

Prediction of Transit Vehicle Arrival Time for Signal Priority Control: Algorithm and Performance

Chin-Woo Tan, Sungsu Park, Hongchao Liu, *Member, IEEE*, Qing Xu, and Peter Lau

Abstract—We develop an algorithm for predicting the arrival times of a transit vehicle at signalized intersections, with a focus on meeting the accuracy requirement associated with signal priority control applications. The algorithm uses both historical and real-time Global Positioning System (GPS) vehicle location data. There are no data from other detectors, such as loops or cameras. The arrival time prediction is formulated as an optimal *a posteriori* parameter estimation problem, where the model is consisted of a historical model and an adaptive model that adaptively adjusts its filter gain based on real-time data. The estimates generated by these two models are fused in a weighted average derived from the solution of the parameter estimation problem. The prediction algorithm adaptively adjusts its weight distribution using error variances obtained from the two models. We include some simulations of field test results and their statistics to demonstrate the performance and convergence of the solution.

Index Terms—Error convergence, historical and real-time adaptive models, intersection arrival time prediction, signal priority, vehicle location data.

I. INTRODUCTION

RECENT advances in Intelligent Transportation Systems and transit vehicle technologies, as well as other customer service innovations, provide major opportunities to improve transit vehicle service and reducing overall traffic congestion. Bus Rapid Transit (BRT) has been viewed as an important application of these technological and operational innovations. A promising BRT concept is a transit signal priority (TSP) system. This is an operational strategy that facilitates the movement of transit vehicles through a signal-controlled intersection by modifying the normal signal operation process [1]–[4]. It aims to minimize travel times through signalized intersections while limiting the impact on the rest of the traffic along the corridor and maintaining pedestrian safety [5], [6]. By reducing the

transit intersection delay time, a TSP system can reduce travel time and improve transit service reliability, thus increasing quality of service.

Signal priority can be implemented in a variety of ways, such as passive priority, early green (red truncation), green extension, actuated transit phase, phase insertion, and phase rotation. A major challenge in designing signal priority control algorithms is that *different* strategies need to be executed to efficiently provide priority with respect to when in the signal cycle the transit vehicle will arrive at the intersection. A critical issue in this design problem is the ability to predict vehicle arrival times at intersections as well as “optimal” times to place priority requests [7]. A more “efficient” or “intelligent” priority scheme involves collection of “better” vehicle location information and execution of “better” controls that adapt to traffic fluctuations. The *intelligence* of a TSP system includes a travel time prediction algorithm that anticipates the arrival of a transit vehicle at a traffic signal and gives priority to the vehicle to minimize its delay at the intersection and impact on nonpriority traffic and ensure pedestrian safety. Efficient adaptive signal control critically depends on the availability and accuracy of vehicle arrival time prediction [7], [8].

Our objective is to design a reliable and accurate real-time *arrival time prediction algorithm* and integrate it with signal priority control for deployment. The *accuracy requirement* associated with signal priority application is that the prediction time error must quickly converge to stay within a bound around zero. Such confidence in the prediction allows a *sufficiently large lead time* for a TSP system to start modifying the signal cycle operation. Our major contribution in this paper is an algorithm design that meets this requirement. We note that the algorithm is “real time” in the sense that the predicted arrival time will be updated as soon as new location data are available. The update rate is typically 1 Hz.

The ability to predict arrival times relies on, to an extent, the detection technologies. For instance, in a *fixed-point location-based* detection system, each transit vehicle is equipped with a transponder, which sends a unique code to the inductive loop that identifies the vehicle. Such a system detects a transit vehicle at a *fixed* time and location, it but does not provide the vehicle’s downstream time-location information, making prediction within *second-level accuracy* not possible. For arrival time prediction, “continuous” *time-based* data are far more useful than fixed *location-based* data. A wide range of Advanced Vehicle Location (AVL) systems has been developed for “continuously” providing location information and communicating it to a traffic management center. Their applications include vehicle scheduling in fleet management and signal

Manuscript received May 5, 2007; revised February 1, 2008, August 1, 2008, and September 19, 2008. First published November 7, 2008; current version published December 1, 2008. The Associate Editor for this paper was H. Dia.

C.-W. Tan is with the Partners for Advanced Transit and Highways (PATH), University of California, Berkeley, Richmond, CA 94804 USA (e-mail: tan@path.berkeley.edu).

S. Park is with the Department of Aerospace Engineering, Sejong University, Seoul 143-747, Korea (e-mail: sungsu@sejong.ac.kr).

H. Liu is with the Department of Civil and Environmental Engineering and the Center for Multidisciplinary Research in Transportation (TechMRT), Texas Tech University, Lubbock, TX 79409-1023 USA (e-mail: hongchao.liu@ttu.edu).

Q. Xu is with PINC Solutions, Berkeley, CA 94709 USA (e-mail: paul_q_xu@yahoo.com).

P. Lau is with the Massachusetts Institute of Technology, Cambridge, MA 02139-4307 USA (e-mail: peterlau@mit.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITS.2008.2006799

priority. Data obtained from our AVL systems are provided by the *Global Positioning System* (GPS). The GPS signal provides the *absolute position coordinate* of a vehicle and its corresponding *coordinated universal time* (UTC). In this paper, there are no data available from other detectors, such as loops or cameras.

Researchers in [9]–[15] have used AVL data to develop models for predicting the arrival time of the next bus. There are also models [16]–[18] for general traffic flow prediction. These models have virtually no sensitivity to operation strategies such as signal priority control and do not meet the *stricter prediction accuracy requirement* associated with signal priority. Commercial systems, such as the SCATS [19] and SCOOT [20] systems, also include a signal priority control module. Bus detectors in SCOOT [20]–[23] are normally placed 70–150 m or 10–12 s before stop line *but* after any bus stop. There is an advantage that these detectors can be placed at optimum points, but prediction is *discretely* updated only when these points are reached. We need “continuous” time-based data rather than fixed location-based data.

We discuss how data are used in Section II. In Section III, a *historical* model is developed for predicting the *time-till-arrival* (TTA) at a signalized intersection *solely* based on historical data. An *adaptive* recursive least-squares (LS) prediction model is developed in Section IV that adaptively adjusts its filter gain based on real-time AVL data. The estimates from the two models are fused to obtain a real-time TTA prediction. This is formulated as an *optimal a posteriori parameter estimation* problem in Section V. The solution is a weighted combination of the TTA predicted by the historical and adaptive models, where the weights are adaptively adjusted using statistical error variances obtained from the two models. The historical model has slow convergence but provides a *good initial estimate* and compensates for the large initial error of the adaptive model. On the other hand, the adaptive model has *fast convergence* and guides the convergence of the fused model. In Section VI, we demonstrate the performance with simulations of field test data. Some statistical characteristics of the error distributions are obtained. The simulations show that the algorithm performs well for the empirical data.

