1 OVERVIEW

Traffic congestion has a substantial impact on both the quality of life and economic opportunities available to the lowest income segments of society. Further, congestion exacerbates pollution and GHG emissions and has a known, substantial negative impact on urban GDP growth. Resource-constrained traffic management agencies are challenged to mitigate congestion, when they do not have access to the sophisticated tools commonly used in advanced economies for monitoring real-time traffic conditions and for collecting and analyzing historic travel time data.

Starting in 2011, the project team worked with the Cebu City Government to develop an open-source platform for collecting, visualizing, and analyzing traffic speed data derived from taxi drivers' smartphones. This pilot project successfully achieved a proof of concept and the platform, Cebu Traffic, won first prize in the 2013 Philippines National E-Governance Competition. With this success, the project team then sought to use these proven methodologies to support the development of a replicable, inexpensive alternative to traditional travel time and congestion data collection and analysis. The goal of this alternative approach would be to empower resource-constrained agencies to make better, evidence-based decisions that previously had been out of reach – decisions about traffic signal timing plans, public transit provision, roadway infrastructure needs, emergency traffic management, and travel demand management.

To this end, the team partnered with GrabTaxi, an on-demand taxi service that generates taxi GPS data in countries the Bank supports, as well as Conveyal, an open-source transport software development firm. With these partnerships and support from the World Bank's Big Data Challenge Innovation Grant of US$65,000, the team improved upon the initial pilot platform, tested it with data from six countries, and deployed the platform in Cebu City for live testing.

This Project Completion Report covers activities completed under the grant implementation period, which was six months (January to June 2015). The report conclusion presents follow up activities undertaken since this period, as well as planned next steps.

2 WHY THIS PROJECT & WHY NOW

Traditionally, to collect data on traffic volumes and flows, transport agencies rely on a combination of manual survey methods and installed physical sensors – underground inductive detector loops, pneumatic tubes, laser-based sensors, cameras with video recognition software, and/or Bluetooth device detectors. These methods share a common trait – they require substantial initial capital outlays, as well as on-going maintenance and technical expertise that lay beyond the capacity of many World Bank counterpart cities. Further, even where these methods are viable, their efficacy is limited by their centrally-managed nature – that is, manual surveys and equipment can only record data in places where they are deployed – select corridors during select time periods.

In recent years, the growing ubiquity of smartphone use (over a third of the world's population is expected to have a smartphone by 2017) [1] has inadvertently created a new source of traffic data, derived from handset GPS signals and Wi-Fi pings. Thinking of smartphones as "traffic probes", we see a sensor network that: (a) is not limited to specific corridors; (b) is continuously updated in real time; (c) does not require any maintenance or upkeep; and (d) provides a level of sampling that is not otherwise achievable through manual methods and not possible for secondary roads not covered by equipment-based sensors.

Also, even more recently, we have seen the rapid rise of international smartphone-based taxi hailing application (app) services, as well as transportation networked companies (TNCs), which provide prearranged transportation services for compensation using an online-enabled app or platform to connect passengers with drivers using their personal vehicles.[2] These companies are relevant, because not only do they maintain databases of millions of GPS points that crisscross urban areas, but each company's individual database may span hundreds of cities, across many countries. This means, for the purpose of developing traffic management applications that rely on GPS data in lieu of installed fixed-location equipment, one application could be applied across hundreds of cities, without the need for additional local data partnerships or equipment installation – or even software installation, since, in theory, a single cloud-based traffic management application could support services in many/all cities simultaneously. If that software were "open-source" – that is, software that can be freely used, changed, and shared by anyone –[3] then the achievement in economies of scale in capturing and analyzing traffic data for use by transport agencies would be unprecedented. And it is this very unprecedented ends that the project team seeks to achieve. Eventually.

3 METHODOLOGY

The project builds upon an initial, successful pilot activity undertaken in Cebu City, where the team established a proof of concept for creating an open-source platform that uses GPS data generated by taxi drivers' mobile phones to derive meaningful traffic statistics for use in planning and analysis. The developer of this pilot platform, Conveyal, with a team led by Mr. Kevin Webb, continued to support the platform's improvement for the Big Data Challenge grant project.

The platform, called Open Traffic,[4] is a graphical user interface for use by government agencies to easily query and visualize stored traffic statistics derived from floating car data, or in our case, GPS data

---

[1] Statistica. 2015. *Smartphone user penetration as percentage of total global population from 2011 to 2018*. Accessed 9-23-2015. http://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/

[2] California Public Utilities Commission. 7-3-2013. *Decision Adopting Rules and Regulations to Protect Public Safety While Allowing New Entrants to the Transportation Industry*. Accessed 9-23-2015. http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M077/K112/77112285.PDF

[3] Open Source Initiative. *The Open Source Definition*. Accessed 9-23-2015. http://opensource.org/osd

[4] Because "Open Traffic" also refers to a larger suite of data processing tools, as well as a newly forming non-profit organization under the leadership of Conveyal, for the purposes of this report, Open Traffic refers to the graphical user interface supported by the grant for use by government agencies to easily query and visualize stored traffic statistics derived from GPS data.

collected from the smartphones of taxi drivers. The open-source code for Open Traffic may be found here: https://github.com/opentraffic.

The following methodology sections describe how the GPS data is collected and processed, as well as how travel time estimates are derived from these data for specific road segments, routes, and a whole road network.

3.1 THE DATA

Under the Big Data Challenge project, the team has formed a partnership with Malaysia-based GrabTaxi, a taxi hailing app company that operates in Malaysia, Singapore, Indonesia, Vietnam, the Philippines, and Thailand. Under this partnership, counterparts in these countries will have access to anonymized traffic statistics generated from GrabTaxi's fleet data, without charge, for at least the duration of the two-year piloting and scaling up phase.

Of GrabTaxi's raw data, the team utilizes the following fields for each recorded GPS point: (a) Vehicle ID; (b) Date and Time (or, "Time Stamp"); (c) Latitude; and (d) Longitude. These data, generated by at least 64,000 vehicles, are updated about every 6 seconds.

During the grant implementation period, these data were selectively queried and stored as city-specific comma-separated value (CSV or .csv) files in Amazon S3 buckets. CSV is a common data exchange format that is widely supported, and among its most common uses is moving tabular data between programs that are otherwise incompatible.[5,6] An Amazon S3 bucket is a cloud storage service offered by Amazon,[7] which facilitates automated uploading and downloading of data directly from and into databases and applications. A drawback of this method is that it required GrabTaxi to set up specialized "buckets" just for our project, and over time, the process and .csv file size became unwieldly.

As of July 2015, at the recommendation of the project team's developer, GrabTaxi began using a more efficient method to aggregate all of their global data as a single "stream", rather than individual files, using the Amazon Kinesis service, which is designed to handle the uploading of real-time data streams from multiple sources.[8]

3.2 DATA PROCESSING

In its raw format, the GPS data stream is difficult to use for meaningful analyses – it is a massive multi-terabyte dataset that continues to grow, and each GPS point may or may not fall on a known road segment.

---

[5] CSV files store tabular data (numbers and text) in plain text. Each line of the file is a data record, and each record consists of one or more fields, separated by commas.
[6] Wikipedia. *Comma-Separated Values*. Accessed 9-23-2015. https://en.wikipedia.org/wiki/Comma-separated_values
[7] Amazon. *Working with Amazon S3* Buckets. Accessed 9-23-2015. http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingBucket.html
[8] Amazon. *Amazon Kinesis*. Accessed 9-23-2015. https://aws.amazon.com/kinesis/

Figure 1: Density of GPS location points captured from participating taxis during Cebu Traffic pilot (2012)

To convert the GPS data into travel times by road segments, the following steps are taken.

