2016

Scalable Methods for Monitoring Limited Access Roadways Using Crowd-Sourced Probe Data

Howell Li  
_Purdue University_

Stephen M. Remias  
_Purdue University_

Aaron Taylor  
_INRIX_

Darcy M. Bullock  
_Purdue University, darcy@purdue.edu_

Follow this and additional works at: [http://docs.lib.purdue.edu/civeng](http://docs.lib.purdue.edu/civeng)

Part of the Civil Engineering Commons

Li, Howell; Remias, Stephen M.; Taylor, Aaron; and Bullock, Darcy M., "Scalable Methods for Monitoring Limited Access Roadways Using Crowd-Sourced Probe Data" (2016). Lyles School of Civil Engineering Faculty Publications. Paper 21.  
[http://docs.lib.purdue.edu/civeng/21](http://docs.lib.purdue.edu/civeng/21)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Scalable Methods for Monitoring Limited Access Roadways using Crowd-Sourced Probe Data
by
Howell Li
Purdue University

Stephen M. Remias
Purdue University

Aaron Taylor
INRIX

Corresponding author:
Darcy M. Bullock
Purdue University
550 Stadium Mall Drive
West Lafayette, IN 47906
Phone (765) 494-2226
darcy@purdue.edu

Word Count: 3,893 words + 10 * 250 words/(figure-table) = 6,393 words

November 3, 2015
ABSTRACT

Commercial crowd-sourced probe vehicle data has been gaining traction in recent years as a ubiquitous and scalable resource for identifying traffic congestion on limited access roadways. It is routinely used in real-time by navigation software that displays color coded maps. However, outside of public agency traffic management centers, there are no factual “big picture” reports on traffic conditions. The media tries to fill this gap, but they either provide descriptions of construction locations, or highly subjective opinions. This paper proposes and illustrates a “big picture” characterization of regional and national traffic conditions using archived and real-time data. Average speeds of vehicles on segments of roadway can be retrieved in near real-time at one-minute intervals to produce performance metrics that measure cumulative miles of congestion per route, per entire Metropolitan Statistical Area (MSA), and on coast-to-coast Interstates using speed profile analysis. Moreover, both real-time and historic archival performance measures can be used for after-action analysis of major traffic events. In this study, the traffic congestion for four MSAs and two Interstates during the week of June 28 to July 6 was used as a case study to illustrate the concepts.

The study found most congestion in the Chicago, Los Angeles, and New York City metropolitan areas occurred during the PM rush on July 2 before the holiday weekend, with at least 20% of all limited access roadways in each area falling below 40 mph between the hours of 4:30 PM and 5:45 PM local time. On a coast-to-coast level, Interstate 80 showed the heaviest congestion eastbound at 5:15 PM EDT with 140 combined miles of congestion across 11 states. Data reduction and aggregation methods using 15-minute medians outlined in this study allow future systems to implement regional congestion graphs, speed profile charts, and temporal congestion graphs for operational and practical uses. This information can be leveraged by local, regional, and state transportation agencies as well as for media dissemination and outreach to inform the public.

Keywords: traffic, crowd-sourced, data, national, congestion, performance, speed, mobility, dashboards.
INTRODUCTION

Performance measurement of high speed, low access highway facilities has often only been possible with the installation of physical infrastructure that measures speeds and manual processing to reduce the data into useful information. Historically, technologies like inductive loops, side-fire radar, license plate readers, and Bluetooth identification matching were used to perform these speed tests (1). While the sample rate was high for small areas of roadway, this was not a scalable approach to assess roadways in very large or aged networks or at the state or national level.

Recently, crowd sourced probe data has established a presence with the increase in hand held technology penetration and ubiquity of GPS applications for roadway navigation. The users of these applications provide meaningful speed information in real-time across the country. These data have become very popular on the state level to assess roadway mobility for individual departments of transportation (2-7), at the corridor level for assessing a route that travels through multiple states (8, 9), and at the metropolitan level to compare the mobility performance of different cities (10). From a national perspective, these data allow for the scalability of in-depth performance measures typically implemented at the smaller local or municipal levels, or at the state level. The current national performance measures using these data (11, 12) are informative, but have stimulated an interest in real-time description of the network. The objective of this research is to develop and enhance national tools that can be used in real-time for the assessment of current national mobility trends. These objectives align with the national goals and milestones put forth by the FHWA in MAP-21 (13).

