Ninth LACCEI Latin American and Caribbean Conference (LACCEI'2011), Engineering for a Smart Planet, Innovation, Information Technology and Computational Tools for Sustainable Development, August 3-5, 2011, Medellín, Colombia.

GPS Capable Mobile Phones to Gather Traffic Data

Nelson R. Gómez-Torres
University of Puerto Rico at Mayagüez

Didier M. Valdés-Díaz
University of Puerto Rico at Mayagüez

ABSTRACT

Mobile phones with integrated GPS are beginning to be utilized to asses traffic conditions in real-time, allowing for improvements in planning and operations processes that will provide more information to users and minimize delays. Using the transportation network effectively is one of the challenges to improve the sustainability of the transportation system. The objective of this paper is to show examples of utilizing GPS capable mobile phones to gather traffic data, show the variables that aid in the identification of travel mode and the limits of those variables. Speed and routes appear to be the most important parameters to identify the transportation mode used. In terms of data acquisition rate, it has been found that gathering data at intervals of 15 seconds seems appropriate to calculate speeds without losing details. When the data is gathered at small time intervals, minor errors in positioning may cause large errors in speed. The data gathered using this type of sensors could be applied to determine the mode usage. The transportation modes that are easier to ID are walking and heavy rail. Cars, trucks, buses, light rail, motorbikes and bikes are harder to identify because they utilize similar routes and differences in speed are small.

Keywords: Mobile phones, traffic studies, GPS, probe filtering, modal split

1. INTRODUCTION

Mobile phones with integrated GPS (Global Positioning System) are beginning to be utilized to asses traffic conditions in real-time. The assessment of traffic conditions in real-time allow improvements in the urban transportation planning and traffic operations processes. It will also provide the support for the development of new routing strategies in real time that can be implemented using Intelligent Transportation Systems (ITS). The development of ITS has expanded the opportunities to provide more traffic information to users and minimize travel time in the transportation system.

There are still various challenges for mobile phones to become a tool to acquire traffic data. Some progress towards the use of mobile phones as a data-gathering tool has been made. Hellinga et al. (2008) described several steps to determine traffic condition from positioning data. Within those steps probe filtering and travel time allocation continue to be a challenge to obtain without user-input. Map matching and path identification steps consist of determining the position of a vehicle or traveler in the transportation network and the possible path utilized to change positions.

These challenges range from the technology itself (its limitations) to the interpretation of the data. This paper presents specific challenges found by the authors and the possible solutions to overcome them in order to be able to begin collecting data from mobile phones and using the data to determine traffic conditions.

Previous research complemented positioning data (from the phones) with input from users (Mobile Millennium, 2009; Winters, et al., 2008). The complexities of avoiding user-input start with the irregular way in which a single person moves. Nevertheless, this technology promises to change the way engineers gather data for the analysis of transportation systems. Using mobile phones as sensors (probes) for several traffic parameters was simulated and obtaining favorable results (Fontaine and Smith, 2005; Ygnace et al., 2000).

Data from mobile phones is not commonly used to optimize the road network operation, neither for planning purposes. Most Departments of Transportation (DOT) in the U.S. rely on historical data in order to design and maintain their traffic control plans. Real-time data is needed to improve the responsiveness of traffic control.

Additionally, mobile phones appear to be a great asset to collect traffic data in real-time to build Dynamic Origins and Destinations (DOD) matrices. DOD matrices are utilized for Dynamic Traffic Assignment (DTA). DTA is used to maximize the network's capacity utilization. Peeta and Ziliaskopoulos (2001) recognized that DTA refers to a broad spectrum of problems, but with some common features such as:

1. The use of time-dependent flows, instead of fixed flows as in static traffic assignment.
2. No universal solution for general networks.
3. The need to effectively represent traffic realism and human behavior.

Therefore, having a reliable, cost-effective method to estimate DOD matrices will help to introduce DTA. DTA, not only can be used to prevent premature drops in roadway Level Of Service (LOS), but can be utilized to improve the LOS of the network. An improved traffic assignment helps the users by reducing congestion, lowering emissions and reducing delays. Improvements on the information available for the users are expected.