1. Calling of the relevant Open Street Map (OSM) tiles.
2. Assigning virtual “detectors” to the OSM tiles.
3. Estimating travel time, based on virtual detector crossings.

3.2.1 CALLING RELEVANT OPEN STREET MAP (OSM) TILES

To support global scaling of the project, the traffic management platform relies on the Open Street Map (OSM), a free, global geographic dataset populated by volunteers.[9] This map is considered “open data” and has no licensing requirements. The OSM may be freely updated and improved by transport agencies and others, per a project’s particular needs, using free and open-source editing tools made available by the OSM Foundation.[10]

Open Traffic relies upon the OSM “Highway” feature, which includes all OSM mapped roads, from unpaved rural roads to expressways. As of time of writing, the OSM includes 87 million segments tagged as roads (“highways”),[11] covering much of the planet.

The Open Traffic platform links average traffic speed calculations to OSM road segments. To achieve this, the first step is to “call” the relevant OSM “tiles” – that is, for the Open Traffic application to download

[9] OpenStreetMap. *About OpenStreetMap.* Accessed 9-23-2015. https://www.openstreetmap.org/about
[10] OpenStreetMap. *Map and Editing Tools*. Accessed 9-23-2015. https://www.openstreetmap.org
[11] Taginfo. *Keys: Highway*. Accessed 9-23-2015. https://taginfo.openstreetmap.org/keys/highway#overview

the relevant portion of the global OSM map for use in calculating average traffic speeds. While it is possible to download the entire global OSM map for these analyses, for the sake of speed and efficiency, only those tiles that are needed to estimate traffic speed for a given set of GPS points are called. This is accomplished using the GPS coordinates from the data stream and calling OSM tiles with a municipal-level zoom level of “11”, which equates to a scale of 1:250,000.[12]

3.2.2 ASSIGNING “DETECTORS” TO THE OSM TILES

The next step is to prepare the OSM tiles by assigning virtual “detectors” to every approach where road segments (or “ways”) intersect. The term “detectors” is used as a nod to the traditional inductive detector loops that are buried under approaches to signalized intersections, used to detect passing vehicles. The use of this method for traffic speed analysis is derived from a paper written by Marco Gruteser Baik Hoh et. al., *Virtual Trip Lines for Distributed Privacy-Preserving Traffic Monitoring* (2008).[13]

The following figure shows how these virtual detectors where automatically generated from the OSM for Cebu City.

Figure 2: Virtual detector generation from the Cebu City Open Street Map.

3.2.3 ESTIMATING TRAVEL TIME, BASED ON VIRTUAL DETECTOR CROSSINGS

Travel time for a single vehicle traversing a road segment across two detectors is calculated as the distance between the two detectors, divided by the time the vehicle spent traveling between the

[12] OSM Wiki. *Zoom Levels*. Accessed 9-24-2015. http://wiki.openstreetmap.org/wiki/Zoom_levels

[13] M.G. Baik Hoh, et. al. 7-2008. *Virtual Trip Lines for Distributed Privacy-Preserving Traffic Monitoring*. International Conference on Mobile Systems, Applications, and Services (MOBISYS), Brekenridge, CO. Accessed 9-23-2015. http://bayen.eecs.berkeley.edu/sites/default/files/conferences/mobisys08.pdf

detectors. So, if it took a vehicle 1 minute to travel 2 kilometers from one detector to the next, then the estimated travel speed for that single trip on that particular segment would be 120 km per hour:

$$2 \text{ km} / (1/60 \text{ hour}) = 120 \text{ km/hour}$$

In cases where the detector analysis is taking place directly on the mobile handset, only timestamps for crossed detectors would be transmitted to the server. But in cases where the Open Traffic application pulls in raw GPS data, additional assumptions are made, since in most cases, the GPS location "pings" may not match directly with the virtual detectors.

For example, take a look at the following figure. To start, we determine trajectory. If GPS points are captured to the west and then east of Detector 1 and then again east of Detector 2, then we know the trajectory of the vehicle is east along the west-east corridor (as opposed to the alternative trajectory of turning south at the intersection, in which case detector 3 would have been tripped). But the GPS pings, denoted by the time-stamped circles, do not match directly with the virtual detectors. How do we assign timestamps to the detectors in this case?

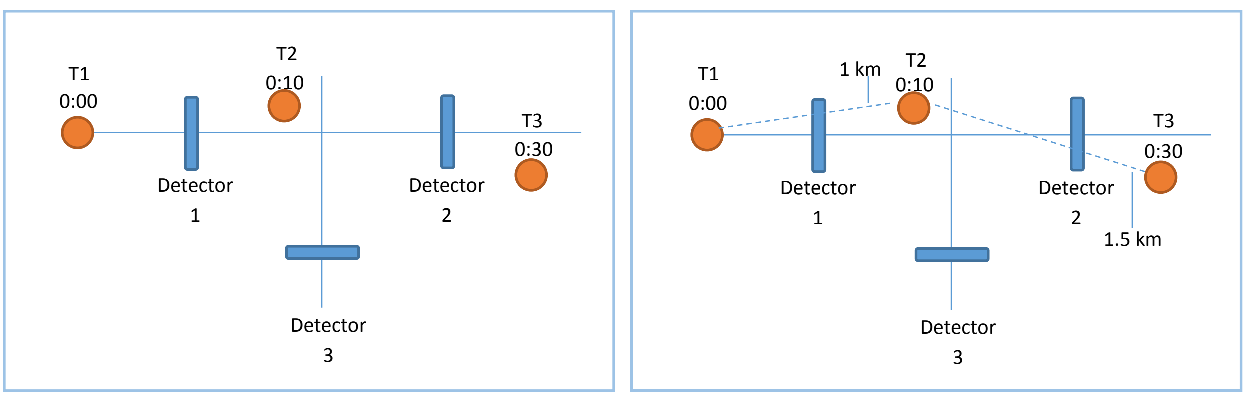


Figure 3: Illustration of how GPS points are used to estimate timestamps for virtual detectors

First, let's look at the GPS pings that occurred at 0:00 and 0:10. We draw an imaginary line between the points, which intersects Detector 1. Let's suppose that Detector 1 bisects the drawn line directly at the midpoint, such that half the line is to the west, and half is to the east of Detector 1. In this case, the timestamp for Detector 1 is:

$$[(\text{Distance from } T_1 \text{ to } D_1 / \text{Distance } T_1 \text{ to } T_2) \times (T_2\text{-}T_1)] + T_1 = D_1$$

or

$$[(0.5 \text{ km} / 1.0 \text{ km}) \times (0{:}10 - 0{:}00)] + 0{:}00 = 0{:}05$$

As for Detector 2, we see that the virtual detector intersects with the drawn line between the GPS points not in the center, but further to the east. So in this case, the timestamp for Detector 2 is calculated based on the proportion of the line that is on the approach side of the detector, from the west headed east:

$$[(1.0 \text{ km} / 1.5 \text{ km}) \times (0{:}30 – 0{:}10)] + 0{:}10 = 0{:}23$$

3.2.3.1 CHALLENGES

Virtual Detector and GPS Data Challenges

The developer team needed to overcome a number of challenges to refine this virtual detector methodology, which was originally conceived of for highways, for use in urban contexts with complicated road geometries and "fuzzy" GPS data.

For example, when a vehicle turns, the line drawn to connect GPS points may cut a corner and miss the virtual detector lines. As a result, the trip would include non-sequential tripped detectors.