METHODOLOGY

SPEED SAMPLING OF ROUTES AND NETWORKS

Crowd sourced probe data samples average speeds of GPS-enabled vehicles and devices at one-minute intervals on roughly half-mile to two-mile segments of roadway throughout the world. These disaggregate segments can be grouped together by route and direction contiguously to give a sense of performance of a road at a particular time or to analyze historic trends. Figure 1a shows a single half-mile segment on I-95 New Jersey Turnpike in the southbound direction in the New York City Metropolitan Statistical (14). The entire length of the New Jersey Turnpike southbound through the MSA is depicted in Figure 1b, with callout i indicating the location of the half-mile segment from Figure 1a. The map in Figure 1c shows all limited access routes in the MSA with route from Figure 1b highlighted in red. The map in Figure 1d shows the location of the MSA region at a national level. The lower 48 states contains over 127,000 miles of limited
access roadway segments as defined by this system, of which 16,224 miles of roadway from 22,434 segments are used for this study.

**DATA REDUCTION AND AGGREGATION**

Due to the scale of the data size nationally and with stochastic variations in the minute-by-minute speeds, a reduction process is taken to aggregate speeds of individual segments into 15-minute median bins. For the eight-day analysis period in this study, over 262 million one-minute segment speed records are recorded on a 3.5 gigahertz, dual-processor server with 128 gigabytes of memory running SQL Server. The data is reduced to 17 million records when aggregated per 15 minutes, with the process taking 4.5 hours of total computation time. The resulting fetch time for the aggregated dataset is 99 seconds per query compared to 28 minutes per query for the one-minute dataset.

To demonstrate an example of the reduction process, Figure 2 illustrates a sample dataset before and after reduction. Figure 2a graphs a series of one-minute speeds for the segment of the New Jersey Turnpike defined in Figure 1a on July 2, 2015 over two hours in the afternoon. The dotted horizontal line at 40 mph indicates the congestion threshold used for this study. For the majority of limited-access roads analyzed, the speed limits are no less than 40 mph and this threshold is used objectively as a measure for congestion in this study. From the graph, the speeds exhibit some noise throughout the period minute-by-minute (callout i), fluctuating at and below the threshold for serval minutes between 1:45 PM and 2:00 PM. After the data is aggregated into 15-minute bins (Figure 2b) using the medians, the noise is reduced while still retaining the overall downward trend in that period with the segment’s speed falling below the threshold at 1:45 PM.

**JULY 4TH CONGESTION**

**AT-A-GLANCE**

For one of the more heavily-travelled holidays in the summer driving season, it is useful for a traffic agency, as well as from a user perspective, to monitor the peak travel patterns on limited access facilities. As an overview, the maps in Figure 3 show at-a-glance the performance of all roads in four MSAs at noon (Figure 3a) and 5:00 PM EDT (Figure 3b) using donut graphs. Figure 3b shows the dramatic shift to congested performance at rush hour for all four areas. The maps in Figure 4 provide a more zoomed-in view of each of the four MSAs from Figure 3b, showing geographically the most congested routes by color. This view is widely accessible from major internet mapping services in real time but often not as a historical playback feature.
WEEKEND PM RUSH

Figure 5 is a series of speed profile graphs that show the top 10 most congested routes in four metropolitan areas at 5:00 PM on Thursday, July 2 in each local time zone. Any route segment operating below 40 mph is considered congested for all areas. For each route, the number of miles of congestion is displayed in four speed categories under 40 mph: 1) light congestion from 30 to 39 mph, 2) moderate congestion from 20 to 29 mph, 3) heavy congestion from 10 to 19 mph, and 4) extreme congestion below 10 mph.

