In review, Transportation Research C. Submitted March 2009. Resubmitted in revised form June 2009

Evaluation of Traffic Data Obtained via GPS-enabled Mobile Phones: the *Mobile Century* field experiment

Juan C. Herrera[a,*], Daniel B. Work[a], Ryan Herring[a], Xuegang (Jeff) Ban[b,1], Quinn Jacobson[c], Alexandre M. Bayen[a,1]

[a] *University of California, Berkeley, CA, 94720-1710, United States*
[b] *Rensselaer Polytechnic Institute (RPI), Troy, NY, 12180-3590, United States*
[c] *Nokia Research Center Palo Alto, Palo Alto, CA, 94304-1003, United States*

Abstract

The growing need of the driving public for accurate traffic information has spurred the deployment of large scale dedicated monitoring infrastructure systems, which mainly consist in the use of inductive loop detectors and video cameras. On-board electronic devices have been proposed as an alternative traffic sensing infrastructure, as they usually provide a cost-effective way to collect traffic data, leveraging existing communication infrastructure such as the cellular phone network. A traffic monitoring system based on GPS-enabled smartphones exploits the extensive coverage provided by the cellular network, the high accuracy in position and velocity measurements provided by GPS devices, and the existing infrastructure of the communication network. This article presents a field experiment nicknamed *Mobile Century*, which was conceived as a proof of concept of such a system. *Mobile Century* included 100 vehicles carrying a GPS-enabled Nokia N95 phone driving loops on a 10-mile stretch of I-880 near Union City, California, for 8 hours. Data were collected using *virtual trip lines*, which are geographical markers stored in the handset that probabilistically trigger position and speed updates when the handset crosses them. The proposed prototype system provided sufficient data for traffic monitoring purposes while managing the privacy of participants. The data obtained in the experiment were processed in real-time and successfully broadcast on the internet, demonstrating the feasibility of the proposed system for real-time traffic monitoring. Results suggest that a 2-3% penetration of cell phones in the driver population is enough to provide accurate measurements of the velocity of the traffic flow.

Key words: GPS-enabled cell phones, traffic monitoring systems, traffic sensors.

---

[*] Corresponding author.
[1] Assistant Professor.

1. Introduction

Before the era of the mobile internet, characterized in particular by the emergence of location based services heavily relying on GPS, the traffic monitoring infrastructure has mainly consisted of dedicated equipment, such as loop detectors, cameras, and radars. Installation and maintenance costs prevent the deployment of these technologies for the entire arterial network and even for highways in numerous places around the world. Moreover, inductive loop detectors are prone to errors and malfunctioning (daily in California, 30% out of 25000 detectors do not work properly [1]).

For this reason, the transportation engineering community has looked for new ways to collect traffic data to monitor traffic. Electronic devices traveling onboard cars are appealing for this purpose, as they usually provide a cost-effective and reliable way to collect traffic data.

Radio-frequency identification (RFID) transponders, such as Fastrak in California or EZ-Pass on the East Coast[2], can be used to obtain individual travel times based on vehicle re-identification [2], [3]. Readers located on the side of the road keep record of the time the transponder (i.e. the vehicle) crosses that location. Measurements from the same vehicle are matched between consecutive readers to obtain travel time. The fundamental limitations of this system is the cost to install the infrastructure (readers), its limited coverage, and the fact that only travel time between two locations can be obtained.

Global Positioning System (GPS) devices found in the market can obtain position and instantaneous velocity readings with a high accuracy, which can be used to obtain traffic information. In [4] the authors addressed some of the key issues of a traffic monitoring system based on probe vehicle reports (position, speeds, or travel times), and concluded that they constitute a feasible source of traffic data. In [5] the authors also investigated the use of GPS devices as a source of data for traffic monitoring. Two tests were performed to evaluate the accuracy of the GPS as a source of velocity and acceleration data. The accuracy level found was good, even though the selective availability[3] feature was still on. The main drawback of this technology is that its low penetration in the population is not sufficient to provide an exhaustive coverage of the transportation network. Dedicated probe vehicles equipped with a GPS device represent added cost that cannot be applied at a global scale. An example of such program at a small scale is HICOMP[4] in California, which uses GPS devices in dedicated probe vehicles to monitor traffic for some freeways and major highways in California. However, as pointed out in [6], the penetration of HICOMP is low and the collected travel times are not as reliable as other systems such as PeMS.

[2]Fastrak and EZ-Pass are electronic transponders used to pay road tolls electronically.

[3]Selective availability is the intentional inclusion of positioning error in civilian GPS receivers. It was introduced by the Department of Defense of the U.S. to prevent these devices from being used in a military attack on the U.S. This feature was turned off on May 1, 2000.

[4]HIghway COngestion Monitoring Program.
http://www.dot.ca.gov/hq/traffops/sysmgtpl/HICOMP/index.htm

Other approaches have investigated the possibility of using dedicated fleets of vehicles equipped with GPS or *automatic vehicle location* (AVL) technology to monitor traffic [7, 8, 9], for example FedEx, UPS trucks, taxis, buses or dedicated vehicles. While industry models have been successful at gathering significant amounts of historical data using this strategy, for example *Inrix*, the use of dedicated fleets always poses issues of coverage, penetration, bias due to operational constraints and specific travel patterns. Nevertheless, it appears as a viable source of data, particularly in large cities.

In the era of mobile internet services, and with the shrinking costs and increased accuracy of GPS, probe based traffic monitoring has become one of the next arenas to conquer by industries working in the field of mobile sensing. Increasing penetration of mobile phones in the population makes them attractive as traffic sensors, since an extensive spatial and temporal coverage could potentially soon be achieved. GPS-enabled cellular phone based traffic monitoring systems are particularly suitable for developing countries, where there is a lack of resources for traffic monitoring infrastructure systems, and where the penetration rate of mobile phones in the population is rapidly increasing. By the end of 2007, the penetration rate of mobile phones in the population was over 50% in the world, ranging from 30-40% in developing countries (with an annual growth rate greater than 30%) to 90-100% in developed countries [10].

Multiple technological solutions exist to the localization problem using cell phones. Historically, the seminal approach chosen for monitoring vehicle motion using cell phones (prior to the rapid penetration of GPS in cellular devices) uses cell tower signal information to identify handset's location. This technique usually relies on triangulation, trilateration, tower hand-offs, or a combination of these. Several studies have investigated the use of mobile phones for traffic monitoring using this approach (see for example [11], [12], [13], [14] and [15]). The fundamental challenge in using cell tower information for estimating position and motion of vehicles is the inherent inaccuracy of the method, which poses significant difficulties to the computation of speed. Several solutions have been implemented to circumvent this difficulty, in particular by the company *Airsage*, which historically developed its traffic monitoring infrastructure based on cell tower information [16, 17, 18]. Based on the time difference between two positions, average link travel time and speed can be estimated. In [19], the authors conduct a field experiment to compare the performance of cell phones and GPS devices for traffic monitoring. The study concludes that GPS technology is more accurate than cell tower signals for tracking purposes. In addition, the low positioning accuracy of non-GPS based methods prevents its massive use for monitoring purposes, especially in places with complex road geometries. Also, while travel times for large spatio-temporal scales can be obtained from such methods, other traffic variables of interest, such as instantaneous velocity are more challenging to obtain accurately.

A second approach is based on GPS-enabled smartphones, leveraging the fact that increasing numbers of smartphones or PDAs come with GPS as a standard feature. This technique can provide more accurate location information, and thus more accurate traffic data such as speeds and/or travel times.

Additional quantities can potentially be obtained from these devices, such as instantaneous velocity, acceleration, and direction of travel. In [14], the authors use cell phone for traffic monitoring purposes, and mention the need of having a GPS-level accuracy for position to compute reasonable estimates of travel time and speed. In [19] and [20], the authors conclude that if GPS-equipped cell phones are widely used, they will become more attractive and realistic alternative for traffic monitoring. GPS-enabled mobile phones can potentially provide an exhaustive spatial and temporal coverage of the transportation network when there is traffic, with a high positioning accuracy achieved by a GPS receiver. Some concerns regarding this technology include the need of a specifically designed handset, and the fact that the method requires each phone to send information to a center [21], [22], which could potentially increase the communication load on the system and the energy consumption of the handset[5]. Another issue is the knowledge of vehicle position and velocity provided by this technology, which needs to be used in a privacy non intrusive way.

The impact of these concerns (communication load, handset energy consumption, and privacy) can be handled with the appropriate sampling strategy. Sampling GPS data in the transportation network can be handled in at least two ways:

- *Temporal sampling*: equipped vehicles report their information (position, velocity, etc.) at specific time intervals T, regardless of their positions.
- *Spatial sampling*: equipped vehicles report their information (time, velocity, etc.) as they cross some spatially defined sampling points. This strategy is similar to the one used by inductive loop detectors, RFID transponders or license plate readers, in which data are obtained at fixed locations. It has the advantage that the phone is forced to send data from a given location of interest.

From a traffic estimation perspective, it is desirable to have a significant amount of information available. Therefore, with a satisfying GPS accuracy, small T or very closely placed fixed measurements would yield more accurate estimates of traffic. However, these objectives conflict with the communication load constraints and privacy preservation.

