How Reliable is the Traffic Information Gathered from Web Map Services?

Alan M. de Lima\textsuperscript{1} and Jorge Campos\textsuperscript{1,2}

\textsuperscript{1}Salvador University – UNIFACS
\textsuperscript{2}Bahia State University – UNEB
Salvador, Bahia, Brazil

alan_empresa@hotmail.com.br, jorge@unifacs.br

Abstract. There are many applications out there monitoring the traffic flow to gain insights and discover relationships about vehicles dynamics. The implementation of traffic monitoring systems, however, requires a considerable amount of investment in equipment and qualified personnel to simply get data. Investments are even greater when it is necessary to process and analyze the data and produce information that supports operational measures, public policies, and investments in infrastructure. Aiming at facilitating and reducing the cost of collecting information about traffic flows, this paper presents a methodology to monitor any area of the road network based on information gathered from free internet services. An experiment was conducted to evaluate the quality of the data obtained through Web mapping services when compared with data obtained from active radars installed along the route. Experimental results show that there are small discrepancies and low latency between these data and that the information about the traffic gathered from web mapping services can be considered as a data source for any application that does not demand high levels of accuracy.

1. Introduction

Intelligent Transportation Systems (ITS) are the collection of advanced application in which information and communication technologies are applied in the field of road transportation networks, including infrastructure, vehicles, and users. ITS can be categorized in five great areas (Aquino et al. 2001), that is, Advanced Public Transport Systems (APTS), which employ information technologies to enhance the safety, efficiency, and effectiveness of public transportation systems; Advanced Traffic Management Systems (ATMS), that attempt to minimize traffic congestion, such as intelligent traffic lights, traffic safety and congestion management; Advanced Traveler Information Systems (ATIS), which make use of navigation and information systems to ensure driver safety and minimize congestion; Commercial Vehicle Operation (CVO), systems that encompass management and operation of commercial vehicles, using the technology to improve the management of cargo transportation services; and Advanced Vehicle Control Systems (AVCS), that aim to improve road system safety, allowing vehicles to interact with drivers, assisting them in various aspects (smart vehicles). Regardless of the area, almost all ITS require information about the traffic flow, more specifically, when and where traffic congestions occurred, is happening, or will occur.
Traffic congestion is a phenomenon associated with urban mobility of growing concern among users and managers of the transportation system. Traffic congestion became one of the most important problems of people living in large cities, because its significantly impacts the economy, the environment, and the health of the citizens. Any initiative aiming at understanding and predicting traffic dynamics requires the analysis and monitoring of the traffic flow. The implementation of traffic monitoring systems, however, requires a considerable amount of investment in equipment and qualified personnel to simply get data. Investments are even greater when it is necessary to process and analyze the data and produce information that supports operational measures, public policies, and investments in infrastructure.

An alternative to gather information about the traffic is the traffic layer of some Web mapping services, such as Google Maps and Bing Maps. These services offer the visualization on a map of the average speed of segments of the transport network. It is also possible to consume this kind of information using a Web application making requests via the Application Program Interface (API) of these services.

The traffic information from Web mapping services has become the preferential source of information of the general media. Radio and TV stations, for instance, use this kind of information to produce traffic bulletins of the main streets of the city. Despite the easiness of gathering this kind information, the lack of metadata about the accuracy and latency, and how this information is computed from the various sources of information can be seen as major obstacles of using this data source by ITS application or other more technical applications. Although not well documented, these services claim that the mean speed of the vehicles is estimated based on sensors installed along the main roads, agreements with local traffic agencies, association with transportation companies and taxi fleets, and, especially and more important, from the voluntary contribution of the multitude of users of the mobile version of these Web mapping services.

This paper presents a methodology to get information about the traffic flow of any area of a road network. The methodology relies on information gathered from free internet services, thus it can be used to obtain information from virtually anywhere at any time. Aiming at gaining insights about this kind of data, we conducted an experiment to compare the data obtained from the traffic layer of Web mapping services with data from active radars installed along the route. Experimental results show that there are small discrepancies and low latency between these data and that the information about the traffic gathered from Web mapping services can be considered as a data source for any application that does not demand high levels of accuracy.

