entry and exit from the system, the time spent in the system can be derived, but typically there is limited to no information on out-of-vehicle trip segments. In systems that are not fully gated and do not require passengers to tap their fare cards upon exiting, information on alighting and transfer stops is missing. In open systems without fare gates, or systems that do not require certain categories of passengers (e.g., pass holders) to tap their fare cards, even the information on boarding stops may not be exact. This makes determining transit OD matrices and deriving travel time distributions challenging, as one has to rely on inferences and limit oneself to the observed trip components.

In this paper, we present a new approach to tracking the trajectories of individual transit passengers and measuring the components of passengers' trips with minimal inferences by combining AVL data with smartphone location traces. Individual-level smartphone location data can be collected from dedicated survey apps or from a variety of commercial apps over long time periods and with low respondent burden. These data allow a high resolution view of individual trips as long as the traveler remains above ground, including out-of-vehicle segments and exact information on stops. The methodology presented here takes as input smartphone location data, automatic vehicle location (AVL) data and static timetable data from the General Transit Feed Specification (GTFS). In developing the system, there were several main objectives. The first was to systematically consider the components of transit travel time from a passenger's perspective, to describe them, and to develop a framework that shows how they can be measured. The second objective was to develop an automated, robust system that (a) can handle low-frequency location data, collected approximately every 30 seconds but sometimes at larger intervals, and (b) will work in dense transit networks where many routes overlap and many vehicles may be in the vicinity of a traveler at any given time. The third objective was to develop a methodology for tracking underground travel in a Metro network and combining it with statistics on transit travel above ground. Furthermore, we present an approach for distinguishing wait times at the origin stop and transfer times from legitimate activities.

2 Literature

To derive passenger-focused reliability metrics or transit OD matrices, passenger trips have to be assigned to transit routes, stops and, if possible, vehicle runs. There is a spectrum of methods published in the literature that make use of AVL, AFC and APC data, in various combinations. As is summarized by Zhao et al (2007), who uses AFC data with entry tags only, a common assumption that is made when only entry data are available is that the stop where passengers board on one trip is the stop where they alighted on the previous trip. In fully gated systems, on the other hand, the exit location is known but the routing or wait times may be unknown Chan (2007). Approaches to inferring various travel time components on fully gated underground systems are presented by Sun and Xu (2012), Frumin and Zhao (2012), and Seaborn et al (2009). For above-ground systems, several authors (e.g., Nassir et al (2011); Munizaga and Palma (2012); Gordon et al (2013)) have focused on connecting passenger trips from bus AFC data to vehicle locations observed via AVL data in an effort to better infer boarding locations and times when the boarding stop was not recorded by the AFC system.

While these aforementioned contributions have been very valuable, they have in common that due to the coarse resolution of the data, researchers could not obtain exact measurements of every travel time component, including out-of-vehicle travel times. Work that attempted to disaggregate total travel time into its individual components did so primarily based on distributions of total travel times.

There has been previous work that utilized passenger smartphone location data, but it was mainly focused on determining the travel mode from data collected through location and other sensors (e.g., accelerometer, microphone). This has been performed by map-matching the location points in GIS (Chung and Shalaby (2005); Gong et al (2012)), by extracting features from location and accelerometer data related to velocity, acceleration, distance traveled, or the proximity of transit stops and vehicles and using those as inputs for mode classification algorithms (e.g., Gonzalez et al (2008); Stenneth et al (2011); Parlak et al (2012)) or with a combination of the two methods (Biagioni et al (2009); Thiagarajan et al (2010).

In summary, research utilizing smartphone and AVL data has so far been focused on mode detection, and no study has combined high-resolution smartphone data to capture the demand side with AVL data to capture the supply side. That is the purpose of the methodology presented in this paper.

3 Problem description and definitions

For every passenger for whom location data are available, we want to derive that person's *personal transit travel diary*, i.e., exactly when and from where to where the person traveled on transit, on board which vehicle, and exactly how much time was spent on board and waiting at the origin stop or at transfer stops. These data allow us to generate high-resolution transit OD matrices, and with repeated observations, to derive time distributions at a segment level. In a second step, we want to be able to compare the observed travel times with scheduled travel times on a per-segment basis to obtain deviation measures. For the latter, we are not specifically interested in whether a given transit vehicle was on time; we are interested in how the overall service received by a passenger differed from the service that passenger could expect based on the timetable.

We begin with two sets of raw data, the phone location data L_p and the AVL data L_v. Each data point includes a time stamp, latitude, longitude and a phone identifier (in L_p) or a vehicle identifier (in L_v). The timetable and network information were obtained from the General Transit Feed Specification (GTFS) file published by the agency. The phone location data were obtained from a survey app.