II. AVL DATA FOR ARRIVAL TIME PREDICTION

The prediction of transit vehicle travel times between signalized intersections and bus stops is challenging, since the travel times depend on a number of unpredictable factors [6]. These factors include stochastic traffic flow uncertainties along the route, queue length in front of a traffic light, route length, ridership variation at bus stops (hence the uncertainties in dwell times), weather conditions, time of the day, statistical fluctuations in historical data (with large standard deviations), and GPS data errors. Furthermore, for signal priority application, the predicted arrival time at an intersection must be within a *required strict level of tolerance* (e.g., within a ± 5 -s error bound) after a priority request is executed. That is, the error between the actual and predicted arrival times computed at the current time t , i.e., $\Delta \hat{t}_g(t)$, satisfies a condition $|\Delta \hat{t}_g(t)| < 5$ s for all $t > T$, where T is the priority request time. This

requirement ensures that the system will have a sufficiently *large lead time* to start modifying its normal signal cycle.

Many prediction algorithms involve a data structure that combines historical and real-time AVL data [12], [15]. Long-term prediction models rely more on historical data and typically require *minute-level accuracy* since they involve longer times and more uncertainties. Short-term prediction models, on the other hand, require *second-level accuracy* and rely more on real-time data and downstream traffic conditions [24]. Our prediction algorithm has a *historical* model that uses historical data to estimate an average travel time. It gives a good initial prediction, but the convergence is slow. This slow convergence is compensated and fine-tuned by an *adaptive* model that uses real-time data. The arrival time prediction is then formulated as a *maximum a posteriori* (MAP) *parameter estimation* problem in which the estimates from the historical and adaptive models are fused in a weighted combination. The weights are adjusted using statistical error variances obtained from the two models.

The real-time GPS position data have an error of ± 15 m as specified by the manufacturer. Thus, the position error has approximately a Gaussian distribution [25] with zero mean and a standard deviation of $\sigma_d = 15$ m. The data accuracy is also affected by uncertainties in communication delays and data packet losses during transmissions between onboard computers and the server. We use wheel speed data to compensate for data inaccuracies and obtain *relative* positions between GPS updates. Thus, they complement the *absolute* positions provided by the GPS. Since there are communication delays and data packet losses, the GPS data were carefully processed to extract accurate location data that matched with the UTC.

The data were collected in collaboration with the San Mateo County Transit District. The test site is a section of the El Camino Real in San Mateo, CA which is about 6 mi south of the San Francisco International Airport. It spans approximately 4 km (or 2.48 mi), including ten northbound and eight southbound bus stops. There are 15 signalized intersections. Bus stops and intersections are called *nodes*, with links connecting them. The GPS coordinates of the nodes, and hence the distances between them, are recorded and stored in a database for simulations and field tests.

The *current* location of a transit vehicle indicates how far it is from its next intersection. Historical data can be used to calculate an “average” travel time to cover the remaining distance between the current location and the next intersection, but are insufficient to obtain a reasonably accurate prediction of the *actual* arrival time. This is evidenced by the statistics shown in Fig. 1. These are the means and standard deviations of the link travel times between successive intersections for northbound traffic. The standard deviations are large, and the percentages of standard deviations over means are also large. Similarly, large second-moment statistical deviations are present in the average link travel times for southbound traffic. Historical data are useful for calculating an average travel time, but this prediction needs to be “continuously” fine-tuned using real-time data as the vehicle travels downstream. This is computed in an *adaptive* model that adapts to the flow (average speed) condition downstream.

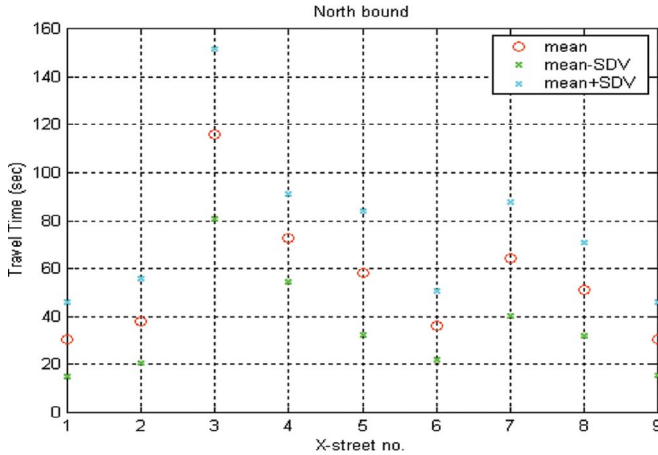


Fig. 1. Mean and standard deviations of northbound link travel times.

III. PREDICTION USING HISTORICAL AVL DATA

In this section, we construct a prediction model that uses only historical AVL data. We call this the *historical model*. Transit vehicle trajectories are separated into drive and stop sections, and we construct a statistical model for each of them. A *drive section* is defined as a *continuous* section of a transit vehicle travel timeline when the vehicle moves at *nonzero speed*. Typically, this is the time when the vehicle travels along a link connecting two neighboring nodes. In general, this is a section of the timeline when the vehicle starts from zero speed and stops again. A *stop section* of a transit vehicle travel timeline is the time when the vehicle stops at *zero speed* at a bus stop or in front of a traffic light. For a drive section, a simple *first-order average speed* model fits the statistical data reasonably well. We first construct historical models for traveled time as a function of traveled distance and for waiting times at traffic lights and bus stops. An algorithm for predicting the TTA using historical AVL data is then presented.

A. Linear Model for Historical AVL Data

A typical space–time diagram for a transit vehicle travelling downstream is shown in Fig. 2, where it takes T_D seconds to travel a drive section of length D meters. The vertical portions of the plot indicate that the vehicle spends some dwell times at bus stops or waiting times in queues. The pair (D, T_D) is for one drive section, with the understanding that the vehicle starts from zero speed at the beginning of this drive section and stops at the end of it. The variable t_{dw} is the dwell time at a bus stop or the waiting time in a queue before a traffic light. Thus, t_{dw} represents a stop section of the vehicle travel timeline. If the vehicle has traveled a distance d from the previous node, the time to travel the remaining distance $D - d$ until it reaches the next node is the TTA t_g . Here, the distance d is the straight-line distance between the current vehicle location and the GPS location of the start node. Note that t_g is referred to as the predicted arrival time computed at the *current time* t or at the *current location* d meters from the start node.