1.1 OBJECTIVE

The objective of this study is to show examples of utilizing GPS capable mobile phones to gather traffic data. Specifically how to gather data utilizing mobiles phones (without user-input) to detect the transportation mode utilized by each user in the transportation system. This paper will show the variables that aid in the identification of travel mode and the capacity of those variables to determine modal split.

1.2 BACKGROUND

Hellinga et al. (2008) sets out that determining traffic conditions from positioning data requires five steps: map matching, path identification, probe filtering, travel time allocation and travel time aggregation. The relation of these five steps with data from cell phones can be seen in Figure 1(next page).

Map matching and path identification steps consist of determining the position of a vehicle or traveler in the transportation network and the possible path utilized to change positions. GPS data is a collection of position and times that may be gathered at a rate at which several paths might be used between data points. Both of these steps can be done with available technology/methods. The challenges lie in probe filtering and travel time allocation.

Probe filtering consists of determining the transportation mode being used. The data have to be analyzed to establish the transportation mode. Speeds and routes are the principal variables to observe.

When several links are used between data points, we need to estimate the travel time on each link, travel time allocation refers to this estimation. Methods to estimate the travel time of each link between data points are also included in this step. In addition, when a change in travel mode occurs, we need to separate the travel time in each mode. This is also related to the time interval between points.

Previous research utilized user-input to perform probe filtering and travel time allocation. In probe filtering the specific topics are: identify travel mode and travel mode change. In travel time allocation the specific topics are: starting and ending trip times, and time of travel mode change. One of the challenges is to perform both of these steps without relying in user-input.

There are two ways of determining positions with mobile phones: point locations and tracking. Point locations may be recorded when a phone moves to another cell and the cell transference is performed. This situation may become a problem because the transferences do not necessarily occur in the same (exact) location each time (Sohn and Hwang, 2008; Sohn and Kim, 2008). Phone tracking is more complicated and requires triangulation or a Global Positioning System (GPS) capable phone. GPS capable mobile phones are more common recently due to the Federal Communications Commission (FCC) wireless Enhanced 9-1-1 (E9-1-1) rules.

Map matching[1]

- This step consists of determining the position of a vehicle *in* the transportation network.

Path identification[1]

- When several paths can be utilized, this step deals with road selection.
- This problem is due to the time interval between data points.

Probe filtering

- Consists of determining the transportation mode being used.
- The data have to be analyzed to establish the transportation mode.

Travel time allocation

- Estimation of travel time in each link.
- When a change in travel mode occurs, this step separates the travel time in each mode.

Travel time aggregation

- The calculation of the average travel time for any link occurs in this step.

Figure 1: Determining traffic conditions from positioning data[1]

1.3 PROBLEM

Taking on the question posed in the article “Data, data, data – Where’s the data?” by Tate-Glass et al. (2000) is the main reason for this study. Research in traffic control is very active, but the methods to obtain data to apply the results of all that research in real world situations are still under development. Gathering the data necessary to improve traffic control, both from the planning perspective and real-time traffic control changes due to unexpected situations, requires the collection of historical data and data in real-time. Figure 2 shows a schematic associating the mobile phones data to traffic control. Historical data alone cannot be used in an optimized traffic control scheme because any discrepancy between the data and the actual road conditions will affect the optimization. Real-Time data alone cannot be utilized because appropriate traffic control requires understanding the traffic pattern in the area.

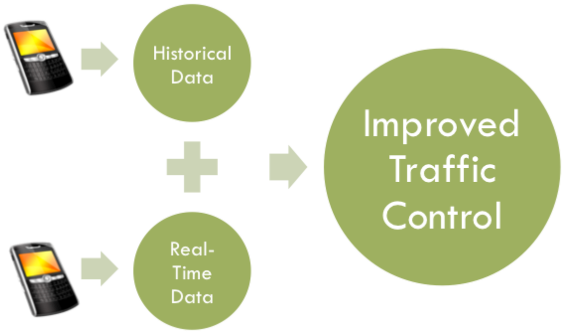

Figure 2: Data combination to improve traffic control

---

[1] Map matching and path identification maps, found in maps.google.com (Oct, 2009)

2. THE DATA AND ITS SOURCES

The data was gathered with three (3) different mobile phones a Nokia E71, a Nokia 5800 and a Pharos Traveler 619 (see Figure 3). All of these mobile phones have integrated GPS and were connected to the T-Mobile network of Puerto Rico.

