

Towards Fine-Grained Urban Traffic Knowledge Extraction Using Mobile Sensing

Xuegang (Jeff) Ban
JEC 4034, CEE, Rensselaer Polytechnic
Institute
110 8th St, Troy, NY 12180
banx@rpi.edu

Marco Gruteser
WINLAB, Rutgers University
671 Route 1 South
North Brunswick, NJ 08902
gruteser@winlab.rutgers.edu

ABSTRACT

We introduce our vision for mining fine-grained urban traffic knowledge from mobile sensing, especially GPS location traces. Beyond characterizing human mobility patterns and measuring traffic congestion, we show how mobile sensing can also reveal details such as intersection performance statistics that are useful for optimizing the timing of a traffic signal. Realizing such applications requires co-designing privacy protection algorithms and novel traffic modeling techniques so that the needs for privacy preserving and traffic modeling can be simultaneously satisfied. We explore privacy algorithms based on the virtual trip lines (VTL) concept to regulate where and when the mobile data should be collected. The traffic modeling techniques feature an integration of traffic principles and learning/optimization techniques. The proposed methods are illustrated using two case studies for extracting traffic knowledge for urban signalized intersection.

Keywords

Urban Traffic Knowledge, Mobile Sensing, Location Traces, Traffic Theory

1. INTRODUCTION AND MOTIVATION

The recent proliferation of Global Positioning System (GPS) equipped vehicles and devices have led to the emergence and rapid deployment of *mobile traffic sensors*, which move with the traffic flow they are monitoring. Mobile sensors can collect detailed location traces of individual persons or vehicles, information that promises great advances in many science and engineering fields, including public health monitoring/diagnostics [19], extraction of personal or social behaviors [7] and mobility patterns [8], and transportation [12, 2].

In the transportation area, mobile sensing has recently motivated two important investigations, namely, *city-scale* transportation knowledge extraction and *fine-grained* urban traf-

fic knowledge extraction. The former concerns with large scale (i.e., city-scale) traffic congestion patterns such as travel times [26], routing [26, 25], social activity patterns [16], urban planning [28], land use [23], human mobility patterns [8], among others. Chapter 5 in [29] provides a summary of city-scale transportation knowledge extraction, focusing on possible patterns that can be extracted from location traces (trajectories). Other related critical issues are also discussed in [29] such as privacy concerns. On the other hand, fine-grained urban traffic knowledge extraction emphasizes on detailed, smaller-scale descriptions of urban traffic flow, such as traffic states and performances, as well as associated traffic operations and control. Applications for fine-grained urban traffic knowledge extraction can be broadly categorized as those for highways and urban arterials. Highway traffic can be generally modeled as continuous flow [11, 12, 24]. Arterial traffic however is often disrupted, e.g., by traffic signals, resulting in discontinuities and kinks in arterial traffic flow states. Such unique features can actually be utilized to reconstruct arterial traffic flow patterns such as delay (or travel times) [2] and queue lengths [1, 6], as well as signal timing information [9].

City-scale and fine-grained urban knowledge extraction represent, respectively, the macro-level and micro-level modeling of urban transportation systems. They are thus equally important for better understanding, describing, and managing the urban transportation systems. They share many commonalities, e.g., their methodologies require integration of data mining tools and some domain knowledge, privacy seems to be a concern to both areas, etc. Their also have distinctive differences, in terms of their specific application domains and the detailed study methodologies. In particular, city-scale applications require wide-area data coverage but may tolerate coarse resolution accuracy in time or space. For example, Taxi GPS data reported in a 2-5 minute interval can be used to mine routing and city-scale traffic conditions [26, 25]. Fine-grained urban knowledge extraction however requires much finer resolution especially in the time domain (second-by-second location traces is usually desirable) and relatively high penetration [1]. On the other hand, fine-grained applications usually need only smaller-area data coverage (e.g., signal performance modeling based on mobile data only needs location traces that cover a single intersection or several intersections).

In this paper, we focus on fine-grained urban traffic knowledge extraction to estimate real time traffic signal perfor-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UrbComp'12, August 12, 2012. Beijing, China.