The first question we investigate is whether there is a relationship between D and T_D . Can we express T_D as a function of D that “best fits” the historical data? We assume that the

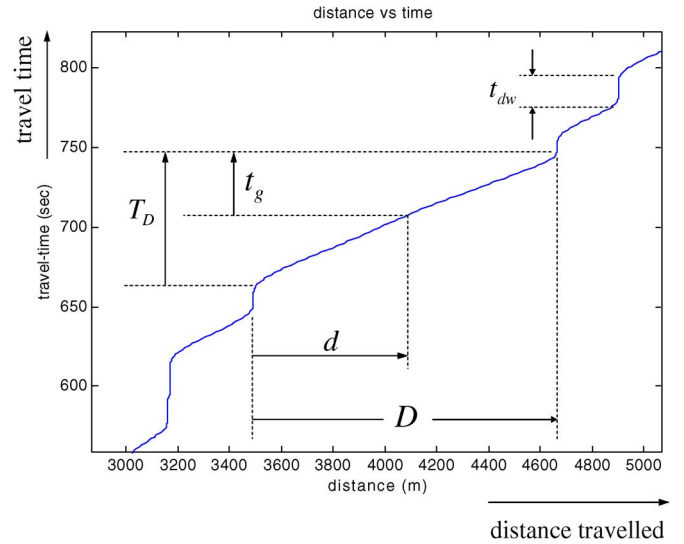


Fig. 2. Typical space–time diagram.

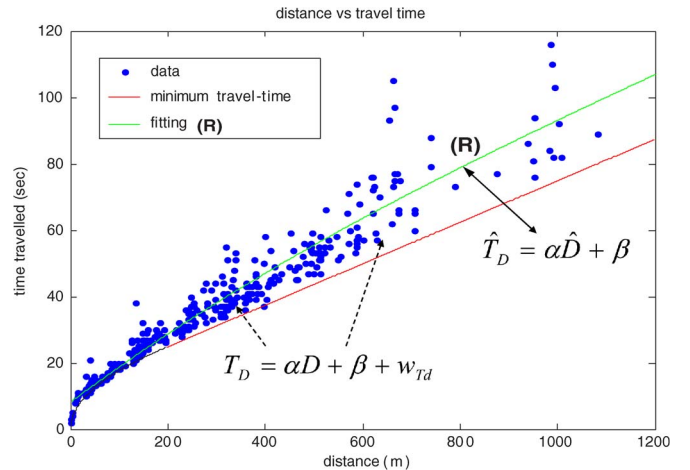


Fig. 3. Linear regression for section lengths and travel times.

driving conditions (e.g., speed limit) are *homogeneous* for the entire test route. If this does not hold, we can divide the route into homogeneous sections and develop a historical model for each section. Thus, the relationship $D \mapsto T_D$, if exists, holds for all the links along the test route. The set of observed data (D, T_D) for all the drive sections for northbound at 7–9 A.M. on weekdays are extracted and plotted in Fig. 3. Since D and T_D are two random variables, the *observations* or *sample points* in Fig. 3 form a joint distribution for the pair (D, T_D) . This distribution might change as more data become available. It is clear that the sample points do not fall on a straight line; thus, we apply *regression* analysis to obtain an approximate functional relationship between the variables D and T_D . This statistical technique determines the values of parameters for a function that cause the function to best fit a set of data. In *linear regression* [26], the function is a linear equation that best predicts T_D from D . A simple *first-order linear average speed model* is a good fit for the historical data. This linear relationship for (D, T_D) has the form

$$\hat{T}_D = \alpha \hat{D} + \beta. \quad (3.1)$$

The “hat” indicates that it is a “best fit” estimate using linear regression. This is the solid line labeled as **(R)** in Fig. 3. The variables D and T_D are the *regressor* (or predictor) and *response* variables, respectively. The parameters α and β are chosen so that the *regression error* $\|T_D - \hat{T}_D\|^2$ is minimized over the region of the regressor variable D in the observed data.

For a drive section length D , the travel time T_D has a certain conditional distribution that depends on D . This distribution can be modeled as a *statistical error* so that for each D , the sample points (D, T_D) are modeled as a linear regression with a distance-dependent error. That is

$$T_D = \alpha D + \beta + w_{TD}. \quad (3.2)$$

The term β is a bias, and w_{TD} is a *Gaussian*-distributed error with zero mean and variance $\sigma_{TD}^2(D)$ that depends on the distance D . We use the notation $w_{TD} \sim N(0, \sigma_{TD}^2(D))$. The variance is determined by the observed data distribution and varies if observations at different times of the day are considered and when more data are collected.

To gain more insight into the linear model, we consider a drive section of length D . The accuracy of the length D is affected by the location uncertainty of bus stop or signalized intersection. This is the GPS data uncertainty. The GPS measurement error has a Gaussian distribution $N(0, \sigma_d^2)$, where $\sigma_d = 15$ m. From (3.1), the *conditional expectation* of the travel time for a given length $D = \hat{D}$ is

$$E(T_D|D = \hat{D}) = \alpha \hat{D} + \beta. \quad (3.3)$$

If we assume that the uncertainty in D and the error process w_{TD} are independent, then the variance of T_D for the given $D = \hat{D}$ is

$$\sigma_T^2(\hat{D}) = \text{Var}(T_D|D = \hat{D}) \approx \alpha^2 \sigma_d^2 + \sigma_{TD}^2(\hat{D}). \quad (3.4)$$

The linear regression model in (3.1), i.e., $\hat{T}_D = \alpha \hat{D} + \beta$, is a line of mean values; that is, the coordinate on the regression straight line (3.1) at any value of D is simply the expected value of T_D for that given D . The slope α is interpreted as the change in the conditional expectation of T_D for a unit change in D . The variance of T_D is determined by the GPS data accuracy and the variance of the error component in (3.2). If the GPS measurement error is smaller, σ_d will also be smaller, thus implying a more accurate estimate of T_D . The linear relationship (3.1) generated by the “Linear-in-the-Parameters Regression” module in MATLAB is the solid line labeled as **(R)** in Fig. 3. The program also calculates the variance $\sigma_{TD}^2(D)$. For northbound traffic between 7 and 9 A.M., this linear relationship is given as follows:

- 1) *northbound, 7–9 A.M.*: $\hat{T}_D = 0.109\hat{D} + 8.0177$;
- 2) *northbound, all day*: $\hat{T}_D = 0.1072\hat{D} + 8.1614$.

A second-order regression model has the form $\hat{T}_D = \alpha_2 \hat{D}^2 + \alpha_1 \hat{D} + \beta$. From the historical data, α_2 is on the order of 10^{-4} . In addition, the regression error for the second-order model is larger than that for the first-order model. Thus, the simple first-order model fits reasonably well for our observed data.