For metropolitan Indianapolis (Figure 5a), relatively low congestion existed on the roadways with only 9 routes in the area exhibiting a total of just under 30 miles operating below 40 mph. This accounts for just 5% of all limited access facilities in the Indianapolis MSA. The outer loop of I-465 was the most congested with 7 miles of congestion and no segments heavily or extremely congested. The only route in the area that had any segment operating below 10 mph was a one mile segment of the inner loop of I-465. By comparison, the Chicago metropolitan area (Figure 5b) exhibited much more congestion with its top 10 routes totaling 141 miles operating below 40 mph. The I-294 Tri-State Tollway southbound had 13 miles operating in heavy congestion and is the most congested in the area, with a total of 24 miles operating below 40 mph. Six out of the top 10 routes had one or more segments in extreme congestion, shown in purple.

The Los Angeles and New York City metropolitan areas, both with more miles of limited access roadway, saw much greater congestion at 5:00 PM. In Los Angeles, the top 10 routes accounted for over 241 miles operating below 40 mph. The most congested route was the I-405 San Diego Freeway, having 40 and 35 miles of congestion for northbound and southbound directions, respectively (Figure 5c, callout i; Figure 5c, callout ii). The congested northbound and southbound segments account for 55% and 48% of the entire length of the freeway, respectively. More significantly, there were 20 miles of roadway in the top 10 of Los Angeles that operated below 10 mph. The slowest-moving segments included over 7 miles on I-405 northbound direction and 5 miles on I-10 Santa Monica Freeway in the eastbound direction. For the New York City metropolitan area (Figure 5d), there were 233 miles of congestion in the top 10 most congested routes, with the Garden State Parkway in the southbound direction being more congested than the rest. The I-495 Long Island Expressway in the eastbound direction had the most number of miles operating below 10 mph, at 6 miles. Over 54% of the route was congested at 5:00 PM. The congestion for many of these routes may perhaps be attributed to users driving away from the central business districts for the holiday.

THE WEEK PRIOR

To put the July 2 congestion in context of the week prior to the holiday, the total miles of congestion on all routes in the Chicago, Los Angeles, and New York metropolitan areas are analyzed over an 8 day period in Figure 6. The peak period in Chicago on July 2 at 4:30 PM local time (Figure 6a, callout i) had significantly
more miles of congestion than any other day of that week, with 343 miles of limited access roads operating below 40 mph. This is equivalent to 26% of the Chicago area limited access network. By comparison, the second-most congested day of that week was Monday, with 246 miles of congestion at 4:45 PM, which was 28% less than Thursday’s peak period. It is also interesting to note that the peak total congested miles in the AM period dwindled as the week progressed: Monday with 228 miles, Tuesday with 133 miles, Wednesday with 105 miles, and Thursday with 67 miles; this phenomena perhaps reflects the increasing number of extended-weekend takers as the holiday approached.

Figure 6b shows the same 8-day window for the Los Angeles area. At its peak, the network had 669 congested roadway miles in the afternoon of July 2 at 4:45 PM local time. This is the equivalent to 41% of the Los Angeles area limited access network. In comparison, the second-most congested day in the same week was on Tuesday (Figure 6b, callout ii), with 586 miles of congestion during its peak at 5:30 PM local time. The ‘holiday rush’ on July 2 had only 12% more congested miles than the second-most congested day in the same week. The AM peaks also did not exhibit the same decreasing trend throughout the week as it had occurred in Chicago.

Shown in Figure 6c, the New York City area had 554 miles of congestion in the network (21% of the area) at 4:45 PM local time on July 2. The second-most congested period in the week was the day prior on July 2 at 5:45 PM with 513 miles at its peak. Comparing with the other two metropolitan areas, New York City had a less dramatic increase during the Thursday’s PM rush. One other observation is that New York’s midday periods from Monday through Thursday of that week had more miles of congestion sustained during the midday (10:00 AM to 3:00 PM local time), with at least 20 miles of roadway operating in heavy or extreme congestion for all four days in that period (dark red-shaded areas in Figure 6). In comparison, Los Angeles fell below 7 miles of heavy or extreme congestion during the midday on Monday and Chicago fell below 1 mile on both Monday and Tuesday.