As suggested in the literature [12], [20], [22], [23], field tests are needed to assess the potential of new technologies such as GPS-enabled mobile phones. Test deployments to assess the potential of traffic monitoring using cell phones go back to the advent of GPS on phones. In particular, the study [24] investigates the deployment of 200 vehicles for an extended period of three months and the potential data which can be gathered from it. As appears in light of [24], one of the main issues in experiments or pilot tests is the problem of *penetration*, i.e. percentage of vehicles equipped vs. total number of vehicles on the road.

[5]With the advent of the 3G network and rapid growth of data and bandwidth intensive applications, this concern has become less important in the last months.

This article presents the results of a large scale field experiment conducted in the San Francisco Bay Area, California, and aimed at assessing the feasibility of a traffic monitoring system using GPS-enabled mobile phones for highways. The specificity of this field experiment is the penetration rate achieved during the test, which the authors believe is representative of upcoming GPS equipped phones penetration in the population within a few months from the experiment. The performance of the system was sustained for a long enough time to show the feasibility of such a monitoring system. In addition to the data gathered, which is among the first in its kind, the article also briefly summarizes the prototype system [25] which was built to gather the data, and which was recently extended for a pilot deployment [26], [27].

The rest of the article is organized as follows. Section 2 describes the system used to collect traffic data, along with the sampling strategy. Section 3 explains the goals of the experiment and its design. Section 4 presents the main results obtained from the data. Finally, Section 5 states the main conclusions obtained from the experiment.

2. System description

2.1. Sampling and Data Collection

As explained earlier, a variety of sampling techniques can be used to collect data from GPS enabled mobile devices. In the case of the Nokia N95, the embedded GPS chip-set is capable of producing a time-stamped geo-position (latitude, longitude, altitude) every three seconds. From this time and position data, the instantaneous velocity is produced by the phone at the same frequency. Over time, this vehicle trajectory and velocity information produces a rich history of the dynamics of the vehicle and the velocity field through which it evolves.

While this level of detail is particularly useful for traffic estimation, it can be privacy invasive, since the device is ultimately carried by a single user. Even if personally identifiable information from the data is replaced with a randomly chosen ID through a process known as pseudo-anonymization, it is still possible to reidentify individuals from trajectory data. For example, pseudo-anonymous trajectories have been combined with free, publicly available data sets to determine the addresses of participants homes [28].

The transmission of high frequency data without regard to location also wastes resources throughout the system, which can pose scalability problems. In addition to disclosing sensitive information, the trajectory information on small roadways near users homes are of lower value to the general commuting public than major thoroughfares such as interstates. Thus, collection of low utility and highly sensitive data should be avoided when sampling using mobile devices.

A variety of methods can be used to address these problems. To manage privacy concerns, in addition to pseudo-anonomization of the trajectory data, the data can be further degraded until a sufficient level of privacy is attained. Common degradation approaches include (i) spatial obfuscation (i.e. blocking data collection from particular regions, such as home), (*ii*) increasing uncertainty in the data through noise addition, and (*iii*) location discretization approaches, which round the measurement to the nearest discrete grid point. The tradeoffs between the measurement utility and privacy under these degradation approaches have been analyzed with experimental data [29] and can be cast as a sampling strategy optimization problem [23].

An alternative sampling strategy which is implemented in this work is based on *Virtual Trip Lines* (VTLs) [25], which act as spatial triggers for phones to collect measurements and send updates. Each VTL consists of two GPS coordinates which make a virtual line drawn across a roadway of interest. Instead of periodic sampling (in time), VTLs trigger disclosure of speed and location updates by sampling in space, creating updates at predefined geographic locations on roadways of interest.

In this sampling strategy, mobile devices monitor their speed and location using GPS and use the locally stored VTLs to determine when a VTL crossing occurs. When the phone intersects a VTL, the device sends an update to a back end server with anonymized position, speed and direction information[6]. The device may also probabilistically send the travel time observed between two consecutive trip lines.

A unique feature of this sampling strategy is that data points are only identified through the ID of the VTL, and not that of the mobile device which generated the update, so no privacy-invasive extended trajectories are collected. Furthermore, measurements can be disregarded by the server to minimize the possibility of correlating VTL measurements at adjacent VTLs, which might still enable the reconstruction of individual trajectories. Through careful placement of trip lines, the system is better suited to manage data quality and privacy than through a uniform temporal sampling interval.

2.2. System Architecture

A prototype system architecture was implemented to test a VTL based sampling strategy (shown in Figure 1). The system consists of four layers: GPS-enabled smartphones in vehicles (driving public), a cellular network operator (network operator), cellular phone data aggregation and traffic estimation (Nokia/Berkeley), and information dissemination (Info Consumers). On each participating mobile device (or client), an application is executed which is responsible for the following functions: downloading and caching trip lines from the VTL server, detecting trip line traversal, and filtering measurements before transmissions to the service provider. To determine trip line traversals, the device checks if the line between the current GPS position and the previous GPS position intersects with any of the trip lines in its cache. Upon traversal, the mobile device creates an encrypted VTL update. The update comprises of a speed reading, timestamp, the trip line ID, and the direction of the trip line

[6]In the system which is currently deployed, as part of a follow up experiment, there is a chance that a phone will cross a VTL and not send data.

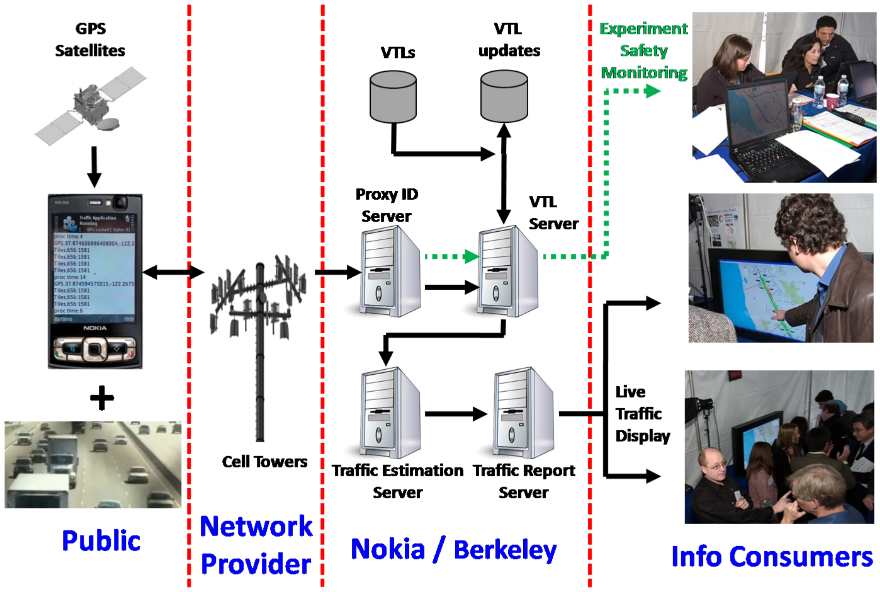

Figure 1: *Mobile Century* system architecture overview. The system consists of vehicles equipped with GPS-enabled smartphones (Nokia N95), a cellular network provider, a data collection infrastructure and a traffic estimation engine, and an information display system. A live tracking infrastructure (shown in dashed green) was also required for the safety of the UC Berkeley student drivers during this experiment, but it is *not* part of the core system (shown in solid black).

crossing. These VTL updates are transmitted to the ID proxy server over a secure channel.

Note that all data packets transmitted from the mobile device, regardless of the application (traffic, email, etc), must contain the mobile device identification information for billing by the network provider. Thus, in the *Mobile Century* system, an ID proxy server is used to first authenticate each client to prevent unauthorized updates, then remove the mobile device identification information from the data packets. It then forwards the anonymized updates to the VTL server. Since the VTL update is encrypted with the VTL server's public key (RSA encryption), the ID proxy server cannot access the VTL update content. It only has knowledge of which phone transmitted a VTL update, but no knowledge of the phone's position or speed information. Thus we prevent any single entity from observing both the identification data required by the network operator, and the sensing data. A more detailed description of privacy protection in VTL based traffic monitoring is available in [25].

The VTL server stores all trip lines in a VTL database and distributes trip lines within a given region to a mobile device upon receiving a VTL download request for that region. The VTL server also aggregates updates from a large number of probe vehicles in a VTL update database and pushes the data to UC Berkeley algorithms for data assimilation (for example [30]), which run on a traffic estimation server. An estimate manager in the traffic estimation server monitors the performance of the various algorithms and transmits the resulting traffic estimates with highest confidence to the traffic report server.

The traffic report server then sends data to information consumers through a mapping interface on a web site. During the *Mobile Century* experiment, large displays were used on the experiment site to show the live traffic estimate. In the current version of the system, the traffic information is now accessible from the mobile devices running the traffic data collection client.

The current VTL implementation generates approximately 1KB of update data for every two minutes per client while driving on a major road. Assuming an average two hours of driving per day on a major road, we expect the total data transfer is 60KB per day. The database servers can easily scale to large number of client updates since the bandwidth and the total data storage demands are rather small by current information industry standards.

2.3. System Safety Monitoring Requirements

In order to address driver safety concerns, the VTL based system architecture described previously was augmented with additional experiment safety monitoring infrastructure. In addition to collecting VTL data, a special build of the client was created which is also capable of transmitting the three second trajectory data and a phone ID in the encrypted update. Each phone ID corresponded to a unique vehicle ID which was prominently displayed on the sides and roof of all vehicles in the experiment. The locations of these vehicles were then monitored by a team of safety officers, through a special mapping application running on laptops at the experiment site. If the vehicle left the experiment site, or appeared disabled, the safety officers were then able to call the drivers via the phone in the vehicle and provide directions or dispatch emergency services (stationed on site) as needed. This infrastructure (shown in green in Figure 1) is not needed or used by the traffic data collection and traffic estimation system, and is not implemented in a non-research build of the system.