B. Prediction of TTA Using Historical Data

The space–time diagram in Fig. 2 roughly reveals the “average” vehicle speed between nodes. Historical data are useful for predicting an average travel time to reach the next traffic light. The empirical data also suggest that for each drive section, the average speed can be modeled as a constant with an additive uncertainty. This is our historical model for predicting the TTA. Let $D = \hat{D}$ be the length of a drive section, the predicted travel time for this section using linear regression is $\hat{T}_D$, where $(\hat{D}, \hat{T}_D)$ satisfies the linear relationship (3.1). We model the average speed along this section as the *constant speed* $\hat{D}/\hat{T}_D$ plus a Gaussian error term, i.e.,

$$v = \frac{\hat{D}}{\hat{T}_D} + w_v, \quad w_v \sim N(0, \sigma_v^2). \quad (3.5)$$

Variations are modeled as first-order perturbations. Then by taking derivatives, the approximate variation in $v = \hat{D}/\hat{T}_D$ at operating point $(D, T_D) = (\hat{D}, \hat{T}_D)$ is

$$\delta v = \frac{1}{\hat{T}_D} \delta D - \frac{\hat{D}}{\hat{T}_D^2} \delta T_D, \quad \text{at } (D, T_D) = (\hat{D}, \hat{T}_D). \quad (3.6)$$

The variation in the drive section length δD is due to the GPS location uncertainty, which has a distribution $N(0, \sigma_d^2)$, where $\sigma_d = 15$ m. The error in the travel time δT_D has a distribution with variance σ_T^2 given in (3.4). The “quality” of an estimate can be measured by its variance. Assume D and T_D are independent, then by (3.6), the variance of v for $D = \hat{D}$ is

$$\sigma_v^2 = \text{Var}(v|D = \hat{D}) \approx \left(\frac{1}{\hat{T}_D}\right)^2 \sigma_d^2 + \left(\frac{\hat{D}}{\hat{T}_D^2}\right)^2 \sigma_T^2. \quad (3.7)$$

The *conditional expectation* of the average speed given that $D = \hat{D}$ is

$$\hat{v}(\hat{D}) = E(v|D = \hat{D}) = \frac{\hat{D}}{\hat{T}_D}. \quad (3.8)$$

A historical model for predicting TTA can now be constructed. We will denote it by t_{gH} to distinguish it from the TTA predicted by a real-time adaptive model that will be developed in the next section. Referring to Fig. 2, if the drive section length is $D = \hat{D}$ and the vehicle has traveled a distance d , the time to travel the remaining distance $\hat{D} - d$ is the TTA. (Our interest is when the end node is an intersection.) Since the speed is modeled as a constant average speed plus an uncertainty, the TTA is thus modeled as the time to complete the remaining distance at the *constant* average speed $\hat{v}$ plus an error term. That is

$$t_{gH}(d) = \frac{\hat{D} - d}{\hat{v}} + w_{tgH}, \quad w_{tgH} \sim N(0, \sigma_{tgH}^2(d)). \quad (3.9)$$

By using the same first-order perturbation analysis for obtaining σ_v^2 in (3.7), the variation in t_{gH} is

$$\delta t_{gH} = \frac{1}{\hat{v}} \delta D - \frac{1}{\hat{v}} \delta d - \frac{\hat{D} - d}{\hat{v}^2} \delta v. \quad (3.10)$$

The variance of t_{gH} for the given $D = \hat{D}$ is given by

$$\sigma_{t_{\text{gH}}}^2(d) \approx \frac{2}{\hat{v}^2} \sigma_d^2 + \left(\frac{\hat{D} - d}{\hat{v}^2} \right)^2 \sigma_v^2 \quad (3.11)$$

where $\sigma_d = 15$ m, and σ_v^2 is given by (3.7). The TTA predicted by the historical model is the *conditional expectation* of t_{gH} given $D = \hat{D}$. This is given by

$$\hat{t}_{\text{gH}}(d) = E(t_{\text{gH}} | D = \hat{D}) = \frac{\hat{D} - d}{\hat{v}} = \hat{T}_D \left(1 - \frac{d}{\hat{D}} \right). \quad (3.12)$$

This is expected since we use a *constant* average speed model. The most important step in developing the historical model is to calculate the variations in the observed data. The variances are σ_T^2 , σ_v^2 , and $\sigma_{t_{\text{gH}}}^2$ obtained in (3.4), (3.7), and (3.11), respectively. The variance $\sigma_{t_{\text{gH}}}^2$ will be used in a weighted combination of a prediction algorithm to be developed in Section V.

We also developed a historical dwell time model by considering the dwell time distribution at each bus stop. The dwell time t_{dw} is modeled by extracting the historical dwell times at the north- and southbound bus stops from the observed data. For each bus stop, we calculate the mean $\bar{t}_{\text{dw}}$ and variance $\sigma_{t_{\text{dw}}}^2$ of the observed dwell times. As suggested by the *law of large numbers*, the accuracies of the mean and variance will improve and converge as more data are available.

IV. ADAPTIVE PREDICTION MODEL

The historical model predicts the TTA in one calculation by using a *constant* speed given in (3.8) for the drive section. It does *not* capture any real-time speed variations. In this section, we develop another prediction model to complement the historical model. This is an *adaptive* model that uses real-time AVL data to adaptively estimate the *downstream average speed*. To correct for the GPS location error due to unknown GPS latency and transmission delay and data packet loss, we use wheel speed data to obtain a more accurate estimate of the current vehicle location and the traveled distance from the start node. Suppose the section length is $D = \hat{D}$ and a transit vehicle has traveled a distance of $d(t)$ from the start node in t seconds. The average speed is roughly $d(t)/t$, subject to statistical errors. We propose an *adaptive* average speed model. If it takes an average speed $a(t)$ to cover a distance $d(t)$ in t seconds, the time to cover the remaining distance, predicted at the current time t , is $(\hat{D} - d(t))/a(t)$. In contrast to the *constant* average speed in the historical model, the average speed $a(t)$ is adaptively tuned to speed variations as the vehicle travels downstream. In this adaptive average speed model, we express the traveled distance as

$$d(t) = a(t)t + b(t) + w_d(t) = H(t)x(t) + w_d(t). \quad (4.1)$$

Here, $a(t)$ is the average speed at time t , $b(t)$ is the distance residual, $H(t) = [t \ 1]$, and $x(t) = [a(t) \ b(t)]^T$ is the state. The traveled distance is calculated using the current real-time

GPS location, which has an error standard deviation of $\sigma_d = 15$ m. Thus, we assume that $w_d(t)$ is an independent identically distributed measurement noise process with zero mean and variance σ_d^2 . The observation $d(t)$ is updated every $\Delta t = 1$ s; thus, $d(k)$ is the k th observation at time $t_k = k\Delta t$. For real-time implementation, the discrete-time adaptive model is

$$\begin{aligned} d(k) &= a(k)k\Delta t + b(k) + w_d(k) \\ &= H(k)x(k) + w_d(k), \quad k = 1, 2, 3, \dots \end{aligned} \quad (4.2)$$

