Mobile Millennium - Participatory Traffic Estimation Using Mobile Phones

Ryan Herring, Aude Hofleitner, Dan Work,
Olli-Pekka Tossavainen, Alexandre M Bayen*

Bio: Alexandre M Bayen is an assistant professor with Systems Engineering program at the Department of Civil and Environmental Engineering, UC Berkeley. He is the principal investigator in the Mobile Millennium project.

Summary: This position paper describes how the mobile internet is changing the face of the transportation cyberphysical system at a rapid pace and what impact this has on urban travel. In the last five years, cellular phone technology has leapfrogged several attempts to construct dedicated infrastructure systems to monitor traffic. Today, GPS equipped smartphones are progressively morphing into a ubiquitous traffic monitoring system where users contribute and receive traffic information in real time. *Mobile Millennium* is a pilot project of such a technology which allows the general public with supported devices to participate. Its relevance for urban traffic and travel in urban environments is of specific interest, since it potentially will be able to unveil traffic patterns previously unobserved with dedicated monitoring infrastructure.

* **email** : bayen@berkeley.edu

1 Introduction

As part of the VOLVO funded research, our team investigated the possibility of integrating mobile probe data into traffic models, based on the assessment that with the advent of the mobile internet, and the emergence of web 2.0 type user generated content, the transportation engineering field would be one of the first to benefit from this new technology. Because of the novelty of this technology, and of the specifics linked with sampling phones, the project was first focused on highways. As part of the work performed for VOLVO, the team created algorithms capable of integrating probe data into highway traffic flow models, as a proof of concept of contributions to come specific to arterial networks.

Following the seed funding given by VOLVO, a team was assembled between a major cellular phone manufacturer (Nokia), the prime mapping company in the US (Navteq), government (California DOT and US DOT). The team built a traffic monitoring system using mobile devices, known as *Mobile Millennium*. The Mobile Millennium project [1], officially launched on November 10, 2008, is an early instantiation of participatory sensing in the form of a traffic monitoring system system which collects traffic data from GPS-equipped mobile phones to estimate traffic conditions in real-time. The traffic conditions are then broadcast back to the users' mobile phones, enabling commuters to make more intelligent route and trip decisions. The deployment area is focused on commuters in Northern California, including the San Francisco Bay Area and Sacramento, which are areas with heavy recurring congestion on many of the roadways. The project is a follow up of the *Mobile Century* experiment, in which 165 UC Berkeley graduate students were hired to drive a 10 mile loop of Interstate 880 in California for a day, demonstrating the feasibility of a real-time traffic estimation service using GPS enabled devices only [2].

Mobile Millennium significantly increases the scale and scope of this work by demonstrating the first real-time permanent monitoring system capable of using GPS data from thousands of mobile devices, as well as existing fixed traffic sensors such as inductive loop detectors embedded in the pavement, to construct velocity fields and travel time estimates. While the previous experiment focused on highway traffic estimation on a single segment of highway, Mobile Millennium aims to estimate traffic on all major highways in and around the target area, as well as on major arterial roads which achieve sufficient user penetration.

2 Mobile Millennium System Architecture

The system architecture which supports this research (shown in Fig. 1) consists of a physical component: GPS-enabled smartphones onboard vehicles (driving public), and three cyber components: a cellular network operator (network provider), cellular phone data aggregation and traffic service provision (Nokia/Navteq), and traffic estimation (Berkeley/Navteq). On each participating mobile device (or client), an application is executed which is responsible for collecting traffic data through a privacy aware spatial sampling technique based on *Virtual Trip Lines* (VTLs) [3], and displaying the current traffic estimates which are produced from the aggregate data of all participants.