3. Experimental design

The experiment was conceived as a proof of concept of the system described in the previous section. It was designed with three fundamental goals:

Goal 1 Assess the feasibility of a traffic monitoring system based on GPS-enabled mobile phones. The system described in Section 2 was shown to provide sufficient and accurate enough data to deliver precise travel time and velocity estimations.

Goal 2 Evaluate speed measurements accuracy from GPS-enabled mobile phones under both free flow and congested traffic conditions. Therefore, the section of highway was chosen to encompass both free flow and congested conditions. A good detector stations coverage was also required for comparison purposes.

Goal 3 Maintain a specific penetration rate of equipped vehicles in the total flow throughout the day. This feature of the experiment is a fundamental difference with previous work, and necessary for the proper testing of traffic flow reconstruction algorithms.

Nicknamed the *Mobile Century* experiment, the February 8, 2008, field experiment involved 100 vehicles carrying GPS-enabled Nokia N95 phones. All rented vehicles were driven by 165 UC Berkeley students in 3-hour shifts. Drivers were instructed to drive as they would normally do, and no other specific instructions were given to them. The vehicles repeatedly drove loops of six to ten miles in length continuously for eight hours on freeway I-880 near Union City in the San Francisco Bay Area, California (see Figure 2). This section of highway has four (and sometimes five) lanes, the leftmost one being a HOV lane. It presents interesting traffic properties, which include alternating periods of free-flow and congestion throughout the day (which thus satisfies the requirements of Goal 2). In particular, the northbound (NB) direction presents a recurrent and severe bottleneck between Tennyson Rd. and CA92 during the afternoon. Moreover, on the day of the experiment, there was an accident during the morning, which activated a non-recurrent bottleneck at this same location. The section is also well covered with existing loop detector stations – 25 between Stevenson Blvd. and Winton Ave. in the NB direction – feeding into the PeMS system [1].

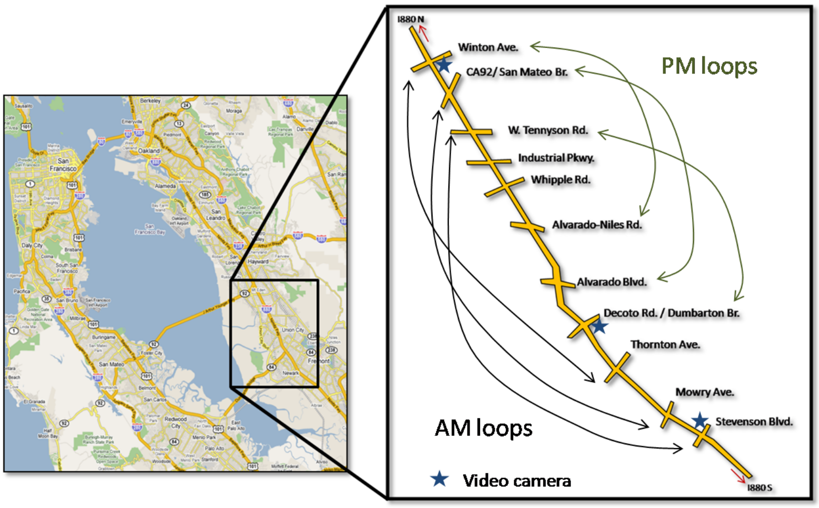

Figure 2: Stretch of highway I-880 CA, used in the *Mobile Century* experiment.

Based on a realistically achievable penetration rate in the near future[7], the goal of the looping behavior is to achieve and maintain a desirable 2%-5% pene-

[7]Analysts predict a rapid increase in the market share of GPS phones in the near future [31].

Table 1: Features of the loops used in the experiment.

Loop type	North end	South end	One-way distance (miles)
AM (long)	Winton Ave	Thorton Rd.	9.4
	CA92 (San Mateo Br.)	Mowry Ave.	8.6
	Tennyson Rd.	Stevenson Blvd.	9.3
PM (short)	Winton Ave.	Alvarado-Niles	4.5
	CA92 (San Mateo Br.)	Alvarado Blvd.	5.2
	Tennyson Rd.	Decoto Rd.	5.4

tration rate of the total volume of traffic on the highway during the experiment (Goal 3). Note that previous studies have reported that data coming from no more than 5% of the total flow are sufficient to obtain accurate estimates of the travel time [4], [11], [19]. Given that the total flow expected on the section of interest is approximately 6000 vehicles per hour (obtained from PeMS) and the number equipped vehicles is 100, the required cycle time to achieve the desired rate is 20 minutes. Knowing the expected speed throughout the day and cycle time is sufficient to determine the length of the loops throughout the day. In the NB direction of the section of interest, free flow conditions are historically expected during the morning until 2-3pm, when the recurrent bottleneck between Tennyson Rd. and CA92 activates. Free flow is expected during most of the day for the southbound direction. For this reason, long loops (or AM) loops were designed during the morning and short (or PM) loops were used during the afternoon. The change was scheduled to start at at 1:30pm. Table 1 presents the main features of the loops used during the experiment, also shown in Figure 2. Three different loops of almost the same length were used during the experiment not to oversaturate any of the ramps being used.

The data were collected in two ways during the experiment. First, each Nokia N95 cell phone was storing its position and velocity log every 3 seconds, which allows for the computation of every equipped vehicle trajectory. This data were gathered locally on the phones for analysis purposes, and is not part of the data gathering process of the system presented in the previous section. It became available only once the experiment was finished, and is useful to test the accuracy of the sampling strategy (Goal 2) a posteriori. Second, the privacy preserving architecture described in Section 2 collected data from the 45 VTLs deployed between Stevenson Blvd. and Winton Ave. (each VTL covers both travel directions). These data were used to produce real-time travel time and speed estimates, and helps to assess the feasibility of the system (Goal 1).

Finally, high resolution video cameras located on Winton Ave., Decoto Rd., and Stevenson Blvd. recorded traffic in the NB direction. This video data is accurate enough to provide exact travel time of individual vehicles through license plate re-identification.

4. Experimental results

This section analyzes the main results derived from the experiment. The analysis is carried out in terms of the three goals of the experiment. Unless otherwise noted, the rest of this section focusses on the highway segment covered by the afternoon loops in the northbound (NB) direction. The section consists of the portion of highway between Decoto Rd. to the south – postmile 21 – and Winton Ave. to the north – postmile 27.5.

Goal 1: Assessment of the feasibility of a smartphone-based traffic monitoring system

The data obtained in the experiment using the system architecture described in Section 2 were processed in real-time. We deployed 30 VTLs during the experiment in the section of interest. Information collected by these VTLs was used to produce real-time travel time and velocity estimates, which were broadcasted for eight hours. Figure 3b illustrates the interface used to broadcast travel time and speed during the day. The figure shows traffic at a time after an accident occurred in the NB direction between Tennyson Rd. and CA92. Figure 3a shows the 511.org traffic display at the same time.

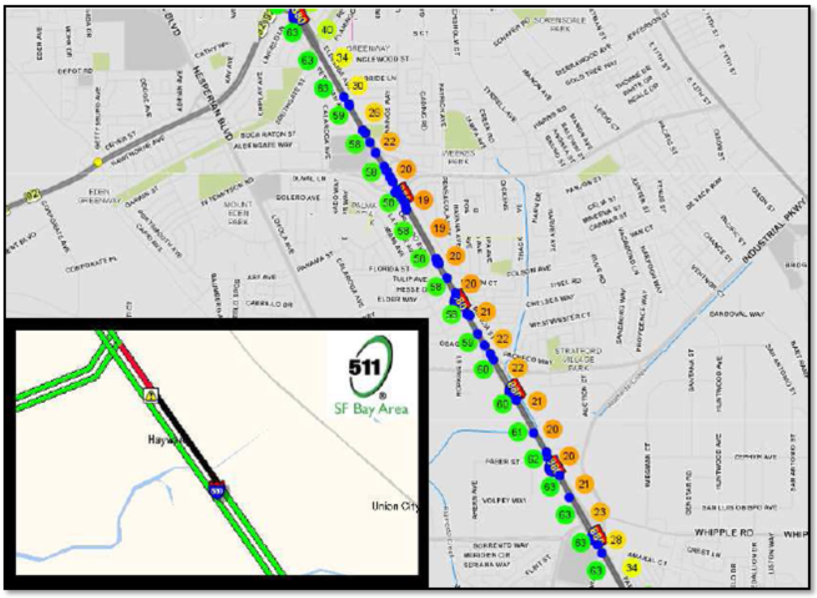

Figure 3: Live traffic feed at 10:52am on February 8, 2008, after an accident on the NB direction of I880 occurred, provided by the proposed system and 511.org (inside). Numbers in circles correspond to speed in mph.