We formulate it as a linear LS estimation problem, where at step k , it is desirable to estimate the state $x(k)$ from the observations $\{d(1), \dots, d(k)\}$ such that the quadratic error is minimized, i.e.,

$$\text{LS estimator } \hat{x}(k) = \arg \min \sum_{j=1}^k |d(j) - H(j)x|^2. \quad (4.3)$$

The LS estimator that minimizes (4.3) is obtained by setting its gradient with respect to x to zero and has a *recursive* form [23] given by

$$\hat{x}(k) = \hat{x}(k-1) + W(k) [d(k) - H(k)\hat{x}(k-1)]. \quad (4.4)$$

In the standard LS estimation, the recursive update estimate $\hat{x}(k)$ is equal to the previous estimate plus a correction term. The correction term consists of a gain $W(k)$ multiplied by the residual, which is the difference between the observation $d(k)$ and the predicted value of this observation from the previous $k-1$ measurements. The filter gain $W(k)$ is

$$W(k) = P(k-1)H(k)^T S(k)^{-1}. \quad (4.5)$$

Here, $S(k)$ is the residual covariance (i.e., covariance of the residual in (4.4)), and $P(k)$ is the covariance of the estimator $\hat{x}(k)$. They can *recursively* be expressed as

$$S(k) = H(k)P(k-1)H(k)^T + \sigma_d^2 \quad (4.6a)$$

$$P(k) = P(k-1) - W(k)S(k)W(k)^T. \quad (4.6b)$$

Next, we develop an adaptive TTA prediction model. Since the optimal state is $\hat{x}(k) = [\hat{a}(k) \ \hat{b}(k)]$ and $\hat{a}(k)$ is the average speed at time $k\Delta t$, we model the TTA as the time to complete the remaining distance at the average speed $\hat{a}(k)$ plus an error term, i.e.,

$$t_{\text{gA}}(d(k)) = \frac{\hat{D} - d(k)}{\hat{a}(k)} + w_{\text{tGA}}, \quad w_{\text{tGA}} \sim N(0, \sigma_{\text{tGA}}^2). \quad (4.7)$$

We approximate variations in t_{gA} by the first-order perturbation method used in deriving $\sigma_{t_{\text{gH}}}^2$ in (3.11), i.e.,

$$\delta t_{\text{gA}} = \frac{1}{\hat{a}} \delta D - \frac{1}{\hat{a}^2} \delta d - \left(\frac{\hat{D} - d}{\hat{a}^2} \right) \delta a. \quad (4.8)$$

The variance of t_{gA} for $D = \hat{D}$ is thus given by

$$\begin{aligned} \sigma_{tgA}^2(d(k)) &= \text{Var}(t_{gA}|D = \hat{D}) \\ &\approx \frac{2}{\hat{a}(k)^2} \sigma_d^2 + \left(\frac{\hat{D} - d(k)}{\hat{a}(k)^2} \right)^2 P(1, 1) \end{aligned} \quad (4.9)$$

where $\sigma_d = 15$ m. The variance of $\hat{a}(k)$ is $P(1, 1)$, i.e., the $(1, 1)$ entry of the covariance matrix $P(k)$. The TTA predicted by this adaptive model is the conditional expectation of t_{gA} given $D = \hat{D}$. From (4.7), this is

$$\hat{t}_{gA}(d(k)) = E(t_{gA}(d(k))|D = \hat{D}) = \frac{\hat{D} - d(k)}{\hat{a}(k)}. \quad (4.10)$$

In the next section, we develop a TTA prediction algorithm that fuses the historical and adaptive models in a weighted combination with weights inversely proportional to the variances σ_{tgH}^2 and σ_{tgA}^2 .

V. ALGORITHM FOR ARRIVAL TIME PREDICTION

Again, we consider a situation when the section length is $D = \hat{D}$ and the vehicle has traveled a distance d from the beginning of the section. The historical model developed in Section III-B predicts the TTA in a single calculation using a *constant* speed model. This is given by $\hat{t}_{gH}(d)$ in (3.12), computed at the current location. The adaptive model developed in Section IV adaptively adjusts some parameters using the past and present data. The adaptive TTA given by $\hat{t}_{gA}(d)$ in (4.10) computes a speed that *adapts* to real-time traffic conditions as the vehicle travels downstream. We can think of $\hat{t}_{gA}$ as an *updated measurement* of $\hat{t}_{gH}$ as the vehicle travels toward the end node of the drive section. Given the measurement $\hat{t}_{gA}$, what is the “best *a posteriori* estimate” of the historical TTA? This “best estimate” will be our predicted TTA. To generalize the TTA prediction problem, we can formulate it as a parameter estimation problem. Consider a measurement t_m of the unknown parameter t_{gH} in the presence of an additive Gaussian measurement noise w_{tg} , where $w_{tg} \sim N(0, \sigma_{tgA}^2)$, and σ_{tgA}^2 is the right-hand side of (4.9). That is

$$t_m = t_{gH} + w_{tg}. \quad (5.1)$$

Note that the process w_{tg} is different from w_{tgA} for the adaptive model in (4.9). That process has a variance that is only approximately given by the right-hand side of (4.9). We want to find the “best *a posteriori* estimate” of t_{gH} when $t_m = \hat{t}_{gA}$. The *prior* information about t_{gH} is that it is Gaussian with mean $\hat{t}_{gH}$ and variance σ_{tgH}^2 given in (3.12) and (3.11), respectively. We assume that t_m and t_{gH} are independent. The *a posteriori* probability density function (pdf) given the measurement t_m is $p(t_{gH}|t_m)$. The MAP estimator [27] is a realization of t_{gH} that maximizes the *a posteriori* pdf, which is defined as

$$\text{MAP estimator} := \arg \max p(t_{gH}|t_m). \quad (5.2)$$

This estimator, which depends on the measurements t_m and, through them, the realization of t_{gH} , is a random variable. It can

be shown that the *a posteriori* pdf of t_{gH} is also Gaussian [27] with mean and variance given by

$$\mu(t_m) = \frac{\sigma_{tgA}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2} \hat{t}_{gH} + \frac{\sigma_{tgH}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2} t_m \quad (5.3a)$$

$$\sigma_{tg}^2 = \frac{\sigma_{tgA}^2 \sigma_{tgH}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2}. \quad (5.3b)$$

The mean $\mu(t_m)$ of the *a posteriori* pdf is therefore the MAP estimator since a Gaussian distribution has the maximum at its mean. The MAP estimator fuses the historical TTA and the measurement t_m . If the measurement is $t_m = \hat{t}_{gA}$, we get

$$\hat{t}_g = \mu(\hat{t}_{gA}) = \frac{\sigma_{tgA}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2} \hat{t}_{gH} + \frac{\sigma_{tgH}^2}{\sigma_{tgA}^2 + \sigma_{tgH}^2} \hat{t}_{gA}. \quad (5.4)$$

This is our TTA prediction algorithm. We note that the MAP estimator is a *weighted combination* of estimates from the historical and adaptive models, and the weights of the prior mean and the measurement are inversely proportional to their variances.