The remainder of this paper is structured as follows: section 2 discusses main technologies for traffic monitoring. Section 3 presents a methodology to get information about the traffic flow from free Internet services and discusses this methodology in the context of an application developed to analyze the impact of events that occur along the road network. Section 4 discusses an evaluation of the accuracy of the data collect from Web mapping services. Section 5 presents conclusions and indicates future work.
2. Technologies for Traffic Monitoring

There is a myriad of technologies used for traffic monitoring and analysis. Among the most common technologies, it is worth mentioning video cameras, active and passive radars, electronic and mechanical sensors, wireless sensor networks, and, more recently, the employment of vehicle and people acting as a sensor.

Samczynski et al. (2011) propose the use of GSM-based passive radars as sensor for speed measurement and traffic monitoring. The proposed solution uses Global System for Mobile Communications (GSM) transmitter as the illuminator of opportunity in a bistatic geometric configuration. Preliminary results show that the system can be used to monitor the average speed of vehicles and road capacity. The results show also that with this technology it is possible to distinguish different-sized objects and to categorize the traffic of vehicles.

The use of video cameras as sensor for determining traffic parameters such as vehicle speed and number of vehicles is also promising. Kiratiratanapruk et al. (2006) present a video traffic monitoring application based on object detection and tracking method. In the detection step, the application uses a gradient-based background subtraction for foreground-background segmentation, which is more robust to lighting changes in outdoor environments and requires significantly less computing resource. Few experimental results show, on one side, good accuracy, and robustness in shadow environments, but, on the other side, suggest difficulty in detecting dark colored vehicles and distinguishing objects near each other.

Tobing (2014) develops an image processing based application for the analysis and monitoring of traffic on highways or toll roads. The application uses different approaches to detect vehicle under diverse lighting conditions. The application uses a method based on the background subtraction method during daytime and proposes a new algorithm using car lights for night conditions. Experimental results show that daytime vehicle detection algorithm achieved a maximum accuracy rate of 86.7%, while the method of vehicle detection at night performs better and achieves an accuracy rate of 96.3%.

Vehicles and drivers acting as sensors of the environment seem to be the dominant technology nowadays. People are everywhere, not only along main roads and intersections. Besides, people can always give their impression and interpretation about a phenomenon, introducing the human dimension in the data collected (Longley et al. 2013). The work of Pham et al. (2015) used data from motorcycles equipped with GPS devices. The authors claim that motorcycles are pervasive in developing countries and are by far the transportation means with the greatest capillarity. The authors developed a mobile application with the main goal of building traffic data with the help of the crowd of motorcycle riders. All data are processed in a control center and converted into useful metrics such as average speed, traffic flow, and the average travel time.

Taxis are another important source of information for traffic monitoring. Li et al. (2009) conducted an experiment in Shanghai, China, to evaluate the flow of the traffic using sensors installed in about 4000 taxis. The data obtained from these vehicles were compared with data collected through video cameras. The results showed that the estimates of traffic status based on these sources were reasonable close and can be used...
Interchangeably. Moreover, the data from taxi fleet have a wide coverage and do not require the installation and maintenance of expensive video camera systems.

Experiments that compare data collected from vehicles in a collaborative fashion and data obtained from usual sensors (e.g. video cameras and radars) are always important to build confidence and boost the use of the former kind of data. The work of Herrera et al. (2010) assesses the average speed of data traffic flow obtained from GPS-enabled smartphones and Loop Detector sensors. Loop Detector is a sensor with the primary function of vehicle passage, presence, count, and occupancy, but it can be used to estimate vehicle speed as well (Hazelton, 2004). In this experiment, the speed of the vehicle carrying the smartphone is recorded only at special places called Virtual Trip Line (VTL). VTL is geographical markers stored in the handset database that probabilistically trigger position and speed updates when the handset crosses them. VTL can be placed anywhere, but they were placed close to every Loop Detector for comparison purpose. The evaluation of the VTL-Loop Detector experiment concludes that it is necessary only 2% or 3% of the total registered cars on the road equipped with GPS to produce an accurate measurement of the traffic flow speed.