In step 1, which is not described in the remainder of this paper but included here for completeness, the raw AVL data must be matched to route shape files to generate runs. A *run* is defined as a single transit vehicle in revenue service on a fixed route traveling one way from one terminal to another terminal. The vehicle position and stop locations are measured as the distance from the departure terminal, called the *milepost*, and each run is described by a trajectory that includes a set of time-milepost tuples. For stops that fall between two AVL data points, the times are inferred via linear interpolation.

In step 2, the phone location data are matched to transit runs. For any given user trip, it must be determined whether or not transit was used, and if so, which runs the user was on, what the boarding, alighting and transfer stops and corresponding times were. This happens in several steps. First, the user's phone traces are matched to runs and the boarding stop and alighting stop are inferred. This is done in two separate processes: One that finds nearby runs for every phone data point ("Initial vehicle matching"), and one that determines whether any segments of the passenger's trip were underground. The initial matching process on the above-ground (surface) data yields so-called candidate runs. Through further processing, a list of surface trip segments with the respective access, egress and transfer stops is determined. In parallel, a list of underground segments is compiled, and at the end, the two are merged. If the traveler cannot be mapped to any surface or underground trips, this step returns an empty result. The time between the boarding and alighting from a single vehicle constitutes a *transit segment*, and in a next step, all segments that are associated with the same trip (i.e., a user traveling from one activity location to another) are assembled to form an uninterrupted, alternating sequence of transit segments and transfers, plus an origin wait at the beginning of the trip. The trip ends if the user stops to carry out an activity somewhere.

In step 3, we then derive the *personal schedule deviation metrics* by comparing the observed wait times and travel times on each segment to what the traveler could have expected from the timetable. This step takes a passenger-centric rather than an operations-centric view of travel time variability, and since no other input is collected from the traveler, it relies on some assumptions about the desired departure time.

4 Detailed description of the phone-vehicle matching procedure

The procedure described in this paper was developed for analyzing travel patterns on San Francisco's bus and rail transit network, commonly called "Muni". The bus system operates entirely above ground, but the rail system operates both above ground and underground. The location of the phones was sampled every 30 seconds, though gaps of one to two minutes were not uncommon, which complicated the determination of access, egress and transfer stops. In the following discussion, we assume that processing step 1 has already been performed and that the individual vehicle runs have been stored in a database. We first describe the matching of phone points to vehicle location points (section 4.1) and the inference of the access and egress stop (section 4.2). The inference of transfer stops is more complex and breaks down into several subcases, which are presented in section 4.3. Following that, we describe the underground matching problem in section 4.4 and the derivation of the final transit travel diary in section 4.5.

4.1 Determining above-ground transit trips

We first take a user-centric view: given the starting and end point of a trip, and the respective series of location points $L_p = \{(lat_0, lon_0), (lat_1, lon_1), ...\}$, we create a three-dimensional search box defined by $\min(lat_n - \Delta)$, $\max(lat_n + \Delta)$, $\min(lon_n - \Delta)$, $\max(lon_n + \Delta)$, t_0, t_{max} and query the database for all vehicle trajectories (runs) that traverse that box. We then use Dynamic Time Warping to calculate the similarity between the phone trajectory and the vehicle trajectories, using the phone location time stamps for the time steps. As a cost function we use the absolute distance between the points (the time distance being zero) to match a linearly interpolated vehicle trajectory location to every phone location. We then define a distance threshold and filter the vehicle trajectories to obtain a set of candidate runs for every phone location point.

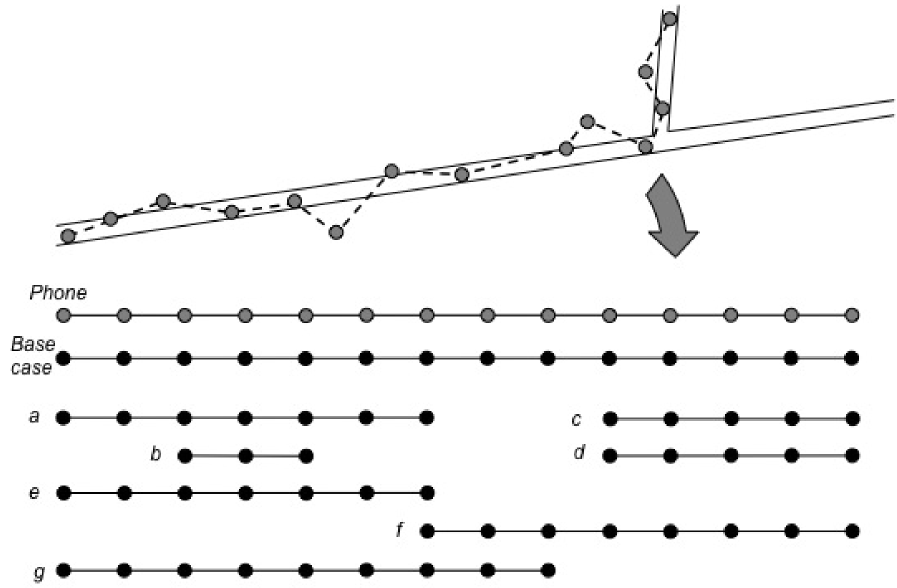

Fig. 1 Possible relationships between candidate runs (a-g).

