

UC Berkeley

International Conference on GIScience Short Paper Proceedings

Title

Constructing a Routable Transit Network from a Real-time Vehicle Location Feed

Permalink

<https://escholarship.org/uc/item/5vm764rp>

Journal

International Conference on GIScience Short Paper Proceedings, 1(1)

Authors

Wessel, Nate

Allen, Jeff

Farber, Steven

Publication Date

2016

DOI

10.21433/B3115vm764rp

Peer reviewed

Constructing a Routable Transit Network from a Real-time Vehicle Location Feed

Nate Wessel¹, Jeff Allen¹, Steven Farber²

¹Department of Geography and Planning, University of Toronto, Sidney Smith Hall, 100 St. George Street, Toronto, ON M5S 3G3, Canada
Email: {nate.wessel; jeff.allen}@mail.utoronto.ca

²Department of Human Geography, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON M1C 1A4, Canada
Email: steven.farber@utoronto.ca

Abstract

Many transit agencies publish open access feeds of real-time fleet locations for use by application developers to deliver mobile navigation services. However, by collecting a transit data feed over time, we demonstrate how to recreate retrospective, routable transit networks that are useful in answering network performance and accessibility research questions. In this research we develop an end-to-end GIS toolchain for 1) downloading and storing a transit data feed, 2) coercing the collected spatiotemporal database into a retrospective transit schedule adhering to the General Transit Feed Specification (GTFS) data standard, and 3) creating a routable transit network with time-dependent travel times in OpenTripPlanner. We further demonstrate how this toolchain can be used to identify discrepancies between scheduled and actual travel times on the network and motivate the usefulness of this approach through an accessibility analysis.

1. Introduction

The emergence of the General Transit Feed Specification (GTFS) data standard, and the publication of GTFS packages by thousands of agencies worldwide has unleashed a flurry of tools development for researching transit networks (Hadas and Ranjitkar 2012), transit travel times (Farber et al. 2014), and time-dependent accessibility metrics (Fransen et al. 2015; Owen and Levinson 2015; Farber et al. 2016). One problem with travel time research based on GTFS schedules is that this format implicitly ignores inaccuracies in travel times due to, among other causes, operational delays, service interruptions or unrealistic schedules. Researchers requiring more accurate measures of transit travel times have relied on field measurements and large-scale simulation models of multimodal network assignments. But these sources of data are expensive to collect, may take years to implement, may not be very accurate, and are difficult to implement on a large network, in continuous time, and in perpetuity.

To account for these shortcomings, we put forward a new methodology that capitalizes on freely available vehicle location data feeds to produce a routable retrospective transit network using open source tools. This network contains actual travel times that are true to observed transit network performance, not the expected performance contained in the official transit schedules. In this paper we will describe the toolchain we developed and demonstrate its use in detecting service that differs from the schedule, as well as the implications of these differences for end-to-end travel times and time-dependent accessibility scores.

2. Data and Methods

This research uses two primary datasets, a current GTFS package for the Toronto Transit Commission (TTC) and a 48-hour extract of all vehicle locations reported by the TTC-NextBus

API. The GTFS package contains the scheduled stop times of every transit vehicle at every stop on every route in the TTC network, structured so as to allow routing between stops across the network. Our general approach is to produce a modified GTFS package with updated trips and stop times based on observations of the actual vehicle locations over the course of a day. The modified GTFS package can then be analysed in the same exact way as the scheduled data. The methodology is summarized in Figure 1.

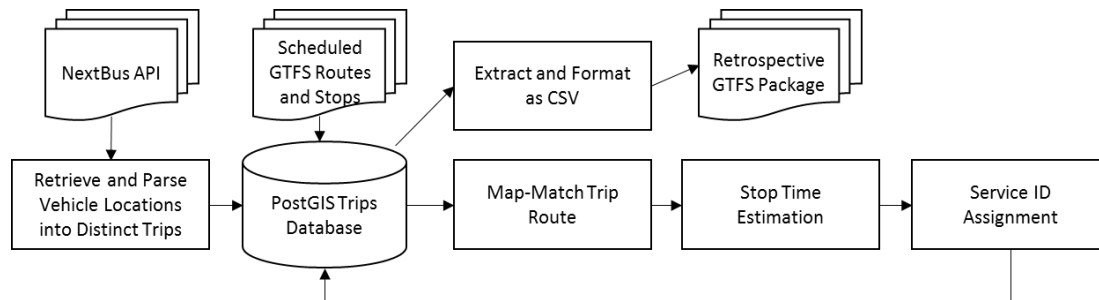


Figure 1. Creation of a Retrospective GTFS Package.