Friday July 3 in all three metropolitan areas saw virtually no congestion on the roads before 9:00 AM local time. The typical PM peak periods shifted to earlier in the day for all three areas: Chicago peaks at 2:15 PM with 111 miles, Los Angeles peaks at 1:15 PM with 281 miles, and New York City peaks at 1:15 PM with 280 miles. July 4 saw very little congestion for all three areas, although interestingly some congestion was captured during the afternoon to late evening hours (Figure 6b, callout iii; Figure 6c, callout iv), possibly due to fireworks displays in each area. Sunday July 5 saw greater afternoon and evening congestion in all three cities: Chicago with 153 miles at its peak at 4:00 PM, Los Angeles with 117 miles at its peak at 3:15 PM, and New York City with 204 miles at its peak at 4:15 PM. Both Los Angeles and New York City had fewer congested miles on Sunday, July 5, than the same day in the prior week. The previous Sunday (June 28) peaked with 188 miles at 2:45 PM in Los Angeles and 226 miles also at 2:45
PM in New York City. Chicago was the exception with 107 miles of congestion at 3:45 PM on June 28, which is 46 fewer miles than July 5 at its peak.

ROUTE DRILL-DOWN

Graphs displaying miles of congestion over 5 days of the July 4 holiday weekend are illustrated in Figure 7 for the I-405 San Diego Freeway (the 405) and the I-495 Long Island Expressway (the LIE) in both directions. The two routes are compared because of their similar lengths (72 miles in each direction) and number of miles congested over the weekend.

The time of the ‘snapshot’ taken from Figure 5 is indicated by the dotted lines in Figure 7. Callout i in Figure 7a shows that at 5:00 PM on July 2, the congestion was already past its peak on the 405 northbound, which peaked at 4:15 PM with 44 miles. In the same figure, callout ii shows a substantial amount of congestion in the late evening hours of July 4, possibly due to fireworks viewing, which accounted for 21 miles of congestion at 11:15 PM local time. Callout iii shows a some amount of northbound congestion Sunday afternoon with 10 miles at 1:15 PM and 1:45 PM operating below 40 mph, while callout iv shows more typical AM and PM peaking characteristics for Monday, July 6.

For the 405 southbound, fewer miles of roadway were extremely congested, or operating less than 10 mph, compared to the northbound direction in the afternoon of July 2. At most three miles were extremely congested in the southbound direction versus 8 miles in the northbound direction. However Figure 7b, callout vi shows more total miles congested during midday on Friday compared with the northbound direction; on July 3 there were 25 miles congested in the southbound at its peak at 1:00 PM compared to only 8 miles at 11:30 AM in the northbound.

The LIE on the other hand shows a much greater directional disparity between the eastbound and westbound directions over the weekend than on the 405, most likely due to the inbound and outbound nature of the route, with New York City at its far western end and the beaches in Nassau and Suffolk Counties on Long Island at its far eastern end. In Figure 7c, callout vii, the number of miles congested had not yet reached its peak at 5:00 PM in the eastbound direction on July 2. The eastbound route peaks fifteen minutes later at 5:15 PM with 40 miles of congestion. Meanwhile, in the westbound direction during the same time period, much fewer miles were congested around 5:00 PM. At the height of the congestion on the westbound LIE, 12 miles operated below 40 mph at 5:15 PM.

The eastbound LIE exhibited peak congestion earlier in the day on Friday than on Thursday (Figure 7c, callout viii), with 20 miles at its height at 1:00 PM. There was essentially no congestion on the LIE westbound from late Thursday through noon on Sunday July 5, with three or fewer miles operating under 40 mph. However, from noon through 8:00 PM on Sunday, the LIE westbound exhibited multiple peak periods with travelers returning to the city (Figure 7d, callout ix). The route saw 6 miles of congestion at
2:45 PM, 14 miles at 4:45 PM, 18 miles at 6:30 PM, 22 miles at 7:15 PM, and 18 miles again at 8:45 PM. During the same afternoon, the eastbound direction had relatively little congestion, peaking at 7 miles between 1:00 PM and 1:30 PM.