Next, the transit runs where the user was most likely on board need to be distilled from the candidate list. This can be challenging in dense networks or corridors with overlapping routes, as the phone might be mapped to multiple vehicles simultaneously. Due to positioning errors, the shortest distance between the phone and a vehicle may not necessarily indicate the correct match. Instead, we switch to a run-centric perspective: Given the set of candidate vehicles for every point, we group the data by run ID and create a candidate run list. In practice, this is an associative array with the run ID as key and a second associative array of run characteristics as value. The run characteristics include a list of phone location points where the phone and the vehicle were matched. This is equivalent to projecting the subsets of runs that are close to the phone location onto the one-dimensional trajectory of the phone, as is illustrated in figure 1. The gray line is the trajectory of the phone, with the dots symbolizing location readings. Below, in black, are the sets of points where the phone location matched a transit vehicle location. As an example, in case (b), the phone was observed to be near this particular transit vehicle in the third, fourth and fifth location point recorded by the phone, but in no others. The base case, where the phone matched the same candidate run over the entire trajectory, is illustrated at the top. If the phone is matched to different candidate runs, the following relationships are possible:

1. They can be disjoint, as (a) and (c).
2. They can be identical, as (c) and (d).
3. One can be a subset of the other, as (b) is of (a).
4. They can share one location point, as (e) and (f).

5. They can share multiple location points, as (g) and (f).

First, we eliminate subsets (case 3). If two candidate runs are identically matched (case 2), the one with the higher average distance between the phone and the vehicle is eliminated. We have now retained only runs where the points are either disjoint or overlapping as in 4 and 5. However, this does not allow us to know whether two runs might overlap in reality, so we query the database for every combination of candidate runs in categories 4 and 5 and request the number of possible transfer stops within the overlapping segment. If there is only one possible transfer stop, that pair is reclassified as being disjoint. With only the "truly" overlapping candidate runs remaining in categories 4 and 5, we can now think of the traveler's trajectory as an undirected graph, represented by an adjacency matrix, with an arbitrary set of unconnected trees that each represent a group of overlapping candidate runs.

The groups must now be processed. We traverse the graph in a breadth-first search to extract overlapping runs and sort them by the time stamp of the first location point they matched with the phone, then recursively do the following:

1. Choose r_0 and r_N, the first and last run in the group.
2. Check whether they are disjoint.
3. If no, discard every run in between.
4. If yes, extract the "inner sequence", i.e. the sequence of runs in between r_0 and r_N, and repeat.

After eliminating subsets as described above, every run overlaps with at most one other run at the beginning and the end.

4.2 Inferring access and egress stops

To determine the access stop, we use (l_1, t_1), the phone's location and time at the first point where it was matched with the transit run r, and the last location/time before that, (l_0, t_0). We query the nearest transit stop served by r to each of those two locations, s_1 and s_0, and the departure times, t_{s_1} and t_{s_0}. If s_1 and s_0 are not identical, one of the two is chosen as follows:

- Define a time window $(t_0 - \Delta, t_0 + \Delta)$, and check if t_{s_0} falls within that window.
- Check that the gap between t_1 and t_0 does not exceed a certain threshold.

If both conditions are met, we designate s_0 as the access stop, otherwise s_1. These conditions guard against gaps that are so large that an unobserved activity aside from waiting may have occurred within it. The calculation of the wait time at the access stop is discussed in section 5.1.

4.3 Inferring transfer stops

If two consecutive transit segments are identified for a user, e.g., run A on route 1 and run B on route 2, we first determine whether the time spent in between