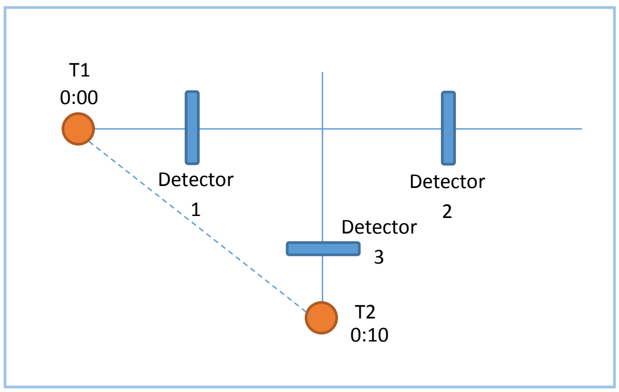


Figure 4: Illustrative example of missed virtual detectors during a turn.

Similarly, there may be stray GPS points that fall on adjacent streets, and some virtual detectors may be missed if a vehicle travels "too fast". Further, in parts of the city with very complicated road geometries, a vehicle may inadvertently trip multiple detectors for adjacent roads or intersections. Also, vehicles may exhibit behavior that could skew travel time estimates, such as sudden U-turns or prolonged stops.

To overcome these issues, the following measures are applied. First, minimum and maximum travel time thresholds between detector pairs, which are derived based on historical data, are assigned. Calculated travel times that fall outside of these thresholds are dropped. With regards to disconnected trip segments, when a missing "link" in a trip is detected (i.e., a vehicle passes Detector 1 misses Detectors 2 and 3 and then hits Detector 4), this link is temporarily stored. If the travel time between Detector 1 and 4 is reasonable and occurs with sufficient frequency (for piloting purposes, 10 or more instances of the same issue), a virtual "bridge" is constructed to link these travel paths over the area where the GPS points seem to "jump" from the expected travel path. In cases where the virtual detector segments cannot be joined, or where the travel time exceeds a reasonable threshold, these GPS points and travel time estimates are dropped.

To minimize the occurrence of these disjointed trips, detector "crowding" on a map is lowered by establishing a minimum segment length between detectors of 20 meters.

It is noted that one of the primary benefits of the virtual detector approach is that it is sufficiently computationally simple that it can run directly on a mobile device, with only the travel speed data being transmitted to a server, thereby substantially reducing data transmission volume and protecting privacy (i.e., raw GPS coordinates are not transmitted – only timestamps for the artificial detectors). What makes this method simple is that it is not necessary to first "snap" an entire GPS trace to a road network – rather, we only need to look at whether a detector is activated – we do not need to know any information about the specific path in between detectors – just the time it takes to reach each one.

Open Street Map (OSM) Challenges

Ironically, one of the most beneficial features of the platform, the leveraging of the OSM, also increases the risk of sub-optimal performance, since the OSM is not always sufficiently complete for the platform to run successfully. What makes this issue particularly challenging is that in many cases encountered during the initial platform tests, the OSM *appeared* complete, with roads mapped and properly labeled, but underneath, these roads (or "ways") were not joined with intersecting roads, leaving virtual gaps in the map. Without these joins, the virtual detector loops cannot be assigned, resulting in the absence of calculated travel times for the missing segments.

While, in the short run, broken OSM links identified by the platform are an issue, in the long run, this is actually an opportunity. Since the platform can identify, with some accuracy, missing links in the OSM network, it would be feasible to write an extension to the platform that would flag these links for correction and upload to the global OSM.

3.3 DATA STORAGE

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 are retained. These travel times are stored as distributions of binned travel times by time of day. The following table provides an illustrative example of how this binning works.

Table 1: Illustrative Example for Data Storage Principles.
Number of Travel Time Observations on Link D_1 to D_2 for January 1, 2015

Hour	*5 km per hour*	*6 km per hour*	*7 km per hour*	…	*200 km per hour*
0	0	10	3	…	0
1	1	11	4	…	0
2	0	15	6	…	0
3	0	30	7	…	0
…	…	…	…	…	…
24	0	6	2	…	0

Thus stored, these travel times can be queried to calculate average traffic speed by different time aggregations (specific day, specific hour over a particular week, etc.) for a single road segment, or cumulatively added together for linked road segments, or routes.

3.4 DATA QUERYING

3.4.1 SYSTEM-LEVEL ANALYSIS

For selected time periods, the Open Traffic platform will query the database of stored travel times by road segment (links between virtual detectors) to generate a map of average travel speeds (see figure, below). The number of secondary roads that appear on the screen and are included within the query increase as the user interface zoom level is increased.

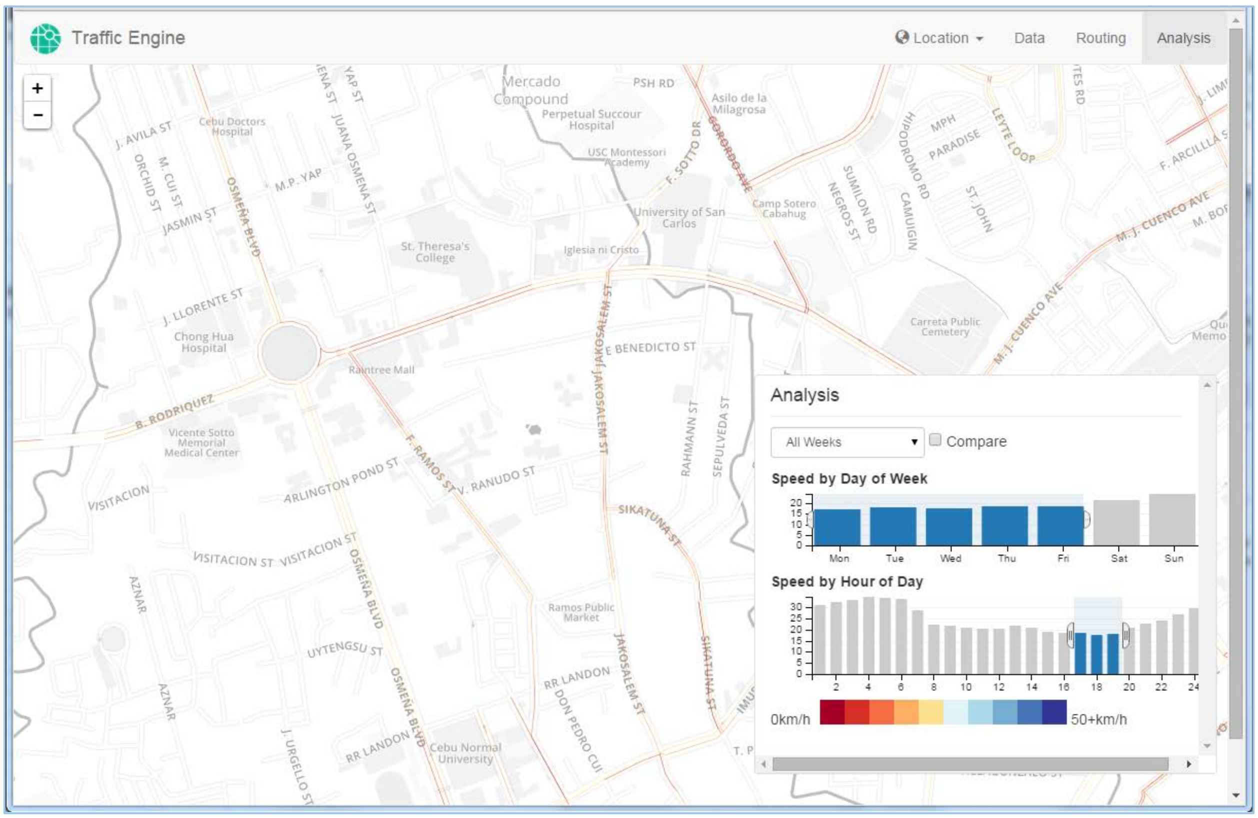


Figure 5: Screenshot from First Iteration of OpenTraffic Platform under the Big Data Challenge Project, with Cebu City System Analysis View

3.4.2 ROUTING QUERIES

The Open Traffic platform facilitates travel time queries between select origin and destination pairs. There are two modes for this analysis: automatic generation of a route between origin and destination (O-D) pairs, and manually defined routes.