The technology of connected and GPS-enabled smartphones can also be used for traffic control. RoadRunner (Gao et al., 2014) is a mobile phone application that uses the 4G telephone network to communicate with a main server and electronically book permission to use a given segment of a transportation network. The system distributes a kind of electronic ticket for the vehicles granting permission to use some segments of the network. The widespread use of RoadRunner ensures that there are not many vehicles (above the road capacity) in a particular section of the road at a given time. Results of a simulation carried out show that RoadRunner has the ability to manage large-scale roads, to improve travel speed by controlling number of vehicles on the road, and to work as an efficient electronic toll system.

Despite the importance of having vehicle and people gathering data for traffic control and analysis, there is only a few number of applications in the ITS domain that explore such kind of information. Tostes et al. (2013), for instance, propose an algorithm to infer the average speed of the traffic flow based on the color scale used by the traffic layer of Web mapping services and uses this information to predict traffic behavior. This strategy, however, get only a rough estimation of the mean speed of the traffic flow. Next section presents an application that use free and open resources to compute the impact of non-recurrent events in the traffic flow. The mean speed of roads segments is obtained from requests made directly to a Web mapping service, which is more precise and eliminates the use of expensive image processing algorithms.

3. Collecting Traffic Information from Web Mapping Services

The motivation to use information from the traffic layer of a Web mapping service came from our needs to measure the impact of non-recurrent events that cause traffic congestions. Considering their cause, traffic congestions can be classified as recurring and non-recurring. Usual and predictable factors such as daily peak times or the occurrence of regular sporting events cause recurring traffic congestion. Unforeseen or atypical factors cause non-recurring traffic congestion, such as traffic accidents and maintenance works along the roads. Whatever the cause of the traffic congestion, the
knowledge of the impact that these events may cause in the vehicle flows it is an indispensable information in the analysis of measures and actions to mitigate their effects.

One strategy to assess the impact of any event that cause traffic congestions is to compare traffic parameters (such as mean speed) right after the occurrence of the event with usual traffic parameters of the place, but without the occurrence of the event. The usual traffic parameters at the event site, for instance, can be historical traffic data from the same region or, in the absence of this information, future data collected during a certain period after the event. This strategy, however, presents some practical problems, especially when dealing with non-recurring events. As the site of a non-recurring event is unpredictable and as the traffic monitoring infrastructure is too expensive to be placed along the entire road network, it becomes virtually impossible to assess the impact of an event that can occur at any place, any time, both comparing with historical or future data.

We have developed a Web application to measure the impact in the traffic flow caused by non-recurrent event. The application itself is not the focus of this paper. We are interested here in discussing the methodology to obtain information about traffic of vehicles from free internet services. For this Web application, we have established the following functional and non-functional requirements: a) an accessible, up to date, comprehensive, and open and free source road network dataset; b) a mechanism for defining the location of non-recurring events as soon as it happens; c) a methodology to define the monitoring points that are likely to be impacted by the occurrence of the event; d) a source of information that provides real-time traffic information at monitoring sites; e) an intuitive graphical interface to present the results of the analysis.

The data of the road network chosen for this project come from the project Open Street Maps (OSM). The OSM project aims to build the existing road network of the world. The information about the road network are collected and submitted by a community of volunteers to a spatial database repository, validated by a group of experts, and made available for the community of users in a free and open source fashion.

The place where the event occurs can be set directly by a user, captured from a social network, or defined by a traffic monitoring system. No matter the source of information, it is important to know the place where the event occurs as soon as possible. In our application, the user manually sets the place of the event.

The algorithm developed to identify monitoring points uses the OSM project database to identify network segments that can potentially be affected by the occurrence of the event. The algorithm requires as input a) the location of the event, b) the radius of influence, c) the penetration level on the road network, d) the duration of the monitoring process, and e) the number of days to repeat the monitoring process.

The location of the event and the radius of influence define a circle or a buffer that includes all routes initially considered for monitoring purposes. Applying traditional buffer algorithms in the road network, however, will return segments of the network that will not be affected by the event. Some segments of the network will be selected only
because they are inside the buffer, but considering the path to be traversed they are far away from the place of the event.