Figure 3: Mobile phones (from left to right Nokia E71, Nokia 5800 and Pharos Traveler 619)

The data gathered with them consisted in locations (latitude and longitude) and times. Each data point consists of a 3-tuple (time, latitude, longitude). Each data point was gathered as the fastest rate allowed by those mobile phones, which turn out to be between 1 and 2 seconds. The data was gathered in GPX files and then converted to XLS files (MS Excel®) for visual review and MAT files for processing in Matlab ®.

Time was provided in the following format "year"-"month"-"day"T"hour":"minute":"seconds"."seconds decimals"Z, for example 2010-05-25T17:43:54.973Z, where the Z at the end means that the time is given for the zero meridian, in other words the time is in Coordinated Universal Time (UTC).

Latitude and longitude was provided in real numbers up to the ninth decimal position where positives represent North or East respectively, for example latitude=18.380916446, notice that it is positive, so refers to latitude 18N.

The data used for this paper was gathered in Puerto Rico's East zone. A GPS application for mobile phones was utilized to gather the data and was saved in the phone itself.

3. EARLY TESTS

The data gathered with one of the mobile phones in May 26, 2010 is shown in Figure 4 and Figure 5. In Figure 4 we can see that while traveling by car from east to west on the left lane, the path (yellow line) fits perfectly when compared with Google Earth images. Nevertheless, in Figure 5 we can see that while below a concrete roof (with no walls); the path (purple line) is erratic. The erratic nature of the data seen in Figure 5 comes from the fact that GPS locations near the building are given but the precision has been compromised. Still Figure 5 shows that the arrival and departure path is correct. For existing GPS technology embedded in current mobile phone devices, data cannot be gathered with precision inside closed concrete buildings.

In the next section we will see several calculations of speeds with data gathered in conditions similar to Figure 4.

Figure 4: GPS data with clear view of sky

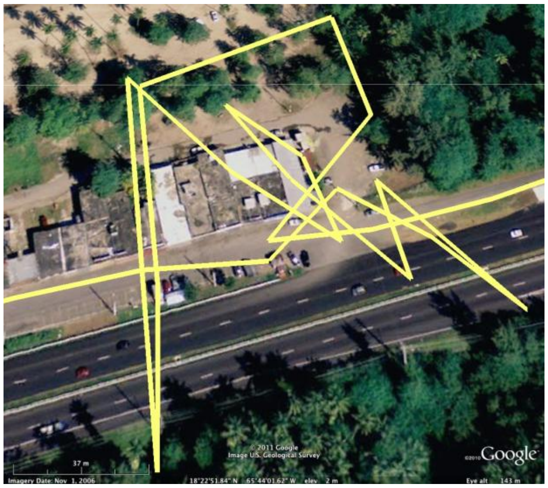

Figure 5: GPS data with obstructed view of sky

4. CALCULATING SPEEDS

Speed is one of the most important parameters to identify the transportation mode used. The speeds here are calculated with 2 points where each point contains the information of latitude, longitude and time. To calculate speeds the distance between points are calculated using the geodetic distance of the earth ellipsoid. The elevations are ignored in the calculations of speed.

Ignoring the elevations becomes a systematic error, but its magnitude is very low. Considering a road segment with a slope of 10% (which is a rather high percentage), a vehicle moving at 100.5 km/h (as measured in an instant) will appear to be moving at 100 km/h if the height is ignored. Also with the current GPS technology available in mobile phones the errors in elevations are high; as a result of both factors it is better to ignore the elevations.

Also, we wanted to determine the rate the data should be gathered without losing important information. To that purpose Figure 6 and Figure 7 were created.