A back end server aggregates data from a large number of mobile devices and pushes the data to UC Berkeley estimation engine for data assimilation, which combines the cell phone data with other information such as loop detectors to produce the best estimate of the current state of traffic. The map data server provides the Navteq Navstreets digital map data which is required for the network based traffic flow models. Multiple estimation algorithms are run in parallel as part of ongoing research, including arterial traffic models. An estimate manager in the traffic estimation server monitors the performance of the various algorithms and transmits the results to the traffic report server. The estimates are integrated with estimates from traffic models provided by Navteq before being transmitted back to the mobile device.

3 Highway Traffic Estimation

3.1 Review of Flow Model Based Traffic Estimation Algorithms

In the past various techniques to combine traffic flow models and the data collected from the highways into a state estimate have been proposed. *Kalman filtering* (KF) has been widely used for traffic state estimation in

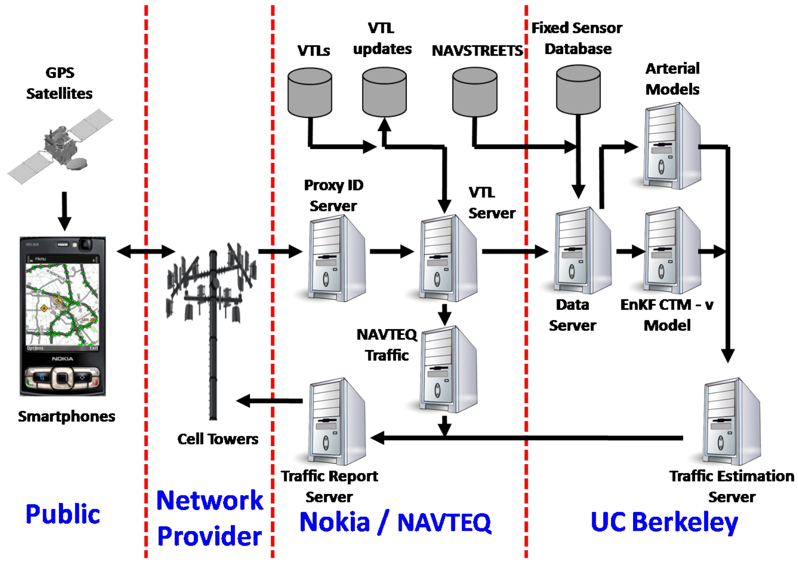

Figure 1: System architecture overview. The system consists of vehicles equipped with GPS-enabled smartphones, a cellular network provider, data collection infrastructure and traffic information provision, and traffic estimation algorithms.

earlier studies in its various forms. In [4], *mixture Kalman filtering* (MKF) was applied to the *Cell Transmission Model* (CTM) [5] to estimate traffic densities for ramp metering. The nonlinear CTM was transformed into a switching state space model, which enabled the use of a set of linear equations to describe the state evolution for the distinct flow regimes on the highway (e.g. highway is in free-flow or congestion). In [6], a Kalman filter was used to incorporate Lagrangian velocity trajectories into a density based CTM for highway traffic. A real–time algorithm for traffic estimation based on the *extended Kalman filter* (EKF) using a second order flow model was used in [7]. A key ingredient of this work is the differentiability of the numerical scheme employed for the second order model of traffic used by the authors, a feature which our model does not possess. Other treatments of traffic estimation include adjoint based control and data assimilation in [8, 9], *unscented Kalman filtering* (UKF) in [10] and *particle filtering* (PF) in [10, 11, 12].

A common feature for CTM based methods [6] described above is that the evolution of traffic state (typically density, not velocity) relies on a set of linearized equations which are needed in order to use the KF or EKF techniques. On the other hand, the PF technique is a nonlinear scheme for solving the Bayesian update problem, but has a higher computational cost. The approach proposed in the present work employs *ensemble Kalman filtering* (EnKF) [13], which enables the use of fully nonlinear evolution equations such as the discretization of the new flow model implemented in this article, while exploiting its linear observation equation. Unlike UKF, which uses a deterministic sampling technique, EnKF uses Monte Carlo integrations to maintain the nonlinear features of error statistics. Furthermore, by employing a fully nonlinear velocity evolution model, no highway mode selection algorithms or simplifications to the equations are needed in this work.