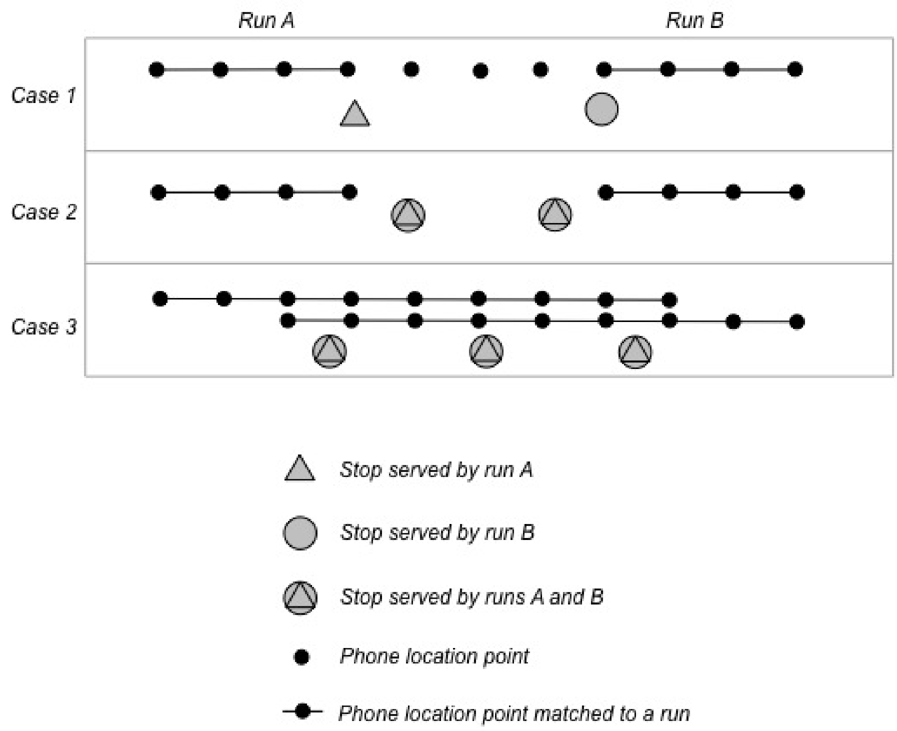

Fig. 2 Scenarios for transfer stop detection.

the two segments was purely a transfer or whether the user stopped at that location to carry out an activity. For that purpose, we check two conditions:

- Whether the user exited a certain perimeter around the transfer location.
- Whether more than a certain number of runs passed on route 2 between the user alighting and boarding.

If any of these conditions is met, the time spent between alighting and boarding is marked as an activity rather than a transfer. If it is determined to be a transfer, the transfer stop may again not be directly observable in the data. Figure 2 illustrates three possible scenarios for the determination of the transfer stop.

In case 1 in figure 2, there are unmatched location points between the two runs, so the access/egress stops for runs A and B can be determined as previously described. In case 2, there are no location points in between, so we query the database for all possible transfer stops between the phone's last sighting on run A and its first sighting on run B. We then eliminate stop combinations that were first served by run B and then by run A. Then, the first stop served by run B after run A is selected as the transfer stop. In case 3, the phone is matched to both runs for a period of time. The database is queried for all possible transfer stops within the segment matching both runs, and again, only the stops are retained that were served by run B after run A. Then, the first stop is chosen where the phone was closer to the vehicle on run B than to the vehicle on run A.

4.4 The underground matching problem

4.4.1 Extracting underground travel segments

Phone location data are generally not available for underground segments, though phones may still generate occasional location readings, either spuriously or in metro stations where a signal is picked up. To identify underground travel, geofences are created around every subway station and tunnel portal in the system, and a ground truth set of typical data collection frequencies from travel above and below ground is collected. The phone location data are then scanned, and every point falling within a geofence is labeled, producing the sequence of geofences traversed. For every pair of geofences, the frequency of the location data collection between them is calculated, and the time spent by the user between the two geofences is classified as above-ground or underground travel based on that frequency. This yields a list of geofence pairs between which the user traveled underground. Sequential in-tunnel segments are combined (e.g., $(A \rightarrow B, B \rightarrow C = A \rightarrow C)$, again using a frequency criterion to detect cases where a person exited at a metro station and re-entered it later.

The location points matched to surface runs and those matched to underground segments are not yet guaranteed to be non-overlapping. As a next step, the relationship between every surface segment and every underground segment is checked, and if the surface segment and the underground segment intersect on one or more points, the surface segment is truncated to exclude the beginning the underground segment. Finally, the routing between the entry and exit point of every underground segment is inferred with a shortest-path algorithm. This approach has limitations in complex networks where no one path between two points is superior, so in other applications, additional information on such as the time spent underground may need to be considered.

4.4.2 Merging underground and surface segments

In the Muni network, rail routes run underground in downtown but emerge to the surface in outer districts. Those tunnel portals were included in the list of geofences. To detect trips where the access was at a surface and the egress at an underground stop or vice-versa, the end points of all remaining surface and underground segments are compared; if an underground segment ends at the tunnel portal where a surface segment begins, or vice versa, the two are combined. At the end of this step, two lists are carried forward:

– A list of surface runs matched to the phone location data, including runs where one of the stations used (access/egress) is underground.

