Virtual Sensors: A Web-based Real-Time Data Collection Methodology for Transportation Operation Performance Analysis

Ender Faruk Morgul, M.Sc. (Corresponding Author)
Graduate Research Assistant, Department of Civil & Urban Engineering, NYU-Poly
Center for Urban Science + Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor, Brooklyn, NY 11201
Tel: 1-(646)-997-0531
Email: efm279@nyu.edu

Hong Yang
Post-Doctoral Fellow, Department of Civil & Urban Engineering, NYU-Poly
Center for Urban Science + Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor, Brooklyn, NY 11201
Tel:1-(646)-997-0548
Email: hong.yang@nyu.edu

Abdullah Kurkcu, M.Eng.
Ph.D. Student, Department of Civil & Urban Engineering, NYU-Poly
Center for Urban Science + Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor, Brooklyn, NY 11201
Tel: 1-(646)-997-0546
Email: ak4728@nyu.edu

Kaan Ozbay
Professor, Department of Civil & Urban Engineering, NYU-Poly
Center for Urban Science + Progress (CUSP), New York University (NYU)
One MetroTech Center, 19th Floor, Brooklyn, NY 11201
Tel: 1-(646)-997-0552
Email: kaan.ozbay@nyu.edu

Bekir Bartin, Ph.D
Assistant Professor, Department of Civil Engineering,
Istanbul Kemerburgaz University, Mahmutbey Dilmenler Caddesi No:26,Bagcilar, Istanbul, Turkey
Tel: 00 1 (212) 604 01 00 Ext. 1411
Email: bekir.bartin@kemerbugaz.edu.tr

Camille Kamga, Ph.D.
Director, Region 2 University Transportation Center, Assistant Professor,
Dept. of Civil Engineering, The City College of New York,
Marshak Hall, Suite J-910, 138th Street & Convent Avenue, New York, NY 10031
Tel: (212) 650-8087
Email: ckamga@utrc2.org

Richard Salloum
Product Manager
Nokia Location & Commerce, 425 W. Randolph St, Chicago, IL 60606
Tel: (773) 554-3697
Email: richard.salloum@here.com

Word count: 5,286 + 3 Tables + 5 Figures = 7,286 words
Abstract: 149 words.

Paper Submitted for Presentation and Publication at the
Transportation Research Board's 93rd Annual Meeting, Washington, D.C., 2014
Re-submission Date: November 15, 2013

ABSTRACT

Recent advances in mobile networks and increase in the number of GPS-equipped vehicles have led to an exponential growth in real-time data generation. Over the last decade, a number of online mapping or vehicle tracking services haves made their data available for third-party users. This paper explores opportunities in utilizing the real-time traffic data provided by online services and introduces a virtual sensor methodology for collecting, storing and processing large volumes of network-level data. In order to assess the validity of the collected data using the proposed methodology, we compare these with data from physical loop detectors and electronic toll tag readers. Statistical analyses show that there is a strong correlation between the travel time measurements from infrastructure based sensors and virtual sensors. We then conduct a travel time reliability analysis using the virtual sensor data methodology and conclude that the results are promising for future research and implementation.

INTRODUCTION

The era of Big Data has arrived. Past practices in data exploration and utilization are changing drastically. Recent developments in mobile networks, cloud computing, and similar technologies have led to an exponential growth in data production and also increased the size of data that can be stored (*1*). It is estimated that 2.5 exabytes of data were created daily in 2012 and this number is expected to double every 40 months (*2*). The abundance of publicly accessible data brings a number of potential opportunities for researchers and scientists to better understand and evaluate real-life problems and to get better guidance in informed decision-making (*3*). Although the benefits of utilizing vast amount of data has shown to yield improvements in productivity in many fields (*4*), it is also important to assure that the quality of any processed data is good enough to avoid biased or unreliable conclusions.

As for many other fields of research, Big Data has very useful implications for transportation studies. For example, rapid increase in GPS-enabled mobile device adoption such as smartphones in recent years provides the opportunity of geo-tracking using the location information of mobile device users (*6*). In city traffic, tens of thousands of smartphone users, traffic sensors, traffic cameras, and computers in cars generate very large data sets of travel time, speed, and location information. The data created by each of these sources can be stored and processed in real time. It is well recognized that the resulting massive amount of traffic-related data will make important contributions to the operations and planning of transportation systems (*7*). The potential uses of Big Data are countless: with the help of Big Data procedures, researchers can make better transportation decisions such as optimizing operations, developing rational infrastructure plans, and examining the distribution and patterns of large public events. Specifically, traffic congestion is one of the important problems frequently revisited by using big data sources in terms of providing high quality traffic information, namely accurate estimates of travel times, incident detection, prediction of time-dependent origin-destination demands, etc. (*8*).

For example, state transportation agencies were used to rely on accident information from costly roadside sensors and traffic cameras. In New Jersey, these surveillance systems cover approximately only 5% of the highway network (*9*). If there is an accident on a roadway, by the time the accident is detected, travelers are alerted and the accident is cleared, several miles of queues could have been formed. However, the impact of an accident can be reduced by faster detection and better information dissemination with the help of traffic maps created using Big Data. According to high-technology experts; the bigger the data, the more ways of harnessing it are going to emerge (*9*).

In this paper we present a “virtual sensor” methodology using open traffic data sources from web-based map providers namely, Bing Maps^{TM} and MapQuest^{TM}. We first give a comprehensive review of existing automated traffic data collection methods along with a discussion about the advantages and disadvantages highlighted by the earlier studies. Next, we summarize the available web-based services that provide real-time traffic data for third-party applications. The next section explains the proposed “virtual” traffic sensing methodology and statistically compares the accuracy of real-time data gathered from online sources with the data from loop detectors and electronic tag readers collected along the New Jersey Turnpike (NJTPK). Finally, the conclusion section presents the summary of results and directions for future research.

BACKGROUND AND LITERATURE REVIEW

Performance measures are defined as indicators of system efficiency. For example, in the context of transportation, travel time variability is an emerging performance measure increasingly used by decision-makers in making many transportation investment decisions. Information on how long it would take to travel between specific points is a vital information for all travelers. Accurate estimation of travel times reflects the system performance based on users' point of view.

In this section we review the existing automated traffic surveillance methods under two major categories: 1) traditional methods that have been utilized for several decades for data collection 2) emerging technologies that use Big Data provided by web-based services for developers. We also present a summary of online services that offer traffic data, the data types they provide and potential applications for performance measure analysis.

Review of Traditional Traffic Surveillance Methods

Traditionally, traffic surveillance is conducted by using two distinct approaches:

- Road based technologies:
    - In-road detectors (inductive loop detectors, magnetometers, piezo electric detectors, pneumatic tubes)
    - Road-side detectors (video image detectors, active/passive infrared detectors, microwave sensors, radar sensors, ultrasonic sensors and passive acoustic sensors) and
- Vehicle-based technologies (probe vehicles, automatic vehicle location (AVL) systems, automatic vehicle identification (AVI) systems, wireless phones).

There are several advantages and disadvantages of these two surveillance technologies. Although vehicle-based technologies are useful in accurate measurement of travel times, they have not been widely used until recent years. This is primarily because of the high implementation costs.

Traffic detectors can be broadly categorized into 3 groups: single-point roadside detectors, multi-point roadside detectors, and area-wide mobile sensors. Each technology has its own strengths and weaknesses. The technical characteristics of the data sources and data types as well as recommended procedures for processing the data streams are summarized in Table 1.