VI. SIMULATIONS AND PERFORMANCE

The performance of the algorithm (5.4) is examined by means of simulation. The simulations are compared with the *actual* TTA calculated from the empirical data. Fig. 4 is a simulation of a sample run between two nodes for northbound traffic during 7–9 A.M. The section length is $D = 1083$ m, and the *actual* travel time is $T_D = 89$ s. The difference between the predicted and actual TTA, i.e., $\Delta \hat{t}_g = \hat{t}_g - t_g$, is the prediction error. We also superimpose the prediction errors from the historical and adaptive models, which are $\Delta \hat{t}_{gH} = \hat{t}_{gH} - t_g$ and $\Delta \hat{t}_{gA} = \hat{t}_{gA} - t_g$. Recall that the predicted arrival time must be within a *required strict level of tolerance* (e.g., within a ± 5 -s error bound) after a signal priority request is executed. We note that $\Delta \hat{t}_{gH}$ is within the error bound when the vehicle is about 200 m from the end node. In terms of travel time, it is about 20 s from the end node. Thus, the convergence of the historical model is slow. The respective cutoffs for $\Delta \hat{t}_g$ are about 650 m and 50 s, allowing priority requests to be made as soon as the vehicle is 650 m from the end node. The TTA predicted by the fused model converges *faster* than that predicted by the historical model alone. Indeed, the adaptive model is crucial in guiding the convergence and accuracy of the fused model. In the weighed algorithm (5.4), the adaptive model improves the convergence rate by including real-time data. The relatively fast rate of convergence allows the system to have a sufficiently long lead time to start modifying its signal operation.

The adaptive algorithm (4.10) tends to have large initial errors, as evidenced in Fig. 4. This is because, initially, the vehicle starts at zero speed; thus, the initial speed $\hat{a}(t)$ is small, and $\hat{t}_{gA}$ is large. A drawback of the LS algorithm is that the initial phase of convergence is not monotonic; thus, the parameter update is initially “not good.” This initial inaccuracy is compensated by the historical TTA in the weighted algorithm.

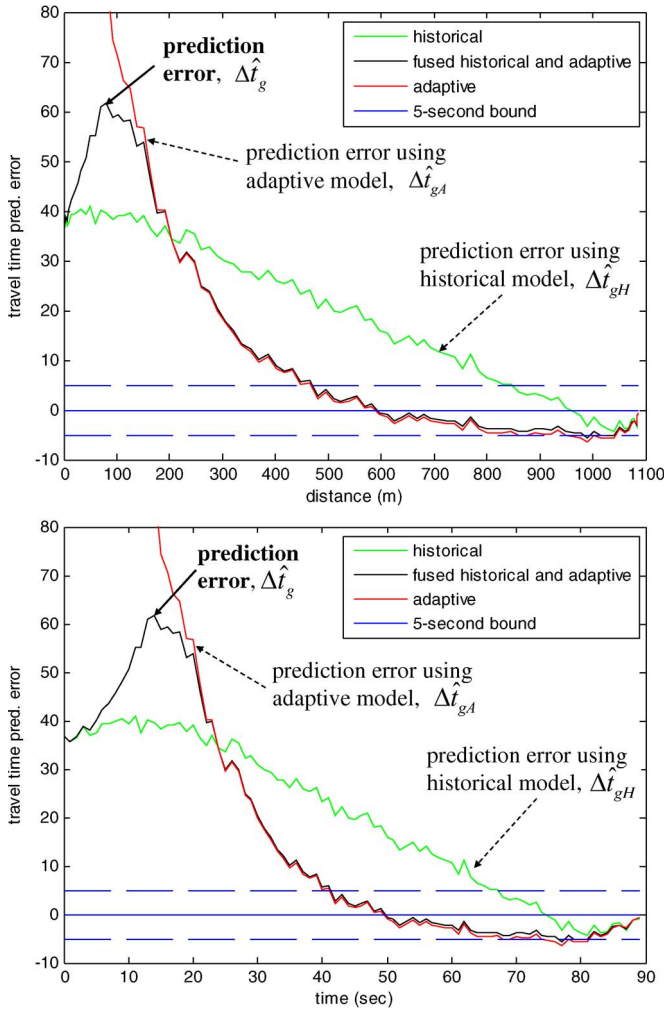


Fig. 4. Comparisons of prediction time errors (top) versus distance traveled and (bottom) versus time elapsed for a sample run.

The historical model provides a fairly good initial estimate of the average flow condition, but the convergence is slow. The adaptive model has good convergence and is crucial in guiding the convergence of the prediction algorithm (5.4).

The simulations presented in Fig. 4 are for one sample run. It will be useful to obtain some statistics for the entire set of field test data. We next obtain some statistics of the prediction error $\Delta \hat{t}_g$ for a link connecting two successive intersections, with no bus stop in between. We consider Barneson and Hobart Avenues (cross street numbers 8 and 9) in the northbound direction. The distance between these two nodes is 257 m. The middle solid line in Fig. 5 shows the mean prediction error versus the time it takes to arrive at the end node. We note that the standard deviation quickly gets smaller as the vehicle travels downstream. Fig. 6 is a 3-D plot of the prediction error histogram. There is a large peak around the “zero coordinate” (TTA = 0 s; $\Delta \hat{t}_g = 0$ s), indicating that the solution almost surely always converges. The prediction error also quickly converges to stay within the ± 5 -s bound. The algorithm has been implemented and integrated with a signal priority control scheme [8]. Field test results show that the algorithm works well with the application.

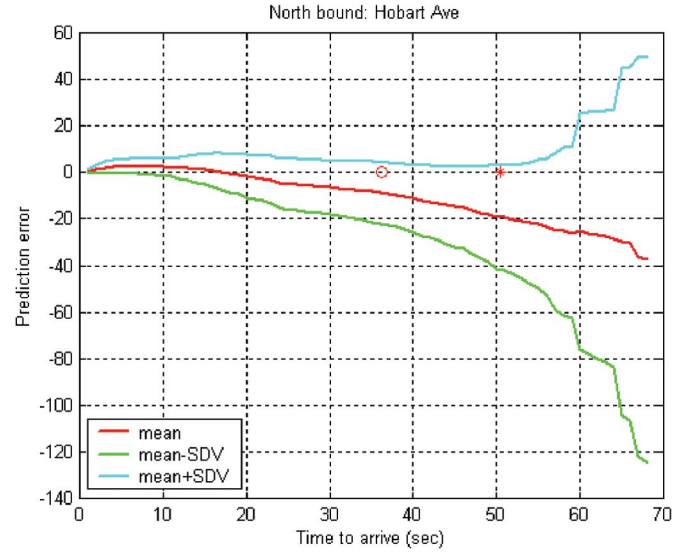


Fig. 5. Mean and standard deviations of the prediction time error for transit vehicles approaching Hobart Avenue in the northbound direction.