We have modified the buffer algorithm to look only for significant roads segments considering the level of penetration in the network (Figure 1). The road where the event occurred has a penetration level of 0 (black line). The streets that intersect the street level 0 have a penetration level 1 (blue line), and so on. This strategy can continue up to a certain level of penetration, after that all roads will have an infinite level of penetration (gray lines). Thus, setting the penetration level to 1 will tell the algorithm to consider only monitoring points along the stretch of the road where the event occurred and along roads that intercept the former.

![Penetration Levels Diagram](image)

**Figure 1.** Modified buffer algorithm considering the level of penetration on a road network to select roads near the event.

The temporal parameters of the monitoring processes are the duration and the number of repetitions. The duration is defined with an absolute value (e.g., one hour). It is also possible to define the duration using relative values, such as during the period in which the average speed of traffic flow is less than a certain threshold. Other relevant information is the definition of the number of days to repeat the monitoring process. Monitoring the same places after the event serves the purpose of establishing the usual traffic parameters for these sites. By default, we assume a repetition period of an entire week. This time window is enough to capture traffic parameters during weekdays and weekends and at the same day of the event one week later.

Once the spatial and temporal monitoring parameters are defined, the process of collecting traffic information begins with the creation and positioning of crawlers around monitoring points. Crawlers are Web robots that periodically make requests to the traffic layer of map services. The frequency that each request will be performed is
also configured in the application. We believe that a request every one or two minutes is sufficient to capture small variations in traffic flow. All information collected by the crawlers is stored in a spatial database for posterior analysis of the impact of the event.

When the monitoring process is over, the user can visualize the observed traffic parameter of the selected site (Figure 2). We monitored, as proof of concept, the impact of a public demonstration held at the vicinity of Bahia Shopping on June 21, 2015 at Salvador, Bahia, Brazil. We have monitored the roads around the event for one hour starting at 16:35 pm. We repeated the monitoring process at the same time and place for every day of the week after that. The penetration level of this monitoring was set to 0, that is, only the road where the event occurs was monitored.

Our application uses graphical and map presentations to show the most significant traffic parameter (Figure 2). On the upper side of the presentation screen there are maps indicating observed stretches of the road network and a graphical representation of the traffic congestion. The thickness of the red lines indicates the level of the traffic congestion. On the upper right side, a gauge indicates the traffic congestion level. The metric used to represent the traffic congestion is the average vehicle delay.

On the lower half of the presentation screen (Figure 2), it is possible to visualize, through charts, the impact of the event in the traffic flow when compared with data of the same place without the occurrence of the event. On the left side, a line chart shows the average speed of the road during the day of the event (red line), the average speed of other days (orange line), and the average operational speed of the observed stretch (blue line). The operational speed of the segment is retrieved from the Web mapping service. On the lower right side of the screen, the same information is shown in a bar chart. In the middle of the screen, a bar chart presents the transit time, that is, the time needed for a vehicle to travel the observed stretch.

![Figure 2. Screenshot of an application showing the result of the impact of a public demonstration.](image-url)
Initially thought as a tool to measure the impact of non-recurrent event, the application can also be used to monitor the impact any event along the road network. This application can be used, for instance, to evaluate the impact of a desired modification in the network before it becomes a permanent change. The installation of a new traffic light, the closing a u-turn, or changing the direction of the flow of some roads segments are examples that can be tested and evaluated before becoming or not definitive.

All traffic information presented by the application comes from the Bing Maps traffic layer. There is no official documentation reporting how this information is computed, what are the sources of information used, and how accurate these pieces of information are. According to ITS metrics, the average operational speed of a road, for instance, is defined as the maximum speed at which 85% of the vehicles travel the observed stretch. We have no clue if Bing uses such definition.

4. Evaluating the accuracy of traffic information of Web mapping services

One alternative to assess the quality of data provided by the traffic services available on the Web is to compare the information from these services with the data captured by other sensors considered reliable by the ITS community. This comparison should be made using a large volume of data to have statistical significance and to give certain degree of confidence about quality and accuracy of information.