Table 1. Summary of Automatic Traffic Detectors and Data Types

Detector	Pros	Cons
Single-Point Roadside Detectors		
Loop Detector	• Accurate estimation of freeway travel times (*10*). • Many existing freeways are equipped with loop detectors (*11*). • Higher accuracy achieved compare to other modes of detectors. • Coverage: highways and major roads • Provides comprehensive data about highway performance (*12*)	• When the sensors are not placed with a small gap, the resolution for incident detection is very low (*13*). • Approximate speed estimates can be made from single loop detectors but the resolution of this speed calculation is low (*13*). • High installation and maintenance cost (*14*, *15*). • Direct measurements are limited to vehicle counts and occupancy. No direct measurement of speed (*16*).

Radar	• Highest in accuracy (*17*). • Can detect vehicle speed, classification, headway and vehicle count (typically less accurate than inductive loop detectors) (*18*). • Ability to detect both moving and stopped vehicles (*19*).	• Vehicle mounted radars cannot detect vehicles properly within cross traffic at intersections. • Measurements have random fluctuations, needs careful filtering. (*19*). • Can experience dead detection zones and "ghost" vehicles (*18*).
Video	• Rich data source (*20*). • Can be used to collect volume, speed, occupancy, density, queue length, headway (*19*).	• High resolution trajectories over a small viewable area, provides unnecessary detail for planning applications. • Shadows from other vehicles and other sources may impact accuracy (*21*). • Environmental conditions may affect the video quality (*19*). • Setup and calibration is more complex and critical (*19*).
Multi-point Roadside Detectors		
License Plate Reader	• Travel time between points • Uses camera technology • Used almost universally since every vehicle is required to have license plates (*22*).	• Travel time observations must be filtered to remove outliers since vehicles detouring or stopping between detectors must be removed. • Challenging and import recognition technique (*23*) • Hard and expensive to implement due to the diversity of plate formats (*24*). • Outdoor conditions during image acquisition may create challenges (*25*).
RFID Transponder	• Cost effective system to identify vehicles (*26*). • Only detects RFID signatures from participants who have a transponder (*26*). • Tags are unique and very simple to install. • Provides good tracking system (*27*).	• Difficult for a reader to read the information in case of tags installed in metal products. • Interferences have been observed. • Damaged tags cannot be tracked. • Good for a specified range only (*27*).
Bluetooth	• High quality and reliable data on highways (*28*). • VMSs or websites can display real time data collected by Bluetooth devices (*29*). • OD pairs and route choice can be estimated using Bluetooth data (*29*). • Low implementation cost (*30*).	• Not suitable for highway segments located very close to local arterials and lane-by-lane data (*28*). • Highway segments with rest area or toll plaza between two sensors create noisy data (*28*). • Pedestrian and bicycle traffic may cause false data measurements for vehicular traffic (*31*). • High speed vehicles between two sensors are sources for outliers (*31*). • Biased results can be provided for slow moving vehicles (*32*). • Poor sampling rate of total traffic volume (*28*).
Wireless sensors	• Energy efficient (*33*). • Low cost and large scale deployment (*33*). • Re-identifies the magnetic signature of	• Cannot accomplish complex processing tasks and store large amount of data (*34*). • Limited mobility pattern due to street shapes, intersections, and vehicle features (*35*).

	vehicles passing over sensors to match vehicles	
Area-wide Mobile Sensors (Probe Vehicles)		
GPS	• Provides high quality data at a relatively low cost and extensive spatial coverage (*36*). • Contributes to monitoring and control of traffic system, including queue and incident detection, dynamic route guidance, real-time multimodal information for travelers, short-time forecasting (*36*). • Taxi-GPS data can be used for high-quality traffic monitoring in urban networks (*37*).	• May not be available in dense urban environments, tunnels, indoor parking lots, forest or underground (*35*). • May be affected or blocked by obstacles (*35*). • Different characteristics of vehicle types affect the measured traffic parameters. For example different acceleration/deceleration profiles of buses can give inaccurate travel time estimations (*38*).

It can be concluded that wide-area traffic data collection using road-side traffic surveillance units comes with higher implementation and maintenance costs. On the other hand, low cost alternatives such as RFID transponders, Bluetooth or wireless sensors have significant problems associated with small number of vehicles transmitting data on a route of interest and other technical difficulties in detecting moving vehicles. GPS-equipped probe vehicles such as taxi cabs, local buses or commercial vehicles offer high quality data streams for traffic monitoring with almost no additional costs. Although some portion of collected data are not made publicly available due to commercial or confidentiality reasons, there are open data sources from services providing real-time travel information such as NextBus™, which are discussed in detail in the following sections.

Emerging Technologies for Web-Based Traffic Surveillance Data

Fortunately, technological developments in Intelligent Transportation Systems (ITS) are opening up new avenues for traffic sensing, computing, and communication methods. Therefore, processing a large amount of real time traffic data has become increasingly manageable for traffic control and traveler information systems. Online mapping technologies have also grown steadily and reached significant popularity during the last decade. Google is the first company that launched Application Programming Interface (API) toolkit for online mapping services in 2005 (*39*). This interface allows software components to communicate with each other and allows users to gather the data offered by the service in different dataset formats (i.e. XML, JSON). Publishing APIs has allowed third-party developers to create an open environment for sharing content and data between each other and applications. In addition, numerous map-based web applications have become accessible via web browsers (*40*).

There are many web-based map APIs that can be embedded directly in websites and mobile applications served by different hosts. In addition to Google Maps™, service providers such as Bing Maps™, MapQuest™, NextBus™, Nokia Maps (Here)™, TomTom™ and many various size map services are available on the internet. In addition to the APIs, there are other companies that help users to create and host their own maps in a way that the style of the maps and applications can match each other. These advances in mapping technologies have resulted in creating new marketplaces such as selling, buying, and sharing geographic data.

Flammia et al. (*41*) suggested that web-based applications offer a promising opportunity for State DOTs and planning agencies to obtain large amounts of data for their transportation systems. In

2003, they developed a mapping website that was able to display county to county worker flows as lines and generate tables for any selected county. This study explored a couple of important technical differences between desktop applications and web-based applications. For example, users could connect to the mapping server simply by using a web browser and there was no need for a Geographic Information Systems (GIS) software or any geographical data installed on users' personal computers. The authors proposed that a web mapping application could carry out a couple of different functions to present graphical and interactive results of transportation studies.

Welch et al. (*42*) presented a system that offers simple, fast, and universal access to location-based traffic data. Their system had a web-based application that allows the viewing, insertion, and management of traffic data. Users only required a web browser and they could query traffic data to quickly access the required records. Data could be stored elsewhere as long as it was accessible by users' web browser in this system.

Thomas et al. (*43*) created system called T.R.A.F.F.I.C. which is an acronym representing Transit Response Analysis for Facilitating Informed Commuters and using Big Data sets to investigate traffic patterns. The system assists users to avoid traffic delays, and provide routing alternatives, and push notifications of traffic information. They used MapQuest API to get the turn-by-turn route directions and TRANSCOM[1] data to locate traffic congestion and events. The framework presented in the study showed promising functionality for the use of Big Data for traffic congestion detection however the system has not been implemented yet.

Inherently, travel time estimation depends on the topology of a transportation network which consists of a set of nodes and links. The travel time estimation procedure is mainly implemented by finding the shortest time from a determined origin to a destination node. In addition, several network elements should also be defined such as one-way streets, turning restrictions or signal timings. Online sources such as Google Maps, Nokia Maps, MapQuest provide useful options to solve the shortest route problem using frequently updated maps based on real-time traffic conditions. Although the information about the calculation of real-time travel duration is not disclosed by these services for confidentiality and commercial reasons, researchers attempted to measure the accuracy of their travel time data. Shortly after Google implemented the API system, Wang et al. (*39*) utilized transportation network data along with the routing algorithm of Google, and compared the reported travel times with those from ArcGIS Network Analyst module. The authors highlighted several advantages of using Google Maps API in travel time estimations such as the reduced necessity of network data preparation and GIS knowledge.

