BIG DATA
*CASE Study*

WORLD BANK GROUP

Urban traffic congestion affects poorer people disproportionately, as they have longer commutes and suffer more from the health effects of pollution

# Open Traffic: Easing Urban Congestion

**Stephanie Debere** with thanks to Task Team Leader **Holly Krambeck**


## SUMMARY

Congestion has known negative impacts on economic growth and can exacerbate urban air pollution and greenhouse gas emissions. Effectively addressing congestion requires accurate traffic speed and flow data, but resource-constrained transport agencies are challenged to collect these data, as modern tools tend to be financially and technically out of reach. In response, the Open Traffic program leverages open-source software and innovative partnerships to substantially reduce the cost of traditional traffic data collection and analysis, while simultaneously improving the quality. The first scalable, open-source program of its kind, the project built on work with the Cebu City Government in the Philippines to develop an open-source platform for collecting, visualizing and analyzing traffic speed data derived from taxi drivers' smartphones. Using GPS data from an on-demand taxi service, Open Traffic successfully analyzed peak-hour congestion, travel time reliability and corridor vulnerability across 10 Southeast Asian cities, and has prepared travel time analyses for select origin- destination pairs. This analysis would not previously have been

possible without substantial time and resources. It shows that the next generation of congestion management solutions will leapfrog the capital-intensive approaches of the past, enabling traffic management agencies to make affordable, evidence-based planning decisions. Open Traffic has now been deployed in Cebu City for live testing.

## CHALLENGE

Urban traffic congestion affects poorer people disproportionately. They generally have longer commutes than the affluent and suffer more from the health effects of higher pollution, as many work outside. Congestion also generates excess greenhouse gas emissions, and it is often the poorest people who live in areas most vulnerable to climate change. Time lost in traffic jams also has a significant negative impact on urban GDP growth.

In many developing countries, decisions about traffic signal timing plans, public transit provision, roadway infrastructure, emergency traffic management and travel demand management are made without observed, quantified congestion or travel-time data. Such data is costly to collect and can also require substantial technical expertise to analyze. This causes avoidable congestion, as well as unnecessary fuel consumption.

In higher-income countries, transport agencies rely on a combination of manual survey methods and installed physical sensors – underground detector loops, pneumatic tubes, laser-based sensors, cameras and Bluetooth device detectors. However, these require initial capital outlays, ongoing maintenance and technical expertise beyond the capacity of poorer cities. They can also record data only in places where they are deployed – select corridors during select time periods. There is urgent need for a viable, inexpensive alternative to traditional travel-time and congestion data collection and analysis. This would allow resource-constrained agencies to make evidence-based decisions to promote traffic flow.

***The project leveraged three trends: Growth in smartphone usage, the emergence of taxi-hailing app companies and increased use of open-source software***

## INNOVATION

The project leveraged three trends to develop a traffic management system reliant on GPS data instead of fixed-location equipment: Growth in global smartphone usage, the emergence of taxi-hailing app companies and increased use of open-source software.

Over a third of the world’s population is expected to have a smartphone by 2017. This has inadvertently created a new source of traffic data, derived from handset GPS signals and Wi- Fi pings. Viewed as traffic probes, smartphones can create a sensor network that is unrestricted to specific corridors, is continuously updated in real time, requires no maintenance and provides a level of sampling unachievable through manual methods or equipment-based sensors.

Recently, international smartphone-based taxi-hailing app services have also emerged. These companies maintain databases of millions of urban GPS points, often spanning hundreds of cities across many countries. The project combined smartphone ‘sensors’ and taxi GPS databases to develop a single cloud- based traffic management application, which could support services in numerous cities simultaneously. By using open-source software, it offers unprecedented economies of scale in capturing and analyzing traffic data.

### *Linking disparate data sources*

The initiative builds on a successful pilot in Cebu City, where the team created an open- source platform that uses GPS data generated by taxi drivers' smartphones to derive meaningful statistics for traffic planning. The platform, called Open Traffic, is a graphical user interface allowing government agencies to easily query and visualize stored traffic statistics derived from GPS data collected from drivers' phones.