1. Retrieve and parse location data. The NextBus API is a publicly available web service designed to serve real-time transit data primarily to mobile phone or web applications. One of its functions is to report the latest locations for all operating vehicles in the fleet. For each vehicle, the API returns information on the vehicle's ID, route, heading, last known location, and the time it reported that location to the server. Vehicles update their location about every 20 seconds.

Every 10 seconds, a Python script requests all newly updated locations from the API, parses the response, and stores the results in a PostGIS database. Each vehicle is taken to be operating a trip which terminates when the vehicle changes its reported route or heading, or when it fails to update its location for more than 60 seconds. Each trip is thus essentially a GPS trace, with a set of ordered points through time and space. To obtain a better spatial resolution, the points of each trip are map-matched to a road/rail network using the Open Source Routing Machine (OSRM) software and detailed network data from OpenStreetMap. Points or trips that could not be plausibly matched to the street network (~2% of trips) were discarded. Many of these trips were also ambiguous or erroneous upon visual inspection, with much of this due to clear GPS error.

2. Estimate Stop Times. Given the heading and route ID, we find the set of stops being serviced from the NextBus API's routeConfig command. From these, stops within 20 meters of the estimated route of the vehicle are matched to a point on that route and the time for each stop is interpolated linearly from the times associated with the vehicle position reports.

3. Create New GTFS Package. With distinct trips, and stop times in sequence for each trip, we have essentially the same data structure as in the GTFS format. Each observed trip becomes a trip in the retrospective GTFS package, and each interpolated time becomes a stop for that trip. A distinct service ID is generated for each day of observed data. All stops with observed times are included, with locations given by the NextBus API. Since real-time data were not available for the TTC's four subway lines, we appended the scheduled GTFS data for these routes in our realtime GTFS package.

All tools we used were open source, and the code we developed is available on GitHub.

3. Accessibility Case Study

We demonstrate the utility of the retrospective transit network by comparing a simple measure of jobs accessibility using the scheduled and observed transit networks. The cumulative accessibility to jobs score for a location i is shown in (1).

$$A_i = |\mathbf{M}|^{-1} \left[\sum_{m \in \mathbf{M}} \sum_{j=1}^J O_j f(i, j, m, T) \right] \quad (1)$$

The inner summation in (1) is an accessibility score for location i at time m . A_i is therefore the mean accessibility score for location i within the time window \mathbf{M} . O_j is a count of jobs at location j and $f(i, j, m, T)$ is an indicator function that equals 1 if the transit travel time from i to j at departure time m is less than some threshold, T . In our specific case, we compute accessibility from census Dissemination Areas (DAs) to the estimated number of jobs at Traffic Analysis Zones (TAZ). We measure average accessibility within the morning rush hour (7:00am-9:00am) using a threshold of 45 minutes, a figure close to the mean one-way transit commute time in Toronto. All travel time calculations were performed using OpenTripPlanner which computed centroid to centroid travel times inclusive of walking, waiting, in-vehicle, and transferring times. We calculated accessibility scores based on the official GTFS package released by the TTC, and compared them to scores based on the observed travel times encapsulated in the retrospective GTFS package.

4. Results

Our primary result is displayed in a map of accessibility differences in Figure 2. Negative values denote locations where accessibility scores using the real-time network are lower than those using the scheduled network, and indicate where observed service (i.e. vehicle speeds and headways) culminated in lower than expected levels of accessibility to jobs. Positive values denote the opposite; these are areas where observed accessibility levels are higher than those obtained by using the scheduled network. This could be due to conservative schedules, expecting delays that weren't actually encountered (Wessel 2015). As expected, the two accessibility scores obtain similar values nearer to the city centre and along the subway lines. This is likely due to a higher dependence of accessibility on walking and subway routes from these locations, and therefore lower levels of sensitivity to street-level transit operations. The reader is reminded that real-time subway data were not available so the real-time network actually contains the scheduled travel times for this mode.