Copyright 2012 ACM 978-1-4503-1542-5/08/2012 ...\$15.00.

mances using mobile sensing especially location traces. We show that the traffic knowledge that can be mined from mobile sensing is much richer than the traffic congestion and mobility pattern knowledge that has been the focus of most research so far for city-scale urban knowledge extraction. For example, it has been a long-standing challenge to collect traffic data in urban environment such as arterial signalized intersections. The primary source of data collection has been traditionally via fixed-location sensors such as loop detectors. However, the deployment of such detectors are limited and the cost to expand their coverage is prohibitive. For example, the New York City has over 95% of its total 12,225 traffic signals as pre-timed and no detector is deployed at most of these intersections [1]. As a result, recently the National Traffic Signal Report Card assigned the grade "F" (the lowest) to the detection and data collection system for traffic signals in the United States (NTOC, 2007). The wide deployment of mobile sensors will eventually allow collecting traces at low cost from a significant fraction of the population, which enables the estimation of detailed traffic system states, i.e., fine-grained urban traffic knowledge extraction, from individual drivers sharing their location/driving information via mobile sensors. This in turn can greatly benefit existing and emerging applications in urban traffic knowledge extraction, such as performance measurement of traffic signals and arterial networks.

Fine-grained urban traffic knowledge extraction using mobile sensors can be considered as a special form of Human Centric Sensing which needs to address a set of challenges [20]. In this paper we focus on two of them that are particularly important to transportation modeling applications: (i) the development of novel modeling methodologies to utilize the unique format of mobile data; and (ii) privacy protection. First, compared with fixed-location sensor data, data from mobile sensors has distinct features. Figure 1 illustrates the difference of the two data types in a time-space coordinate system at a signalized intersection. We show the red time duration and green time duration at the location of the intersection. Short line segments represent the fixed-location sensor data collected by the loop detector system which consists of a presence detector at the stop line and an advanced detector in the upstream link. Two long thick curves in the figure represent the trajectories obtained from the mobile sensors. As shown in the figure, fixed-location sensors collect traffic flow measures, such as volume, density, and speed, for all vehicles, but only at spatially discrete locations where sensors are deployed. Mobile sensors on the other hand can reveal detailed behaviors and provide (almost) spatially continuous trajectories of vehicles, but only for a sample of the traffic flow. Since part of the traffic flow is hidden, we cannot obtain accurate aggregated measures, such as traffic volume or density, from mobile sensors. As a result, existing modeling methods that work well for fixed-location sensors may not be directly applied to mobile data.

The unique characteristics of mobile data thus call for novel modeling approaches. Unfortunately, existing research on using mobile data (often called "probe data") in transportation is limited by very low penetration of such data. As a result, mobile data have been used mainly as a supplement to fixed location sensor data. Pure statistical analysis has been the main method to deal with mobile data [14], focusing on

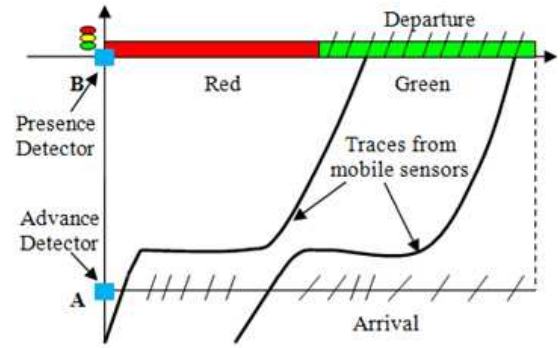


Figure 1: Comparison of mobile data and fixed-location-sensor data

applications to estimate average traffic flow or travel times [18]. Those previous approaches cannot fully capture and utilize the unique features of mobile data shown in Figure 1. In this paper, we focus on relatively large penetration of mobile data (e.g., larger than 10%) since we believe this will be the future trend. Under relatively high penetration, mobile data will play a dominate or at least equally important role (compared with fixed-location sensor data) for traffic data collection and modeling. This has the potential to transform current practice of urban traffic knowledge extraction. Privacy violation is another issue to concern. As mobile sensors can potentially reveal the complete traces of travelers that could contain sensitive location related information, the medical conditions, political affiliations or commercial secret may be inferred. Today the privacy issues of mobile devices are well realized by the public; see e.g., [22].

These two challenges call for an integrative framework to simultaneously consider modeling data needs and privacy preservation, i.e., to *co-design* privacy algorithms and modeling techniques so that the needs for urban traffic knowledge extraction and privacy protection can be both satisfied. This paper summarizes the authors' recent work in this area. We present the privacy protection scheme first in the next section. The new traffic modeling methods using privacy preserving mobile data are then presented and illustrated using case studies for fine-grained urban traffic knowledge extraction. We conclude the paper by discussing several important future research directions.