Open API services have also been used for research fields other than transportation. With the help of the powerful API technologies, it was possible to create a local-scale interactive web-based public health data system. Data on the system is easily accessible for a wide variety of communities such as policy makers, planners and non-profits who conduct research. Highfield et al. (*44*) updated the Community Health Information System (CHIS) and launched a web mapping application in Greater Houston Area. Their system conveys "query on the fly" technology that means the data are not created until the query is generated in the system.

[1] TRANSCOM is a coalition of 16 transportation and public safety agencies in the New York – New Jersey – Connecticut metropolitan region. It was created in 1986 to provide a cooperative, coordinated approach to regional transportation management.

On the other hand, APIs offered by social networking services such as Twitter™ have also been utilized for research purposes. User tweets may contain valuable information for tracking or forecasting events while the useless tweets might take much more space in Twitter. Signorini et al. (*45*) used Twitter, micro-blogging service, to track down public interest and used the data for forecasting future activity in monitoring influenza related events in the U.S. The results showed that Twitter data traffic can be effectively used to estimate possible disease-related activity in real time and proposed methodology can decrease the estimation duration by a couple of weeks compared to the current forecasting practice.

Available Online Services for Traffic Data

Web mapping APIs deliver several HTTP web services such as static map, directions, distance matrix, elevation, geocoding, and places. Web mapping applications provide efficient methods to visualize large amounts of datasets, such as supplying real time traffic data to users. Responses from the mapping services are usually delivered in XML or JSON formats which can be easily processed in almost any computer language. Services such as Google Maps and Bing Maps have very comprehensive real-time traffic coverage around the world. Table 2 gives a summary of selected web-based services that offer traffic information for developers through APIs.

Nokia (Here) Maps

Nokia launched its traffic services in 2002. After purchasing on-board navigation company Navteq™ in 2007, Nokia has been one of the major suppliers of online map-based data. While Nokia provides APIs for web browsers, it is mostly focused on providing traffic content to a variety of customers across many industries and channels, including automotive, navigation devices, mobile phone companies, public sector agencies, etc. The company currently provides real-time data in 34 countries around the world and also historic traffic data in 77 countries with with new countries added each year. It started supplying both traffic and map content to Bing Maps in 2012, and to Yahoo Maps since 2011.

Nokia owns and manages several thousand sensors on roadways all over the country. Sensor data is one key input to traffic reporting, along with GPS probe data and incident data. Nokia identifies and creates incident data such as accidents, construction, and road closures. They also perform quite a lot of ground truth testing to ensure that the speeds are as accurate as possible. Nokia is currently processing approximately 20 billion individual probe points globally per month, with 4 billion of those coming from solely the USA road network. Probe sources are a mixture of both passenger and commercial vehicle and the probe vehicle volume has been growing at a very significant pace.

Nokia developer APIs are available under three different categories; JavaScript API and REST API, and Mobile HTML5 framework (*46*). The JavaScript API allows users to build web applications and it consists of libraries of classes and methods with different functionalities. Nokia Map Image API or Representational State Transfer (REST) API is a web service API that provides easy and quick access to map images. Users can formulate a request that combines the URL and set of parameters such as location and zooming to retrieve a map image. Mobile HTML5 framework is also a JavaScript framework designed for creating location-based mobile web apps and provides several features that users need for creating web applications, including, instantaneous map rendering, routing and directions for transit.

Google Maps

The Google Maps API is the first API system that made embedding maps into several web applications very convenient for researchers, even for those with a limited knowledge in programming. More than 1 million Maps API sites and mobile applications use Google Maps APIs, which constitute one of the most popular JavaScript libraries on the web. The Google Maps JavaScript API allows users to insert maps into external web pages and overlay any kind of data onto maps. The latest version (i.e. version 3) is designed to be faster and more applicable for both mobile devices and desktop browser applications (*48*). The API offers a number of functions for controlling maps and adding content to the map through a variety of services.

Microsoft Bing Maps

Bing Maps was launched by Microsoft in 2005 as MSN Virtual Earth. It offers specifications such as locations service, route service, imagery service, and traffic service, as well as geocoding and searching options. Bing Maps Spatial Data Services delivers services for batch geocoding, hosting, managing and querying customers own point data within Bing.

There are three different API protocols in Bing Maps. Firstly, AJAX/JavaScript APIs are primarily for web browsers and web applications. They are similar to other services in that there are no additional plug-ins or applications needed beyond an internet browser. However, JavaScript lacks advanced features such as graphic design, transparency, and animation. Secondly, SOAP/XML APIs are server sided calls. Users can develop .NET applications that support XML and take the advantage of advanced features such as graphic design. Finally, Silverlight APIs run inside the user's internet browser but they use .NET based APIs. These interfaces have rich graphic design and high performance display.

MapQuest

MapQuest has been providing maps for nearly 15 years on the web. MapQuest is the only company that provides licensed and open data at simultaneously, and gives an option to choose the version users need (*49*). Licensed Data offers businesses and developers map resources, industry expertise for building rich maps, web and desktop applications, and developer tools. The data, including traffic data, is automatically updated for licensed data owners. Users are able to upload and search custom data. On the other hand, Open Data, which uses the free editable map service OpenStreetMap and free satellite and aerial data, is being updated every day by users.

All the map tiles are updated 15 minutes after OpenStreetMap data has been edited (*49*). There are separate JavaScript and Flash/Flex APIs for different types of data. Geocoding, The Traffic Web Service and Search Web Services options only exist on Licensed Data.

Table 2. Summary of Selected Web-based Services that Offer Traffic Information for Developers

	API	Geocoding Service	Transit Integration	Live Traffic Info.	Directions	Distance Matrix	Map Data Providers	Offered Services	Mobile app.
Nokia Maps™ (*46*)	JavaScript, REST, Mobile HTML5	2,500 daily limit (Base Plan) / 10,000 daily lmit (Core Plan)	Yes	Yes	2,500 daily limit (Base Plan) / 10,000 daily lmit (Core Plan)	2,500 daily limit (Base Plan) / 10,000 daily lmit (Core Plan)	Navteq	Positioning, Routing, Traffic	Yes
Bing Maps™ (*47*)	AJAX. WPF, WP, Android, iOS, Silverlight, REST, SOAP, Win 8 (.NET, JS)	Yes	Yes	Yes	Yes	Yes	Navteq, Intermap, Pictometry International, NASA	Geocode, Imagery, Route, Search, Common Classes and Enumerations	Yes
Google Maps™ (*48*)	Javascript, iOS SDK, Android SDK	Request per day 2,500 (free license) / 100,000 (business license)	Yes	Yes	Request per day: 2,500 (free license) / 100,000 (business license)	100 elements per query (free license) / 625 elements per query (business license)	MAPIT, TeleAtlas, DigitalGlobe, MDA Federal	Directions, Distance Matrix, Elevation, Geocoding, Maximum Zoom Imagery, Street View.	Yes
MapQuest™ (*49*)	JavaScript, AS3/Flex, SDK,iOS	No preset limit (Open data) / 5,000 cals/day (Licensed Data)	Yes	Yes (Only for License d Data)	No preset limit (Open data) / 5,000 cals/day (Licensed Data)	No preset limit (Open data) / 5,000 route pairs /day (Licensed Data)	Navteq, OpenStreetMap user contributions	Directions, Geocoding, Search, Route Matrix, Traffic	Yes
NextBus™ (*50*)	Yes	No	Yes	No	No	No	Cubic Transportation Systems, Inc.	Real-Time Passenger Information	Yes
TomTom™ (*51*)	Yes	Yes	No	Yes	Yes	No	TomTom International BV, Whereis	Maps, Routing, Geocoding, Traffic	Yes