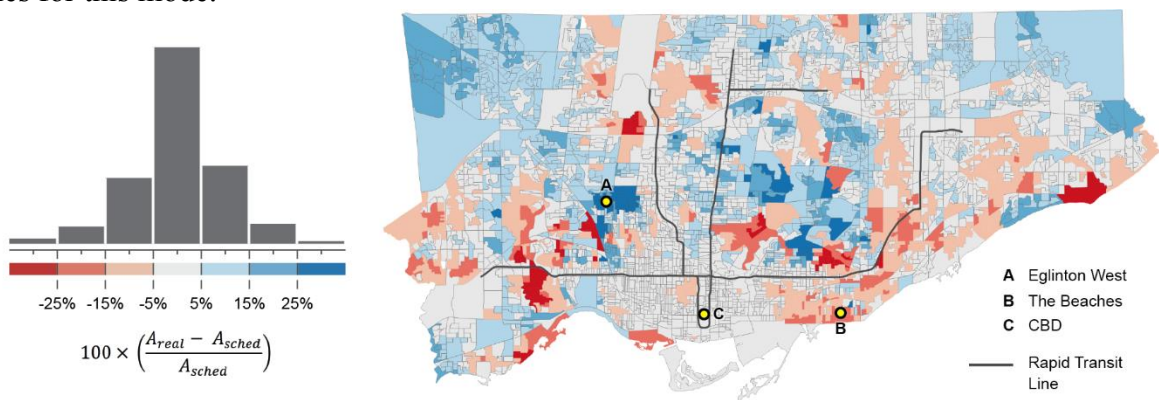


Figure 2. Differences Between Scheduled and Real-Time Access Scores.

Perhaps surprisingly, there are about as many areas exhibiting relative declines in accessibility as there are increases. These differences in accessibility are clustered in certain neighbourhoods and along specific routes where transit operations differ substantially from the published schedule. This is further visualized in Figure 3 with two comparative minute-by-minute travel time plots, each from a residential neighbourhood (The Beaches and Eglinton West) to Toronto's Central Business District (CBD). The left plot is an example of where transit operates with greater headways and results in longer commutes compared to the published schedule. Presumably there were less transit vehicles in operation than scheduled during this period or there was severe vehicle bunching. Conversely, the plot on the right is an example of where travel times from the real-time network are on average less than those from the scheduled network. This could be because the TTC overestimated operational delays when scheduling their service.

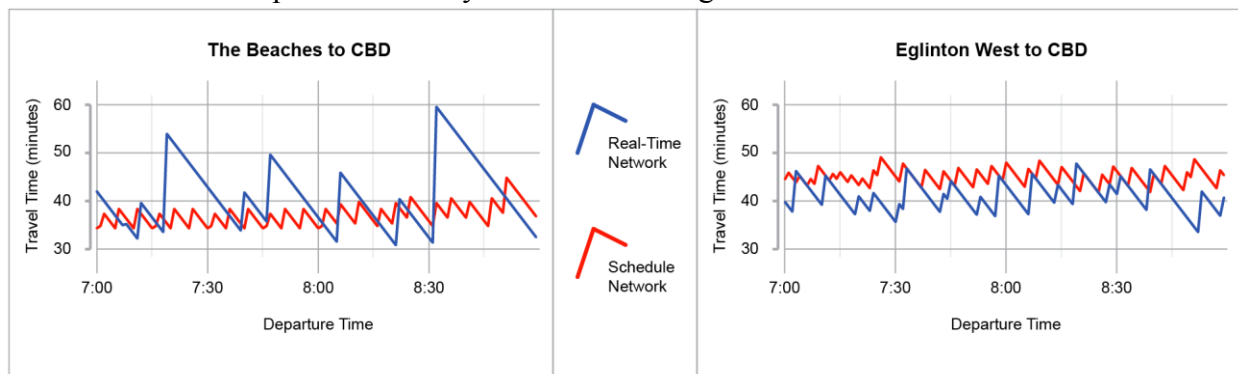


Figure 3. Scheduled and Real-Time Travel Times Resulting in Accessibility Differences.

5. Conclusions

This paper presented a new methodology for constructing a routable retrospective transit network based on freely available data and open-source software. The utility of the tool was demonstrated in a case study of accessibility to jobs, a measure of transit benefit with a wide range of applications. To our knowledge, this is the first time that accessibility on observed transit service has ever been computed and compared to scheduled accessibility metrics. Such results, especially if averaged over longer time periods, could be used to provide more realistic accessibility measures than what is available from GTFS schedule data alone. Future work will extend the analysis to include real-time subway travel times and to explore the causes for, and implications of, any systematic differences between scheduled and observed travel times and accessibility levels.

References

- Farber S, Morang MZ and Widener MJ, 2014, Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, 53:149–159.
- Farber S, Ritter B and Fu L, 2016, Space–time mismatch between transit service and observed travel patterns in the Wasatch Front, Utah: A social equity perspective. *Travel Behaviour and Society*, 4:40–48.
- Fransen K, Neutens T, Farber S, De Maeyer P, Deruyter G and Witlox F, 2015, Identifying public transport gaps using time-dependent accessibility levels. *Journal of Transport Geography*, 48:176–187.
- Hadas Y and Ranjitkar P, 2012, Modeling public-transit connectivity with spatial quality-of-transfer measurements. *Journal of Transport Geography*, 22:137–147.
- Owen A and Levinson DM, 2015, Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, 74:110–122.
- Wessel N, 2016, *Discovering the Space-Time Dimensions of Schedule Padding and Delay from GTFS and Real-Time Transit Data*. Masters Thesis, University of Cincinnati.