To calculate the speeds shown in Figure 6 and Figure 7 the following nomenclature is used. The first data point will be called DP1 and the second DP2, therefore the n data point will be called DP(n). This test was performed May 24, 2010. In Figure 6 the curve marked as 1 denotes speeds calculated with each data point where the first speed marked is calculated with the 2-tuple (DP1, DP2) and will be denoted as S1. The second speed marked (S2) is calculated with (DP2, DP3), therefore in general the following relation applies:

$$\left(DP(n), DP(n+1)\right) \rightarrow S(n); \forall n \qquad (1)$$

In Figure 6 the curve marked as 10 denotes speeds calculated with each data point where the first speed marked is calculated with the 2-tuple (DP1, DP11) and will be denoted as S1. The second speed marked (S2) is calculated with (DP2, DP12), therefore in general the following relation applies:

$$\left(DP(n), DP(n+10)\right) \rightarrow S(n); \forall n \qquad (2)$$

Therefore, in Figure 6, the relations for the curves marked as 30 and 60 appear in equations (3) and (4).

$$\left(DP(n), DP(n+30)\right) \rightarrow S(n); \forall n \qquad (3)$$

$$\left(DP(n), DP(n+60)\right) \rightarrow S(n); \forall n \qquad (4)$$

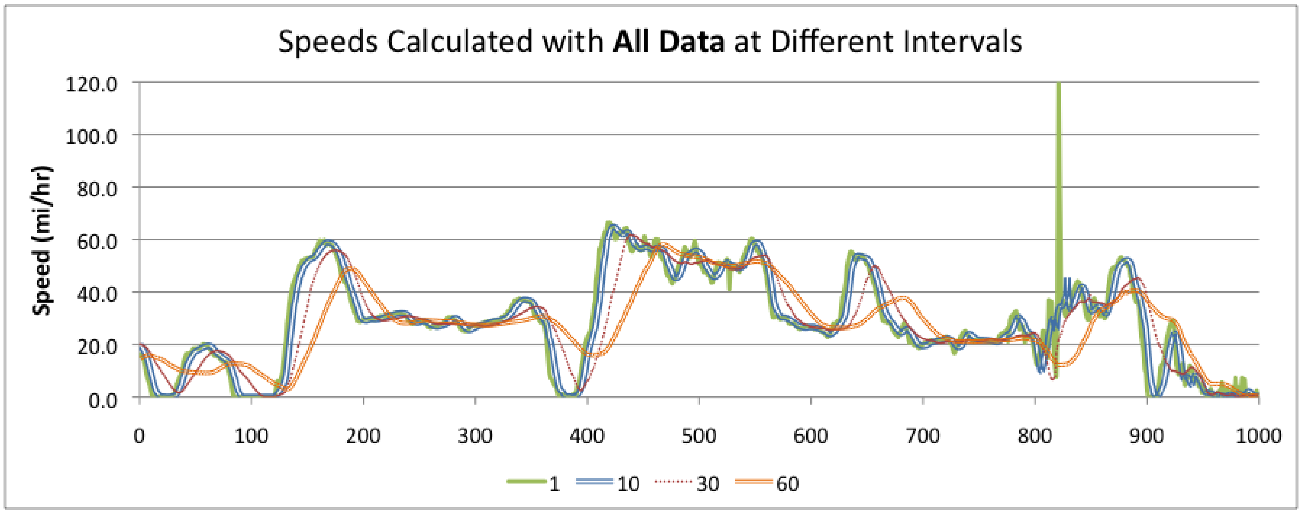

Figure 6: Speeds calculated using all the data

For Figure 7 the relations are similar, but not all data points are used. The curve marked as 1 is exactly the same as the curve marked 1 in Figure 6. The first speed (S1) of the curve marked as 10 is calculated with the 2-tuple (DP1, DP11), the second speed (S2) with the 2-tuple (DP11, DP21); therefore the following relation applies:

$$\left(DP(n), DP(n+10)\right) \rightarrow S(n); n = \{1,11,21,31\ldots\} \qquad (5)$$

Therefore, in Figure 7, the relations for the curves marked as 30 and 60 appear in equations (6) and (7).

$$\left(DP(n), DP(n+30)\right) \rightarrow S(n); n = \{1,31,61,91\ldots\} \qquad (6)$$

$$\left(DP(n), DP(n+60)\right) \rightarrow S(n); n = \{1,61,121,181\ldots\} \qquad (7)$$

Comparing Figure 6 and Figure 7, we can see that at different intervals not all the data is needed. The shapes of both speed charts show that we will be able to select greater intervals without losing critical information. In the Figure 7 the curve marked as 10 (skipping 10 data points) seems to be adequate; therefore, gathering data at an interval of between 10 and 20 seconds seems to be appropriate (remember that each data point is gathered every 1 to 2 seconds) Even gathering data at an interval of 30 data points (every 30 to 60 seconds) is possible without losing most of the speed details.

