Impacts of the Mobile Internet on Transportation Cyberphysical Systems: Traffic Monitoring using Smartphones

Daniel B. Work and Alexandre M. Bayen

Abstract— **This article describes how the mobile internet is changing the face of the transportation cyberphysical system at a rapid pace. 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 an ubiquitous traffic monitoring system, with the potential to provide traffic information in real time for the entire transportation network. Traffic information systems are one of the first instantiations of the potential of participatory sensing for large scale cyberphysical infrastructure systems. While mobile device technology is very promising, fundamental challenges remain to be solved, in particular in the fields of modelling and data assimilation.**

I. Traffic monitoring at the era of mobile internet services

Smartphones as sensors of the built environment. The convergence of communication and sensing on multimedia platforms such as smartphones provides the engineering community with unprecedented monitoring capabilities. Smartphones such as the Nokia N96 now include a video camera, numerous sensors (accelerometers, light sensors, GPS, microphone), wireless communication outlets (GSM, GPRS, WiFi, Bluetooth, infrared), computational power and memory. This phone can be used to listen to the radio, to watch digital TV, to browse the internet, to do video conferencing, to scan barcodes, to read pdfs, etc. The rapid penetration of GPS in smartphones is enabling device geopositioning and context awareness, which in turn is causing an explosion of *Location Based Services* (heavily relying on mapping) on the devices. For example, *Nokia Maps* display theaters and museums near the phone, *Google Mobile* provides driving directions from the phone location, and the *iPhone Travelocity* shows hotels near the phone. Due to their portability, computation, and communication capabilities, smartphones are becoming useful for numerous applications in which they act as sensors moving with humans embedded in the built infrastructure. Large scale applications include everything from population migration tracking and traffic flow estimation to physical activity monitoring for assisted living.

The competition for probe traffic data collection as a proxy for the larger war to conquer the mobile internet. In recent months, there have been increased levels of competition between cell phone manufacturers, network providers, internet service providers, computer and software manufacturers, and mapping companies. Following the transition from desktops to laptops to smaller and more portable devices, top companies in these industries are redefining themselves to remain relevant as the internet goes mobile. In the context of traffic monitoring, the examples below are eloquent and show the importance of information technology for transportation systems. In late 2007, Google made a move towards the phone industry with the launch of the Open Handset Alliance and the Linux-based Android platform (leading to the T-Mobile G1 Google phone). In part because of the pressure to use open platforms enhanced by the Google OS, Nokia, who manufactures 40% of the cell phones in the world, purchased Symbian, which licenses the operating system running on more than half of the smartphones in the world. Nokia then established the Symbian Foundation, with the intention of unifying the platform and making it open-source (Apple also partially opened its iPhone OS to software developers with the release of a software development kit). To strengthen its own mapping capabilities, Nokia bought Navteq, which is the largest mapping company in the world, following personal navigation device manufacturer TomTom's purchase of Tele Atlas, Navteq's chief competitor. Navteq in turn owns Traffic.com, one of the leading traffic data collection and broadcast companies. Its competitors include Inrix, which provides traffic data to Microsoft's web, desktop, and mobile applications.

New sources of traffic data for the transportation network. Highways have traditionally been monitored using static sensors, which include loop detectors built in the pavement, radars and cameras along the road, and more recently toll tag readers (such as FasTrak or EZ-pass), which can serve as probes wherever such infrastructure exists. While this infrastructure has proved to be efficient for highways, the costs of deployment, communication, and maintenance for such an infrastructure in the arterial network make it prohibitive for public agencies or companies to deploy on a global scale. To alleviate possible communication bottlenecks, on October 21, 1999, the *Federal Communications Commission* allocated 75MHz of spectrum as part of the US *Department of Transportation*'s (DOT) *Intelligent Transportation Systems* (ITS) US-wide program, with mostly traveler safety, fuel efficiency and pollution in mind. The first industry-government supported standard followed on August 24, 2001, when ASTM's E17.51 Standards Committee voted 20-2 to base *Dedicated Short Range Communication* (DSRC) on a modification of the IEEE 802.11a specification now named IEEE 802.11p.[1] At the same time, the US DOT launched a plan which included the deployment of around

[1] Source: Professor Raja Sengupta.