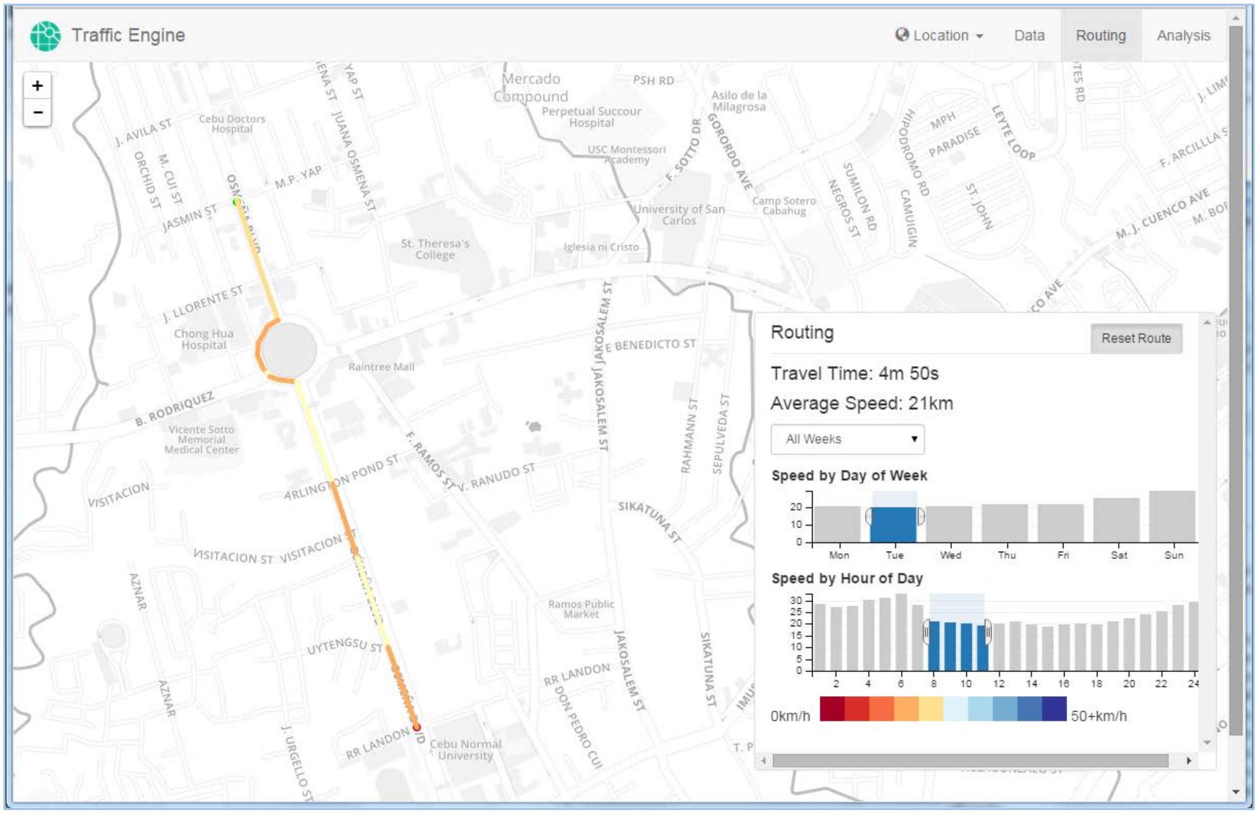


Figure 6: Screenshot from First Iteration of Open Traffic Platform under the Big Data Challenge Project, with Sample Cebu City Route Query

The automatically derived routing between O-D pairs is based on the shortest path, on the basis of time (calculated using all available data in the database). The algorithm for this problem is the same as that used in Open Trip Planner,[14] an open-source trip planning application that works with the OSM. Open Trip Planner uses the A* (pronounced "A star") algorithm,[15] which plots an efficient, traversable path between multiple points (or, in our case, virtual detectors). Relative to other shortest path algorithms, such as the well-known Djikstra's algorithm, A* has the advantage of taking into account not only shortest path to the next node, but also the *total* distance (or time) already traveled, as the algorithm works its way from origin to destination. To improve performance, the platform also uses "Contraction Hierarchies",[16] available as open source software, which through some form of magic make the process of finding the shortest path across a particularly large network more efficient.

[14] Github. *OpenTripPlanner*. Accessed 9-24-2015. https://github.com/opentripplanner/OpenTripPlanner.
[15] Wikipedia. *A* Search Algorithm*. Accessed 9-24-2015. https://en.wikipedia.org/wiki/A*_search_algorithm
[16] Wikipedia. *Contraction Hierarchies*. Accessed 9-24-2015. https://en.wikipedia.org/wiki/Contraction_hierarchies

For all route queries, the platform sums the average travel speeds stored in the server for each link (distance between virtual detectors) in the route, for the specified time period. A level of confidence indicator is provided, based on the raw number of observations used to derive the travel time and average travel speed estimate. In future iterations of the platform, it is envisioned further statistics indicating variance will also be applied.

4 INITIAL TEST RESULTS

The core methodology for translating GPS points into average travel speeds by road segment using virtual detectors was implemented and tested for the pilot platform, Cebu Traffic, in 2013. At that time, a simulation test framework was developed, which proved that the methodology is sound and can produce valid results with a sufficient sample.[17]