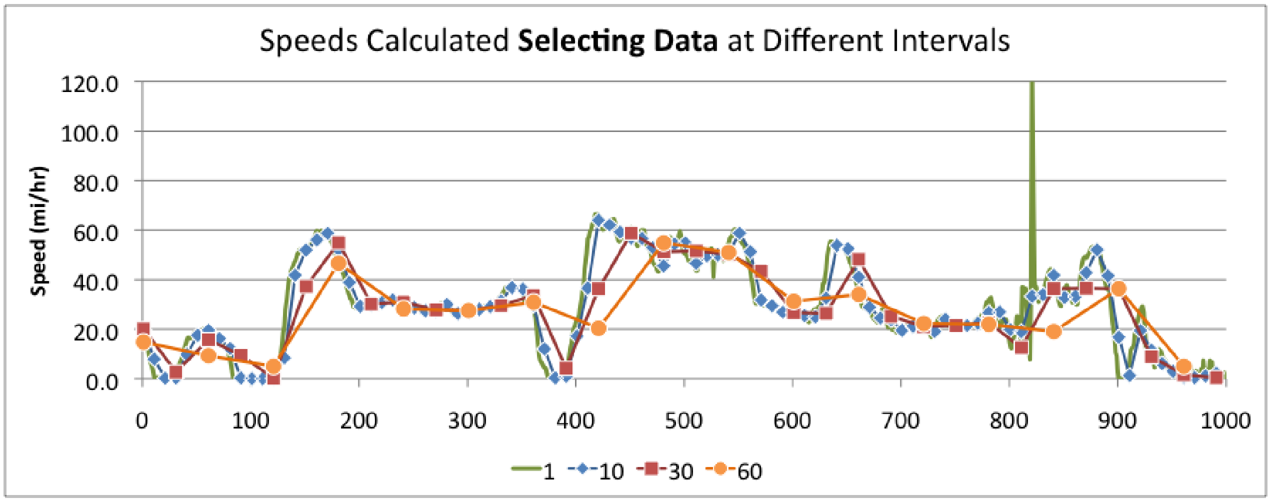

Figure 7: Speeds calculated with a selection of data

A systematic error was found when data is gathered at a high rate (between 1 and 2 seconds) regarding the calculation of speeds. When the data is gathered at small time intervals, minor errors in the positioning may cause large errors in the calculation of speed. A normal walking speed of 1.4 meter per second (5 km/h) can easily turn into 15 km/h, if the first data point has an error of 1 meter to the East and the second data point has an error of 1 meter to the West and the mobile is moving from East to West. Each position may have up to 3 meters of error; therefore for data gathered each second, the calculated speed can be 6 meters per second (20 km/h) for an object at rest. If data is gathered every 15 seconds the same error will become 6 meters per 15 seconds or 0.4 m/s (1.5 km/h). The Figure 8 shows an example of that type of error for 2 consecutive data points. As shown, not only the speed will be affected, but the direction of the speed vector will be affected too. When the sampling rate decreases, the inherent smoothing of the data tends to decrease the effect of this type of error.

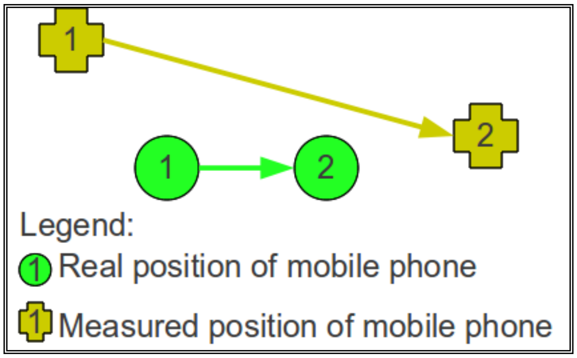

Figure 8: Speed calculation error example

5. IDENTIFYING TRANSPORTATION MODE

A single round trip might include several transportation modes. Also, different users tend to use different transportation modes. An example of a round trip from home to work can be seen in Figure 9