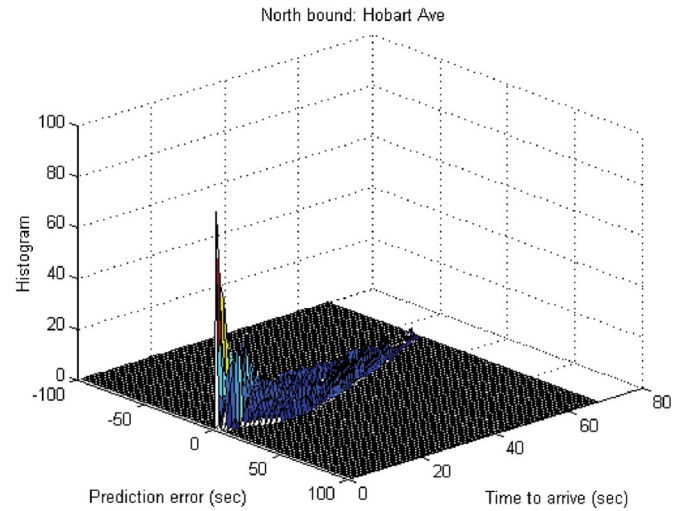


Fig. 6. Prediction time error histogram for transit vehicles approaching Hobart Avenue in the northbound direction.

VII. CONCLUDING REMARKS

A critical issue encountered in implementing TSP control is to predict the time it takes for a transit vehicle to arrive at the next signalized intersection if the current distance from the intersection is known. This paper has addressed this issue and has proposed a prediction algorithm that can be implemented and integrated with signal priority control. Specifically, it requires the arrival time prediction error to be no more than ± 5 s after a priority request is made. This is necessary so that a TSP system has a large lead time to start modifying its normal signal cycle.

The problem in question is the time it takes to travel the distance between the current location and the next signalized intersection. A plot of the historical data for the traveled distance versus its corresponding travel time in Fig. 3 suggests that a linear regression model fits the data reasonably well. This is given by (3.1), i.e., $\hat{T}_D = \alpha \hat{D} + \beta$, where it is interpreted that the *average* time to travel a distance of $D = \hat{D}$ is $\hat{T}_D$.

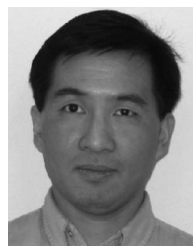
The historical model assumes a *constant average speed* for the entire drive section, where the constant speed is given in (3.8), and the predicted arrival time is given by $\hat{t}_{gH}$ in (3.12). The estimate $\hat{t}_{gH}$ uses a constant speed for the entire drive section and does not capture any fluctuations in downstream traffic flow and speed that might affect the actual arrival time. An adaptive model is developed to complement the historical model. This is an *adaptive average speed* model that uses real-time AVL data to adaptively estimate the downstream average speed. It is formulated as a linear LS problem in Section IV, and the predicted arrival time is given by $\hat{t}_{gA}$ in (4.7).

The predicted arrival time is obtained by fusing the estimates from the historical and adaptive models. The adaptive estimate tunes the historical estimate; thus, we think of $\hat{t}_{gA}$ as an updated measurement of $\hat{t}_{gH}$ as the vehicle travels downstream. This is formulated as a parameter estimation problem in which we seek the “best *a posteriori* estimate” of $\hat{t}_{gH}$ given the estimate $\hat{t}_{gA}$. This MAP estimator is the predicted arrival time $\hat{t}_g$ given in (5.4). It is a weighted combination of the estimates $\hat{t}_{gH}$ and $\hat{t}_{gA}$, where the weights are inversely proportional to the variances σ_{tgH}^2 and σ_{tgA}^2 , respectively. We have included simulations of field test data to demonstrate the performance of the solution. The algorithm has since been implemented and integrated with a signal priority control scheme discussed in [8].

The derivations of the historical and adaptive estimates in Sections III–IV assume a drive section of length D . If the transit vehicle stops after it has traveled a distance $d < D$, the algorithm restarts with the new drive section length $D - d$ when the vehicle moves again. This occurs, for example, when a transit vehicle joins the back of a queue and stops. In these situations, the error might not converge to stay within the ± 5 -s bound fast enough to allow the system to modify the normal cycle. Thus, the algorithm probably might not perform as good in situations of heavily congested stop-and-go traffic. It will be useful to include models of queue length and queue discharge rate so that the prediction includes the additional time the vehicle might spend in a stop-and-go traffic. We agree that travel time predictions that deal with great uncertainties in downstream traffic conditions are more complex. However, we emphasize that we have designed an algorithm that works well for the test site and under most traffic conditions. In general, stochastic systems with highly fluctuated and unpredictable behaviors, such as a stock market, are difficult to analyze.