Earlier studies have specifically approached the highway traffic estimation problem using cell phone network data. In [14], an EKF was applied to a second order model of vehicle density and velocity, and validated in simulation. In practice, the modeling assumption that network providers can accurately provide both density and flow of the cellular phones currently on the highway of interest is limiting, especially in dense roadway networks. The work [15] uses a fully nonlinear particle filter to assimilate the mean velocity of a vehicle traveling between cell tower hand-off points, but also suffers from the same practical limitations in dense road networks.

3.2 Mobile Millennium Approach

The velocity field $v(x,t)$ on a highway segment $x \in [0,L]$ is a distributed parameter system in space, see Fig 2. Vehicles labeled by $i \in \mathbb{N}$ travel along the highway with trajectories $x_i(t)$, and measure the velocity $v(x_i(t),t)$ along their trajectories. These measurements are used to reconstruct the function $v(x,t)$, in a process referred to as *data assimilation* or *inverse modeling*. The technique used to perform data assimilation with this sampling is based on *Ensemble Kalman Filtering* (EnKF), which is applied to a discrete velocity evolution equation. Field experiments have been used to validate this method [2, 16].

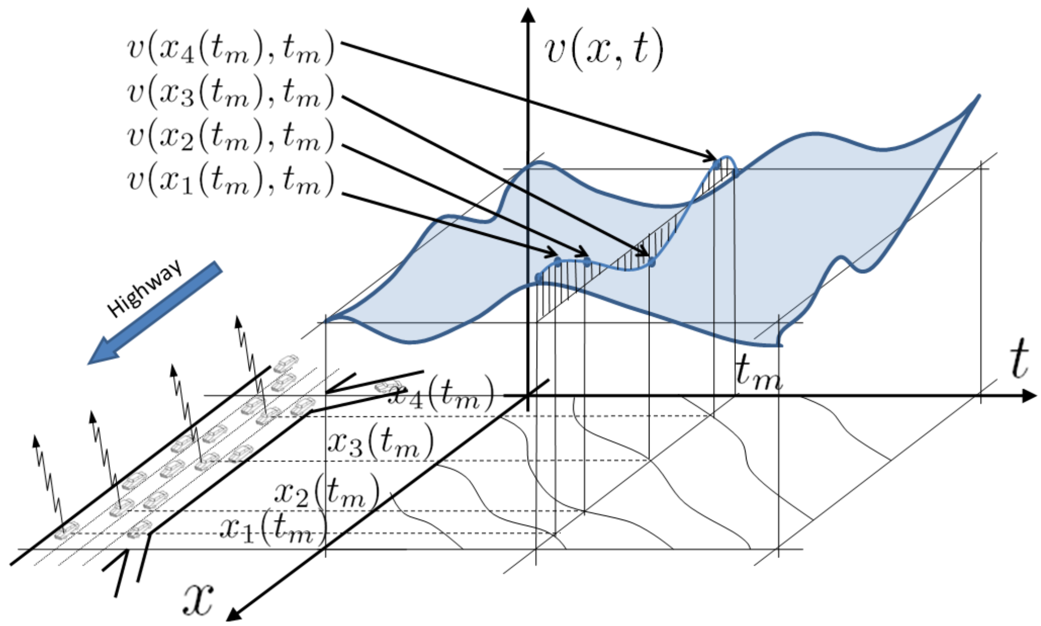

Figure 2: Illustration of the distributed velocity field $v(x,t)$ to be reconstructed from Lagrangian samples. Four samples $v_i(x_i(t),t)$ are shown at $t = t_m$, from vehicles i transmitting their data (indicated by up-arrows above the vehicles).