As can be seen from the two subfigures in Figure 3, the extent of congestion estimated by our algorithm[8] and based on the GPS data only match closely the 511.org display, which uses a combination of data sources for velocity and travel time calculation including loop detectors, FasTrak-equipped vehicles, and speed radars. However, 511.org only provides speeds in discrete increments (e.g. the black color represents "stop and go" and red means "heavy traffic"), while our algorithm generated speeds with a finer scale, which is important because it allows a more accurate identification of the limit between zones with different traffic states (i.e. the location of the shockwave). Comparisons with the 511.org speed map at other times during the experiment showed similar results, which confirm that the GPS cell phone based technique and the system described in Section 2 can produce reasonable speed estimates for the section of interest, at least for the experiment day.

Goal 2: Assessment of the accuracy of the probe data

This sub-section analyzes the data stored in each phone and the type of information that can be collected by the system described in Section 2. By nature of the test site, it provides an assessment of GPS data quality in suburban freeways, and of these data's values for highway traffic estimation.

Trajectory data

Each phone stored its position (latitude and longitude) and a velocity log every 3 seconds. We refer to this data as *trajectory* data since vehicle trajectories can be reconstructed from them.

Trajectory data were processed after the experiment, in order to conduct a more detailed analysis of the quality of the data collected by the GPS-enabled smartphones. Figure 4 shows 50% of the gathered trajectories between Stevenson Blvd. (postmile 17) and Winton Ave. (postmile 27.5) in the NB direction. The transition from the AM loops to the PM loops that occurs at 1:30pm can be clearly seen in the figure, as well as the fact that different vehicles were using different ramps to get in and out of the highway (as shown in Figure 2). The propagation of the shockwave generated by the accident is clearly identified from this plot as well.

The red lines in Figure 4 represent the approximate propagation of the shockwaves generated by the accident, and are drawn by hand. The information about the propagation of shockwaves can be used to infer parameters of the fundamental diagram (assuming triangular relationship), as well as flows and densities that mobile sensors are not able to capture directly (see Appendix A). This information can be useful in the absence of loop detectors, since it relates the sample that provides GPS data with the total driving population.

Using these trajectories, a velocity field can be reconstructed and compared with the PeMS velocity field using data from the 17 loop detector stations de-

[8]The algorithm to estimate real-time travel times and velocities is described in [30], and is out of the scope of the present article.

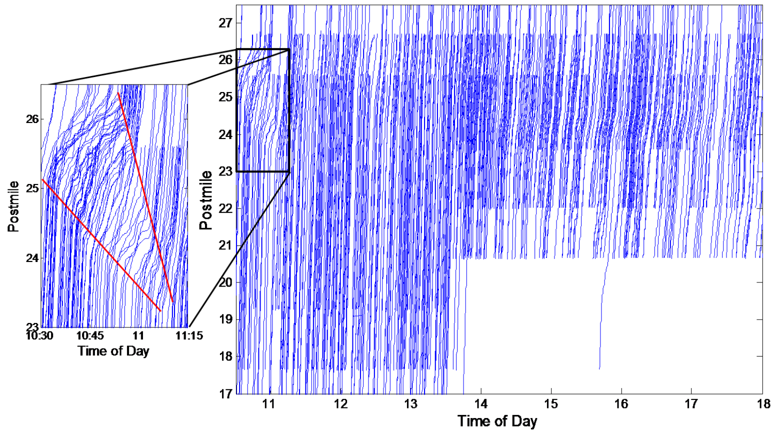

Figure 4: Vehicle trajectories in NB direction extracted from the data stored by 50% of the cell phones. The propagation of the shockwave from the accident can clearly be identified from this plot. The red lines in the subfigure were drawn by hand by fitting a line through the points where trajectories change slope.

ployed in the section of interest[9] (loop detector locations are shown in Figure 5). Loop detectors compute the *temporal mean speed* (TMS) every 30 seconds (considering all lanes); every 5 minutes the average of the 30-second observations collected during this time is computed. The PeMS velocity field is shown in Figure 6a. The method used in PeMS associates an influence area with each detector station. The assumption is that measurements for this area are provided by the corresponding detector station. The size of the influence area depends on the proximity of neighbor detector stations. Therefore, the closer the neighbor detectors are, the smaller the influence area and the better the estimates that can be obtained using this method.

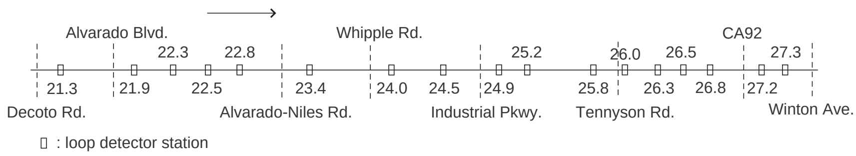

Figure 5: Loop detector locations for the NB direction. Numbers indicate postmiles. Traffic flows from left to right, and loop detectors have been numbered sequentially from 1 (upstream) to 17 (downstream).

Since equipped-vehicle trajectories are known, the velocity field is computed

[9]Loop detectors on lanes 1 and 2 at detector station 3 and lane 5 at detector station 6 were not working properly as reported by PeMS.

using Edie's generalized definition, in which "*the speed of a traffic stream in a given space-time domain is the aggregate distance traveled divided by the aggregate time spent by all vehicles traversing it*" [32]. The corresponding result is shown in Figure 6b. The qualitative agreement between subfigures a) and b) is evident – in terms of bottlenecks location, and their spatial and temporal extent. Note that less than 5% of the total trajectories are enough to provide a spatio-temporal coverage qualitatively comparable to the one accessible from 17 detectors for this section of highway.

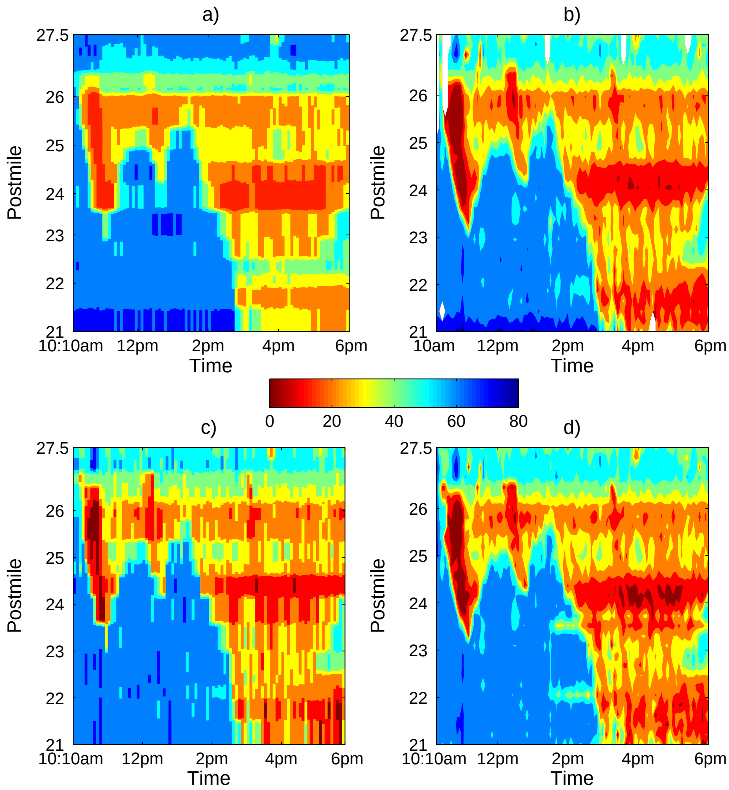

Figure 6: Velocity field in (mph) using a) 17 loop detector stations, b) vehicle trajectories and Edie's speed definition, c) 17 VTLs at the loop detector locations, and d) 30 VTLs equally spaced.

VTL data

In addition to the *trajectory* data stored by each phone, VTL data were collected during the experiment using the system architecture described in Section 2. As mentioned earlier, 30 VTLs were deployed during the experiment in the section of interest. Note that since all vehicle trajectories can be reconstructed, it is possible to artificially recreate VTL data off-line at different locations. This proves to be very useful because it allows a better analysis of the VTL data by not restricting its locations to the 30 locations deployed during the experiment.

By placing VTLs on existing loop detector locations (17 in total), velocity measurements collected by a loop detector every 5 minutes can be compared to the ones provided by a VTL at the same location. For comparison purposes, VTL measurements are also aggregated in 5-minute periods, and the TMS is computed for each period. Using data from the 17 VTLs, the velocity field is reconstructed using the same method described before for the loop detectors (see Figure 6c). The velocity map exhibits the same main features captured by the loop detector velocity field. Even though both sensors provide qualitatively similar information, there is some discrepancy in the velocity value they report (suggested by the difference in colors observed at certain times and locations). The field in Figure 6d was constructed using the 30 VTLs deployed during the experiment, and it is shown here just for comparison. The different level of granularity among the plots is explained by the different number of detector stations deployed in each case (17 loop detectors/VTLs versus 30 VTLs).

Since ground truth velocity is not known, accuracy of VTL velocity measurements cannot be directly assessed. Note that loop detector measurements are usually considered as ground truth. However, it is known that they include – sometimes significant – errors. For this reason, we decided not to use them as ground truth. Instead, travel times between Decoto Rd. and Winton Ave. from 10:45am to 5pm are extracted from high definition video cameras using license plate reidentification. A total of 4789 vehicles were matched between 10:40am and 5pm, but only 4268 of them were considered to correspond to vehicles staying on the freeway all the time (the other 521 matches correspond to vehicles exiting the freeway between Decoto Rd. and Winton Ave. and re-entering later between Decoto Rd. and Winton Ave. as well, resulting in unusually high travel times). These vehicles represent at least 10% of all the vehicles traveling the entire section between 10:40am and 5pm.