REFERENCES

- [1] K. N. Balke, C. L. Dudek, and T. Urbanik, II, “Development and evaluation of an intelligent bus priority concept,” *Transp. Res. Rec.*, vol. 1727, pp. 12–19, 2000.
- [2] R. Baker *et al.*, *An Overview of Transit Signal Priority*, Jul. 11, 2002. Prepared by ATMS and APTS Committees of ITS America.
- [3] N. Hounsell and F. McLeod, “Automatic vehicle location and bus priority: The London system,” in *Proc. 8th WCTR*, Antwerp, Belgium, Jul. 1998, pp. 279–292.
- [4] J. Wu and N. Hounsell, “Bus priority using pre-signals,” *Transp. Res., Part A Policy Pract.*, vol. 32, no. 8, pp. 563–583, Nov. 1998.
- [5] F. Dion, H. Rakha, and Y. Zhang, “Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial,” *J. Transp. Eng.*, vol. 130, no. 3, pp. 294–303, May/Jun. 2004.
- [6] V. Ngan, T. Sayed, and A. Abdelfatah, “Impacts of various parameters on transit signal priority effectiveness,” *J. Public Transp.*, vol. 7, no. 3, pp. 71–93, 2004.
- [7] H. Liu, W.-H. Lin, and C.-W. Tan, “Operational strategy for advanced vehicle location system-based transit signal priority,” *J. Transp. Eng.*, vol. 133, no. 9, pp. 513–522, Sep. 2007.
- [8] M. Li, Y. Yin, K. Zhou, W.-B. Zhang, H. Liu, and C.-W. Tan, “Adaptive transit signal priority on actuated signalized corridors,” in *Proc. 84th Transp. Res. Board Annu. Meeting*, Jan. 2005.
- [9] D. J. Dailey and L. Li, “An algorithm to estimate traffic speed using uncalibrated cameras,” *Transp. Res. Rec.*, no. 1719, pp. 27–32, 2000.
- [10] D. J. Dailey, S. D. Maclean, F. W. Cathey, and S. R. Wall, “Transit vehicle arrival prediction: An algorithm and a large scale implementation,” *Transp. Res. Rec., Transp. Netw. Modeling*, vol. 1771, pp. 46–51, 2001.
- [11] R. Kalaputapu and M. J. Demetsky, “Modeling schedule deviations of buses using automatic vehicle location data and artificial neural networks,” *Transp. Res. Rec.*, vol. 1497, pp. 44–52, 1995.
- [12] W.-H. Lin and J. Zeng, “An experimental study of real-time bus arrival time prediction with GPS data,” *Transp. Res. Rec.*, no. 1666, pp. 101–109, 1999.
- [13] R. Mishalani, S. Lee, and M. R. McCord, “Evaluation of real-time bus arrival information systems,” *Transp. Res. Rec.*, vol. 1731, pp. 81–87, 2000.
- [14] A. Shalaby and A. Farhan, “Prediction model of bus arrival and departure times using AVL and APC data,” *J. Public Transp.*, vol. 7, no. 1, pp. 41–61, 2004.
- [15] Z. Wall and D. J. Dailey, “An algorithm for predicting the arrival time of mass transit vehicles using automatic vehicle location data,” in *Proc. 78th Transp. Res. Board Annu. Meeting*, Jan. 1999.
- [16] F. Yang, Z. Yin, H. X. Liu, and B. Ran, “Online recursive algorithm for short-term traffic prediction,” *Transp. Res. Rec.*, no. 1879, pp. 1–8, 2004.
- [17] M. Carey, Y. E. Ge, and M. McCartney, “A whole-link travel-time model with desirable properties,” *Transp. Sci.*, vol. 37B, no. 1, pp. 83–96, 2003.
- [18] K. Ashok and M. Ben-Akiva, “Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows,” *Transp. Sci.*, vol. 34, no. 1, pp. 21–36, Feb. 2000.
- [19] *Sydney Coordinated Adaptive Traffic System (SCATS) Developed by the Road and Traffic Authority, New South Wales, Australia*. [Online]. Available: <http://www.rta.nsw.gov.au>
- [20] *Split Cycle Offset Optimisation Technique (SCOOT) Adaptive Traffic Control System Website*. [Online]. Available: <http://www.scoot-utc.com>
- [21] G. T. Bowen, “Bus priority in SCOOT,” *Transp. Res. Lab., Wokingham, U.K.*, TRL Rep. 255, 1997. [Online]. Available: <http://www.trl.co.uk>
- [22] N. B. Hounsell, F. N. McLeod, K. Gardner, J. R. Head, and D. Cook, “Headway-based bus priority in London using AVL: First results,” in *Proc. 10th IEEE Int. Conf. Road Transp. Inform. Control*, 2000, pp. 218–222.
- [23] N. B. Hounsell, F. N. McLeod, and B. P. Shrestha, “Bus priority at traffic signals: Investigating the options,” in *Proc. 12th IEEE Int. Conf. Road Transp. Inform. Control*, Apr. 2004, pp. 287–294.
- [24] X. Zhang and J. A. Rice, “Short-term travel time prediction,” *Transp. Res., Part C Emerg. Technol.*, vol. 11, no. 3/4, pp. 187–210, Jun.–Aug. 2003.
- [25] F. van Diggelen, “GNSS accuracy: Lies, damn lies, and statistics,” *GPS World*, pp. 26–32, Jan. 2007.
- [26] D. C. Montgomery, E. A. Peck, and G. Vining, *Introduction to Linear Regression Analysis*, Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley, 2001.
- [27] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. Hoboken, NJ: Wiley, 2001.



Chin-Woo Tan received the M.A. degree in mathematics and the Ph.D. degree in electrical engineering from the University of California, Berkeley (UC Berkeley).

He has been with the Partners for Advanced Transit and Highways (PATH), UC Berkeley, Richmond, CA, for more than ten years and has consulted with Siemens TTB and PINC Solutions, both in Berkeley. He has taught electrical engineering courses at UC Berkeley, The University of California at Davis, and San Jose State University, San Jose, CA. His research

interests include signal processing, estimation, navigation, intelligent transportation, and nonlinear dynamics.



Sungsu Park received the B.S. and M.S. degrees in aerospace engineering from Seoul National University, Seoul, Korea, in 1988 and 1990, respectively, and the Ph.D. degree in mechanical engineering from the University of California, Berkeley, in 2000.

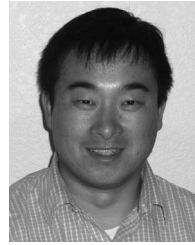
He is currently an Associate Professor with the Department of Aerospace Engineering, Sejong University, Seoul. His research interests include estimation theory, adaptive control, and robust control, with applications to MEMS and aerospace systems.



Hongchao Liu (M'08) received the M.A. degree in transportation engineering from Tsinghua University, Beijing, China, in 1996 and the Ph.D. degree in transportation engineering from the University of Tokyo, Tokyo, Japan, in 2000.

He is currently an Assistant Professor with the Department of Civil and Environmental Engineering and the Center for Multidisciplinary Research in Transportation (TechMRT), Texas Tech University (TTU), Lubbock. Prior to joining TTU in 2004, he spent three years as a Postdoctoral Researcher with

the Institute of Transportation Studies, University of California, Berkeley. He has more than 15 years of experience in the field of intelligent transportation systems, traffic operation and control, and traffic simulation.



Qing Xu received the M.S. degree in mechanical engineering, the M.S. degree in electrical engineering, and the Ph.D. degree in mechanical engineering from the University of California, Berkeley (UC Berkeley), in 2000, 2005, and 2005, respectively.

He is currently a Senior Research Engineer with PINC Solutions, Berkeley. He was a Summer Intern with Daimler-Chrysler Research and Technology Center North America and a Researcher with the Partners for Advanced Transit and Highways (PATH), UC Berkeley, Richmond, CA. His research

interests are in transportation, vehicular ad hoc networks, active vehicle safety systems, and data-fusion technologies applied to navigation systems and RFID.

Peter Lau received the B.S. degree in electrical engineering from the University of California, Berkeley, in 2007. He is currently working toward the master's degree in transportation engineering at the Massachusetts Institute of Technology, Cambridge.

His research interests include system design and software development in intelligent transportation systems.