NextBus

NextBus is a web-based service that provides real-time bus trip information for passengers, in particular estimating what time a bus is expected to arrive to a given bus stop. The system is based on GPS tracking technology and real-time routing of transit vehicles which enables accurate estimations with instantaneous updates considering external factors such as heavy traffic conditions or delays (*50*). The location data acquired from partner transit-agencies' buses are made available publicly for developers without a number of request limit. The data can be gathered by the minute in XML format for selected transit routes. The geographical location of bus stops, estimated time to next stop and schedule delay information are also provided by the system. A combination of these data fields can be used for estimating the travel times on bus routes, taking into consideration the effect of vehicle specific acceleration/deceleration profiles while approaching and leaving a bus stop.

INRIX

INRIX™ monitors traffic flows across more than 260,000 miles of U.S. and Canadian highways, provides real-time traffic information for 32 countries across North America and Europe as well as information that comes from 800,000 vehicles equipped with GPS devices (*52*). In addition, INRIX receives information from road sensors located in about 9,000 miles of highways. It is the only crowd-sourced traffic network and it receives the information from commercial fleets – taxi cabs, delivery vans and long-haul trucks, and mobile devices. INRIX also reports incidents and unique local variables (*53*). INRIX offers developers real-time traffic and routing information using API access. Kim et al. (*54*) evaluated the accuracy of travel times based on Bluetooth sensors, TRANSMIT[2] (Electronic toll tag) readers, and INRIX data. They compared the travel times with the ground truth data and worked on the study segment of I-287 in New Jersey. They concluded that the speeds of probe vehicles are closer to the estimated speed using Bluetooth sensors than the INRIX data. In addition, INRIX data showed some latency issues.

VIRTUAL SENSOR METHODOLOGY

The purpose of this section is to investigate the opportunity to develop a "virtual sensor" methodology using Big Data provided by web-based mapping services described above. In particular, we are interested in the real-time traffic information that can be used as a substitute for traditional data collection technologies such as loop detectors. If the validity of the Big Data obtained from the virtual sensor concept can be shown statistically via comparison with physical detectors' data, then it could be concluded that similar web data based frameworks can be used for monitoring sections of a transportation network without any fixed traffic surveillance units. This is a very desirable feature for traffic monitoring especially on local roads for non-recurrent congestion conditions. Travel time reliability which is another important traffic performance measure can also be extracted using real-time traffic information from the proposed virtual sensor methodology.

[2] TRANSMIT readers are stationary roadside vehicle detectors that detect vehicles equipped with E-ZPass tags as anonymous probes to determine traffic speeds and travel times between readers. The TRANSMIT system and the readers are operated by TRANSCOM.

We propose a web-based "virtual sensor" methodology where real-time traffic information can be obtained using online data stream from Bing Maps and MapQuest. Figure 1 shows the proposed methodology of the virtual sensor concept. The coding of this methodology is performed using the JavaScript programming language. The two requirements for data collection are an internet connection and a web browser. The system is easy to use even for novice computer users. Geographical coordinates of origin destination (OD) pairs for the sections to be monitored needs to be defined by the user within the code. Bing Maps' REST API and MapQuest Open Data Map API services are utilized for extracting real-time traffic based routing information. Routing requests for each of the defined OD pairs are sent every five minutes automatically using the developed JavaScript code with a web browser and the responses are received from the server in XML format. There is no limit for the number of requests per day using Bing Maps API service or MapQuest Map API although some other services (i.e. Google) may put a limit on maximum number of daily requests for free user accounts. XML Document Object Model (DOM) platform is used to retrieve the traffic data from the response information. Users define the mode of travel in the code by choosing driving, walking and transit. In the proposed methodology we use routing for driving since we are interested in performance measures on highway networks. The response data includes routing information including point-to-point total travel duration, total distance and whether there exists an incident on the given route. Incidents are also placed into groups such as work zones, accidents or road constructions. Data such as the severity of the incident, whether the road is closed or not, and if the incident is verified by responsible transportation agencies are provided with detailed description of each item. The real-time response information is automatically saved in CSV or any other desired data format[3].

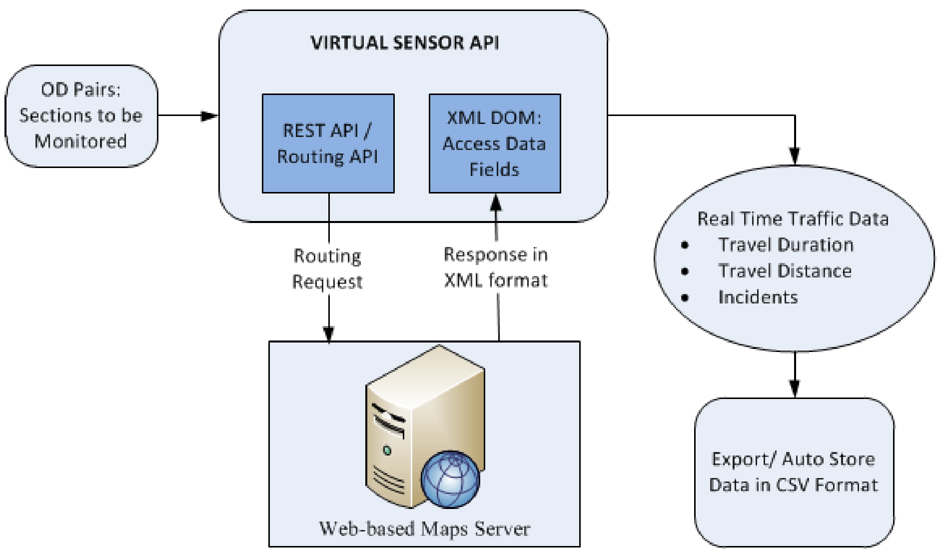

Figure 1. Virtual Sensor Methodology.

CASE STUDY

To demonstrate the functionality of the proposed virtual sensor concept we tested our methodology on the New Jersey Turnpike (NJTPK), for real-time travel time estimation. For our first case study we used 4 OD pairs to test the accuracy of traffic data for different cases including both recurrent

[3] The source code for virtual sensor methodology is available upon request.

and non-recurrent traffic conditions. Traffic data was investigated between interchanges 7 to 7A and 8 to 8A in both directions. The selected sections were already equipped with remote traffic microwave sensor (RTMS) stations which are used as reference data source for this analysis. Figure 2 depicts the locations of the eleven RTMS sensors between interchanges 7 and 7A. It can also be observed that there are only two virtual sensors, which are sufficient to gather both travel time and incident information for the same section. In particular, virtual sensors are reference points in the map which reflect the start and end of the section under study. It is noteworthy to mention that real-time traffic data assumes fixed travel speeds for certain sections depending on the congestion level. Since the virtual sensor data is generated by mainly through probe vehicles, collected space mean data also incorporates local disturbances.

