

Uber Movement: Travel Times Calculation Methodology

Uber Movement's Travel Times solution provides cities average travel time calculations between two 'zones' in a region for a given time and date. Zones defined for a region are commonly census tracts, traffic analysis zones (TAZs), or neighborhoods.

This document describes how zone-to-zone travel times are synthesized from GPS trace pings from cars on Uber's network, and the steps taken to ensure we're preserving rider-driver privacy throughout the Movement tool.

Since all Uber driver-partners use a smartphone to handle the logistics of their trips, anonymized and aggregated travel data can be used to measure a region's transportation infrastructure. This approach is particularly well-adapted to deliver accurate data where it is most scarce: smaller roads and arterials with limited commercial traffic and where fixed sensor infrastructure is a costly and inefficient investment.

The Uber Partner app, while on trip, records latitude, longitude, and a timestamp (Date/time) every 4 seconds. These GPS trace pings are commonly used to provide navigational routing, fare calculations, match partners with riders, and user experience elements, such as plotting the position of the car in the Uber Rider app. When aggregated, these GPS trace pings can also be used to derive average travel times between the zones in a given region. Data precision is aided by the large number of GPS trace pings Uber is able to aggregate for a given stretch of road, thereby providing reliable estimates.

Uber Movement processes these GPS trace pings using the following high-level steps:

STEP 1 - Zone Assignment: For each trip, unsorted GPS trace pings are assigned an appropriate zone as defined by a shapefile.

STEP 2 - Mean Epoch: For each zone a trip passes through, the mean GPS ping within that zone is computed. After this step, the overall trajectory is lost but we do know the average timestamp within each zone a trip passed through.

STEP 3 - Zone To Zone Travel Time: The elapsed time from each mean GPS ping to all subsequent GPS pings is measured, thereby providing zone-to-zone travel times for each trip.

STEP 4 - Aggregate Trips: Zone-to-zone travel times are aggregated from all trips. After this step, trip level information is lost and we only know statistical measures of Zone-to-zone travel times aggregated from many trips.

STEP 5 - Privacy Constraints: Travel time statistics are removed for zone pairs that either a) do not meet a minimum number of trips or b) the minimum count of unique riders necessary to preserve rider privacy. (This step is implemented in tandem with Step 4 but listed as a separate step for ease of understanding)

STEP 6 - Release: Zone-to-zone travel time averages are made available via Movement's interactive travel times solution, including several available CSV export options.

Travel Times Processing in Detail

The following provides a more detailed walkthrough of the steps above. A simplified extract of unprocessed GPS trace ping data appears as the following:

```
(Trip1, Latitude1, Longitude1, Date/time1)
(Trip1, Latitude2, Longitude2, Date/time2)
(Trip1, Latitude3, Longitude3, Date/time3)
(Trip1, Latitude4, Longitude4, Date/time4)
...
(Trip2, Latitude1, Longitude1, Date/time1)
(Trip2, Latitude2, Longitude2, Date/time2)
(Trip3, Latitude3, Longitude3, Date/time3)
...
```

1. Zone Assignment:

This step adds the zone Id column to the data shown above. We geocode every GPS trace ping to a travel zone (ie, census tracts, TAZs etc.)