2. PRIVACY PROTECTION SCHEME

Our research pursues a privacy-by-design approach [5] to balance traffic modeling data needs and privacy protection when dealing with mobile data. This approach seeks to minimize the collection of personally identifiable information. To this end, we anonymize information by omitting any identifiers (such as names, equipment serial numbers, etc.) and further restrict the collection of location traces to limit re-identification risks. One aspect of restricting collection is to only record data in those locations where it is actually required for providing the service. To achieve this, we have introduced together with other colleagues Virtual Trip Lines (VTLs) [13], which define point locations on roadways where data should be collected, and VTL zones, which are areas along roadways where data is needed. Figure 2 shows an example VTL zone for an intersection monitoring application.

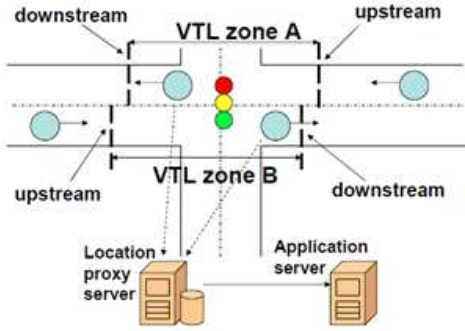


Figure 2: VTL mobile data collection system (source: [22])

It shows the VTL locations for the east-west street through dashed bold lines. Not shown are the VTL locations for north-south street, which would be part of a general solution. These VTL locations can be stored on the mobile sensing device. The device monitors its own location using GPS, for example, and will only report its speed and location trajectory to the location server between the upstream and downstream VTLs that demarcate the VTL zones. Notice that the exact VTL locations can be adjusted depending on application requirements. In this example VTL1 is placed farther away from the intersection than VTL2 in order to capture the queuing process of vehicles approaching the intersection. Since anonymous data can be often re-identified by correlating it with other data sources, the application server can apply further cloaking algorithms that filter the data to reduce risk as described in [13] or [27].

The use of these different types of privacy filtered data has been tested for both freeway modeling [11, 24] and urban arterial modeling [2, 1]. Sun et al. [21] further showed that the mobile data (such as travel times, short trajectories) collected via such a system can be properly used for traffic modeling applications such as real time estimation of vehicular queue lengths at a traffic signal. In this paper, we will illustrate this with the following two case studies that use the travel times of vehicles equipped with mobile sensors (called sample vehicles).

3. CASE STUDY I: DELAY PATTERN

Delay caused by traffic signals is the major source of delay in urban environment. Intersection delay pattern here refers to the experienced delay of a vehicle arriving at the intersection at any time. As shown in Figure 3, what we can actually measure in real world is the discrete delays or travel times from individual vehicles (shown as the circles on the piecewise linear curve in the bottom of the figure). Delay pattern is thus a continuous approximation of such discrete measurements. In fact, when people talk about intersection delays at a certain time (say 8 am), they never care if there is a vehicle actually arriving at the intersection at that particular time; indeed, they refer to the delay pattern of the intersection. To show how delay pattern can be estimated using mobile data, we show in Figure 3 a signalized intersection with two VTLs deployed upstream (VTL1) and downstream (VTL2). Under certain assumptions, we can use the bold solid triangles (or trapezoids) in the figure

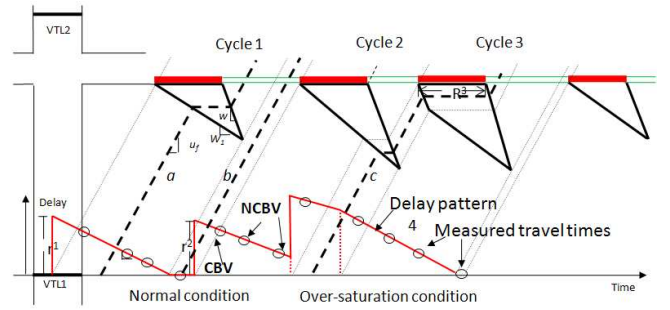


Figure 3: Intersection delay pattern

to represent how queue forms and dissipates based on the traffic shockwave theory [15, 17, 2]. The horizontal part of a triangle represents the duration of red time. As shown by the trajectories of vehicles (dashed lines), if a vehicle approaches the intersection in red time or if the queue length is not zero (e.g., trajectory a in the figure), the vehicle will join the end of the queue first and thus be delayed. The delay encountered by the vehicle is the horizontal part of the trajectory. Otherwise, if a vehicle arrives during green time and there is no queue (e.g., trajectory b), the vehicle will pass the intersection with no delay. By analyzing the geometry of the triangles, we can construct the theoretical delay pattern curve as shown in the bottom of Figure 3. The curve is piecewise linear and contains *critical points*, i.e., discontinuities and non-smoothness. Discontinuities indicate traffic signal changes (such as the start of the red time) and non-smoothness indicate traffic state changes (such as a queue is fully discharged).