250,000 road side DSRC radios, but only led to around 100 radios deployed for the entire US to this day. This example highlights the difficulty of creating a dedicated system for the transportation network. At the same time, the need of monitoring traffic remains unsolved: if traffic information was available at a global scale including the arterial network, several problems could potentially be solved: (1) real-time congestion estimation of the arterial network; (2) re-routing of the highway traffic into arterial networks where beneficial; (3) optimized travel time, fuel efficient or emission optimal routes for commuters.

II. Impact of the mobile internet on the transportation CPS

Smartphones: a transformation from dedicated infrastructure to market-driven technology. The scale at which cell phones are produced, and the rate at which they integrate new technology is dramatic. The total number of cell phones worldwide exceeds three billion, with some European countries with a penetration of more than 150% (150 cell phones for 100 people). Nokia alone produces more than 17 phones a second, which means with the increasing penetration of GPS in the cellular phone fleet, cell phones will soon constitute one of the major traffic information sources available to the public. In North America and Europe, the overwhelming majority of commuters have a cell phone, potentially populating the entire arterial network with probe traffic sensors. Obviously, the use of cellular devices as traffic sensors has numerous benefits. (1) It is possible to leverage the market driven communication infrastructure already in place. (2) The spatio-temporal penetration of cell phones in the transportation network is increasing at an extremely fast pace. (3) The use of cell phones as traffic probes is device and carrier agnostic, leading to faster penetrations. (4) Major car manufacturing companies already have cradles and interfaces with cell phones (for example BMW and the iPhone) in their new cars so the sensing information gathered by modern cars can also be sent to such monitoring system.

Lagrangian vs. Eulerian information. While cellular phones provide an ideal bridge between the physical world (vehicle flows and dynamics on the road) and the information world (software systems monitoring the network), there is one major difference between the data collected by cell phones and traditional data, commonly used to estimate traffic in real time: the data collected by phones in cars is *Lagrangian*, i.e. gathered along cars trajectories, and not *Eulerian*, i.e. control volume based. This poses major challenges in building an information system for a cyberphysical infrastructure such as the transportation network. While a static loop detector or a camera (both Eulerian) can easily capture *all* vehicles going through the space monitored by the sensor, and therefore infer from it exhaustive quantities (flows, counts, local speed), a *Lagrangian* sensor can only monitor quantities following the vehicle, which does not give direct access to flows, counts, etc. In addition, measurements are only available where participating vehicles / phones are located. These are not predictable, and the local penetration of devices in the network might vary. This list of problems opens numerous research avenues with direct impact on technology development for traffic monitoring.

Modeling and computational challenges for monitoring the transportation CPS. As indicated in the name *cyberphysical*, the two key components of cyberphysical systems are "information" (cyber) and "constitutive laws" (modeling the physics of the system). Monitoring cyberphysical systems such as the transportation network poses two major challenges:

- *Distributed models for the transportation network.* Because GPS enabled phones sense velocity, or travel time between two consecutive GPS readings, constitutive models used to describe the evolution of the system need to incorporate these reading and bypass quantities which cannot be measured (density, flows, counts). The development of such *flow models*, for highways and arterials is still at its infancy. Techniques used for this include partial differential equations, queuing systems, and hybrid systems models of flow equations.
- *Machine learning models to circumvent lack of geographical information.* Mapping the entire US with an automated traffic monitoring system prevents the use of accurate knowledge of signage and traffic light presence, let alone cycle information. The presence of stops, lights, and their effect on traffic is not available from databases on a US-wide scale. Furthermore, they change too often to be incorporated in flow models. This difficulty has to be circumvented by machine learning algorithms capable of learning the flow features without knowledge of the detailed infrastructure, using for example clustering analysis.

Spatially aware sampling and privacy. At the heart of such a system, privacy by design sampling techniques must be used to prevent privacy invasion. In addition to proper anonymous data collection and encryption, sampling the vehicles at locations which are privacy safe is key to ensuring ongoing participation of the public which is needed for such a system. One possible architecture for preserving privacy is to collect data using a concept known as *Virtual Trip Lines* (VTLs), which are virtual geographic line segments deployed across roadways in the transportation network, triggering phones to collect and transmit data to the system. Defining optimal sampling strategies, which are privacy preserving is still a relatively unexplored field.