Aiming at getting insights about the data obtained from Web mapping services, we conduct an experiment to evaluate the level of accuracy of these data. The experiment compares the data obtained from Microsoft Bing Maps with the data collected from three intelligent traffic lights. Intelligent Traffic Light is a device that measures, among other features, the number and the speed of every vehicle passing by. The data of the traffic lights were provided by the city of Salvador transit authority (Transalvador). Transalvador provided 14 days of data of the traffic lights number 86, 106 and 107 between November 30 and December 12, 2015. For the sake on confidentiality, Transalvador provides only the time stamp and the speed of the vehicles on the road. We use this information to compute the average speed in a one-minute interval.

It is worth mentioning how average speed at the traffic lights sites were acquired from Bing Maps. Bing Maps does not offer the average speed of a point on the road network. Instead, the API gives the transit time of a given segment of the road. In this way, we need to specify a segment of the road with the position of every intelligent traffic light in the middle of the segment (Figure 3). After running exhaustive tests with the Bing Maps API, we find out that the smallest segment of the road that always return a valid transit time is 30 meters. When segments are smaller than that value, the API sometimes returns null transit times. The documentation of the API does not mention this characteristic.

Our experiment captures data from Microsoft Bing Maps at a rate of one reading at every minute, during the same period provided by Transalvador. We have used the Web application introduced earlier in this paper to collect data from three events (one at each traffic light). For this experiment, we have set the penetration level to 0, and
configure the temporal parameters to retrieve data along a 24 hours period, and repeat that for 14 days.

![Traffic lights # 86](image)

**Figure 3.** Segment the road with the traffic light number 86 in the middle.

We combine the speeds obtained from Bing Maps and from intelligent traffic lights in a single table with one record for every minute. Our first strategy to gain some insights was to conduct a visual analysis of the evolution of the average speed recovered from each source during the entire period of observation. Figure 4 depicts a 24-hour period of these data. To improve the presentation of the graphic, we have grouped the average speed in bins of 5 minutes intervals. A visual inspection of the graphic shows a salient difference on the pattern of Bing maps versus traffic lights data during the period between 8pm to 6am (nighttime period) and from 6am to 8pm (daytime period). Thus, we have decided to analyze these periods separately.

![Graphical presentation of the average speed](image)

**Figure 4.** Graphical presentation of the average speed collected by Intelligent Traffic Light number 107 and retrieved from Bing Maps.
During nighttime periods, Bing Maps’ data shows little or no variation. This fact is represented by the prevalence of flat red line segments during this period. Moreover, the average speeds gathered from Bing Maps are close to the maximum operating speed allowed for the route. One possible explanation for this fact has to do with the number of vehicles transiting on the route during this period and with the role of intelligent traffic lights on the monitoring and traffic control system. On one hand, there is a significant reduction in the number of vehicles on the track during nighttime. Thus, the source of data to compute traffic parameters used by Web mapping services is greatly reduced. Perhaps, the lack of information coming from users of the Web service, forces the service provider to inform only the operational speed of the track. On the other hand, the location of Intelligent Traffic Lights is widely known by most users of the system. As these devices are used for monitoring traffic violations, such as speeding, it is reasonable to assume that drivers, even without traffic congestion, will respect the speed limit, especially in places close to Intelligent Traffic Lights.

During daytime periods, the data from Bing Maps shows some fluctuations. A visual inspection of the graphic shows that the average speed obtained from the traffic lights have the same fluctuation (i.e., both lines have almost the same shape), but with an almost constant gap between them (Figure 4). Considering only the shape of the curves, it can be verified that Bing Maps traffic service has a low latency, that is, the time needed to the Web service to reflect a change in the average speed as it is collected from a sensor placed on the road is relatively short. At this point, we verified that the latency is not greater than a five minutes’ interval for periods with a reasonable number of vehicles on the road. Considering the gap between the two samples, it can be verified that average speeds retrieved from Bing Maps service do not differ from traffic lights’ data more than 10 km/h most of the time. This empirical and expedite evaluation can be tested with some formal method to evaluate the statistical significance of this statement.