– A list of independent underground segments where both the access and the egress station are underground.

4.4.3 Inferring runs from underground-only data

For underground-only segments, we have underground AVL data but lack phone positioning data. In the Muni network, underground travel involves at most one transfer, and we have only two data points to match the phone to a run: The last time the phone was "seen" inside the origin station geofence, t_{orig}, and the first time inside the destination station geofence, t_{dest}.

– If there was no underground transfer, we assume that all unobserved wait time was incurred at the origin station. After defining a minimum egress time for the given station k, $t_{egress,k}$, we query the database for the run between the origin and the destination station that arrived most recently before $t_{dest} - t_{egress,k}$, then assign the user to it.
– If there was an underground transfer (e.g., from run A to run B), it is unknown whether the unobserved wait time was incurred at the origin or the transfer stop. We arbitrarily assume it was at the transfer stop. Run B is determined as described above, and run A is determined by adding an access time, $t_{access,j}$, to t_{orig}, querying the database for the first run between the origin and the transfer stop that departed after $t_{orig}+t_{access,j}$, and assigning the user to it.

Given the uncertainties associated with underground segments, and the fact that phones can sometimes report an old location if they are not able to acquire a new one, the system should contain checks to ensure that the inferred underground routing is feasible.

4.5 Deriving the final transit travel diary

After ordering the combined list of above-ground and below-ground transit segments by time, we obtain a diary of all transit travel by the user. At this point, the list may still contain segments identified as being on board transit even though the phone user was inside a car or on a bicycle traveling alongside a transit vehicle. Identifying these false positives is difficult based on location data alone and would require additional sensor data, e.g., from the accelerometer or microphone. With no additional sensor data, a simple heuristic approach was chosen by eliminating transit segments that were either below a minimum distance or where less than a given fraction of phone location points were within a minimum distance of the vehicle. This assumes that due to speed differentials between cars, bicycles and transit vehicles, the majority of spurious matches would be short.

In the transit travel diary, a *trip* is series of transit segments between two activity locations. At the beginning of the trip is an *origin wait*, and between each subsequent segment is a *transfer*. This terminology is used in the following sections, where the derivation of reliability measures is described.

5 Comparisons with timetable

With each user's personal transit travel diary in hand, we compare the travel times experienced by the user with what that person could have expected based on the timetable. For that, we first need to calculate the experienced wait times at origin stops and transfer stops for surface segments.

5.1 Characterization of the origin wait

The at-stop wait time is calculated by observing how long a person dwelled within a certain radius around the origin stop before boarding. The challenge is that, especially in dense cities, time spent in the immediate vicinity of the stop before boarding may belong to an activity and not represent wait time. Therefore, we query the database for all runs that served the passenger's access and egress stops while the passenger was observed to be near the access stop. Time spent near a stop is only classified as wait time if:

- At most one run passed without being boarded by the passenger.
- A maximum wait time was not exceeded.

In order to compare a passenger's observed wait time with a wait time that the passenger could have expected based on the timetable, we need to infer an *intended scheduled departure* based on the observed departure time. Many transit planning sites and apps base their information on GTFS, and thus on the static timetable, so a traveler who consults these products can be considered to be cognizant of the timetable. It should be emphasized that this is *not* an operational assignment, i.e., an assignment of observed departures to scheduled departures, but rather, it is an inference about the scheduled departure the passenger *intended* to take. Figure 3 illustrates different cases.

In case A, a departure was scheduled at *s1*, *s2* and *s3*. All three were served by an observed departure, and since the user arrived at the stop between *s1* and *s2*, we infer that the user was intending to take *s2*. Due to crowding, the user was not able to board *d2* (which is counted as a "pass-up") and instead boarded *d3*, so the passenger experienced a deviation of *d3*-*s2*. Case B is very similar, except that the passenger boarded *d2*, which departed slightly before *s2*. In this case, the experienced deviation is *d2*-*s2*. In case C, neither *s1* nor *s2* correspond to an observed departure. The passenger took *d2*, but we do not know whether the passenger was intending to take *s1* and got delayed, or whether the passenger intended to take *s2*, but realized after consulting real-time information that there was no departure at that time. *d3* may have been too late, so the passenger may have hurried to the transit stop to catch *d2*. As an approximation, we assign the user's intended departure to *s1* and *s2* with equal probability and calculate the deviation as an average. The same holds for case D. Lastly, in case E, both *s1* and *s2* correspond to an observed departure, so we conclude that the passenger did not intend to take either of those. We assign a deviation of zero to this case.