Velocity fields constructed using 17 VTLs and 17 loop detector stations can be integrated to compute travel time[10], which can be used to assess which velocity measurements are more likely to be closer to ground truth. Figure 7 shows the 4268 travel times obtained by reidentifying vehicles at Decoto Rd. and Winton Ave, and also the travel times computed by integrating both the VTL and loop detector velocity fields. The travel times shown in the figure correspond to those experienced by a vehicle entering the section at the corresponding time in the x-axis. Note that at 3pm, the left most lane becomes a HOV lane, which explains the points traveling the section faster than the rest of the traffic after 3pm.

Both estimates replicate the main trend observed in the travel time during the day. The VTL estimates, however, also adequately reproduce the value

[10]This a-posteriori travel time estimation method is also known as dynamic travel time or walk the speed matrix method.

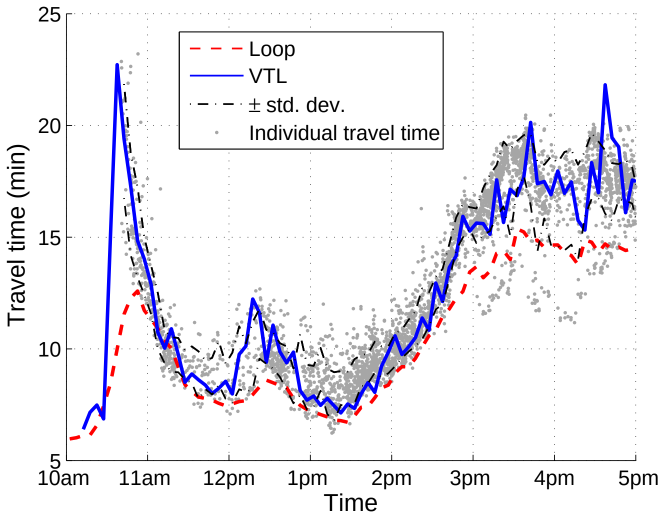

Figure 7: Travel time (in minutes) between Decoto Rd. and Winton Ave. Dots correspond to individual vehicle travel times (4268 in total), collected manually using video cameras at the ends of the section of interest. The time in the x-axis is the entry time to the section of interest.

of travel times. Loop detector estimates tend to underestimate travel times, implying that they tend to overestimate velocities. In fact, the VTL estimates are almost always within one standard deviation of the average travel time obtained from the video cameras in 5-minute windows (represented by the two black dash-dotted lines in the figure), while the opposite occurs with the loop detector estimates.

Travel times computed with the VTL velocity field are in better agreement with real travel times experienced by the flow during the day than loop detector travel times. This suggests that the VTL velocity field is more likely to be closer to the actual velocity experienced by the vehicles, and therefore more accurate, than the loop detector velocity field. That is, accuracy of this technology is such that a low proportion of equipped vehicles can often provide more accurate measurements of velocity than loop detectors – which sample (eventually) all vehicles. This has to be kept in mind when loop detector data are considered as ground truth, especially for an assessment of alternative data sources.

Because of the previous considerations, loop detector measurements are not considered as ground truth in this study. A data analysis is carried out only to observe the main features of both types of measurements, and not to determine accuracy of measurements.

For each location, velocity measurements have been classified into three classes according to the VTL or loop detector velocity reported in the corresponding 5-minute period (v_{VTL} or v_{loop}, respectively): congested conditions ($v_{(\cdot)} \leqslant 40$ mph), free flow conditions ($v_{(\cdot)} \geqslant 55$ mph), and the transition between both ($40 < v_{(\cdot)} < 55$). Tables 2 and 3 show the mean velocity, its standard deviation, and the number of measurements used to compute these quantities in each class, for each location, and for both VTL and loop detector measurements, respectively.

The means and standard deviations computed for both sensors show substantial differences in some cases. For a specific number of observations (more than 20), loop detectors tend to provide higher velocity measurements – suggesting certain bias on the detector –, but with a smaller variability than VTLs – which may be explained by the 5-minute aggregation performed in each case. As expected, higher variability in velocity is experienced during congestion.

Given the differences observed in the mean and variability, the distribution of velocity measurements is not always similar between VTL and loop detector. Figure 8 shows velocity histograms at four different locations – detectors 3, 6, 8, and 12. Histograms at the top correspond to VTL velocity measurements, while the ones at the bottom are loop detector velocity measurements. Both the free flow and congested mode can be identified at some locations.

Some of the 17 VTLs deployed generate similar velocity profiles as loop detectors, but some others exhibit significant differences. Figure 9 shows a time-series of loop detector and VTL velocity measurements for four different locations with changing proportion or penetration rates during the day[11] (see subfigure e). Locations on the figure correspond to detectors 1, 4, 7, and 17.

Both loop and VTL measurements differ from each other, and the level of discrepancy varies with time, location, penetration rate, and traffic conditions (i.e. velocity). Differences between both measurements can be explained by several factors, among them:

- Loop detectors and VTLs compute instantaneous velocity in different ways, and they have different measurement errors. While loop detectors use the travel time between dual coils to compute the speed for each vehicle, VTLs use GPS-computed velocity. Also, different aggregation methods explain some of the differences in the variability of the measurements (loop detectors compute the TMS every 30 seconds, and the ten values computed during 5 minutes are then averaged again to obtain the 5-minute average velocity; the VTL compute the TMS using the data collected during each 5-minute period).

- VTLs collect velocity from a proportion of all vehicles crossing that location, while loop detectors collect data from (eventually) all the vehicles. If this proportion is too small, it might not be statistically representative of the entire population.

- Selectivity bias in the sample chosen[12]. Drivers hired for the experiment are not necessarily a proper statistical sample of the population. The 165

[11] Penetration rate is the proportion of GPS equipped vehicles in the total flow. The next subsection describes how this rate is computed.

[12] This reason applies only for this experimental case.

Table 2: Mean and standard deviation (both in mph) of 5-minute VTL velocity measurements (v_{VTL}). Velocity classification is defined according to VTL velocity (v_{VTL}). The number of observations (i.e. 5-minute periods) on each class is also provided.

Location	$v_{\mathrm{VTL}} \leqslant 40$ mph			$40 < v_{\mathrm{VTL}} < 55$ mph			$v_{\mathrm{VTL}} \geqslant 55$ mph		
	Mean	Std.Dev	# obs.	Mean	Std.Dev	# obs.	Mean	Std.Dev	# obs.
1	20.7	8.6	33	47.7	5.6	4	65.2	3.4	57
2	17.4	7.1	39	48.9	-	1	65.1	2.5	54
3	24.5	6.1	39	52.3	4.1	3	63.7	2.9	52
4	30.0	4.5	38	45.4	0.8	2	65.4	2.3	54
5	28.7	7.4	29	46.2	4.3	13	65.4	2.3	52
6	28.5	5.8	39	50.2	5.3	6	64.6	2.9	49
7	17.1	7.7	52	49.1	1.7	4	63.7	2.4	38
8	14.2	5.6	64	48.1	-	1	64.8	2.5	29
9	28.6	7.6	69	47.9	4.6	5	60.0	2.7	20
10	28.1	8.5	62	44.5	3.9	22	60.4	3.1	10
11	22.9	6.6	86	44.0	4.1	4	59.2	2.4	4
12	20.3	7.1	90	-	-	0	58.9	1.1	4
13	29.9	6.4	91	50.2	6.0	2	60.5	-	1
14	32.7	9.5	43	42.4	2.4	50	60.5	-	1
15	29.2	7.9	16	45.8	2.6	77	60.5	-	1
16	36.6	1.5	2	51.0	2.5	44	58.9	3.9	48
17	32.0	7.8	7	49.2	3.5	71	58.0	3.5	16
All	24.6	9.0	799	47.0	4.2	309	63.6	3.7	490

Table 3: Mean and standard deviation (both in mph) of 5-minute loop detector velocity measurements (v_{loop}). Velocity classification is defined according to loop detector velocity (v_{loop}). The number of observations (i.e. 5-minute periods) on each class is also provided.