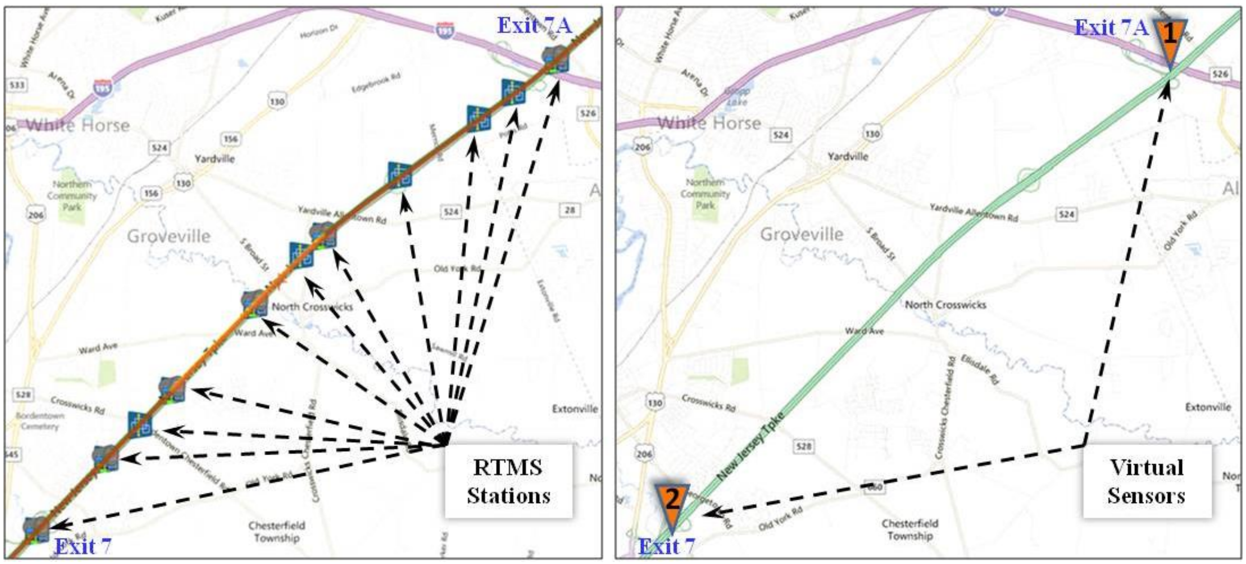

Figure 2. Sensor Locations between Interchanges 7 and 7A.

Travel time derived from virtual sensors on both Bing Maps and MapQuest were compared with the reference data and example results are shown in Figure 3. Observation of the graphs make it apparent that, travel time estimation is highly correlated between use of a virtual sensor and derivation from RTMS data throughout the entire day. As depicted in these figures, virtual sensors also capture fluctuations in travel time due to collisions, disabled vehicles, work zones, and recurrent congestions similar to RTMS. It should be mentioned that all incidents (i.e. crash, work zone, disabled vehicle) information shown in the figures was obtained from Bing Maps.

The Wilcoxon signed-rank test was applied to statistically examine the two-paired samples from RTMS stations and virtual sensors one of the maps (The null hypothesis H_0 is that the two datasets are equivalent). As shown below in Table 3, most of the corresponding p-values were less than 0.05 indicating that the results from the web-based virtual sensors and RTMS are not statistically different within a 95% confidence interval. Thus it can be concluded that the travel time measurements derived from the virtual sensors are consistent with the RTMS data. Therefore, similar to the RTMS data, virtual sensor measurements can also help to capture incident information that may disrupt normal traffic conditions. For the time period RTMS data was collected, continuous data flow was disturbed several times possibly due to malfunctioning or equipment shut downs for maintenance. For some of the sections we analyzed, missing data rate were as high as 10% of all collected data during the analysis period whereas with virtual sensor data almost flawlessly through the same time period. This is a promising conclusion for the future of web-based data as a possible additional data source to supplement traditional data collection methods.

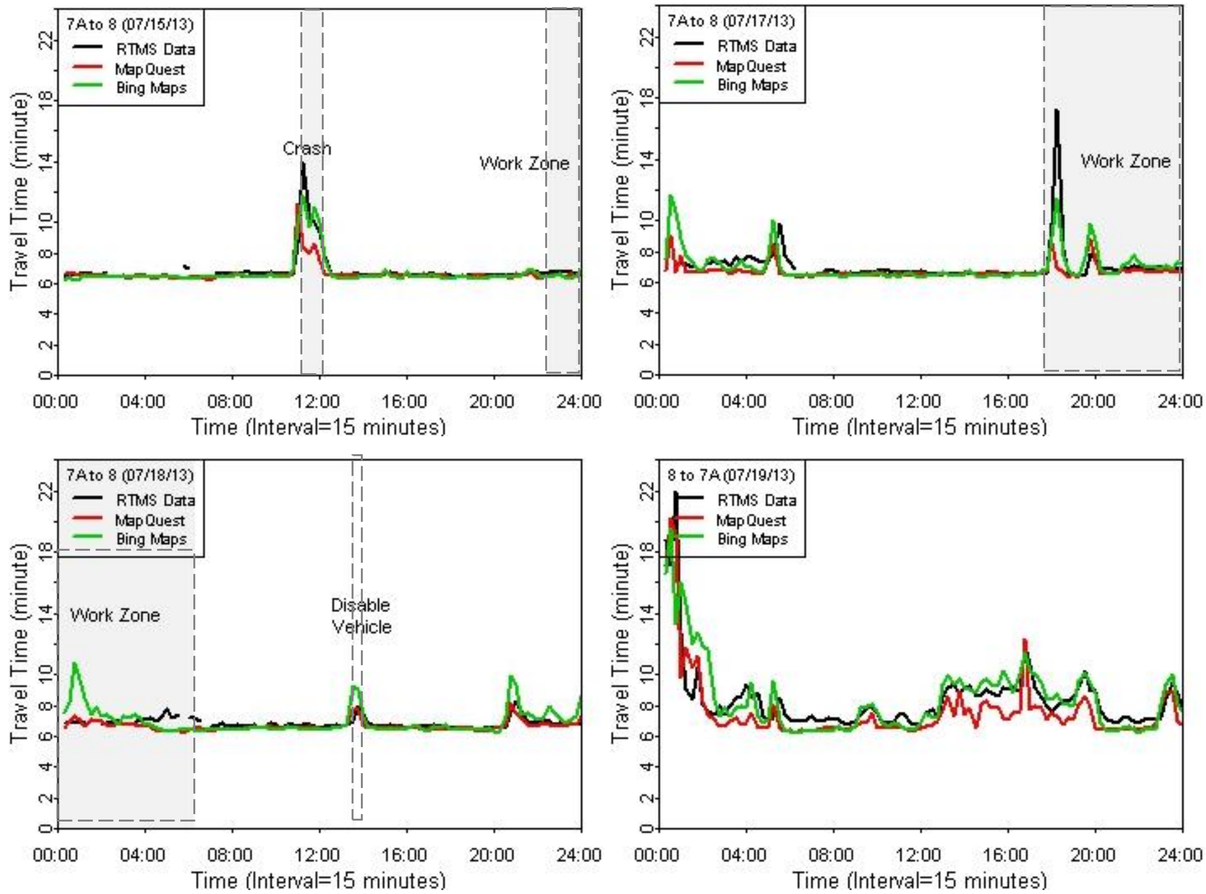

Figure 3. Travel Time Comparison: Virtual Sensors vs. RTMS.