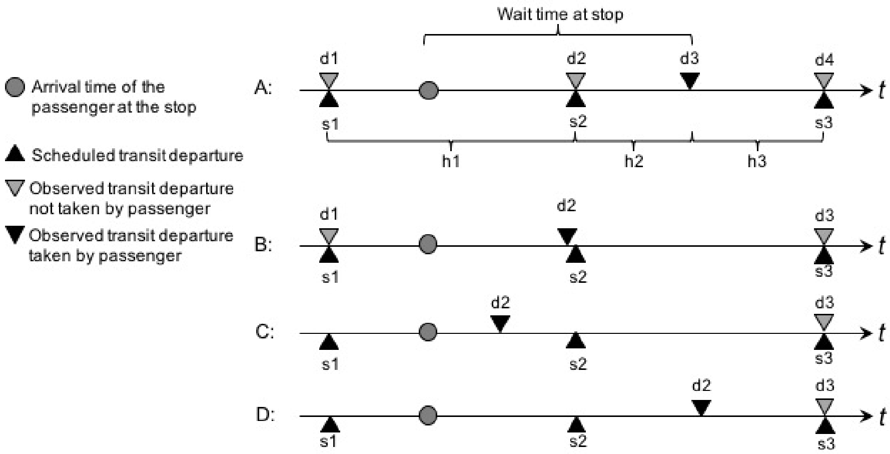

Fig. 3 Components of the origin wait time from a passenger's perspective and inference of intended departure time.

Case A in figure 3 also includes the observed headways, *h1* through *h3*. In addition to pass-ups and the experienced deviation, we track the observed headways preceding and following the departure taken by the passenger and compare them to the scheduled headways during that time period. In case A, the observed headway would be *h2* and *h3*, but since the passenger was unable to board *d2*, we can consider the preceding headway to be *h1+h2*.

5.2 Characterization of the in-vehicle travel time

Once the passenger has boarded a vehicle, the passenger will experience an in-vehicle travel time (IVTT). To calculate the deviation between the experienced IVTT and what the passenger could have expected based on the schedule, we must find a scheduled departure time that is close to the passenger's observed departure time. In cases A or B as described above, the assignment to an intended scheduled departure time has already been made. In the other cases, we assign it to the nearest scheduled departure based on simple heuristic rules. This assignment has no behavioral significance; we only use it to query the scheduled travel time between the passenger's boarding and alighting stop, $ivtt_{sched}$.

For the next steps, we need to know when the passenger could have expected to arrive at the transfer stop or the destination based on the schedule. This can be calculated in different ways, leading to the following definitions:

- The *projected arrival time* is the observed departure time, d, plus the scheduled travel time.
- The *scheduled arrival time* is the scheduled departure time, s, plus the scheduled travel time.

5.3 Characterization of the transfer time

Given a transfer between two runs, we can calculate the observed transfer time. To compare it to the schedule and calculate any deviations, we use the definitions of arrival times at the end of the previous section. Suppose the passenger first used route 1. It was scheduled to depart at *s1*, but was delayed and did not depart until *d1*. It was again delayed while the passenger was on board, adding IVTT. Following the previous section, the scheduled arrival time at the transfer stop would have been *a1*, the projected arrival time would have been *a2* and the observed arrival time was *a3*, with *a1 ¡ a2 ¡ a3*. Note that *d1-s1 = a2-a1*.

At the transfer stop, the passenger transferred to route 2. There are two scheduled departures, at *s2* and later, *s3*. Both are delayed, so *s2* departs at *d2* and *s3* departs at *d3*. Assume that *a1 ¡ a2 ¡ s2 ¡ d2 ¡ a3 ¡ s3 ¡ d3*. The passenger catches the connection at *d3*. Thus:

- The *scheduled* transfer time is *s2-a1*.
- The *projected* transfer time is *d2-a2*.
- The *observed* transfer time is *d3-a3*.

This assumes that neither the scheduled nor the projected transfer times are less than the minimum transfer time for that stop.

Based on these definitions, we can calculate two deviation measures for the observed transfer time: One with respect to the scheduled and one with respect to the projected transfer time. This information is carried forward to generate the scheduled and projected arrival times of the following segment, which is required if a trip involves multiple transfers.

As with the origin wait, we are not only interested in the deviation of the transfer time, but also in missed departures and pass-ups. Using the definitions in this section, we get:

- The number of scheduled and observed departures that occurred between the scheduled arrival time at the transfer stop and the observed arrival time (*a3-a1*). These are connections that the passengers could have made if the first segment(s) of the transit trip had departed on schedule and had not been delayed en route.
- The number of scheduled and observed departures that occurred between the projected arrival time at the transfer stop and the observed arrival time (*a3-a2*). These are connections that the passengers could have made if the first segment(s) of the transit trip had departed as it did, but had not been delayed en route.
- The number of scheduled and observed departures that occurred during the passenger's wait time at the transfer stop (*d3-a3*).