Location	$v_{\text{loop}} \leqslant 40$ mph			$40 < v_{\text{loop}} < 55$ mph			$v_{\text{loop}} \geqslant 55$ mph		
	Mean	Std.Dev	# obs.	Mean	Std.Dev	# obs.	Mean	Std.Dev	# obs.
1	31.7	3.5	32	46.4	5.5	6	70.9	2.3	56
2	26.1	4.0	39	-	-	0	64.8	1.5	55
3	36.3	2.0	34	43.4	3.9	6	64.0	1.3	54
4	38.4	1.0	3	48.0	3.1	37	67.7	1.3	54
5	31.1	3.7	29	47.1	4.0	13	62.0	1.6	52
6	29.7	3.1	37	44.8	5.2	6	67.2	3.6	51
7	18.8	6.4	57	49.3	4.0	4	60.3	1.5	33
8	21.9	4.2	61	46.9	2.0	5	67.5	3.2	28
9	32.8	5.5	62	45.4	3.8	10	64.2	3.3	22
10	31.3	5.4	73	45.6	4.0	12	60.6	2.9	9
11	26.9	4.3	90	42.8	1.5	2	60.0	0.1	2
12	27.4	4.1	92	-	-	0	60.7	0.6	2
13	30.2	6.8	98	50.0	2.7	81	57.1	3.1	5
14	37.3	2.3	10	46.3	2.3	83	55.2	-	1
15	-	-	0	53.9	0.5	8	58.2	2.0	86
16	-	-	0	54.9	-	1	63.2	3.5	93
17	39.1	-	1	47.3	5.0	3	64.2	5.0	90
All	28.2	6.5	628	47.8	3.6	277	64.0	4.3	693

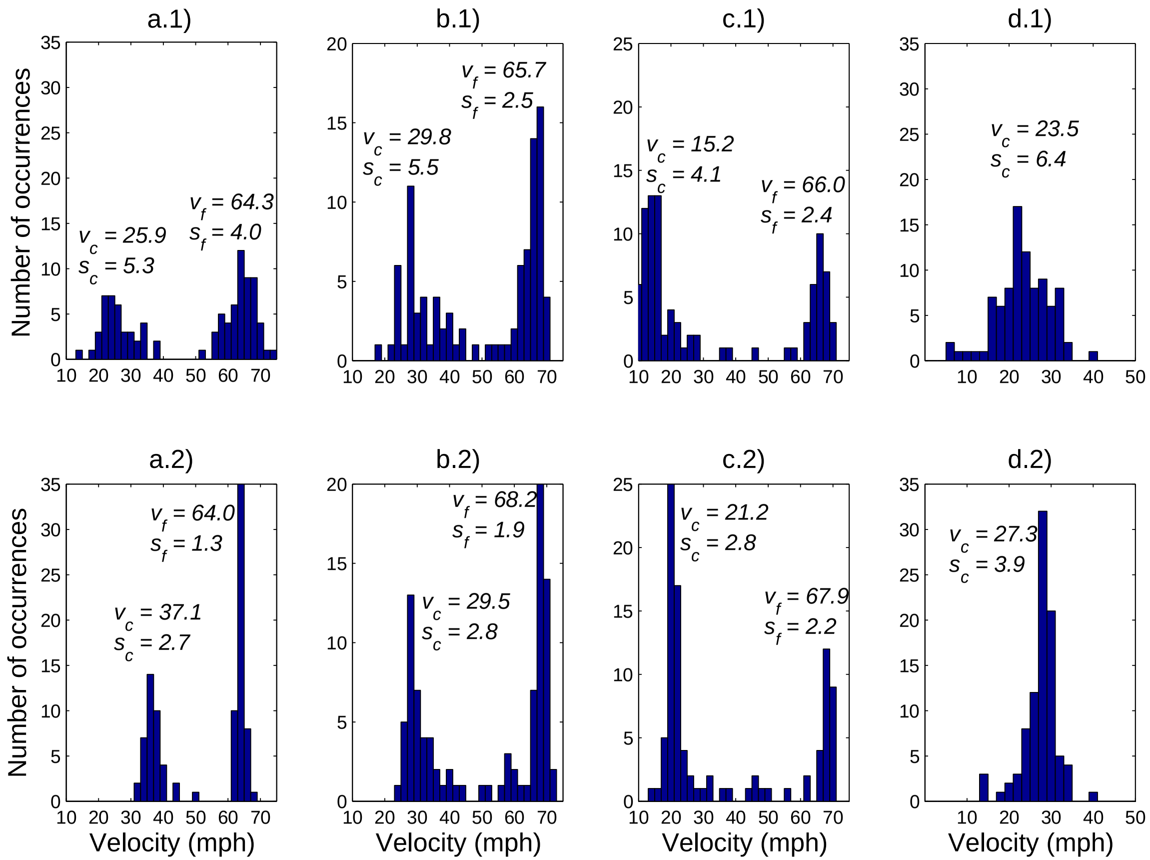

Figure 8: Velocity histograms at four different locations: a) detector 3, b) detector 6, c) detector 8, and d) detector 12. Graphs at the top (x.1) correspond to VTL measurements. Graphs at the bottom (x.2) correspond to loop detector measurements. v_f and s_f are the mean and standard deviation for free flow observations, while v_c and s_c are those for congested observations.

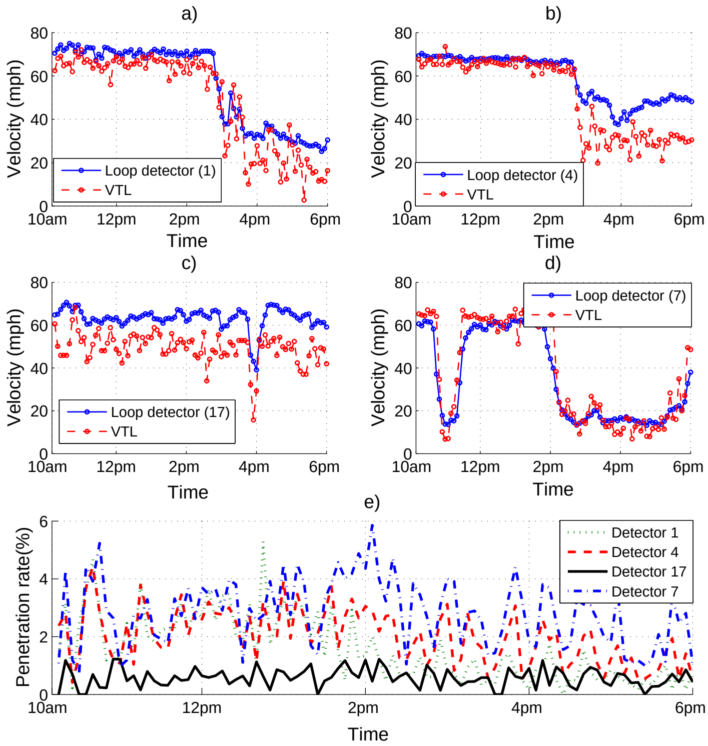

Figure 9: Loop detector versus VTL velocity data collected at detectors a) 1, b) 4, c) 17, and d) 7. Subfigure e) shows the penetration rate at these four locations during the day.

drivers were UC Berkeley students over 21, which may constitute a biased sample of the driving population. In addition to this, the driving behavior may be biased with respect to the rest of the traffic for other reasons, including fatigue and gained knowledge of the location and driving conditions (which may be similar to the expertise gained by regular commuters). A specific bias can be observed at some locations close to the off-ramps used during the experiment, where VTL velocity measurements are always lower than the loop detector velocity measurements (Figure 9c). However, this bias is not observed at some other locations (Figure 9d). Therefore, the bias is most likely due to *i*) bias in the detector, or *ii*) test driver dynamics before exiting the mainline of the highway.

Table 4 shows the mean and standard deviation of the absolute difference between loop detector and VTL velocity measurements. For each location, measurements have been classified as before (and according to the VTL velocity v_{VTL} reported).

Discrepancy between VTL and loop detector measurements is higher for

Table 4: Mean and standard deviation of the absolute difference between 5-minute VTL and loop detector velocity measurements. Velocity classification is defined according to VTL velocity (v_{VTL}). The number of observations (i.e. 5-minute periods) on each class is also provided.

Location	$v_{\mathrm{VTL}} \leqslant 40$ mph			$40 < v_{\mathrm{VTL}} < 55$ mph			$v_{\mathrm{VTL}} \geqslant 55$ mph		
	Mean	Std.Dev	# obs.	Mean	Std.Dev	# obs.	Mean	Std.Dev	# obs.
1	12.2	5.8	33	9.2	5.8	4	5.9	3.5	57
2	9.4	5.2	39	17.0	-	1	2.0	1.7	54
3	12.3	5.3	39	6.2	3.0	3	2.1	2.2	52
4	17.0	5.9	38	8.5	2.2	2	2.5	1.7	54
5	5.8	5.3	29	3.5	5.2	13	4.1	2.7	52
6	4.0	3.6	39	4.8	1.9	6	3.8	2.3	49
7	4.3	3.4	52	13.0	3.0	4	6.1	6.0	38
8	8.9	4.4	64	2.1	-	1	4.3	3.6	29
9	6.0	4.4	69	8.4	7.7	5	4.7	3.1	20
10	4.8	3.7	62	6.0	4.3	22	5.0	4.0	10
11	5.0	3.6	86	6.2	4.2	4	10.9	12.5	4
12	7.5	5.0	90	-	-	0	13.2	14.7	4
13	18.6	5.0	91	1.4	1.6	2	2.2	-	1
14	11.1	6.4	43	4.4	2.5	50	5.3	-	1
15	27.2	8.9	16	12.3	2.7	77	4.1	-	1
16	20.9	5.1	2	12.4	4.0	44	5.1	3.2	48
17	22.8	7.8	7	14.7	4.3	71	6.5	2.9	16

lower velocities. Differences observed among locations suggest that some detectors are either biased or not computing the velocity properly.

Finally, Figure 10 plots the VTL versus loop detector velocity measurements for all the observations collected at the 17 locations.