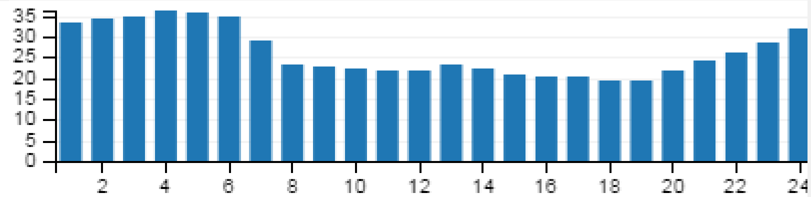
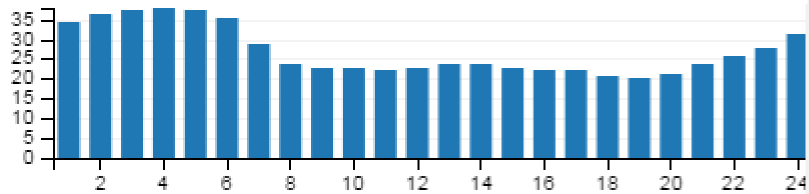
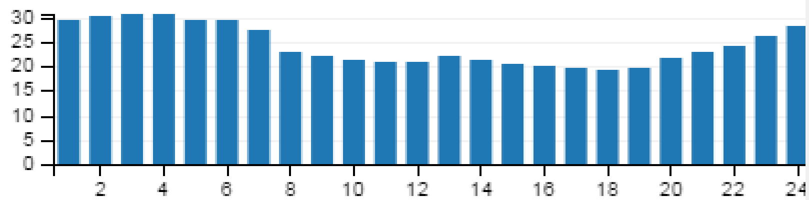
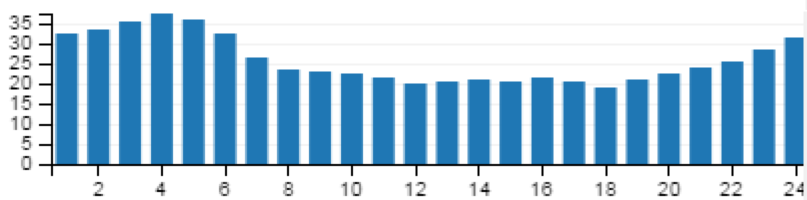
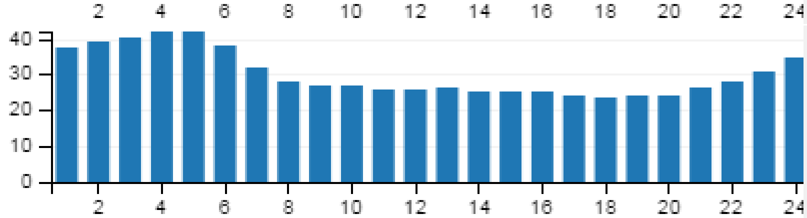
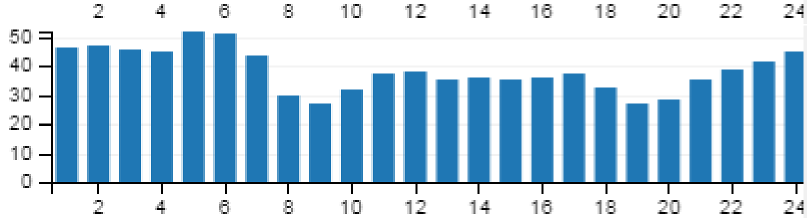
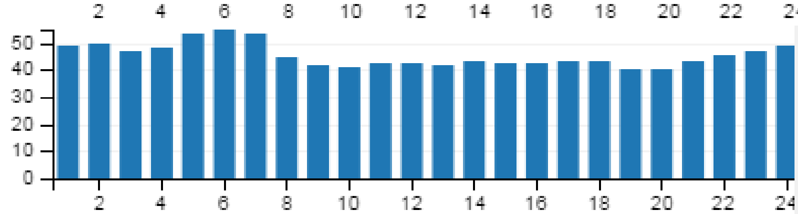
The following initial tests were conducted to determine whether the results produce by Open Traffic make intuitive sense, as well as its appropriateness as a tool for examining congestion period duration and variation, as well as for conducting travel time surveys and understanding how externalities and traffic interventions can impact traffic speed.

4.1 PEAK HOUR ANALYSIS

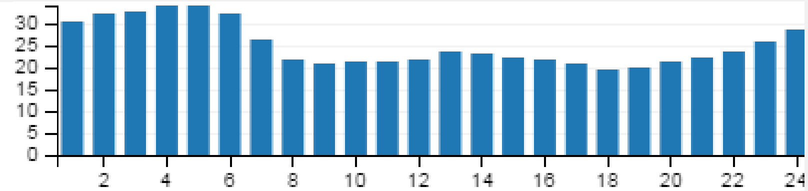
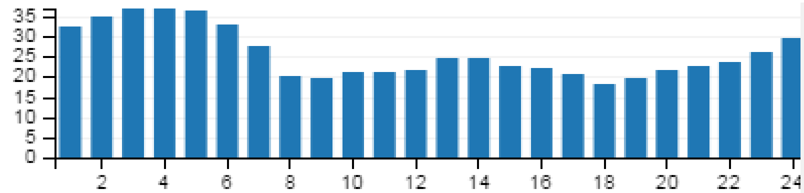
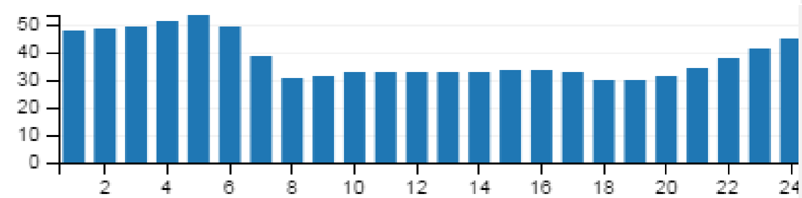
As a preliminary check to determine whether the analytical results derived from the Open Traffic platform make sense, the team used the system analysis tool to observe weekday peak and non-peak travel patterns (see table, below).

[17] The simulation used known O-D pairs and travel times to generate simulated GPS points (with some spatial error to reflect real-world conditions). These points were fed back into the Cebu Traffic platform to generate travel speeds, which were compared to the original speeds. For GPS data points with a 10 second frequency, and +/- 15m spatial error, the inference engine was able to calculate journey times within +/- 5% of the average expected values.

Table 2: Average Weekday Traffic Speed on Key Corridors in Select SE Asian Cities (2015)[18]

City	Average weekday travel speed across downtown major roads (kph)	Min Avg. kph	Max Avg. kph	% Diff.
Cebu City		20	36	80%
Manila		20	38	90%
Davao		19	31	63%
Surabaya		19	37	95%
Jakarta		23	43	87%
Kuala Lumpur - Petaling Jaya		28	52	86%
Singapore		40	55	38%

[18] Note: Data date ranges and number of operating vehicles vary by city – thus, these results may not be considered directly comparable, as they represent different sample sizes. Further statistical work is required to produce a comparable dataset, which may be accomplished after data export functionality is introduced to the platform.

City	Average weekday travel speed across downtown major roads (kph)	Min Avg. kph	Max Avg. kph	% Diff.
Ho Chi Minh City		20	39	95%
Hanoi		18	37	106%
Bangkok		30	53	77%

For the most part, these peak hour graphs reflect what we intuitively know about urban traffic, that traffic speeds will be highest when most people are sleeping and slowest during weekday morning and evening commuting times. Thus, we can say that the traffic speed derived from the GPS data appear to make intuitive sense.

That said, what is striking about the peak hour analysis is that our traditional notion of “peak hour” traffic does not seem to really exist in a number of major Southeast Asian cities. In the data, we see a decline in travel speeds during morning commuting hours that generally persists until late in the evening, with only a modest couple hours’ reprieve in the middle of the day. This phenomenon is most prevalent in Jakarta, where congestion along main downtown corridors begins on weekdays at 7:00 a.m. and persists until about 9:00 p.m. Also of note is Singapore, where there is virtually no congestion on key corridors, with overnight free-flow speeds very close to peak hour speeds.

After the platform’s data export functionality is initiated, further study will be undertaken on these traffic patterns and their implications. For example, one key objective of traffic management is to reduce the duration of congestion peak hours – to this end, Open Traffic could be deployed as a tool for monitoring the efficacy of citywide and corridor-specific mitigation measures.

4.2 CONGESTION VARIABILITY

When evaluating congestion, in addition to examining the duration of congestion, we also look at the “reliability” or predictability of congestion. That is, along key corridors, how consistent are expected travel times between O-D pairs? If a resident leaves their home at 7:00 a.m. for work, what is the likelihood that they will arrive at their office around the same time each day?

The team tested the platform's ability to analyze congestion variability by comparing average travel times in Manila along a 7.4 km stretch of EDSA, the primary north-south arterial, on Mondays from 6:00 p.m. to 6:59 p.m. (May 18 through August 31).