4 Arterial Traffic Estimation

4.1 Review of Arterial Travel Time Estimation Algorithms

In the past, various models have been developed to estimate arterial travel times, with the focus on signalized intersection delays. For example, statistical methods are proposed in [17, 18, 19], in which travel times are modeled as a linear combination of occupancy, flow, and signal parameters. Xie et al. [20] treat arterial link travel time as the summation of "cruise time" (*i.e.* free flow travel time) and signal delay. The cruise time is computed using detector speeds and the signal delay is estimated using a simplified intersection queuing diagram which requires basic signal parameters (such as cycle length, effective green time, flow/capacity ratio etc). An improved speed-flow relationship is developed in [21] which is shown to be effective to calculate arterial link travel times [22]. The above models are mainly for estimating average (or static) arterial travel times, and recent attention has been focused on estimating dynamic (or time-dependent) arterial travel times [23, 24]. In [23], link travel time is modeled as the summation of free flow travel time and signal delay, while the latter consists of three components: single vehicle delay, queuing delay, and over-saturation delay. In particular, the calculation of signal delay requires 30-second traffic volume and detailed signal timing parameters. By utilizing high-resolution (second-by-second) traffic signal events data (such as phase/timing changes) and vehicle actuation data, "virtual" vehicle trajectories are constructed in [24], which make it possible to estimate accurate dynamic arterial travel times. Finally, [25] formalizes the "intersection delay function" under free flow and over-saturation conditions. This function describes vehicle travel time through an intersection as a function of the time entering the road segment of interest.

In [26], the authors demonstrate how to estimate traffic conditions on arterial roads using GPS traces. This means that for some subset of the vehicles driving on the road, the position (latitude/longitude) is recorded every n seconds ($4 \leq n \leq 10$ in [26]). If full trajectory information is known about even a relatively small subset of the vehicles, then the authors show that traffic conditions can be estimated with very high accuracy. However, full trajectory information will likely never be available due to privacy concerns.

[27] is one of the few papers to address arterial traffic conditions in a machine learning framework. The authors use a Bayesian estimator and a neural network model to estimate the travel time along each link in the road network. Data from dual-loop detectors along each link of the network are used to learn the patterns along those links as well as the correlations between links. Dual-loop detectors record speed, flow, and occupancy using all vehicles traversing the link, which is data that we assume will not be available.

4.2 Mobile Millennium Approach

Arterial traffic is modeled in a machine learning framework that allows the system to learn the parameters of traffic while providing an inference algorithm at the same time. The arterial traffic estimation algorithm blends VTL data collected from the phones with Navteq historical data collected from fleet vehicles. At any given time, the real-time measurements cover only a fraction of the road network. These sparse measurements are aggregated over time and a probabilistic model is constructed to recognize traffic patterns. The real-time system then uses any current VTL measurements and the correlations between road segments to produce an estimate of the current travel time along all segments (including those with no current measurements). Road features are used to classify roads in order to reduce the number of distinct probability distributions required to be determined. Maximum likelihood estimation is used to determine the relevant weights for various features, which can then be used to infer the most likely state of the system given the real-time data.

5 Outcomes and future of the project

An example of the services provided by the project can be seen in Fig. 3. Traffic information is provided for both arterials and highways using the developed system. Currently the highway model covers the highways from south of San Jose, CA, to Sacramento and Lake Tahoe. Major highways that encounter severe congestion in this area are for example: Interstate 880, Interstate 80 and State Route 101. Arterial models cover the major roads (class 2 to 4 under the *Navteq* classification) of the San Francisco Bay Area. For privacy issues, small residential roads are not modeled. Given the multiple possible routes to drive from location A to location B within an arterial network, the display of severely congested intersections and bottlenecks enable the user to adapt his itinerary to the traffic condition and helps equilibrate the traffic flow over the network to reduce congestion, and thus reduce gas and time spent in traffic.