6 Deployment, validation and example analyses

The system was deployed in a real-world, large scale study as part of the *San Francisco Travel Quality Study* (SFTQS), which ran from October to December 2013 and is described in Carrel et al (2017). 756 of the study participants downloaded a survey app and consented to having their phone location tracked during the study. The system described here identified approximately 7700 transit segments on Muni, of which 675 (8.8%) involved a Muni-to-Muni transfer according to this paper's definition.

6.1 Validation

To test the system, ground truth data were collected on a total of 103 transit passenger trips in spring 2014, consisting of 86 above-ground bus or light rail trips and 17 underground rail trips. The evaluation showed that 92 (89.3%) of all runs were correctly identified and, for a further 4 underground trips, the origin/destination stations and times were correctly identified, but due to missing AVL data, the passengers were mapped to different trains. We include those 4 in the list of correctly identified trips since the error was due to the AVL data, and we conclude that the system identified transit trips with 93% accuracy. Out of those 96 trips, the boarding and alighting stops were identified exactly in 83 (86%) cases, and in 95 (98%) cases, the boarding and alighting stops identified were within one stop of the true boarding and alighting stop. These numbers describe the true positive rates. A limitation of this validation procedure is that no ground truth data on non-transit trips was available, so it was not possible to assess the false positive rates.

6.2 Example analysis: Contribution of origin wait and transfer time to total travel time

To demonstrate the value of the disaggregate travel time data, we present a high-level analysis of travel times experienced by passengers in a sample of 322 trips collected as part of the SFTQS. Only trips containing two transit segments and one transfer are included in this analysis, and thanks to the methodology presented here, out-of-vehicle wait times could be directly observed. Figure 4 shows the contribution of origin wait time, transfer time and IVTT to the total travel time - defined as the time between arriving at the origin stop and alighting at the destination stop - experienced by the traveler. As total travel time increases, the contribution of the origin wait time remains between 8% and 14% of total travel time, but the contribution of transfer time appears to increase, from approximately 17% for the shortest trips to 35% for travel times between 50 and 60 minutes. For example, passengers who made a 45-minute trip spent, on average, 31 minutes in a vehicle and 14 minutes waiting. Of those 14 minutes, approximately 9.5 were transfer wait time, which was likely to have caused more disutility to the passenger than origin wait time.

In a more in-depth analysis of 533 trips, travel time variability experienced by a passenger between boarding at the origin stop and alighting at the destination stop was attributed to either IVTT variability, variability of transfer

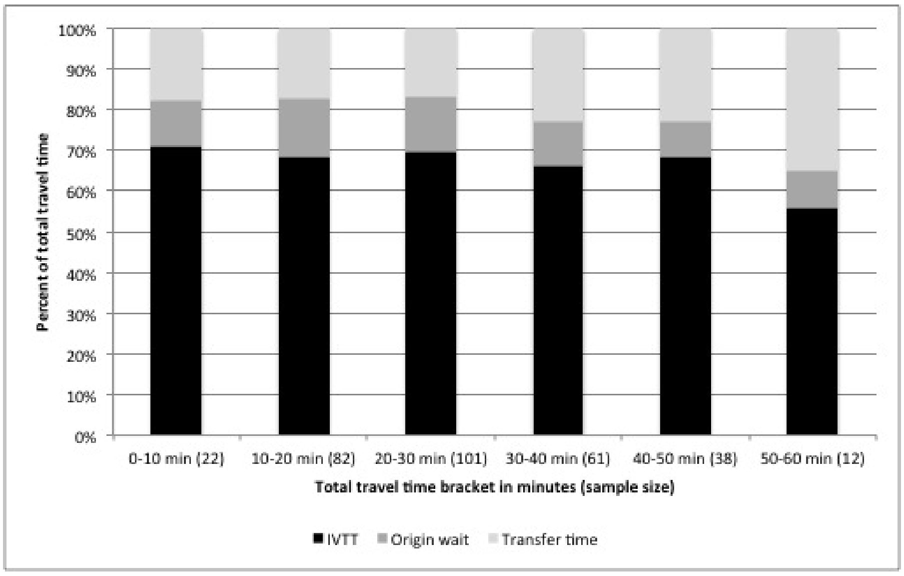

Fig. 4 Contribution of trip segments to overall travel time.

times, or both, according to a methodology described in Carrel et al (2015). Variability was defined as the deviation of the experienced travel time from the scheduled travel time, as defined in sections 5.2 and 5.3. It was found that 26% of the observed deviations from scheduled travel time could be attributed to IVTT variability, whereas 74% could be attributed to transfer time variability. This finding underscores the importance of capturing and tracking out-of-vehicle travel times, as they are a major source of unreliability experienced by travelers.