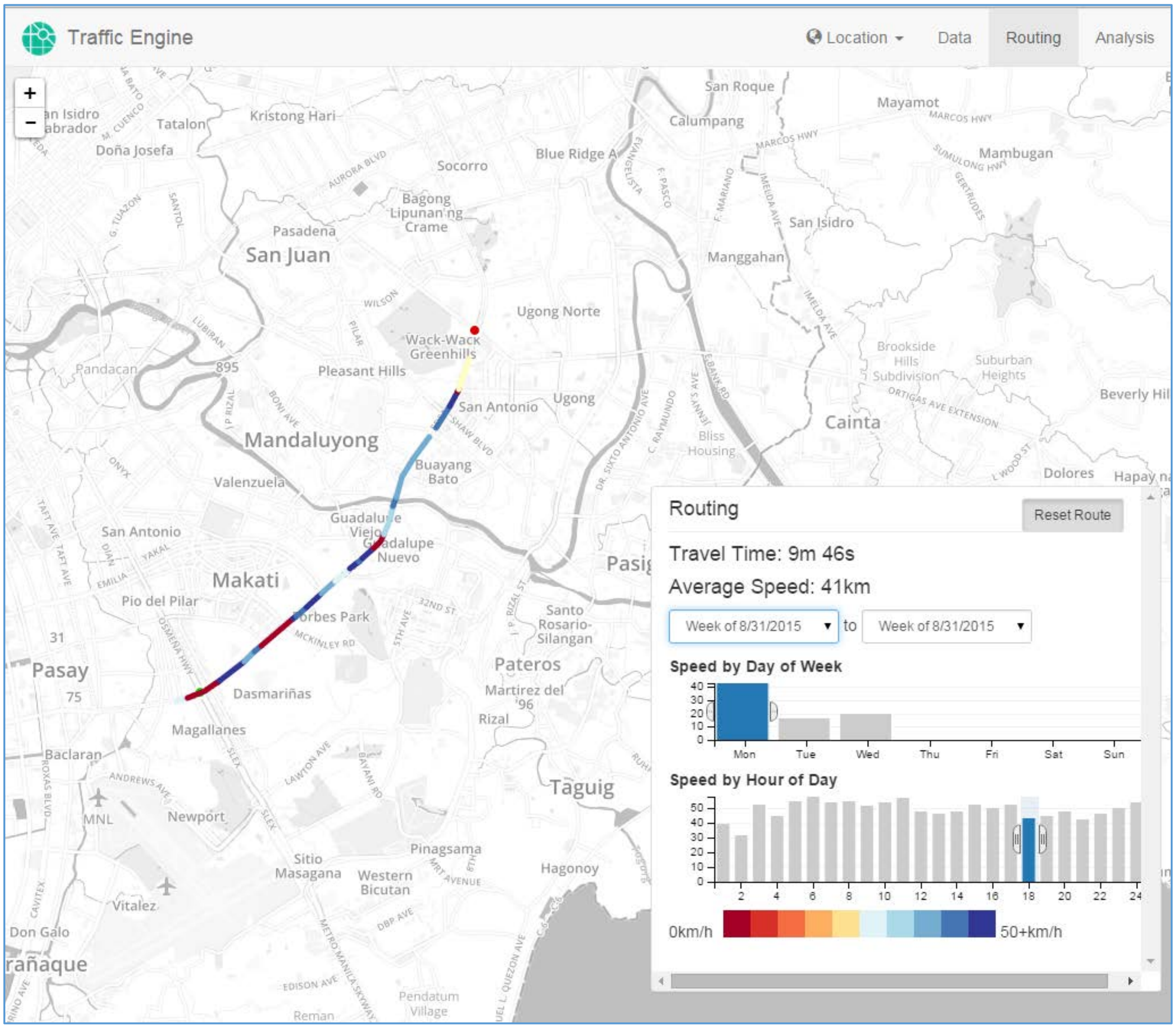


Figure 7: Screenshot from First Iteration of OpenTraffic Platform under the Big Data Challenge Project, with EDSA 7.4 km Stretch

Results show that the average travel time of this trip is 65 minutes, with a variance of 354 and standard deviation of 25 (to compare, an 8.4 km trip taken at the same time along a similarly critical arterial in Singapore – the Pacific Island Expressway – takes an average of 6.5 minutes, with a variance of only 26 and standard deviation of 15).[19] On a national holiday, the trip has taken as little as 9 minutes, but on a rainy day and/or a day with traffic accidents, the same trip has taken as long as 80 minutes. While Manila congestion is well-known, it is poorly understood, especially the impact of this high variability on the economy.

[19] Note that the Manila sample only includes 14 days of observations and the Singapore sample only includes 4 days. These are preliminary results, to be re-calculated as more data is collected through the Open Traffic platform.

These initial test results are promising. After the platform's data export functionality is initiated, the team will prepare a more detailed analysis of congestion variability across Southeast Asian cities, as well as identification of specific road segments in select cities that appear to contribute the most to this variability.

4.3 VULNERABILITY ANALYSIS

So far, we have looked at using the data and platform to support work on reducing congestion period duration, increasing average travel speeds on key arterials, and improving the "predictability" of travel speed. What we have not discussed yet is reducing vulnerability of the transport network to traffic incidents and weather events. To accomplish this task, we need to identify the portions of the transport network that are most severely impacted by accidents and weather events, for the most vehicles. While this type of modeling work is outside the scope of the grant project, the team has done some preliminary work to see whether the Open Traffic platform could eventually support such work.

For example, the following figure presents a typical northbound trip (Fridays, at 6:00 p.m.) along a key 7.4 km stretch of EDSA, one of two primary north-south corridors in Metro Manila. On average, the trip takes just over 17 minutes, with an average speed of 23 km per hour.

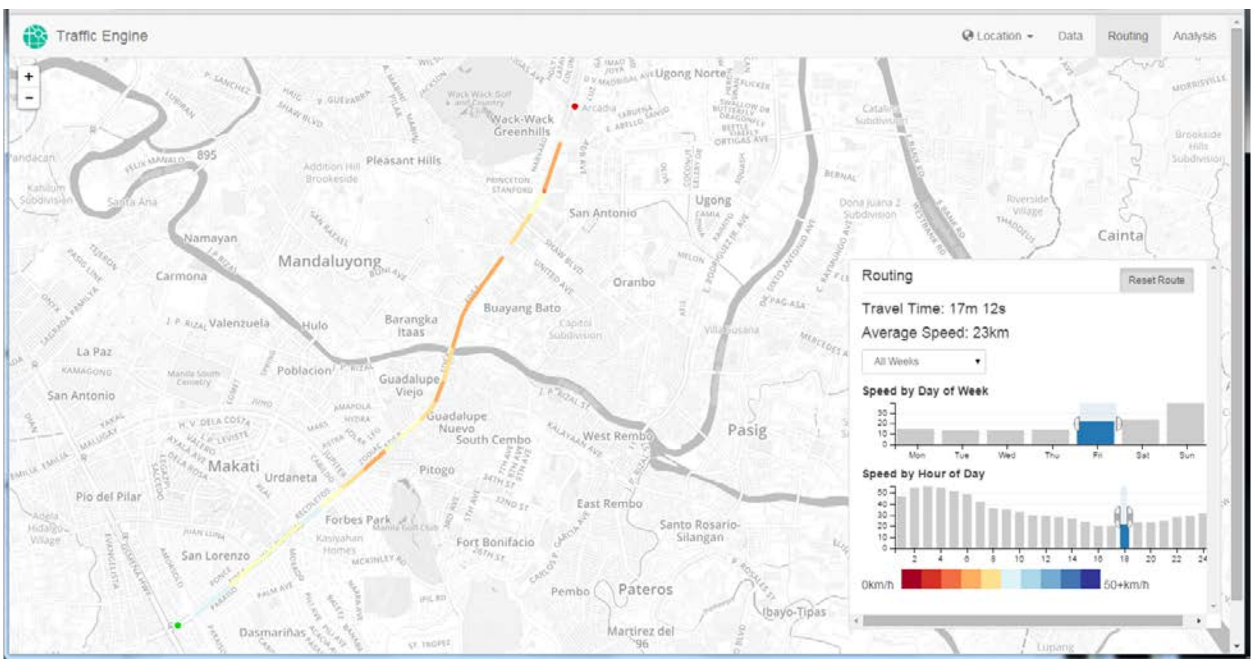


Figure 8: Northbound Trip - Average Travel Time and Speed on Fridays at 6:00 p.m.