Table 3. Wilcoxon Signed-Rank Test to Compare Virtual Sensors and RTMS

Link	Test Sensor	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
1(7 to 7A)	MQ	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**
	BM	<0.001**	<0.001**	0.047**	0.257	0.679	<0.001**	0.041**
2(7A to 7)	MQ	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**
	BM	<0.001**	<0.001**	<0.001**	<0.001**	0.684	0.433	0.035**
3(8 to 8A)	MQ	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**
	BM	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**
4 (8A to 8)	MQ	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**
	BM	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	0.001**	0.175

** : 95% confidence interval MQ: MapQuest
* : 90% confidence interval BM: Bing Maps

Travel time estimations using RTMS data is conducted for the mainline sections of the highway. Therefore traffic delays at toll booths are not captured within the comparison analysis. Another method of measurement is through the use of tag reader data for exit-to-exit travel times. We conducted a second analysis to compare virtual sensor travel times with ETC travel times which depicts the average of observed travel times for all vehicles passing through toll plazas. Selection of the study sections is mainly affected by the sample size distribution throughout the day. Since some sections have very low recordings during certain time periods (i.e. mostly at night) average travel time estimation using ETC data may result in unrealistic estimations. Similarly, rest/service areas in some sections result in comparably high travel times for those travelers who make stops along the way. For comparison analysis we selected the two sections that contained the least number of outliers between interchanges 9 to 10 in the northbound direction and interchanges 11 to 10 in the southbound direction. Virtual sensors were located directly at the toll plazas for this case. Figure 4 depicts the comparison between virtual sensor data sources and ETC data. It can be seen that unexpected changes in travel time due to incidents can be captured as well as travel times under regular traffic conditions. It should be noted that for certain time intervals ETC data suffers from a low number of observations simply because very few vehicles entered and exited between the selected toll plazas. For these cases outliers as shown in Figure 4, can be observed in the ETC data where the virtual sensor data displays comparably less fluctuations.

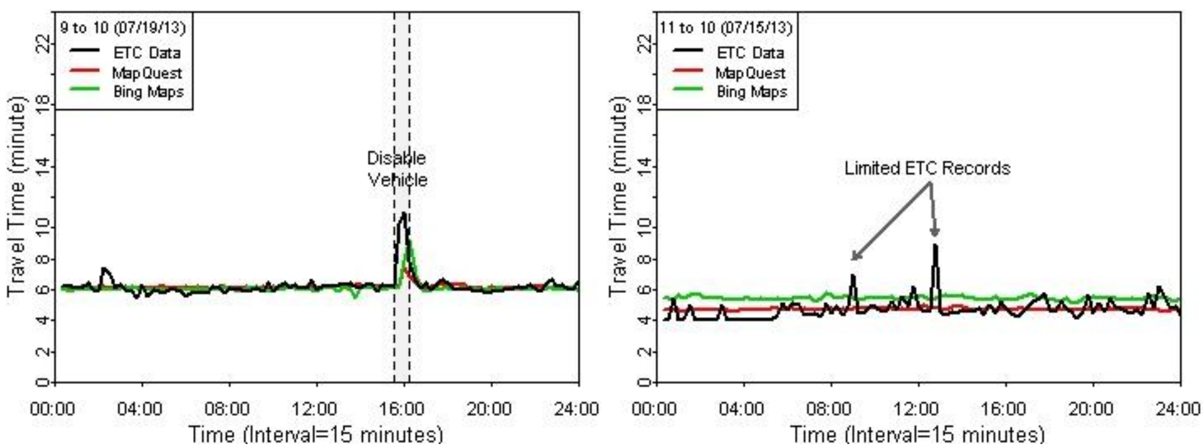

Figure 4. Travel Time Comparison: Virtual Sensors vs. ETC Tag Data.

We further investigated the travel time reliability on a 78-mile section of the NJTPK between interchanges 2 and 11 using virtual sensor data. There are eleven sub-sections which are exit-to-exit segments along facility and for each sub-section reliability measures were calculated separately. Buffer index (BI) is used as the reliability parameter, which represents the extra time (or time cushion) most travelers add to their average travel time when planning trips to ensure on-time arrival. It is calculated as the ratio of the difference between the 95th percentile travel time and mean travel time to the mean travel time (MTT). For a specific route trip and time period, the buffer index may be computed using the following equation:

$$\text{Buffer Index}(\%) = \frac{95^{th}\text{ percentile travel time} - \text{average travel time}}{\text{average travel time}} \quad (1)$$

Figure 5 depicts the results from the one-week collection of real-time data using virtual sensor methodology from both Bing Maps and MapQuest and as well as the reliability parameters comparisons using heat maps. The vertical axis displays the road sections and the horizontal axis the time-of-day. Buffer index is a parameter between zero and one with increasing values representing lower reliability or higher variability in travel times. As seen in both northbound and southbound segments, real-time data from both sources yield similar results in terms of subsections' travel time reliability. The northbound segment observed the most unreliable travel times during the afternoon peak hours (i.e. between 4:00 pm and 8:00 pm) while the southbound segment, in addition to afternoon peak hours, observed unreliable travel times during the midday period (i.e. around 12 pm). Sub-sections with higher travel time variability may also be identified using the proposed methodology, for example, subsections 4 to 8 have comparably higher travel time variability during afternoon peak for the northbound segment.

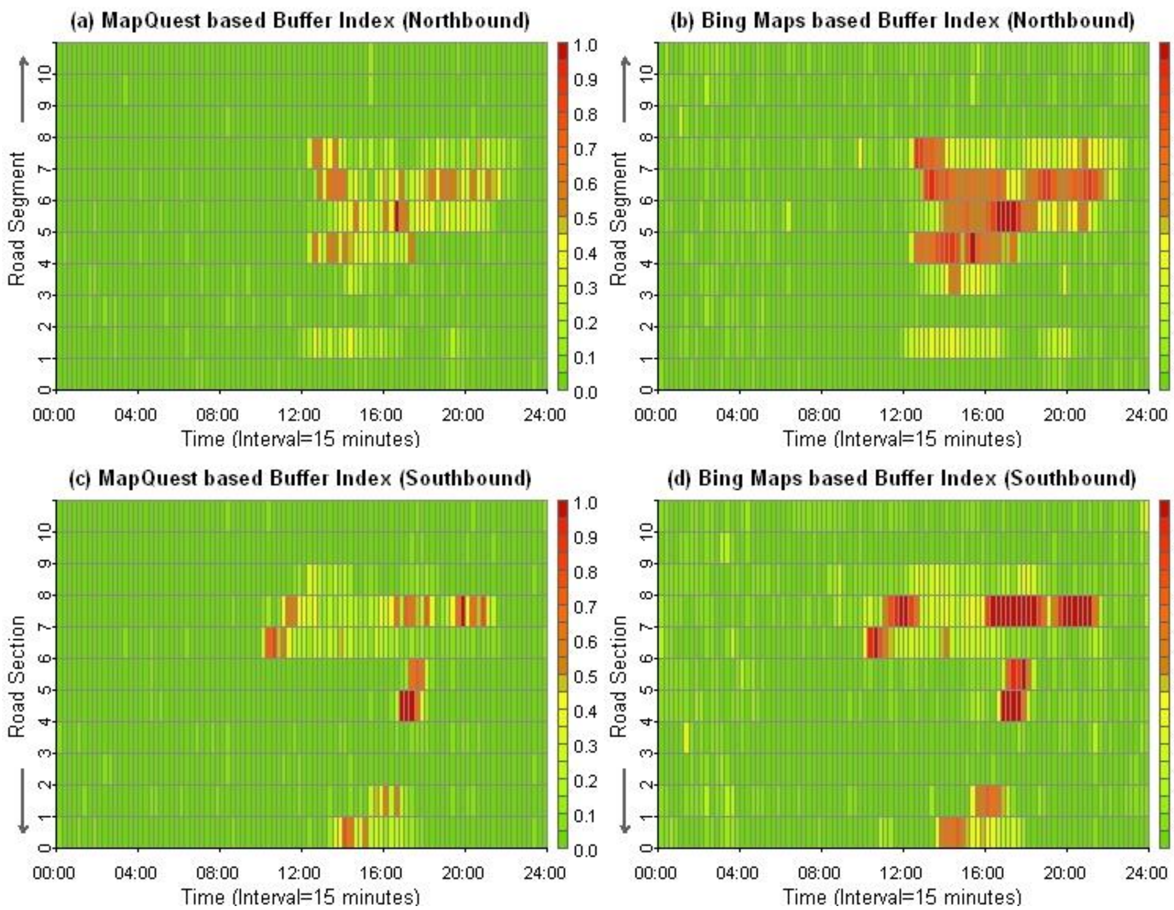

Figure 5: Buffer Index Analysis using Virtual Sensor Data.