Mobile data however cannot be used to construct directly the delay pattern; rather they provide samples of intersection delays, shown as circles along the delay curve in the figure. These sample delays, under proper penetration, can be used to identify the critical points of and further to estimate the delay pattern curves. In [2], this is done via a least square estimation algorithm to fit the sample travel times to the piece wise linear curves after grouping the samples into different cycles. Figure 4 shows the results of applying the estimation algorithm to a field test in the Bay Area in California [1]. In the figure, the asterisks along the piecewise lines (i.e., the estimated delay pattern) are the observed travel times (delay plus a constant minimum traverse time from VTL1 to VTL2) and the plus signs at the bottom are the errors. It is clear that the delay pattern can match well the observed samples. Knowing the pattern will help identify traffic conditions, e.g., over-saturation (i.e., vehicles cannot be fully discharged within a cycle) as indicated in the figure, or to estimate real time queue lengths [1].

The above analysis underlines the most salient feature of the new modeling method, i.e., a proper integration of traffic principles and learning/optimization techniques. For the delay pattern estimation here, the *knowledge* is based on the traffic flow theory that describes the delay pattern as piecewise linear curves whose critical points (discontinuities and non-smoothness) have clear physical meanings. On the other hand, the *learning/optimization techniques* is least square estimation that helps estimate the key parameters of such

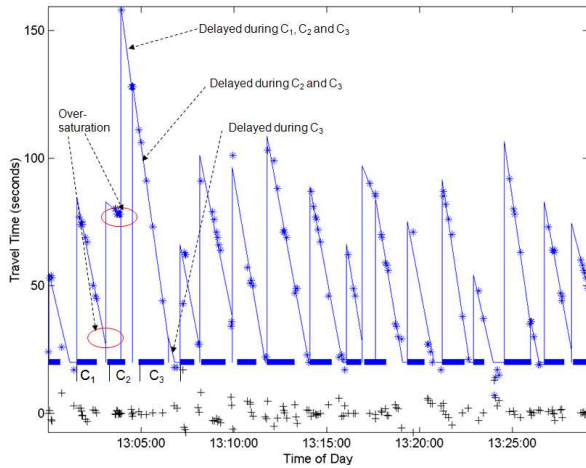


Figure 4: Delay pattern estimation of a field test (Source: [2])

patterns from mobile data and ultimately reconstruct the patterns. These two components need to be integrated in a holistic manner, which however may be different for different applications, as shown in the next case study.

4. CASE STUDY II: SIGNAL TIMING

Mobile data, especially intersection travel times, can be used to estimate the timing parameters of urban traffic signals (such as the cycle length, number of phases, and cycle by cycle red and green times), which are important for traffic signal operations and signal/arterial performance measurement. It has been for long assumed that such parameters should be available input, e.g., from transportation management agencies such as departments of transportation, to traffic models. In fact, collecting signal timing parameters directly from the agencies is probably trivial for small scale data collection (such as for a few signals). However, collecting such information this way for large areas (such as a region or nation-wide) can be very challenging and time consuming due to many possible technical and institutional hurdles. On the other hand, many traffic information providers have started to collect increasingly large amount of mobile data. Therefore an alternative way is to infer the signal timing information directly from the data that have already been collected such as travel times, probably with the help of limited (and easily obtained) knowledge about traffic signals.

[9] developed a robust signal timing estimation method based on intersection travel times. The method is again featured by a combination of traffic flow theories and learning/ optimization methods, which can estimate the exact cycle start/end times. The method contains three major steps: cycle breaking, exact cycle boundary detection, and effective red (or green) time estimation. Cycle breaking determines whether a new cycle starts by applying the support vector machine (SVM) to identify travel time samples that indicate the starts of red times. The exact cycle boundary estimation detects the exact cycle start/end times. It can be formulated as a nonlinear program by assuming that the cycle length is constant (the effective red and green times may