But notice in the next figure, that same northbound trip took an average of more than 151 minutes on Friday, July 24, at 6:00 p.m., when heavy rains combined with a series of accidents along EDSA reduced travel time starting at 10:00 a.m., with delays persisting till midnight.

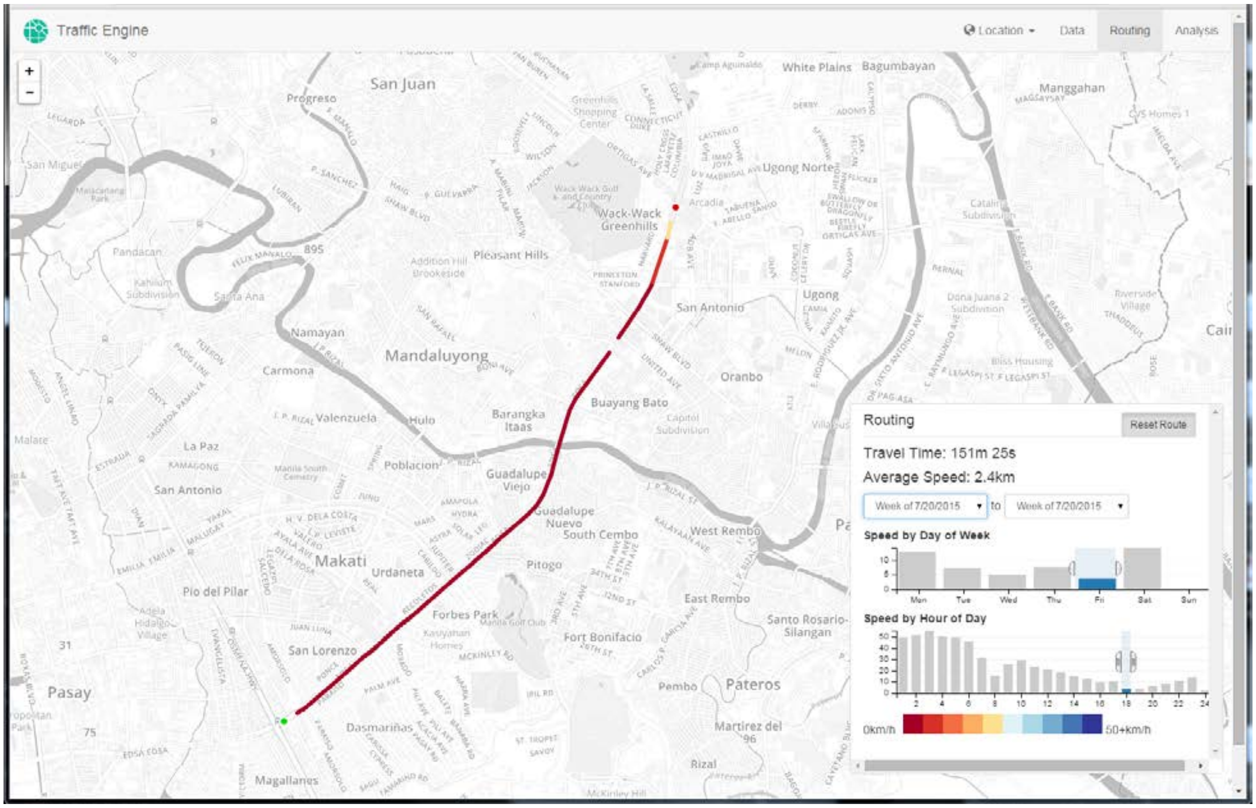


Figure 9: Northbound Trip – Average Travel Time and Speed for Friday, July 24, 2015

What is most remarkable about the impact of these accidents on the northbound route, is the remarkably little impact they had on the southbound trip. While northbound traffic took more than 2.5 hours to cover 7.4 km, travelers headed southbound at that same time spent only 14 minutes – see figure, below.

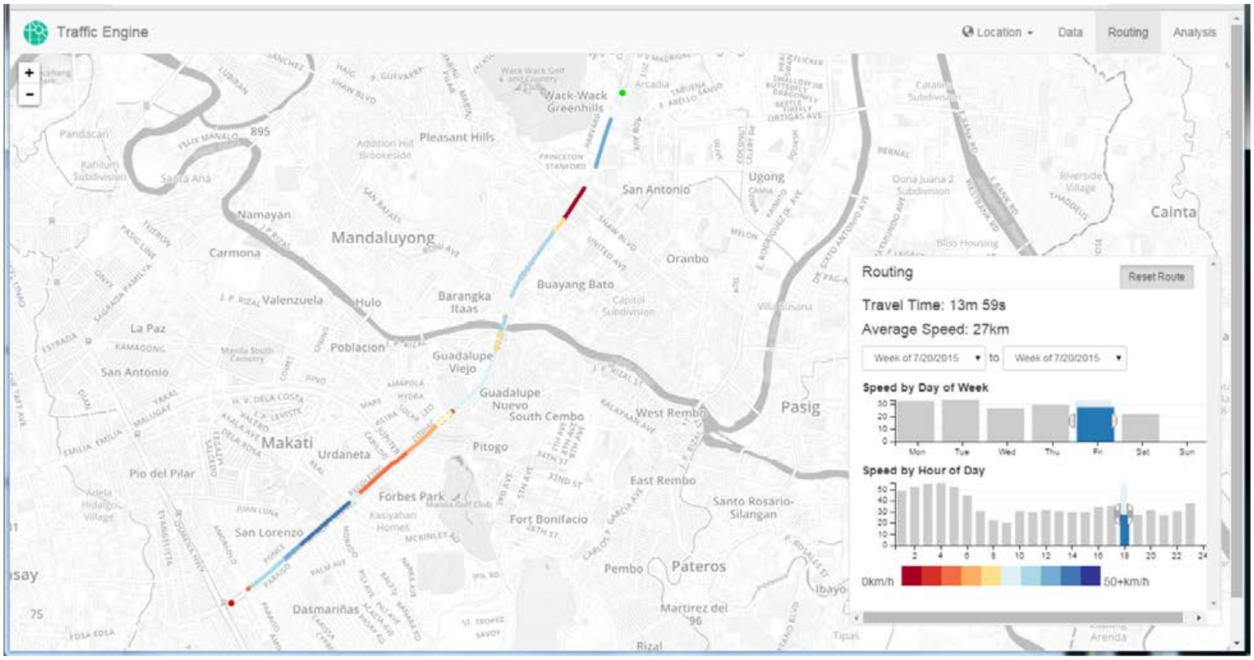


Figure 10: Southbound Trip – Average Travel Time and Speed on Friday, July 24, at 6:00 p.m.

What these data reveal is that a single hiccup on EDSA can have disastrous effects on traffic for an unusually long period of time. When these data are combined with traffic count data, we can estimate what the cost of this single incident was to the city, as well as the cost of all similar incidents. These costs may be used to justify the necessary investments needed to make EDSA less vulnerable – such as modern ITS, improvement of alternative corridors, more efficient accident response practices, and flexible lanes for diverting directional traffic to a contraflow lane.

4.4 TRAVEL TIME SURVEY

In addition to identifying priority areas for congestion management, the team envisions that the Open Traffic platform could also be used to generate inputs for traditional transportation planning analyses, such as travel time surveys.