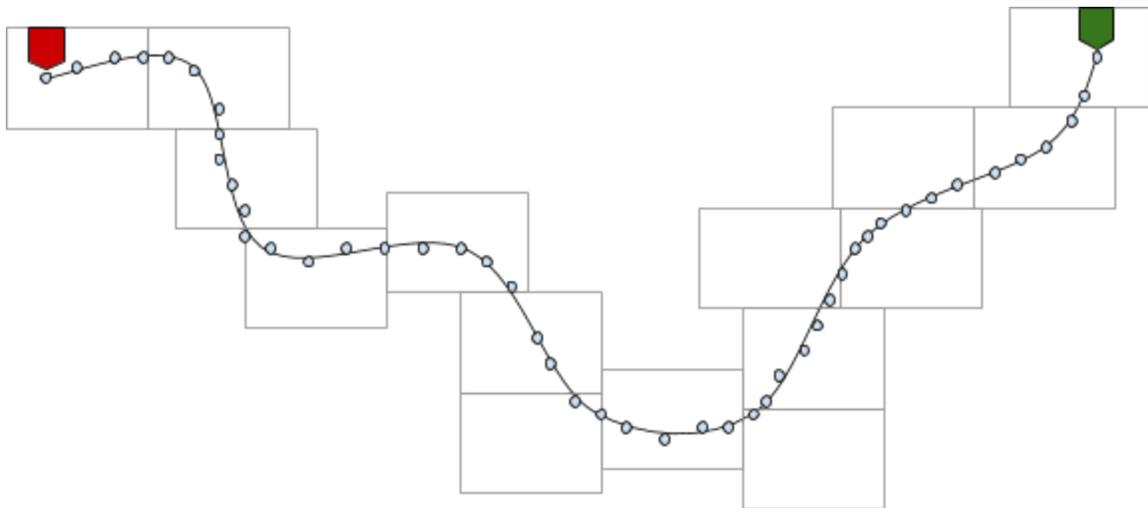


Fig. 1

In Fig 1, above:

- The start and end points are indicated by the colored markers. They correspond to the pickup and dropoff locations
- The trip trajectory is represented by the solid curve
- The shaded circles along the solid curve represent GPS readings from the driver's phone. Each GPS reading has a (lat, long, timestamp) associated with it
- The rectangles are boundaries in a city, as defined by a shapefile (ie, census tracts, TAZs etc.). Only the zones intersecting with the trip trajectory are shown in the figure. Also in practice the zones don't have to be rectangles, rather they are complex polygons

The output of this step is a table which conceptually looks like below:

```
(Trip1, Latitude1, Longitude1, Date/time1, Zone1)
(Trip1, Latitude2, Longitude2, Date/time2, Zone2)
(Trip1, Latitude3, Longitude3, Date/time3, Zone3)
(Trip1, Latitude4, Longitude4, Date/time4, Zone4)
...
(Trip2, Latitude1, Longitude1, Date/time1, Zone1)
(Trip2, Latitude2, Longitude2, Date/time2, Zone2)
(Trip3, Latitude3, Longitude3, Date/time3, Zone3)
...
```

2. Calculating Average Timestamp In Each Zone:

In this step, we calculate the average timestamp (aka epoch) in each zone. In Fig 1, for instance, this would give us 15 numbers for the trip above for the 15 intersecting zones (rectangles). After this step, the GPS trajectories are lost and all we are left with is average timestamp of each trip in each zone. The output of this step has following columns:

Trip Id, Zone Id, Average Timestamp

3. Calculate Zone to Zone Travel Time Per Trip:

We use the table in the previous step to then compute the zone to zone travel time for every (source,destination) zone pair per trip. For example, continuing the above example based on Fig. 1, the zone to zone computation would result in ${}^{15}C_2 = 105$ combinations out of the 15 readings (ordering the combinations such that date/time1 < date/time2). The output of this step would have following columns:

Trip Id, Source Zone, Destination Zone, Travel Time = Destination Epoch - Source Epoch

4. Aggregate Zone to Zone Travel Time:

This step will compute aggregated statistics like sum, sum of squares, sum of log, sum of squares of log for each (source zone, destination zone, date, hour of day). After this step, the individual trips are lost and all we are left with is the aggregated statistics. The output of this step would have following columns:

Source Zone, Destination Zone, Date, Hour of Day, Travel Time Statistics

5. Ensuring rider/driver privacy:

Travel time data are removed for zones that either a) do not meet a minimum number of trips or b) the minimum count of unique riders and drivers necessary to preserve rider-driver privacy.

6 Data added to Uber Movement:

The calculated travel time averages are made available through a web-based interactive visualization, including CSV export options. The resulting data available via Movement includes:

- Date range
- Display Name for Zone
- Origin and Destination zone geometry
- Average Travel Time (seconds)
- Average Range - Lower Bound (seconds)
- Average Range - Upper Bound (seconds)