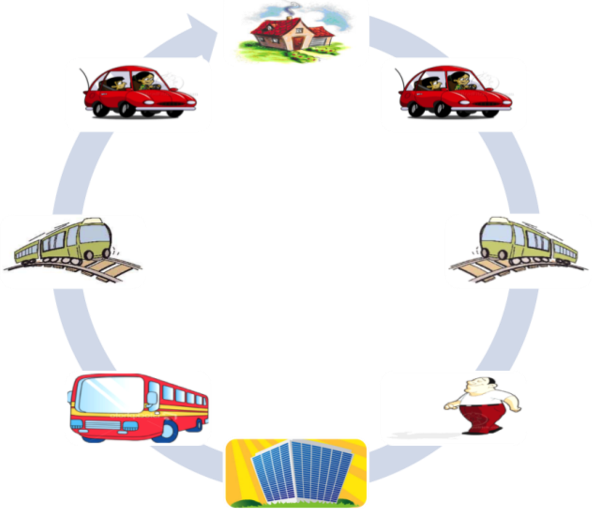

Figure 9: Round trip example

One of the main problems is that some modes of transportation are harder to identify than others. The transportation modes that are expected to be easier to ID are walking and heavy rail. Walking will always be slower than all the others alternatives making it easier to ID. Heavy rail will always use the same route (that is typically segregated from other modes) and that characteristic will make it easier to identify.

Therefore cars, trucks, buses, light rail, motorbikes and bikes are harder to identify because they utilize similar routes and differences in speed are small. Buses might be ID with the aid of Bus-stops. To ID light rail trains, a combination of routes, stops and speeds is needed.

6. COMPLEXITY

The complexity of utilizing mobile phones lies in the nature of the problem, the capacity of the technology and the sampling rate. To utilize mobile phone data, we need to be *able* to do so, but at the same time we do not want to collect more data than needed. Still, the biggest challenge is the need for an algorithm capable of determining the transportation mode. After getting over the challenges, transportation agencies will be able to gather the much-needed data in order to improve transportation planning and operations.

REFERENCES

FCC, Federal Communications Commission. (2009). Enhanced 9-1-1 - Wireless Services, http://www.fcc.gov/pshs/services/911-services/enhanced911/Welcome.html, (January 2009).

Fontaine, M.D. and Smith, B.L. (2005). “Probe-based traffic monitoring systems with wireless location technology: an investigation of the relationship between system design and effectiveness.” *Transportation Research Record.* No. 1925, pp. 3-11.

Hellinga, B., Izadpanah, P., Takada, H. and Fu, L. (2008). “Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments”. *Transportation Research Part C: Emerging Technologies*, Vol. 16, No. 6, pp 768-782.

Mobile Millennium. (2009). The Mobile Millennium project http://traffic.berkeley.edu/theproject.html. (July, 2009)

Sohn, K. and Hwang, K. (2008) "Space-Based Passing Time Estimation on a Freeway Using Cell Phones as Traffic Probes". Intelligent Transportation Systems, IEEE Transactions on, Vol. 9, No.3, pp 559-568.

Sohn, K. and Kim, D. (2008). “Dynamic Origin–Destination Flow Estimation using Cellular Communication System”. IEEE Transactions on Vehicular Technology, Vol. 57, No. 5, pp 2703-2713.

Tate-Glass, M.J., Bostrum, R. and Witt, G. (2000). “Data, data, data – Where’s the data?” *Transportation in the New Millennium.* pp 7.

Winters, P., Barbeau, S. and Georggi, N. (2008). “Smart phone application to influence travel behavior (trac-it phase 3),” National Center for Transit Research. University of South Florida. Florida Department of Transportation. http://www.nctr.usf.edu/pdf/77709.pdf, February 2008 [Online]. (July, 2009).

Ygnace, J., Drane, C., Yim, Y.B. and de Lacvivier, R. (2000). “Travel time estimation on the San-Francisco bay area network using cellular phones as probes.” University of California, Berkeley, California Partners for Advanced Transit and Highways (PATH), Working Papers: Paper UCB-ITS-PWP-2000-18.

Authorization and Disclaimer