Aiming at achieving a more reliable knowledge of the difference between the average speeds obtained from Bing Maps and Intelligent Traffic lights we used the statistic method T-test. The T-test assesses whether the means of two groups are statistically different from each other. This analysis is appropriate whenever you want to compare the means of two groups. Our null-hypothesis was that the difference between the two speeds is less than a given value. Our goal is to identify the minimum value of the difference between the two speeds that satisfies the null-hypothesis at a 95% confidence level.

Our sample contains 60,480 records, considering the three sites together. We have decided to run the statistic T-test for the daytime period and another for the entire sample. Based on our initial analysis, we believe that the statistic evaluation of the nighttime periods alone will lead to fallacious assertions.

Table I shows the result of our T-test for each Intelligent Traffic Lights. In general, we can infer that most of the time the difference between the average speeds retrieved from Bing Maps does not differ from average speeds measured by traffic lights sensors by 10 km/h. Considering only day times periods, this difference drops to 9 km/h.
The amount of data used in our experiment was not enough to make strong assertions or to extrapolate to the entire road network. This small sample, however, put some light in the data and suggests that the use of information coming from Web mapping services can be used in lieu of expensive technologies to measure the speed of the traffic flow.

5. Conclusion and Future Work

This paper presents a methodology to monitor any area of the road network based on information gathered from free internet services. The methodology was discussed in the context of a Web application developed to evaluate the impact of non-recurrent events on some segments of the road network. The same methodology, however, can be used by any ITS application that needs information of the traffic flow dynamics in place that there are no available sensors to measure such information. Few municipalities and government agencies have the resources to deploy the necessary infrastructure for collecting traffic information in a broad and comprehensive fashion. Thus, the use of free and open source technologies becomes a viable and interesting approach to develop ITS solutions.

Regardless of the ease of gathering information about the traffic dynamics from most Web mapping services, there is little documentation about the methodology used to compute this kind of information. These services claim that the mean speed of the vehicles is estimated based on sensors installed along the main roads, agreements with local traffic agencies, association with transportation companies and taxi fleets, and, from the voluntary contribution of their users. The lack of knowledge about the accuracy and latency, and how this information is computed from the various sources of information, however, can be seen as a major obstacle of using this data source by ITS application or other more technical applications.

Aiming at gaining inside and knowledge about the traffic layer of Web Mapping services, we conduct an experiment to evaluate to compare this information with information obtained from active radars installed along the route. Experimental results show that there are small discrepancies and low latency between these data and that the information about the traffic gathered from Web mapping services can be considered as a data source for any application that does not demand high levels of accuracy.

Table 1. Difference between Speeds Retrieved from Bing Maps and Intelligent Traffic Lights

<table>
<thead>
<tr>
<th>Traffic Light ID (#)</th>
<th>Difference between Bing Maps and Traffic Light Speed (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daytime Only</td>
</tr>
<tr>
<td>86</td>
<td>7</td>
</tr>
<tr>
<td>106</td>
<td>9</td>
</tr>
<tr>
<td>107</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 1. Difference between Speeds Retrieved from Bing Maps and Intelligent Traffic Lights

Proceedings XVIII GEOINFO, December 04th to 06nd, 2017, Salvador, BA, Brazil. p 234-245.
We are planning to conduct a more broad and comprehensive experiment, including more sensors and other Web mapping services (e.g., Google Maps). The city of Salvador has, at the time of this writing, more than 180 intelligent traffic lights scattered throughout the city. The main hindrance at this point is that it is necessary a commercial key to make request about the traffic layer using Google Maps API and Microsoft Bing Maps API does not allow too many requisitions using a free account. Thus, comparing data from lots of sensors and from major Web mapping services for a prolonged period will improve all knowledge and confidence about this kind of information.

References


Pham, D.T., Hoang, B.A.M., Thanh, S.N., Nguyen, H. and Duong, V., 2015, September. A Constructive Intelligent Transportation System for Urban Traffic Network in Developing Countries via GPS Data from Multiple Transportation Modes. In Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference (pp. 1729-1734). IEEE.