The case studies presented in this section show the potential power of web-based data acquisition for future transportation studies. The presented virtual sensor methodology comes with almost no additional cost while the quality of obtained data is shown to be quite satisfactory compared to physical sensors. Virtual sensor concepts can collect vast amounts of high-quality data for nearly any highway section in any region of the United States to supplement the existing data sources with minimal effort and cost.

CONCLUSION

Travel times can be collected from a large number of sources. Conventionally, fixed detectors such as inductive loops embedded in the pavement have been used to measure vehicle flows and estimate speeds. Recent technological advances and widespread deployment of Global Positioning Systems (GPS) in consumer devices make mobile data sources a promising and potentially cost-effective way to monitor transportation systems. GPS integrated into cellular phones and in-vehicle navigation devices are now ubiquitous, and numerous technologies exist to identify vehicles or the signals from smart devices in them using roadside sensors (e.g., RFID or Bluetooth). It is now possible to measure location, speed, and travel time from probe vehicles, vehicles that are already in the traffic stream, in order to estimate traffic conditions in real-time across large networks. Real-time traffic data is often collected and archived so that a continuous historical record of traffic conditions can be used for analysis.

Problems associated with the traditional traffic surveillance methods were discussed in several studies in the literature. Major concerns are the high implementation and maintenance costs for fixed equipment and low penetration rates for the probe vehicle data. In this study we propose a different and novel approach that takes advantage of the emerging web-based Big Data sources available for transportation research. Web-based services already compile a large amount of “crowdsourcing” data from probe-vehicles or mobile devices and then make their real-time data available to developers and researchers. We first give a review and a summary of the online services that provide real-time or historical traffic data using their API services with the main goal of shedding light to the possible uses of these “new” data sources.

The success of any traffic management and planning application depends on several factors, the most important of which is the “quality of information”. In this study, we present a virtual sensor methodology based on Bing Maps API and MapQuest Map API traffic data. Data quality is tested through the comparison of travel time estimations from virtual sensors with physical loop detector and electronic tag reader data for different sections of the NJTPK. The results of these statistical comparisons are promising for future research especially in terms of travel time reliability analysis.

With the advances in data collection technologies, more and more data are being generated every day and the future of web-based virtual sensor concept is encouraging because it offers low cost and high-quality data for research and deployment purposes. Although with the increasing size, data management has also become a growing challenge for transportation professionals and researchers, emerging technologies such as cloud computing services can help handle and process Big Data. Moreover the presented methodology can be an attractive alternative traffic surveillance method for transportation agencies with governmental budget constraints.

ACKNOWLEDGMENTS AND DISCLAIMER

The contents of this paper only reflect views of the authors who are responsible for the facts and accuracy of the data and results presented herein. The contents of the paper do not necessarily reflect the official views or policies of sponsoring agencies. The authors appreciate the Region 2 University Transportation Research Center (UTRC) for partially funding this study.

REFERENCES

1. Bollier, D. The Promise and Peril of Big Data. The Aspen Institute, Washington D.C., 2010. http://www.emc.com/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf.

2. McAfee, A., and E. Brynjolfsson. Big Data: The Management Revolution. *Harvard business review*, Vol. 90, no. 10, October 2012, pp. 60-66.

3. Brown, B., M. Chui, and J. Manyika. Are You Ready for the Era of 'Big Data'? *McKinsey Quarterly*, Vol. 4, 2011, pp. 24-35.

4. Bughin, J., M. Chui, and J. Manyika. Clouds, Big Data, and Smart Assets: Ten Tech-enabled Business Trends to Watch. *McKinsey Quarterly*, Vol. 4, 2010, pp. 26-43.

5. Boyd, D., and K. Crawford. Critical Questions for Big Data: Provacations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, Vol. 15, no. 5, June 2012, pp. 662-679.

6. Laurila, J. K., D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The Mobile Data Challenge: Big Data for Mobile Computing Research. in *Proceedings of the Workshop on the Nokia Mobile Data Challenge, in Conjuction with the 10th International Conference on Persasive Computing*, 2012, pp. 1-8.

7. Keller, A., S. E. Koonin, and S. Shipp. Big Data and City Living - What Can It Do for Us? *Significance*, Vol. 9, no. 4, August 2012, pp. 4-7.

8. Lane, N. D., E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A Survey of Mobile Phone Sensing. *Communications Magazine, IEEE*, Vol. 48, no. 9, 2010, pp. 140-150.

9. Ovide, S. Tapping 'Big Data' to Fill Potholes. *Dallas Mobility*, June 2012. http://dallasmobility.org/Newsroom/6-12-12%20WSJ%20-%20Tapping%20big%20data%20to%20fill%20potholes.pdf. Accessed July 23, 2013.

10. Petty, K. F., P. Bickel, M. Ostland, J. Rice, F. Schoenberg, J. Jiang, and Y. Ritov. Accurate Estimation of Travel Times from Single-Loop Detectors. *Transportation Research Part A: Policy and Practice*, Vol. 32, no. 1, January 1998, pp. 1-17.

11. Guo, H., and J. Jin. Travel Time Estimation with Correlation Analysis of Single Loop Detector Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1968, no. 1, January 2006, pp. 10-19.

12. Jia, Z., C. Chen, B. Coifman, and P. Varaiya. The PeMS Algorithms for Accurate, Real-time Estimates of g-factors and Speeds from Single-Loop Detectors. in *Proceedings of the 4th Intelligent Transportation Systems Conference*, Oakland, CA, 2001.

13. Naranjo, J. E., F. Jiménez, F. J. Serradilla, and J. G. Zato. Floating Car Data Augmentation Based on Infrastructure Sensors and Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, no. 1, March 2012, pp. 107-114.

14. Zhang, J., F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen. Data-Driven Intellgient Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, no. 4, December 2011, pp. 1624-1625.

15. Zheng, F., and H. V. Zuylen. Urban Link Travel Time Estimation based on Sparse Probe Vehicle Data. *Transportation Research Part C*, Vol. 31, April 2012, pp. 145-157.

16. Guo, J., J. Xia, and B.L. Smith. Kalman Filter Approach to Speed Estimation Using Single Loop Detector Measurements under Congested Conditions. *Journal of Transportation Engineering*, Vol. 135, December 2009, pp. 927-934.

17. Advanced Traffic Analysis Center. Non-Intrusive Traffic Detection Comparison Study Final Report. North Dakota Department of Transportation, North Dakota, 2009.