The team partnered with Malaysia-based Grab, the largest taxi-hailing app company in Southeast Asia, to further develop and pilot the smartphone 'traffic sensor' approach. Through the partnership, traffic management agencies in Malaysia, Singapore, Indonesia, Vietnam, the Philippines and Thailand will have access to anonymized traffic data generated by 250,000 vehicles in Grab's fleet, free of charge for at least the two-year pilot and scaling-up phases of the project.

The platform uses Open Street Map (OSM), a global geographic dataset populated by volunteers without cost or licensing requirements. The map may be freely updated and improved by transport agencies and others, using open-source editing tools. Its 'Highway' feature includes all OSM mapped roads, from unpaved rural tracks to expressways, covering much of the planet. Drawing on the OSM Highway, Open Traffic links average traffic speed calculations to OSM road segments via several steps:

- Open Traffic downloads the relevant portion of the global OSM map.
- It prepares the map sections by assigning virtual 'detectors' to every approach where road map segments intersect.
- It calculates the travel time for a single vehicle traversing a road segment across two detectors, as the distance between the two detectors divided by the time the vehicle spent traveling between them.

### *From raw data to travel times*

The estimated travel time for each road segment on a given trip is stored on a server. Neither raw GPS data nor information associated with a particular vehicle is retained. Data are stored as the number of travel times for every hour of the day (how many travel times of, for example, five, six or seven kilometers per hour, etc.). These travel times can be queried to calculate average traffic speed for different time specifications (a specific day, a specific hour each day, etc.) for single or multiple road segments.

Open Traffic can query the database of stored travel times by road segment to generate a map of average travel speeds for selected time periods. It also facilitates travel-time queries between select origin and destination pairs, either for automatically generated routes (based on the shortest path) or manually defined ones. A 'confidence indicator' is provided, based on the number of observations used to derive the travel time and average speed estimates.

## RESULTS

Using GrabTaxi's data from 10 major Southeast Asian cities, the team tested whether Open Traffic's analytical results made intuitive sense. They used the platform to observe weekday peak and non-peak travel patterns in each city. These peak-hour graphs mostly reflected expectations about urban traffic, with travel speeds highest at night and slowest during commuting times. The results meant Open Traffic could be used to monitor the efficacy of congestion mitigation measures.

The team also tested the platform's suitability for examining peak-traffic duration and variation, conducting travel-time surveys and understanding how externalities and traffic interventions affect traffic speed. Tests successfully examined the predictability of congestion, checking the consistency of expected travel times between origin and destination pairs along key corridors.

The Open Traffic platform could also be used to generate inputs such as travel-time survey data for traditional transportation planning, without the cost of fieldwork, encoding and analysis. The team compared manual survey data in Cebu City to Open Traffic data, and found that the Open Traffic results provide less variation between road segments. This is unsurprising, as the traditional survey represents only a single sample, whereas the Open Traffic dataset represents thousands of samples over the same time period.

These results show that the next generation of congestion management solutions will leapfrog the capital-intensive approaches of the past. They will enable traffic management agencies to make better, evidence-based decisions about traffic signal timings, public transport, road infrastructure, emergency traffic management and travel demand management.

The Open Traffic platform has now been deployed in Cebu City for live testing. Next steps include development of a methodology for optimizing traffic signal timing plans using GPS data instead of traditional sensors, as well as a standardized methodology for estimating the cost of congestion (in terms of fuel usage, greenhouse gas emissions and economic impact). Discussions are underway with Grab on launching the platform in other cities.

## LESSONS LEARNED

Open Traffic illustrates that much of big data's potential lies in combining existing disparate sources of information in unprecedented and innovative ways.

- ***Keep seeking the potential in new combinations of data***
  By bringing together existing independent data sources, the team was able to mine a rich new source of information. To exploit big data fully, it will be important to remain alert to the potential of combining disparate and even seemingly unrelated sources of data for analysis.

- ***Think about data storage from the beginning***
  Big data analytics can generate unwieldy amounts of data. Open Traffic data was initially stored as city-specific files in Amazon 'buckets', a cloud-storage service which facilitates automated uploading and downloading of data directly from and into applications and databases. However, this required Grab to set up specialized buckets for the project, and the process and file sizes soon became unwieldly. On the team's recommendation, Grab began aggregating its global data as a single stream, rather than individual files, using Amazon's Kinesis service, which can upload real-time data streams from multiple sources.

**The Open Traffic platform can generate travel time survey data without the cost of manual fieldwork and analysis**