In a separate analysis of origin wait times, we confirmed that the independence of observed origin wait times from the trip distance, as shown in figure 4, could also be observed in other contexts. The observed wait times were not found to differ strongly between different times of day or between infrequent (headways ¿ 12 minutes) and frequent transit services, suggesting that thanks to real-time information, passengers tended to time their arrivals at the stop consistently and largely independently of the time of day or service type.

7 Conclusions

This paper introduced a system to extract the personal transit travel diary of passengers by matching smartphone location data to AVL data. It described the various components of a transit passenger's trip and how they can be measured, focusing on the problems that can arise when phone location data are sparse and when the phone is near multiple vehicles. Furthermore, it described an approach to detect underground travel on metro networks when AVL data but no phone location data are available. The procedure presented here is of a general nature and can be applied to a variety of different systems; the steps described can be understood as a blueprint. Because location data can be collected from virtually any app, a researcher could use data from third-party apps such as route planners, and the system described here requires no further input from participants.

Detecting the origin, destination and transfer stops with phone location data can support the derivation of OD matrices in systems where only incomplete information, or none at all, is available from AFC data. The sample application presented in this paper is intended to demonstrate the potential value of such data, but there are numerous further applications where the out-of-vehicle travel times, access and egress routes and times, and observed travel time reliability derived from this matching procedure could be valuable.

Although the methodology has its limitations, we hope to address those in future research. Notably, additional sensor data would help reduce false positive rates caused by a person traveling alongside a transit vehicle and for better distinguishing activities from wait times at stops, and a methodology is needed to determine passenger routing in complex underground networks with multiple possible paths. Overall, however, the approach outlined in this paper is a promising step toward delivering data that will enhance the planning and management of transit systems.

References

Biagioni J, Agresta A, Gerlich T, Eriksson J (2009) Transitgenie: A context-aware, real-time transit navigator. In: 7th ACM Conference on Embedded Networked Sensor Systems, ACM, pp 329–330

Carrel A, Lau PS, Mishalani RG, Sengupta R, Walker JL (2015) Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. Transportation Research Part C 58:224–239

Carrel A, Sengupta R, Walker JL, et al (2017) The san francisco travel quality study: Tracking trials and tribulations of a transit taker. Transportation 44(4):643–679

Chan J (2007) Rail transit od matrix estimation and journey time reliability metrics using automated fare data. Master's thesis, Massachusetts Institute of Technology

Chung EH, Shalaby A (2005) a trip reconstruction tool for gps-based personal travel surveys. Transportation Planning and Technology 28(5):381–401

Frumin M, Zhao J (2012) Analyzing passenger incidence behavior in heterogeneous transit services using smartcard data and schedule-based assignment. Transportation Research Record 2274(1):52–60

Gong H, Chen C, Bialostozky E, Lawson CT (2012) A GPS/GIS method for travel mode detection in new york city. Computers, Environment and Urban Systems 36(2):131–139

Gonzalez P, Weinstein J, Barbeau S, Labrador M, Winters P, Georggi NL, Perez R (2008) Automating mode detection using neural networks and assisted GPS data collected using GPS-enabled mobile phones. In: 15th World Congress on Intelligent Transportation Systems

Gordon JB, Koutsopoulos HN, Wilson NH, Attanucci JP (2013) Automated inference of linked transit journeys in london using fare-transaction and vehicle location data. Transportation Research Record 2343:17–24

Munizaga MA, Palma C (2012) Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from santiago, chile. Transportation Research Part C 24:9–18

Nassir N, Khani A, Lee SG, Noh H, Hickman M (2011) Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system. Transportation Research Record 2263:140–150

Parlak S, Jariyasunant J, Sengupta R (2012) Using smartphones to perform transportation mode determination at the trip level. In: Transportation Research Board 91st Annual Meeting, 12-3606

Seaborn C, Attanucci J, Wilson NH (2009) Analyzing multimodal public transport journeys in london with smart card fare payment data. Transportation Research Record 2121(1):55–62

Stenneth L, Wolfson O, Yu PS, Xu B (2011) Transportation mode detection using mobile phones and gis information. In: 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, pp 54–63

Sun Y, Xu R (2012) Rail transit travel time reliability and estimation of passenger route choice behavior. Transportation Research Record 2275(1):58–67

Thiagarajan A, Biagioni J, Gerlich T, Eriksson J (2010) Cooperative transit tracking using smart-phones. In: 8th ACM Conference on Embedded Networked Sensor Systems, ACM, pp 85–98

Zhao J, Rahbee A, Wilson NH (2007) Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering 22(5):376–387