18. Minge, E. Evaluation of Non-Intrusive Technoliges for Traffic Detection. Minnesota Department of Transportation, Minneapolis, Final Report MN/RC 2010-36 , 2010.

19. Middleton, D., and R. Parker. Vehicle Detector Evaluation. Texas Transportation Institute, Austin, Texas, FHWA/TX-03/2119-1, 2002.

20. Wang, Y., Y. Zou, H. Shi, and H. Zhao. Video Image Vehicle Detection system for Signaled Traffic Intersection. in *Ninth International Conference on Hybrid Intelligent Systems*, Shenyang, 2009, pp. 222-227.

21. Kranig, J., E. Minge, and C. Jones. Field Test for Monitoring of Urban Vehicle Operations Using Non-Intrusive Technologies. Minnesota Department of Transportation, Minneapolis, Final Report FHWA-PL-97-018, 1997.

22. Fleming, P. R. Methods and Systems for Enhancing Read Accuracy in Automated License Plate Reader Systems. US 2011/0084126 A1, April 14, 2011.

23. Wang, W. License Plate Recognition Algorithm Based on Radial Basis Function Neural Networks. *2009 International Symposium on intelligent Ubiquitous Computing and Education*, May 2009, pp. 38-41.

24. Zheng, D. N., Y. N. Zhao, and J. X. Wang. An Efficient Method of License Plate Location. *Pattern Recogit. Lett.*, Vol. 26, no. 15, November 2005, pp. 2431-2438.

25. Wen, Y., Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen, and P. Shi. An Algorithm for License Plate Recognition Applied to Intelligent Transportation System. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, no. 3, September 2011, pp. 830-845.

26. Mittal, A. K., and D. Bhandari. A Novel Approach to Implement Green Wave System and Detection of Stolen Vehicles. in *2013 IEEE 3rd International Advance Computing Conference*, Ghaziabad, 2013, pp. 1055-1059.

27. Kalbande, D. R., N. Deotale, Singhal, and S. Shah. An Advanced Technology Selection Model using Neuro Fuzzy Algorithm for Electronic Toll Collection System. *International Journal of Advanced Computer Science and Applications*, Vol. 2, no. 4, 2011, pp. 97-104.

28. Haghani, A., M. Hamedi, K. F. Sadabadi, S. Young, and P. Tarnoff. Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record*, Vol. 2160, 2010, pp. 60-68.

29. Haseman, R. J., J. S. Wasson, and D. M. Bullock. Real Time Measurement of Work Zone Travel Time Delay and Evaluation Metrics. *Transportation Research Record: Journal of Transportation Research Board*, Vol. 2169, 2010, pp. 43-50.

30. Pulugurtha, S. S., V. R. Duddu, S. Imran, P. Najaf, and Z. Reza. Commercial Remote Sensing and Spatial Information (CRS & SI) Technologies for Reliable Transportation Systems. Draft Synthesis Report on Literature Review. *Research and Innovative Technology Administration (RITA), USDOT*, 2013.

31. Van Boxel, D., W. H. Schneider IV, and C. Bakula. Innovative Real-Time Methodology for Detecting Travel Time Outliers on Interstate Highways and Urban Arterials. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2256, 2011, pp. 60-67.

32. Malinovskiy, Y., Y. J. Wu, Y. Wang, and U. K. Lee. Field Experiments on Bluetooth-Based Travel Time Data Collection. *Presented at the 90th Annual Meeting of Transportation Research Board*, 2010.

33. Yousef, K. M., J. N. Al-Karaki, and A. M. Shatnawi. Intelligent Traffic Light Flow Control System Using Wireless Sensor Networks. *Journal of Information Science and Engineering*, Vol. 26, 2010, pp. 753-768.

34. Buttyan, L., and T. Holczer. Traffic Analysis Attacks and Countermeasures in Wireless Body Area Sensor Networks. in *2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, San Francisco, CA, 2012, pp. 1-6.

35. Dogru, N., and A. Subasi. Traffic Accident Detection By Using Machine Learning Methods. in *Third International Symposium on Sustainable Development*, Sarajevo, Bosnia and Herzegovina, 2012, pp. 468-474.

36. Rahmani, M., H. N. Koutsopoulos, and A. Rnganathan. Requirements and Potential of GPS-based Floating Car Data forTraffic Management: Stockholm Case Study. in *2010 13th International IEEE Annual Conference on Intelligent Transportation Systems*, Madeira Island, Portugal, 2010, pp. 730-735.

37. Morgul, E. F., K. Ozbay, S. Iyer, and J. Holguin-Veras. Commercial Vehicle Travel Time Estimation in Urban Networks Using GPS Data from Multiple Sources. in *Transportation Research Board 92nd Annual Meeting*, Washington, D.C., 2013.

38. Uno, N., F. Kurauchi, H. Tamura, and Y. Iida. Using Bus Probe Data for Analysis of Travel Time Variability. *Journal of Intelligent Transportation Systems*, Vol. 13, 2009, pp. 2-15.

39. Wang, F., and X. Yanqing. Estimating O-D Travel Time Matrix by Google Maps API: Impelemtation, Advantages, and Implications. *Annals of GIS*, Vol. 17, no. 4, December 2011, pp. 199-209.

40. Wu, Y. J., and Y. Wang. An Interactive Web-based System for Urban Traffic Data Analysis. *International Journal of Web Applications*, Vol. 1, no. 3, September 2009, pp. 115-125.

41. Flammia, G., and A. Rabinowicz. Turning Transportation Planning Data Into Effective Web Sites. in *In Proceedings of the Ninth TRB Conference on the Application of Transportation Planning Methods*, Baton Rouge, 2003.

42. Welch, T. J., K. Tufte, R. S. McCourt, R. L. Bertini, and A. Snook. Implementing a User-Oriented Web-based Traffic Data Management and Archive System. *TRB 2007 Annual Meeting*, November 2007.

43. Thomas, D., J. Lawrence, and J. Rettig. *T.R.A.F.F.I.C.*, 2012. http://www.jacobrettig.com/sites/default/non-drupal/668/G1Gps/TRAFFIC.pdf. Accessed July 20, 2013.

44. Highfield, L., J. Arthasarnprasit, C. A. Ottenweller, and A. Dasprez. Interactive Web-based Mapping: Bridging Technology and Data for Health. *International Journal of Health Geographics*, Vol. 10, no. 69, 2011.

45. Signorini, A., A. M. Segre, and P. M. Polgreen. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS ONE*, Vol. 6, no. 5, May 2011.

46. Nokia (HERE) Maps. *HERE*, http://developer.here.com/documentation. Accessed July 23, 2013.

47. Microsoft Corporation. *Bing Maps*, http://www.bing.com/dev/en-us/dev-center.

48. Google. *Google Maps*, https://developers.google.com/. Accessed July 23, 2013.

49. *MapQuest*, developer.mapquest.com. Accessed July 23, 2013.

50. NextBus Incorporated. *NEXTbus*, April 4, 2013. http://www.nextbus.com/xmlFeedDocs/NextBusXMLFeed.pdf. Accessed July 23, 2013.

51. *TomTom*, http://developer.tomtom.com/. Accessed July 23, 2013.

52. Inrix. http://www.inrix.com/devzone.asp. Accessed July 23, 2013.

53. INRIX, I. *INRIX*, 2013. http://www.inrix.com/differentiate.asp. Accessed July 19, 2013.

54. Kim, K., S. I.-J. Chien, and L. Spasovic. Evauation of Technologies for Freeway Travel Time Estimation: A Case Study of I-287 in New Jersey. in *The 90th Annual Meeting Transportation Research Board*, Washington D.C., 2010.