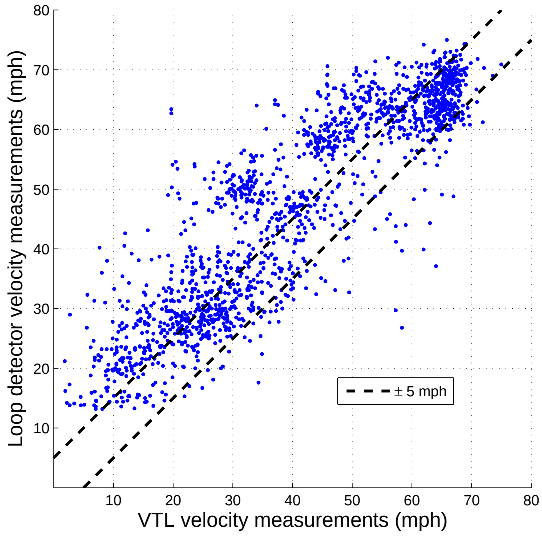

Figure 10: Loop detector vs. VTL velocity measurements (all locations). Dotted lines are the ± 5 mph thresholds.

For velocities below 40 mph, 31% of those observations have an absolute difference of less than 5 mph. This number reaches 70% for high velocities (over 55 mph). As mentioned before, loop detector velocity measurements tend to be higher than VTL measurements, which explain the smaller travel times computed with the loop detector velocity field – shown before in Figure 7.

Goal 3: Enforcement of a specific penetration rate of equipped vehicles during the experiment

Penetration rate of equipped vehicles refers to the proportion of equipped vehicles in the total flow. This proportion can be computed by placing VTLs on each of the 17 existing loop detector locations and dividing the VTL count by the loop detector count every 5 minutes.

During the experiment, penetration rate changes over time and space, as shown in the *penetration rate map* in Figure 11. Locations that are traveled by vehicles from the three loops – such as between Decoto Rd. (postmile 21) and Tennyson Rd. (postmile 26) in the morning and between Alvarado-Niles Rd. (postmile 23.3) and Tennyson Rd. in the afternoon – experience the highest proportion of equipped vehicles during the day, while locations at the ends of the section – such as between CA92 and Winton Ave. during the whole day

– are traveled by only one third of the equipped vehicles and thus present the lowest proportions during the day.

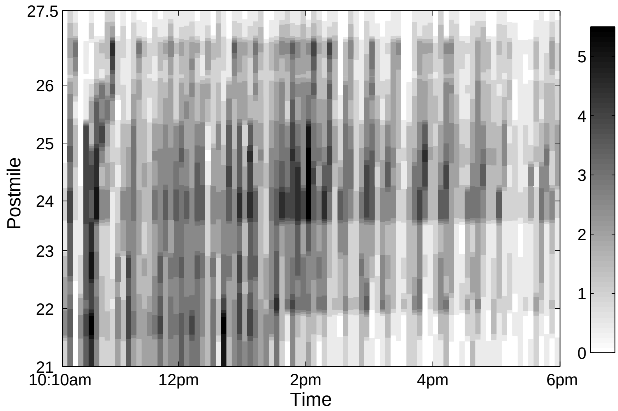

Figure 11: Penetration rate map (in %).

The penetration rates for locations between Decoto Rd. and Winton Ave. can be seen in Figure 12. Circles in part a) and c) of the figure represent the average penetration rate along the section of interest during the morning and the afternoon, respectively. The range corresponds to one standard deviation below and over the mean. The histograms in part b) and d) include the 17 locations for the morning and afternoon periods, respectively.

During the morning, less than 3% of the 5-minute periods have no observations, and in the afternoon that number goes down to less than 1%. In addition, 50% of the periods in the morning have a penetration rate of at least 2%, while in the afternoon only 35% of the periods meet this condition. This suggest that a continuous flow of equipped vehicles was achieved, which makes most of the 5-minute periods to contain at least one vehicle crossing each location.

5. Conclusions

The *Mobile Century* field experiment presented in this article was conceived as a proof of concept for a traffic monitoring system based on GPS-enabled mobile phones. The prototype system exploits the extensive coverage provided by mobile phones and the high accuracy in position and velocity measurements provided by GPS units. The sampling strategy proposed is based on the use of VTLs, and provides enough data for traffic monitoring purposes while managing the privacy of participants.

The experiment demonstrates the feasibility of the proposed system for real-time traffic monitoring, in which GPS-enabled mobile phones can be used as traffic sensors, providing their velocity at different points on the freeway.

The way in which the experiment was conceived allows the comparison of the velocity measurements collected by both VTLs and loop detectors, as well as the computation of the penetration rate achieved during the day.

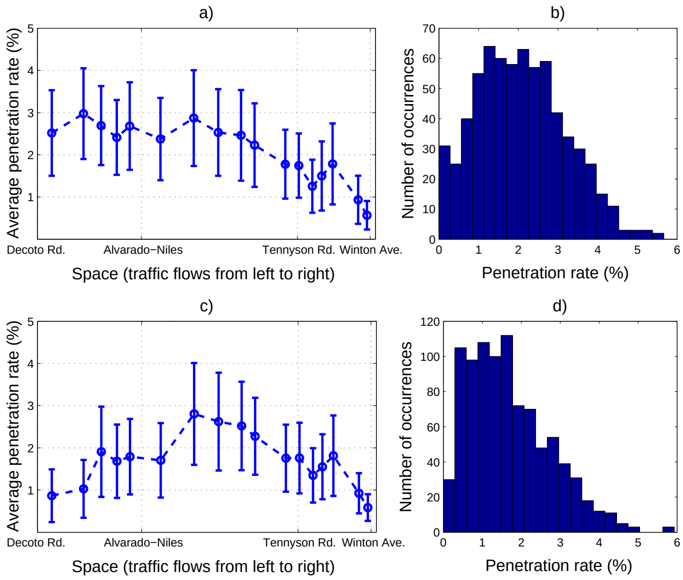

Figure 12: a) and c): Average penetration rate over time at existing detector station locations during the morning and the afternoon. The range is one standard deviation below and over the mean. Traffic flows from left to right. b) and d): Histogram of the penetration rate including all the 17 locations during the morning and the afternoon.

Ground truth is not known for this setting. Furthermore, the notion of speed, travel time is not unequivocal due to the heterogeneity of driver behavior on the highway. However, a comparison of travel times generated by VTL and loop detector velocity measurements suggests that VTL measurements are more likely to be closer to the actual velocity observed on the field. For this reason, loop detector data were not used as benchmark, and only a comparison with travel times was carried out to assess accuracy of the data. The comparison suggests the presence of some bias in the velocity estimation for some loop detectors, showing sometimes significant differences with the VTL measurements. Because of the different 5-minute aggregation methods used, VTL measurements exhibit more variability than loop detector measurements.

An average penetration rate between 2-3% was achieved during the experiment, which is viewed as realistic in the near future, considering the increasing penetration of GPS-enabled cellular devices. It is expected that GPS-enabled cell phones will penetrate the market rapidly in the near future, and the quality of measurements will increase with the evolution of GPS technology itself, thus opening new opportunities for smartphone-based monitoring systems.

In addition to the higher accuracy achieved with this technology, the proposed traffic monitoring system has other advantages over current systems based on loop detectors. From the standpoint of transportation agencies, the system comes at almost no installation and maintenance cost. Thus, a traffic monitoring system based on GPS-enabled mobile phones is particularly appropriate for developing countries, where there is a lack of resources and monitoring infrastructure, and the penetration of mobile phones in the population is significant[13] (and rapidly increasing).

Moreover, since the sensors are moving over the transportation system, a sufficient penetration of mobile phones would achieve an extensive spatio-temporal coverage of the network. Nokia, Navteq and UC Berkeley have now proceeded with a field operational test which extends this system to the urban network. The field operational test in the initial phase of the development – called *Mobile Millennium* – consists of the free distribution of traffic software such as the one presented earlier in this article to regular commuters, and the collection of traffic data (travel times mainly) during months, and will principally cover Northern California [26] in its initial phase.

A system that fuses both static (loop detectors) and mobile sensors (GPS-enabled mobile phones) is expected to provide a more accurate estimation of traffic than each of them individually, as suggested in [11]. Besides real-time traffic monitoring, the data collected could also be used for traffic state estimation and/or planning purposes. Eventually, if the amount of data received is large, modeling assumptions can be relaxed and replaced by data.

Finally, note that no processing was done to the raw data presented in this article beyond usual techniques to provide meaningful statistical features and displays. A traffic information system such as *Mobile Millennium* includes inverse modeling and data assimilation algorithms aimed at circumventing the potential deficiencies of data sets. Therefore, the potential errors, inaccuracies, and/or biases observed in the data will be addressed to compute travel time estimates or other features extracted from it as clearly as shown for the raw data, with the proper flow models of highway traffic and corresponding inverse modeling techniques.

Acknowledgements

The authors wish to thank Ken Tracton, Toch Iwuchukwu, Dave Sutter, and Murali Annavaram at Nokia Research Center Palo Alto, Baik Hoh and Marco Gruteser of Winlab at Rutgers University, and Christian Claudel of UC Berkeley for their invaluable contributions to develop, build, and deploy the traffic monitoring system implemented as part of the Mobile Century experiment. We thank the staff of the California Center for Innovative Transportation for the Mobile Century logistics planning and successful implementation. This research was supported by the Federal and California DOTs, Nokia, Center for Future Urban Transport at UC Berkeley (a Volvo International Center of Excellence), the Center for Information Technology Research in the Interest of Society (CITRIS), the Finnish Funding Agency for Technology and Innovation (Tekes), the

[13]The penetration rate of GPS phones will vary by country and the setting. However, emerging economies such as China and India are expecting rapid adoption of GPS technology due to cheap GPS enabled mobile phones [33].