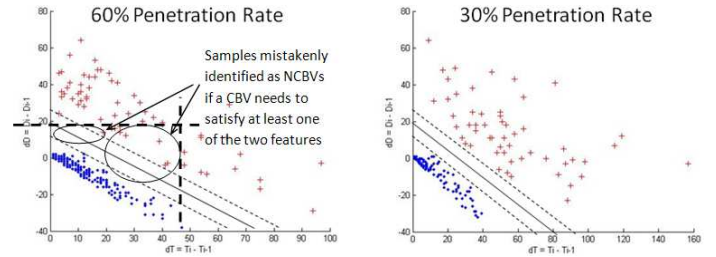


Figure 5: Cycle breaking (source: [7])

vary from cycle to cycle, which covers a large portion of existing traffic signals, e.g., in the US). The method can further detect the number of missing cycles using sample delays and the SVM results. The effective red (or green) time estimation calculates the duration of effective red (or green) times. This is done via using delay patterns by investigating when non-smoothness in the delay pattern happens. Due to space limitation, we only present in this paper how the cycle breaking is done via the use of SVM.

We first define, as shown in Figure 3, a cycle breaking vehicle (CBV) as the first sample vehicle in a cycle. The other vehicles in this cycle are defined as non-cycle breaking vehicle (NCBV). Note that the CBV of a cycle is not necessarily the first vehicle actually arriving at the signal in the cycle if the penetration is not 100% (in this case, the first vehicle may not be sampled). The CBVs usually have higher delays as shown in Figure 3. In [2], this feature is used to detect whether a new cycle starts by defining a threshold: if the delay increase from one vehicle to the next vehicle exceeds this threshold, a new cycle starts. The results however are not reliable due to oscillation and noise in measurements, and especially low penetration rate of mobile data. In [9], the SVM model applies two features: the arrival time difference $t_i - t_{i-1}$ and the delay difference $\tau_i - \tau_{i-1}$ between two consecutively sampled vehicles. Here t_i is the i th sample vehicle's arrival time at VTL1 and τ_i is the intersection delay of the i th sample vehicle. The second feature is exactly what was used in [2]. Figure 5 depicts these two features for a field test under 60% and 30% penetration rate of travel time data [9]. In the figure, dots are for NCBVs and plus signs are for CBVs. We can see that there is a clear margin of separation between CBVs and NCBVs using these two features. Using either feature or a simple combination of the two features however is not effective. In Figure 5, the vertical dashed bold line indicates the threshold in delay increase; the horizontal dashed bold line indicates the threshold in arrival times. The figure shows that even both measures are used (e.g., a CBV needs to satisfy at least one of the two measures), there will be still large errors for mis-identification, as those indicated by the circles.

SVM can combine the two features in a more intelligent way. To show how the SVM model can be developed, let the historical travel time data be denoted by $(x_i, y_i), i = 1, \dots, M$, where $x_i = (t_i - t_{i-1}, d_i - d_{i-1})^T$ is a data point and $y_i = 1$ is the corresponding label ($y_i = 1$ for CBV and $y_i = -1$ for NCBV). SVM divides the data set into two groups: one for $y_i = 1$ and the other for $y_i = -1$. It can further produce two support planes (lines in the R^2

Simulation			
	Observation	Estimation	Absolute error
Start of red time in the first cycle (s)	55904.250	55905.720	1.470
Cycle length (s)	55.000	55.003	0.003
Effective red time length (s)	31.000	30.210	0.790
RMSE by the proposed method (s)	1.632		
RMSE by method of [2] (s)	9.593		
NGSIM			
	Observation	Estimation	Absolute error
Start of red time in the first cycle (s)	98.000	99.916	1.916
Cycle length (s)	100.000	100.370	0.037
Effective red time length (s)	68.000	70.210	2.210
RMSE by the proposed method (s)	3.077		
RMSE by method of [2] (s)	8.249		

Figure 6: Cycle breaking results: Simulation and NGSIM datasets (source: [7])

space) for such separation as depicted in Figure 5. Let $w = (w_1, w_2)^T \in R^2$ and b be a scalar. If w and b are properly selected, we will have $wx_i - b \geq 1$ for $y_i = 1$ and $wx_i - b \leq -1$ for $y_i = -1$. Then the two support lines are $wx_i - b = 1$ and $wx_i - b = -1$. The distance between these two lines can be shown as $2/||w||$ with $||w||$ denoting the norm of w . If we aim to maximize the distance between these two support planes, (w, b) can be determined by solving the following SVM problem [9]:

$$\min_{w,b} \quad 1/2||w||^2 + G \left(\sum_{i=1}^M \varepsilon_i \right) \quad (1)$$