Future steps in this project include the creation of tools which will leverage the traffic estimation engine. In particular, we are interested in computation of guaranteed travel time, robust travel time and travel time reliability metrics. These computations will serve as the basis for routing engines which will run to provide optimal routing, guaranteed routing and robust routing in the network. Finally, the problem of trends (or short term forecast) will be investigated both for arterials and for highways, using machine learning techniques.

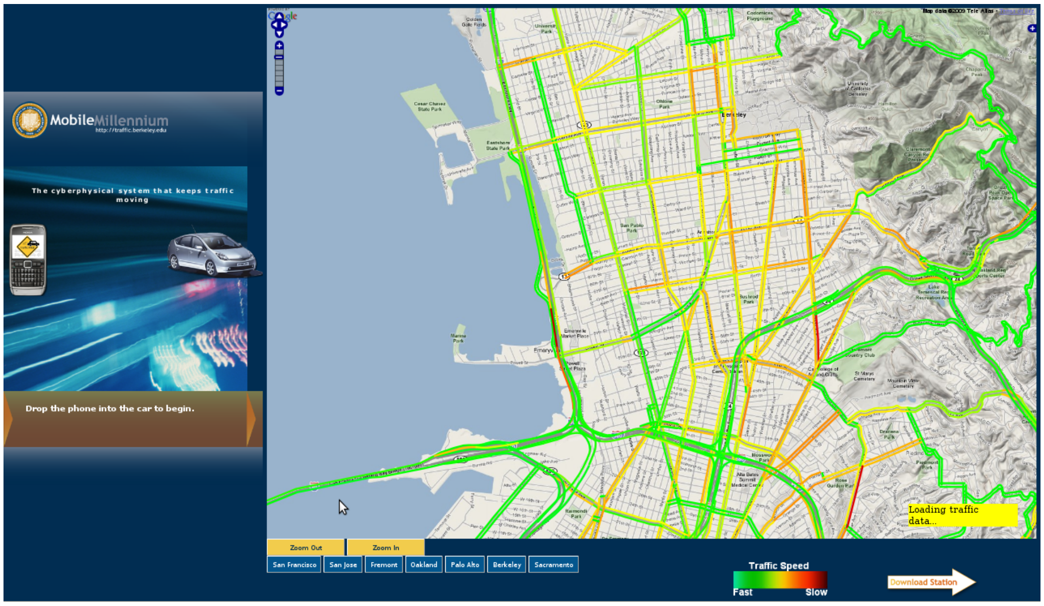

Figure 3: Screenshot of the Mobile Millennium Traffic Viewer.

References

[1] http://traffic.berkeley.edu/.

[2] 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*, (Cancun, Mexico), pp. 2141–2147, December 2008.

[3] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. 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*, (Breckenridge, CO), pp. 15–28, June 17-18 2008.

[4] X. Sun, L. Munoz, and R. Horowitz, "Mixture Kalman filter based highway congestion mode and vehicle density estimator and its application," in *Proc. of the American Control Conference*, vol. 3, (Boston, MA), pp. 2098 – 2103, 2004.

[5] C. F. Daganzo, “The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory,” *Transportation Research Part B*, vol. 28, no. 4, pp. 269–287, 1994.

[6] J.-C. Herrera and A. Bayen, “Traffic flow reconstruction using mobile sensors and loop detector data,” in 87^{th} *TRB Annual Meeting*, (Washington D.C.), Transportation Research Board, Jan. 12-17 2008.

[7] Y. Wang and M. Papageorgiou, “Real-time freeway traffic state estimation based on extended Kalman filter: a general approach,” *Transportation Research Part B*, vol. 39, no. 2, pp. 141–167, 2005.

[8] D. Jacquet, C. Canudas de Wit, and D. Koenig, “Traffic control and monitoring with a macroscopic model in the presence of strong congestion waves,” in *Proc. of the 44th IEEE Conference on Decision and Control, and European Control Conference*, (Sevilla, Spain), pp. 2164–2169, 2005.