The following Monday also saw atypical weekday activity on the LIE, with generally fewer miles and less severe peaks of congestion for both eastbound and westbound. Interestingly in the westbound direction, some of the more congested periods occurred early in the morning similar to the Thursday before the holiday. There were 10 and 12 miles of congestion recorded at 6:30 AM and 7:45 AM, respectively for the westbound direction on Monday (Figure 7c, callout x), but was not comparable to Sunday evening. This may be due to some of the travelers taking extended vacation through the beginning of the week. In comparison, the 405 in Los Angeles exhibited more typical AM and PM peaking patterns on Monday.

**COAST-TO-COAST INTERSTATES**

The methods used to identify congestion on routes in metropolitan areas can be applied nationally for coast-to-coast Interstates. Figure 8 illustrates the number of miles of roadway operating below 40 mph in a map with each state shaded by congestion severity. Both I-80 and I-10 at 5:00 PM Eastern Daylight Time on July 2 are shown for all states. In Figure 8a, the bulk of the congestion is located in California with 30% (42 miles) of the total congestion on I-80 nationally in both directions. Indiana, Pennsylvania, New Jersey, and Illinois share much of the remaining congestion with 16% (23 miles), 14% (21 miles), 12% (17 miles), and 11% (15 miles), respectively. Figure 8b shows I-10 during the same period with California on top with 41% (83 miles) of the total congestion, followed by Texas (26%, 53 miles) and Louisiana (21%, 44 miles). Figure 9 illustrates an example of a state-level drill-down of congestion by county on I-80 in Pennsylvania for 5:00 PM on July 2. As depicted by the darkly-shaded area, Clarion County had the most congestion with 11 miles operating below 40 mph.

The information depicted in Figure 8 can be viewed over time in a stacked-area graph in Figure 10. The congested miles of each state on each of the routes is represented by color. Figure 10a, callout i shows the same PM rush on Thursday on I-80 eastbound with California having the most miles congested from noon through midnight Pacific Daylight Time. Figure 10b, callout ii shows most of the I-80 westbound congestion dominated by Pennsylvania and New Jersey from morning through midnight Eastern Daylight Time on Thursday. Figure 10a, callout iii and Figure 10b, callout iv shows a similar congestion pattern on I-80 eastbound for California, and I-80 westbound for Pennsylvania and New Jersey on Friday. Figure 10a, callout v and Figure 10b callout vi shows somewhat of a reversal of the Thursday and Friday trend with Pennsylvania being congested in the eastbound direction and California being congested in the westbound direction. Additionally, Ohio adds to some of the congestion eastbound while Utah, Nebraska, and most
significantly Indiana adding to the congestion westbound. The congestion on I-80 eastbound reached 140 miles at its peak at 5:15 PM EDT coast-to-coast.

Figure 10c and Figure 10d illustrates the same 5-day period for I-10 through all states for the eastbound and westbound, respectively. Figure 10c, callout vii and Figure 10d, callout viii shows California, Texas, and Louisiana accounting for the majority of the congestion Thursday afternoon and evening. Figure 10c, callout ix shows I-10 eastbound through Louisiana significantly congested Friday afternoon, while during the same period Texas was congested in the other direction (Figure 10d, callout x). On Sunday, the bulk of the congestion going eastbound was also in Texas with 37 miles at 5:15 PM EDT. In the westbound direction, Alabama tops the list with 26 miles congested at 4:15 PM EDT. In contrast, Monday afternoon and evening had less congestion coast-to-coast than Sunday in three out of the four routes.