$$\text{subject to} \quad y_i(wx_i - b) \geq 1 - \varepsilon_i, i = 1, \dots, M, \quad (2)$$

$$\varepsilon_i \geq 0, i = 1, \dots, M. \quad (3)$$

Here ε_i is the error term for cases where the classes cannot be perfectly divided. G is the weight factor assigned to the error terms in the overall objective function; a larger G assigns a larger penalty to the error. Solving the above SVM model is usually not computationally demanding since it is a convex, quadratic program; see [4, 3]. After solving the SVM model, the (w, b) pair, in particular, the two planes (lines): $wx_i - b = 1$ and $wx_i - b = -1$, can be used to identify whether a given data sample x_j is a CBV or NCBV. Figure 6 shows the results of applying the SVM model to break cycles using the datasets of simulation and NGSIM (next generation simulation, widely used for traffic research) [9], based on which to estimate the cycle boundaries and lengths. Compared with the previous method in [2], the root mean square error (RMSE) of the results is much improved: from 9.6 seconds to 1.6 seconds for the simulation data, and from 8.2 seconds to 3.0 seconds for the NGSIM data.

Figure 7 shows the signal timing estimation results by applying the three step method for the NGSIM dataset. There are in total 9 cycles in the dataset and is no sample for the second and third cycles. The figure shows that the two

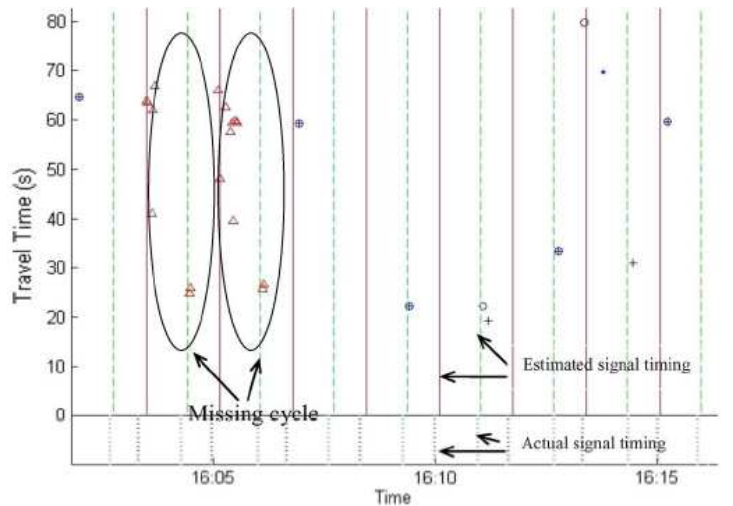


Figure 7: Signal timing estimation results: NGSIM dataset (source: [7])

missing cycles can be correctly detected by the three step method. The estimated start times of red are plotted using vertical solid lines and the estimated start times of green are plotted using vertical dashed lines. The actual starts of red and green times are also shown in the figure using dotted lines. We can see that the estimated cycle boundaries and red and green times match well with those from field observations, indicating that the method works reasonably well. Same conclusion can also be obtained for the simulation data which is omitted here.

The above analysis underlines again the unique feature of the new modeling method that integrates traffic principles (in this case, how delay changes within a cycle and across cycles) with learning/optimization methods (in this case, SVM and nonlinear programming). It also highlights the importance of learning techniques in such a process: the knowledge (i.e., the two features for cycle breaking) has to be used intelligently since simple use of them may not work well (as illustrated by the plot to the right in Figure 5). On the other hand, traffic knowledge (such as the signal cycle length is fixed and the vehicle delay after signal turns red will “jump”) is also critical. After all, we are dealing with a physical traffic system with control devices and vehicles/drivers. Traffic knowledge usually represents some important physical phenomena or characteristics of the system and thus needs to be properly respected and integrated into the learning/optimization models to produce meaningful results.

5. DISCUSSIONS AND FUTURE RESEARCH

We presented in this paper our recent work on co-designing privacy protection algorithms and traffic modeling methods for fine-grained urban traffic knowledge extraction using location traces from mobile sensing. The privacy protection scheme is based on the VTL concept to regulate the collection of mobile data. The traffic modeling method is featured by a combination of traffic principles and learning/optimization techniques. Two case studies were also presented to demonstrate the proposed methods.