[9] D. Jacquet, M. Krstic, and C. Canudas de Wit, “Optimal control of scalar one-dimensional conservation laws,” in *Proc. of the 25th American Control Conference*, (Minneapolis, MN), pp. 5213–5218, 2006.

[10] L. Mihaylova, R. Boel, and A. Hegyi, “Freeway traffic estimation within recursive bayesian framework,” *Automatica*, vol. 43, no. 2, pp. 290–300, 2007.

[11] J. Sau, N. El Faouzi, A. Ben Assa, and O. De Mouzon, “Particle filter-based real-time estimation and prediction of traffic conditions,” *Applied Stochastic Models and Data Analysis*, vol. 12, 2007.

[12] A. Hegyi, L. Mihaylova, R. Boel, and Z. Lendek, “Parallelized particle filtering for freeway traffic state tracking,” in *Proc. of the European Control Conference*, (Kos, Greece), pp. 2442–2449, July 2007.

[13] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*. Berlin Heidelberg: Springer-Verlag, 2007.

[14] A. Alessandri, R. Bolla, and M. Repetto, “Estimation of freeway traffic variables using information from mobile phones,” in *Proc. American Control Conference the 2003*, vol. 5, (Denver, CO), pp. 4089–4094, June 2003.

[15] P. Cheng, Z. Qiu, and B. Ran, "Particle filter based traffic state estimation using cell phone network data," in *Proc. IEEE Intelligent Transportation Systems Conference ITSC '06*, pp. 1047–1052, 2006.

[16] D. Work, O.-P. Tossavainen, Q. Jacobson, and A. Bayen, "Lagrangian sensing: Distributed traffic estimation with mobile devices." Accepted for publication, *ACC American Control Conference, 2009*, Saint Louis, MO.

[17] H. Gault and I. Taylor, "The use of output from vehicle detectors to access delay in computer-controlled area traffic control systems," Tech. Rep. Research Report No. 31, Transportation Operation Research Group, University of Newcastle upon Tyne, 1977.

[18] V. Sisiopiku and N. Rouphail, "Travel time estimation from loop detector data for advanced traveler information system applications," tech. rep., Illinois University Transportation Research Consortium, 1994.

[19] H. Zhang, "A link journey speed model for arterial traffic," *Transportation Research Record*, vol. 1676, pp. 109–115, 1998.

[20] X. Xie, R. Cheu, and D. Lee, "Calibration-free arterial link speed estimation model using loop data," *Journal of Transportation Engineering*, vol. 127, no. 6, pp. 507–514, 2001.

[21] A. Skabardonis and R. Dowling, "Improved speed-flow relationship for planning applications," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1572, pp. 18–23, 1997.

[22] H. Xiong and G. Davis, "Travel time estimation on arterials," in *Proceedings of the 87th Annual Meetings of Transportation Research Board (CD-ROM)*, 2008.

[23] A. Skabardonis and N. Geroliminis, "Real-time estimation of travel times on signalized arterials," in *16th International Symposium on Transportation and Traffic Theory*, pp. 387–406, 2005.

[24] H. Liu and W. Ma, "A virtual probe approach for time-dependent arterial travel time estimation," *Presented at the 87th Annual Conference on Transportation Research Board, and Submitted for publication*, 2008.

[25] X. Ban, R. Herring, P. Hao, and A. Bayen, “Delay pattern estimation for signalized intersections using sampled travel times,” in *2009 Transportation Research Board Annual Meeting*, (Washington, D.C.), 2009.

[26] J. Yoon, B. Noble, and M. Liu, “Surface street traffic estimation,” in *MobiSys ’07: Proceedings of the 5th international conference on Mobile systems, applications and services*, (New York, NY, USA), pp. 220–232, ACM, 2007.

[27] T. Park and S. Lee, “A bayesian approach for estimating link travel time on urban arterial road network,” *Lecture notes in computer science*, p. 1017, 2004.