National Science Foundation under contract CNS-0615299, and the US Department of Transportation through the Dwight David Eisenhower Transportation Fellowship Program.

References

[1] `http://pems.eecs.berkeley.edu/Public/`.

[2] J. Wright and J. Dahlgren. Using vehicles equipped with toll tags as probes for providing travel times. California PATH Working Paper UCB-ITS-PWP-2001-13, Institute of Transportation Studies, University of California, Berkeley, CA, 2001.

[3] X. Ban, Y. Li, A. Skabardonis, and J.D. Margulici. Performance evaluation of travel time methods for real time traffic applications. In *11th World Conference on Transport Research*, Berkeley, CA, June 2007.

[4] K.K. Sanwal and J. Walrand. Vehicles as probes. California PATH Working Paper UCB-ITS-PWP-95-11, Institute of Transportation Studies, University of California, Berkeley, CA, 1995.

[5] R. Zito, G. D'Este, and M. Taylor. Global positioning systems in the time domain: how useful a tool for intelligent vehicle-highway systems? *Transportation Research C*, 3(4):193–209, 1995.

[6] J. Kwon, B. McCullough, K. Petty, and P. Varaiya. Evaluation of PeMS to improve the congestion monitoring program. California PATH Research Report UCB-ITS-PRR-2007-6, Institute of Transportation Studies, University of California, Berkeley, CA, 2007.

[7] J. Moore, S. Cho, and A. Basu. Use of Los Angeles Freeway Service Patrol vehicles as probe vehicles. Technical report, Berkeley, CA, 2001.

[8] A. Schwarzenegger, D. E. Bonner, W. Kempton, and R. Copp. State Highway Congestion Monitoring Program (HICOMP), Annual Data Compilation. Technical report, Caltrans, Sacramento, CA, June 2008.

[9] R. Bertini and S. Tantiyanugulchai. Transit buses as traffic probes: empirical evaluation using geo-location data. *Transportation Research Record: Journal of the Transportation Research Board*, 1870:35–45, 2004.

[10] International Telecommunication Union. ITU World Telecommunication/ICT Indicators Database. http://www.itu.int/ITU-D/ict/statistics/ict/index.html, accesed on 09-23-2008.

[11] M. Westerman, R. Litjens, and J-P. Linnartz. Integration of probe vehicle and induction loop data – estimation of travel times and automatic incident detection. PATH Research Report UCB-ITS-PRR-96-13, Institute of Transportation Studies, University of California, Berkeley, CA, June 1996.

[12] J-L. Ygnace, C. Drane, Y.B. Yim, and R. de Lacvivier. Travel time estimation on the San Francisco Bay Area network using cellular phones as probes. California PATH Working Paper UCB-ITS-PWP-2000-18, Institute of Transportation Studies, University of California, Berkeley, CA, 2000.

[13] D. Lovell. Accuracy of speed measurements from cellular phone vehicle location systems. *Journal of Intelligent Transportation Systems*, 6(4):303–325, 2001.

[14] M. Fontaine and B. Smith. Investigation of the performance of wireless location technology-based traffic monitoring systems. *Journal of Transportation Engineering*, 133(3):157–165, March 2007.

[15] H. Bar-Gera. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research C*, 15(6):380–391, 2007.

[16] H. Liu, A. Danczyk, R. Brewer, and R. Starr. Evaluation of cell phone traffic data in Minnesota. In *Proceeding of the Transportation Research Board (TRB)* 87^{th} *Annual Meeting*, Washington D.C., January 22-26 2008.

[17] Smart Travel Laboratory. Wireless Location Technology-Based Traffic Monitoring Demonstration and Evaluation Project, Evaluation Final Report. Technical report, Center for Transportation Studies, University of Virginia, Blacksburg, VA, 2006.

[18] Airsage Inc. Airsage traffic data server application program interface. Technical Report Document No. /SDD 4001, Center for Transportation Studies, University of Virginia, 2006.

[19] Y.B. Yim and R. Cayford. Investigation of vehicles as probes using global positioning system and cellular phone tracking: field operational test. California PATH Working Paper UCB-ITS-PWP-2001-9, Institute of Transportation Studies, University of California, Berkeley, CA, 2001.

[20] Y.B. Yim. The state of cellular probes. California PATH Working Paper UCB-ITS-PRR-2003-25, Institute of Transportation Studies, University of California, Berkeley, CA, 2003.

[21] G. Rose. Mobile phones as traffic probes: practices, prospects, and issues. *Transport Review*, 26(3):275–291, 2006.

[22] Z. Qiu, P. Chen, J. Jing, and B. Ran. Cellular probe technology applied in advanced traveler information. *International Journal of Technology Management*. Accepted for publication.

[23] A. Krause, E. Horvitz, A. Kansal, and F. Zhao. Toward community sensing. In *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, St.Louis, MO, April 2008.

[24] A. Demers, G. F. List, W. A. Wallace, E. E. Lee, and J. M. Wojtowicz. Probes as path seekers: a new paradigm. *Transportation Research Record: Journal of the Transportation Research Board*, 1944:107–114, 2006.

[25] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J. Herrera, A. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *6th International Conference on Mobile Systems, Applications, and Services*, pages 15–28, Breckenridge, CO, June 17-18 2008.

[26] `http://traffic.berkeley.edu/`.

[27] D. Work and A. Bayen. Impacts of the mobile internet on transportation cyberphysical systems: Traffic monitoring using smartphones. *National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation, & Rail*, November 18-20, 2008.

[28] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46, 2006.

[29] J. Krumm. Inference attacks on location tracks. In *Fifth International Conference on Pervasive Computing (Pervasive 2007)*, Toronto, Ontario, Canada, May 2007.

[30] D. Work, O.-P. Tossavainen, S. Blandin, A. Bayen, T. Iwuchukwu, and K. Tracton. An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices. In *Proc. of the 47th IEEE Conference on Decision and Control*, pages 2141–2147, Cancun, Mexico, December 2008.

[31] Tony Judge and Matt Lewis. Gps mobile phones: the privacy and regulatory issues. Technical report, ARCchart Ltd, 2008.

[32] L. Edie. Discussion on traffic stream measurements and definitions. In *Proc. of the Second Int. Symp. on the Theory of Traffic Flow*, pages 139–154, Paris, France, 1965.

[33] World gps market forecast to 2013. Technical report, RNCOS, 2009.

A. Infering parameters from shockwave speed

This appendix shows how the information about the propagation of shockwaves – presented in Figure 4 – can be used to infer parameters of the fundamental diagram (assuming triangular relationship), as well as flows and densities that mobile sensors are not able to capture directly.

We start by assuming that a vehicle spans s_J feet when stopped at a traffic jam. That is, the jam density is $k_J = \frac{5280}{s_J}$ vpmpl. Because of physical considerations, this can be seen as a standard value for jam density. For instance, $s_J = 26$ ft (8 meters) yields a jam density around $k_J = 200$ vpmpl. The other two parameters needed to fully characterize the triangular fundamental diagram correspond to the free flow speed v_f and the shockwave speed w, which are obtained from the data.

The free flow speed corresponds to the speed of the vehicles before or after the incident ($v_f = 65$ mph in Figure 4). The shockwave speed is the speed of the second wave traveling upstream in Figure 4, which is $w = -15.6$ mph (the slope of the red line in the figure). With this information, we can conclude that the maximum flow and the critical density are around $q_{\max} = 2570$ vphpl and $k_C = 40$, respectively.

From the data, velocity in the queue can also be obtained. Most of the speeds range from 3 mph to 7 mph, although few vehicles with speed in the order of 12 mph can be found. The difference in the speed among vehicles can be explained by the lane used by each vehicle. An average value of $v_{\text{queue}} = 6$ mph can be used for the speed in the queue. Using the triangular fundamental diagram obtained before, the speed in the queue is sufficient to characterize the traffic state in the queue, which in this case correspond to $q_{\text{queue}} = 867$ vphpl and $k_{\text{queue}} = 144$ vpmpl. The flow is very close to the flow reported by PeMS using loop detectors, which is 850 vphpl.

This information can be used to infer the flow before the accident occurred using the Rankine-Hugoniot condition, which relates the speed of the shockwave u_s (inferred from the data) with the flow and density difference before and after the passing of the shockwave. Noting that the state after the passage of the shockwave corresponds to the state of the queue obtained before, we have:

$$u_s = \frac{q_{\text{queue}} - q_{\text{before}}}{k_{\text{queue}} - k_{\text{before}}} \tag{1}$$

Both the flow and density before the passage can be obtained using equation 1 and knowing that $q_{\text{before}} = k_{\text{before}} \cdot v_f$. In this case, and speed of the first shockwave is $u_s = -3.6$ mph, which yields a traffic state with a flow $q_{\text{before}} = 1300$ vphpl and $k_{\text{before}} = 20$ vpmpl. The flow obtained in this way is similar to the flow of 1100 vphpl collected with loop detectors before the accident.

Finally, if different values of jam density are tried, the results will change. For instance, for $k_J = 175$ vpmpl, $q_{\text{queue}} = 760$ vphpl and $q_{\text{before}} = 1140$ vphpl; for $k_J = 225$ vpmpl, $q_{\text{queue}} = 975$ vphpl and $q_{\text{before}} = 1465$ vphpl. However, the flows are still reasonably close to the flows measured with loop detectors. Considering that the flows are obtained using data from an unknown proportion of the total flow, the information is valuable.