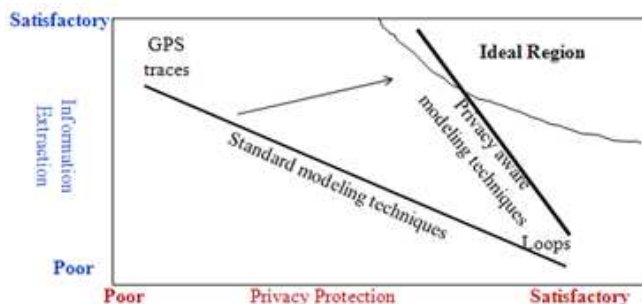


Figure 8: Privacy-aware modeling paradigm

The development of the VTL concept is based on a close collaboration between location privacy experts and transportation researchers, with the aim to properly balance the needs of both traffic modeling (i.e., traffic knowledge extraction) and privacy protection. It clearly indicates that for many traffic/transportation applications, data needs for modeling do not necessarily have to be compromised to ensure privacy protection. An application-aware design of privacy algorithms can retain features important for the application, while still achieving privacy by removing features that are less important. The key is the close collaboration between the two research communities of privacy protection and transportation to simultaneously consider privacy protection and application needs: being aware of the effects of applying privacy schemes to data when developing modeling methods, and being aware of data needs when designing privacy preserving mechanisms. As the future paradigm will likely shift from traditional sensors (such as loops) to mobile sensors as shown in Figure 8, such collaboration is critical to transform the current modeling techniques with no or little privacy consideration to a new paradigm for privacy-aware modeling techniques to satisfy requirements of both data needs for modeling and privacy protection.

The proper integration of transportation principles and data-mining tools is important to develop new mobile-data-based traffic modeling methods, especially when the penetration of mobile data reaches certain “critical mass.” In this case, the mobile data can present patterns that reveal traffic system control or state changes (such as red time starts or queue disappears, as we discussed in the delay pattern estimation section). Knowledge about these patterns (many of which are indeed well-known in the application domain) are crucial since they can provide guidance to the learning/optimization models to focus on the most relevant, important features. Applying advanced learning/optimization techniques on the other hand is also critical to obtain accurate, robust estimation of the patterns, and further system control and states.

Fine-grained urban traffic knowledge extraction using mobile sensing is an emerging area and many challenges still remain. Below we summarize the limitations of our current research and highlight some challenges for future research.

(a) Our current focus is on urban intersections which are a crucial component for urban traffic. The next step is to expand it to model arterial corridors (consist of a number of intersections) and networks. The important

issue for such expansion is to capture the interactions among different intersections, which can be revealed by vehicle platooning. The key challenge therefore is to study how vehicle platoons form and disperse at intersections using mobile data.

- (b) The presented models in this paper are deterministic. Since traffic is random in nature, we need to capture such stochasticity in the modeling process. This is particularly important when modeling urban corridors or networks (such as to study the vehicle platooning process). We recently developed Bayesian Network based statistical learning methods to model arterial traffic flow [10]. The method will be expanded for large scale arterial applications.
- (c) The VTL based privacy protection algorithm is suitable for modeling urban intersections. Our research also indicates that specific privacy techniques may need to be developed for different urban applications such as origin-destination demand estimation, urban congestion pricing, among others. For this, a comprehensive framework is needed. It should contain a suite of privacy techniques that can be used/tailored depending on specific urban applications. The authors are working on this topic and results will be reported in subsequent papers.

6. ACKNOWLEDGMENTS

The authors thank the comments of two anonymous reviewers which improve the quality of an early version of this paper. The authors greatly appreciate their collaborators and students who worked with them on this topic over the last several years. Many technical details related to the case studies presented in this paper can be found in the joint papers with those collaborators. This research is partially funded by National Science Foundation grants: CMMI-1031452, CMMI-1055555, CMMI-1031400, CNS-0845896; US Department of Transportation (USDOT); California Department of Transportation (Caltrans); and New York State Department of Transportation (NYSDOT). Any opinions, findings, and conclusions or recommendations in this paper are those of the authors and do not necessarily reflect the views of any of the above funding agencies.