CONCLUSIONS AND FUTURE WORK

Fifteen-minute aggregated crowd-sourced probe data was used to analyze traffic congestion during the week of June 28 to July 6 over the Independence Day holiday. The study found peak congestion occurred in the afternoon of July 2 at four Metropolitan Statistical Areas (MSAs):

- 30 miles (5%) of limited access roads in the Indianapolis area at 5:45 PM EDT;
- 343 miles (26%) of limited access roads in the Chicago area at 4:30 PM CDT;
- 669 miles (41%) of limited access roads in the Los Angeles area at 5:30 PM PDT;
- 554 miles (21%) of limited access roads in the New York area at 4:45 PM EDT;

Nationally, two coast-to-coast routes were analyzed during the July 4 holiday weekend. California accounted for 42 miles (30%) of the total congestion on I-80 and 83 miles (41%) of the total congest on I-10 nationally during the PM rush on July 2. The congestion on I-80 reached a peak in the eastbound direction at 5:15 PM EDT on July 6 with 140 miles of congestion combined across all states.

One-week scalability in these data analytics have been demonstrated on over 16,000 miles of roadway across the United States. With greater computing power, it is feasible to leverage robust software application implementations to enable wide-area, informative roadway performance visualizations during normal or extraordinary travel periods. Future developments of these systems would enable transportation agencies to leverage real time and historic traffic data to make better data-driven operational decisions, as well as providing a source of information for users who are planning trips on busy holidays through widespread media dissemination.
ACKNOWLEDGEMENT

The data source used for this paper was provided by INRIX. This work was supported in part by the Joint Transportation Research Program administered by the Indiana Department of Transportation and Purdue University. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.
REFERENCES


FIGURES AND TABLES

Figures

Figure 1. Scalability of Probe Data........................................................................................................ 13
Figure 2. 15-minute aggregate speed data for segment of I-95 in Figure 1a.................................. 14
Figure 3. Speed Summary for major metropolitan areas on July 2, 2015 at 5:00 PM local time. 15
Figure 4. Spatial representation of traffic speeds on July 2, 2015 at 5:00 PM local time .......... 16
Figure 5. Top congested routes in four metro areas, July 2, 2015......................................................... 17
Figure 6. Routes in Los Angeles and New York over Independence Day, 2015. ...................... 19
Figure 7. Comparing week prior to Independence Day weekend for three metropolitan areas,
2015.................................................................................................................................................... 18
Figure 8. Map highlighting miles of congestion per state at 5:00PM Eastern Daylight Time. .... 20
Figure 9. Map highlighting miles of congestion per county in Pennsylvania on I-80 in both
directions at 5:00PM Eastern Daylight Time........................................................................................ 21
Figure 10. Congestion on I-80 over Independence Day weekend in all through states............. 22
a) Segment Level (n = 1, length = 0.5 miles)

b) Corridor Level (n = 64, length = 53.8 miles)

c) Metro Level (n = 4750, length = 2,583 miles)

d) National Level (n ~170,000, length ~ 127,000 miles)

Figure 1. Scalability of Probe Data
Figure 2. One-minute and 15-minute aggregate speed data for segment of I-95 in Figure 1a.
Figure 3. Speed overview for major metropolitan areas on July 2, 2015.
a) Indianapolis metropolitan area.

b) Chicago metropolitan area.

c) Los Angeles metropolitan area.

d) New York City metropolitan area.

Figure 4. Spatial representation of traffic speeds on July 2, 2015 at 5:00 PM local time
(a) Indianapolis metropolitan area at 5:00PM Eastern Daylight Time.

(b) Chicago metropolitan area at 5:00PM Central Daylight Time.

(c) Los Angeles metropolitan area at 5:00PM Pacific Daylight Time.

(d) New York City metropolitan area at 5:00PM Eastern Daylight Time.

Figure 5. Top congested routes in four metro areas, July 2, 2015.
Figure 6. Comparing week prior to Independence Day weekend for three metropolitan areas, 2015 (local time).
Figure 7. Routes in Los Angeles and New York over Independence Day, 2015.
Figure 8. Map highlighting miles of congestion per state at 5:00PM Eastern Daylight Time.
Figure 9. Map highlighting miles of congestion per county in Pennsylvania on I-80 in both directions at 5:00PM Eastern Daylight Time.
Figure 10. Congestion on I-80 over Independence Day weekend in all through states.