Real-time, online and robust availability. Unlike the more permanent Eulerian detectors, to which data quality, reliability and performance indices can be easily attributed, the penetration of cell phones at a given location and time is highly variable. Furthermore, in the coming few years before this type of monitoring becomes the standard, the participation of the public will be spatially and temporally varying. This means that the algorithms used for estimating the traffic must be robust to variability in penetration.

Inverse modeling and data assimilation. At the age of massive data collection, one of the most fundamental theoretical challenges associated with the reconstruction of traffic using mobile data will be the proper use of techniques to incorporate data into flow models or statistical models. The development of these techniques in fields such as oceanography or meteorology is relatively mature. For cyberphysical systems, in particular large scale infrastructure systems, the state of modeling, model inversion and computation is still at its infancy and promises significant breakthroughs in the near future.

III. Case study: Mobile Millennium

In order to study these new challenges, researchers from Nokia Research Center Palo Alto, Navteq, and UC Berkeley, with support from Caltrans and US DOT, have built a traffic monitoring system using mobile devices, known as *Mobile Millennium*. Spanning 18 months, the Mobile Millennium project will implement a state-of-the-art system to collect traffic data from GPS-equipped mobile phones and estimate traffic conditions in real-time. The study will focus on commuters in the San Francisco Bay Area, with specific emphasis on monitoring the heavy congestion experienced by travelers to and from the Lake Tahoe Ski Resorts one hundred fifty miles away. 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 I880 in California for a day, demonstrating the feasibility of a real-time traffic estimation service using GPS enabled devices only. *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 to construct velocity fields and travel time estimates, using the VTL sampling strategy. While the previous experiment focused on highway traffic estimation on a single segment of highway, *Mobile Millennium* will estimate traffic on all major highways in and around the Bay Area, Sacramento, and Lake Tahoe. New estimation algorithms will also be implemented for determining congestion on major arterial roads which have sufficient user penetration. Because this system will require public participation (with an initial target of 10,000 participants), the system must be scalable, require low bandwidth, and protect the privacy of its participants, while displaying accurate estimates on a software client which should be user friendly. The deployed system will be broadly characterized by five major components: GPS-enabled smartphones in vehicles (driving public), a cellular network operator, cellular phone data aggregation, traffic estimation and traffic service provision. On each participating mobile device, an traffic application is executed (see Fig. 1) which is responsible for both collecting traffic data through VTL sampling, and displaying the current traffic estimates which are produced from the aggregate data of all participants.

Fig. 1. *Mobile Millennium traffic client.* Traffic is displayed directly on the phone screen, with traffic information (construction, accidents, etc.), and audio traffic reports. For drivers who want to opt in, the phone can also send information to the system (Lagrangian measurements).

A back end server will aggregate data from a large number of mobile devices and push the data to UC Berkeley estimation engine for data assimilation, which will combine 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 will provide the Navteq Navstreets digital map data which is required for the network based traffic flow models. Multiple estimation algorithms will be run in parallel as part of ongoing research, including arterial traffic models. Finally, estimate managers at UC Berkeley and Navteq monitor the performance of the various algorithms and select the best estimate to transmit back to the mobile device.

Biographies

Daniel Work is a Systems Ph.D. Student in the Lagrangian Sensors Systems Laboratory at the University of California, Berkeley (Civil and Environmental Engineering). His research is focused on estimation, control, and optimization of transportation cyberphysical systems.

Alexandre Bayen received the Engineering Degree in applied mathematics from the Ecole Polytechnique, France, in July 1998, the M.S. degree in aeronautics and astronautics from Stanford University in June 1999, and the Ph.D. in aeronautics and astronautics from Stanford University in December 2003. He was a Visiting Researcher at NASA Ames Research Center from 2000 to 2003. Between January 2004 and December 2004, he worked as the Research Director of the Autonomous Navigation Laboratory at the Laboratoire de Recherches Balistiques et Aerodynamiques, (Ministere de la Defense, Vernon, France), where he holds the rank of Major. He has been an Assistant Professor in the Department of Civil and Environmental Engineering at UC Berkeley since January 2005.