7. REFERENCES

- [1] X. Ban, P. Hao, and Z. Sun. Real time queue length estimation for signalized intersections using sample travel times from mobile sensors. *Transportation Research Part C*, 19(6):1133–1156, 2011.
- [2] X. Ban, R. Herring, P. Hao, and A. Bayen. Delay pattern estimation for signalized intersections using sampled travel times. *Transportation Research Record*, 2130:109–119, 2009.
- [3] K. Bennett and C. Campbell. Support vector machines: Hype or hallelujah? *SIGKDD Explorations*, 2(2):1–13, 2000.
- [4] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.
- [5] A. Cavoukian. Privacy by design. Technical Report Internet Link: <http://www.privacybydesign.ca/content/uploads/2010/03/PrivacybyDesignBook.pdf>

- (Accessed on June 20, 2012), Information and Privacy Commissioner of Ontario Canada, 2009.
- [6] Y. Cheng, X. Qin, J. Jin, and B. Ran. An exploratory shockwave approach to estimating queue length using probe trajectories. *Journal of Intelligent Transportation Systems*, 16(1):12–23, 2012.
 - [7] T. Choudhury, G. Borriello, and et al. The mobile sensing platform: an embedded system for activity recognition. *IEEE Pervasive Magazine*, 7(2):32 – 41, 2008.
 - [8] M. Gonzalez, C. Hidalgo, and A. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(Jun 5):779–782, 2008.
 - [9] P. Hao, X. Ban, K. Bennett, Q. Ji, and Z. Sun. Signal timing estimation using intersection travel times. *IEEE Transactions on Intelligent Transportation Systems*, 13(2):792–804, 2012.
 - [10] P. Hao, D. Guo, X. Ban, Q. Ji, and Z. Sun. Vehicle index inference for signalized intersections using sample travel times. In *Proceedings of the 91st Transportation Research Board Annual Meeting*, Washington, DC, U.S.A., January 2012.
 - [11] J. Herrera and A. Bayen. Incorporation of lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B*, 44(4):460–481, 2009.
 - [12] J. Herrera, D. Work, R. Herring, X. Ban, and A. Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: the mobile century field experiment. *Transportation Research Part C*, 18(4):568–583, 2010.
 - [13] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J. Herrera, and A. Bayen. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *The Sixth Annual International conference on Mobile Systems, Applications and Services (MobiSys 2008)*, Breckenridge, U.S.A., June 2008.
 - [14] 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, U.S.A., 2008.
 - [15] M. Lighthill and J. Whitham. On kinematic waves. i: Flow movement in long rivers; ii: A theory of traffic flow on long crowded roads. In *Proceeding of Royal Society A*, volume 229, pages 281–345, 1955.
 - [16] S. Networks. <http://www.sensenetWORKS.com/products/macrosense-technology-platform/citysense>.
 - [17] P.I.Richards. Shock waves on the highway. *Operation Research*, 4:42–51, 1956.
 - [18] C. A. Quiroga and D. Bullock. Travel time studies with global positioning and geographic information systems: an integrated methodology. *Transportation Research Part*, 6(1-2).
 - [19] J. Sriram, M. Shin, T. Choudhury, and D. Kotz. Activity-aware ecg-based patient authentication for remote health monitoring. In *Proceedings of the International Conference on Multi-Modal Interfaces (ICMI-MLMI)*, pages 297–304, November 2009.
 - [20] M. Srivastava, T. Abdelzaher, and B. Szymanski. Human-centric sensing. *Phil Trans.R.Soc.*, 370, ser. A:176–197, 2012.
 - [21] Z. Sun, B. Zan, X. Ban, M. Gruteser, and P. Hao. Evaluation of privacy preserving algorithms using traffic knowledge based adversary models. In *Proceedings of the 14th IEEE conference on Intelligent Transportation Systems*, Washington, DC, U.S.A., 2011.
 - [22] N. Y. Times. <http://www.nytimes.com/2011/04/21/business/21data.html>.
 - [23] J. Toole, M. Ulm, D. Bauer, and M. Gonzalez. Inferring land use from mobile phone activity. In *In Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, 2012.
 - [24] D. Work, S. Blandin, O. Tossavainen, B. Piccoli, and A. Bayen. A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress*, 2010(1):1–35, 2010.
 - [25] J. Yuan, Y. Zheng, X. Xie, and G. Sun. Driving with knowledge from the physical world. In *In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 316–324, 2011.
 - [26] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. T-drive: Driving directions based on taxi trajectories. In *In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 99–108, 2010.
 - [27] B. Zan, Z. Sun, M. Gruteser, and X. Ban. Vtl zone-based path cloaking algorithm. In *Proceedings of the 14th IEEE conference on Intelligent Transportation Systems*, Washington, DC, U.S.A., 2011.
 - [28] Y. Zheng, Y. Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *In Proceedings of the 13th international conference on Ubiquitous computing*, pages 89–98, 2011.
 - [29] Y. Zheng and X. Zhou. *Computing with spatial trajectories*. Springer, 2011.