For the World Bank-financed Cebu Bus Rapid Transit Project (P119343), five key corridors were surveyed in 2009, which took at least 10 days of field work (two directions per corridor, with each origin beginning at midnight), plus another 10 to 15 man-days for encoding and analysis. With the Open Traffic platform, this entire job could have been done by one person – an employee from the traffic bureau -- in about 10 minutes, with a much higher degree of accuracy. The traditional method employed in Cebu provides a single travel time estimate during free-flow conditions on a single day, for each corridor. In contrast, the GPS dataset includes, literally, thousands of samples for the same corridors, not just for free-flow periods, but also for all times of day and days of week, as well as during anomalous periods, such as holidays or weather events.

To put the gain in efficiency in monetary terms, if we assumed a cost of US$300 per man-day, then for one observation per surveyed segment, the five-corridor survey cost about US$55,000. To produce a sample as robust as that that could be obtained with the crowd-sourced GPS data, the cost would be, well, incalculable, as it simply would not be possible to mobilize the man-power to collect the level of up-to-the-minute data that is possible with the GPS data.

Although we would expect some difference between data collected in 2009 and 2015, since the 2009 survey data were collected beginning on a weekday midnight, we would expect to see some similarities. For example, along one surveyed route, from Cebu South Road to the Mactan Airport, we would expect travel times to be slower towards the western end of the route, where the route traverses traffic signal-controlled corridors in downtown Cebu City, than towards the east of the route, with very few signalized junctions. We would also expect that travel times between segments, at that time of night, would not vary significantly. The chart, below, shows a comparison of these original travel times collected on Monday, June 8, 2009, from 12:00 a.m. to 1:00 a.m., against the travel times estimated by Open Traffic for the same time period (12:00 a.m. to 1:00 a.m.) on all Mondays from February 23 through September 28, 2015. Note that this chart has been presented as a line chart, so that the level of variation in travel times along corridor segments can be readily understood from the chart.

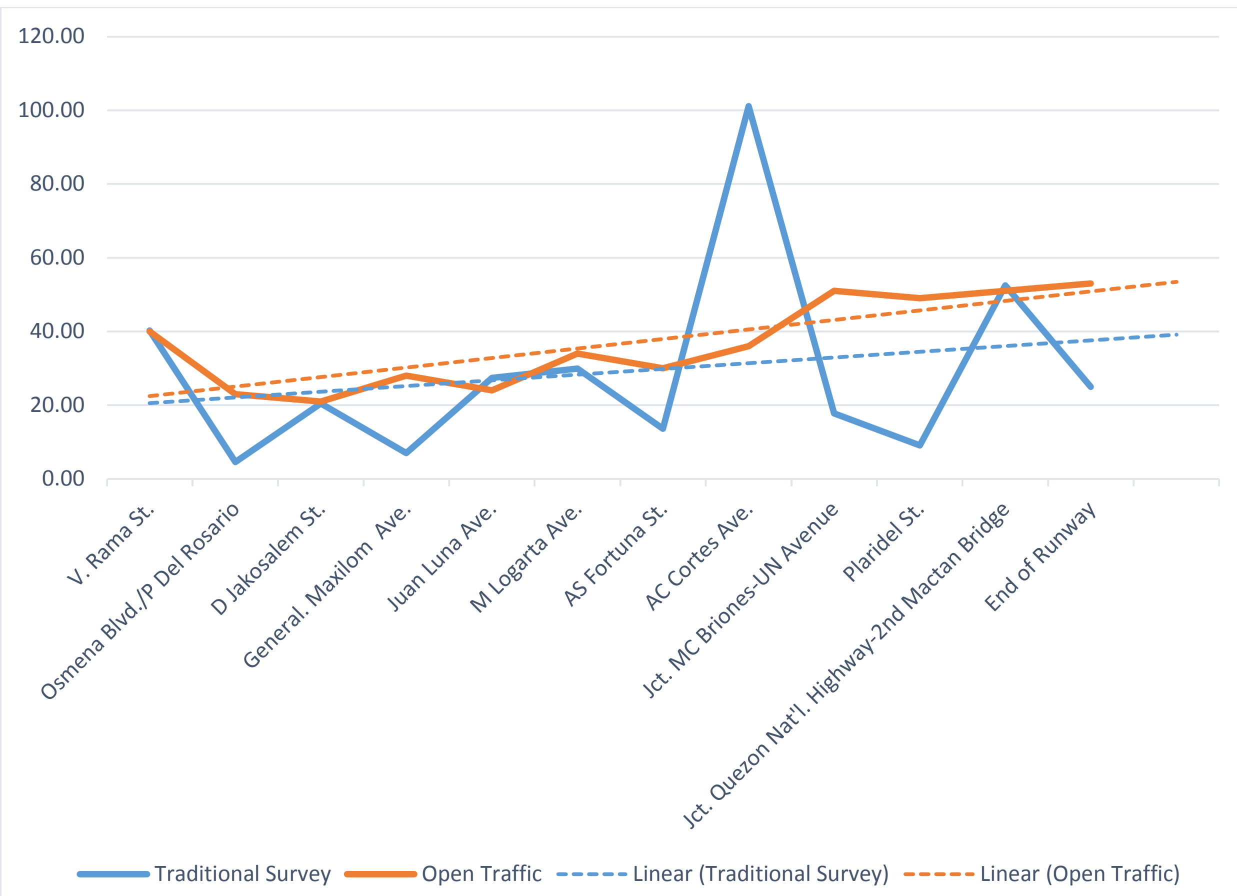


Figure 11: Travel Time Survey Comparison for Cebu South Road to Mactan Airport: 2009 Traditional Travel Time Survey vs. Open Traffic Output (kph)

Just by visual inspection, we see that compared to the traditionally-conducted travel time survey, the Open Traffic results are more "smooth", with less variation between road segments (a variance of 142 and standard deviation of 12, compared to a variance 712 and standard deviation of 26). Although the overall trend derived from the traditional survey data is close to what we would expect – increasing travel speeds on the eastern part of the route -- the individual segment results are highly varied and the overall travel time is almost double from what we found through the Open Traffic platform. These results are not surprising, however, since the traditional survey represents only a single sample, whereas the Open Traffic dataset represents thousands of samples over the same time period. Variations over a single run could stem from traffic signal timing plans, behavior of other vehicles, or human error from trying to record times at night at a precise location using an imprecise methods. In the Open Traffic analysis, such variations are smoothed out.

When support for downloading traffic speed and observation count data is integrated into the Open Traffic platform, the project team can derive a more precise estimate of the difference in cost and accuracy between traditional methods and the Open Traffic method for estimating travel times for use in planning analyses.

5 NEXT STEPS

The first iteration of the platform was initially introduced and tested in Cebu City in July 2015, and as of time of writing, based on feedback provided by the Cebu City Transportation Office and the project team, a second iteration is under development, to be completed by October 2015. With this second iteration, the team will proceed with further research work that is being funded by the Korean Green Growth Trust Fund (KGGTF) and the Energy Sector Management Assistance Program (ESMAP). Under these programs, the team is developing and testing a methodology for optimizing traffic signal timing plans using GPS data in lieu of traditional sensors, as well as a standardized methodology for estimating the cost of congestion (in terms of fuel usage, GHG emissions, and economic impact) that can be applied across multiple cities and countries. The team is also planning to develop a methodology that combines the traffic speed data with accident data through a parallel initiative for optimizing ambulance stations and dispatch.

In preparation for the completion of the platform, the World Bank Team is in discussions with GrabTaxi and EAP country office teams on launching the platform in other counterpart cities.

In parallel, Conveyal, with support from MapBox, MapZen, and the World Bank, is leading a broader initiative, seeking to link data providers such as GrabTaxi with a non-profit organization, Open Traffic. Open Traffic organization's primary function would be to collect, store, and support distribution of open traffic data for the world's use.

Since these activities fall outside of the grant implementation period, they will be presented in a future report.