Transportation Research Part C 16 (2008) 768–782

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments

Bruce Hellinga [a,*], Pedram Izadpanah [a], Hiroyuki Takada [b], Liping Fu [a]

[a] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1
[b] Hanshin Expressway Company Limited, 4-1-3 Kyutaro-machi, Chuo-ku, Osaka-shi, Osaka 541-0056, Japan

ARTICLE INFO

Article history:
Received 1 February 2007
Received in revised form 14 April 2008
Accepted 17 April 2008

Keywords:
Probe vehicles
Link travel time
Mobile phone
Travel time allocation

ABSTRACT

In probe-based traffic monitoring systems, traffic conditions can be inferred based on the position data of a set of periodically polled probe vehicles. In such systems, the two consecutive polled positions do not necessarily correspond to the end points of individual links. Obtaining estimates of travel time at the individual link level requires the total traversal time (which is equal to the polling interval duration) be decomposed. This paper presents an algorithm for solving the problem of decomposing the traversal time to times taken to traverse individual road segments on the route. The proposed algorithm assumes minimal information about the network, namely network topography (i.e. links and nodes) and the free flow speed of each link. Unlike existing deterministic methods, the proposed solution algorithm defines a likelihood function that is maximized to solve for the most likely travel time for each road segment on the traversed route. The proposed scheme is evaluated using simulated data and compared to a benchmark deterministic method. The evaluation results suggest that the proposed method outperforms the bench mark method and on average improves the accuracy of the estimated link travel times by up to 90%.

1. Introduction

Monitoring traffic conditions over large road networks has been a significant challenge for many transportation authorities due to large capital expenditures required by most existing traffic monitoring technologies. One possible solution is the use of systems that are able to anonymously track vehicles – such as cellular phone based traffic monitoring systems (Cayford and Yim, 2006).

Anonymous tracking systems, such as cellular phone based traffic monitoring systems, are not restricted to obtaining information from only a set of dedicated vehicle probes. Rather, they anonymously sample from the total population of units. In the case of cellular phone based traffic monitoring systems, the positions of a sample of the cell phones within a specified geographical area are tracked over time. This process is called location referencing. The location referencing process is usually carried out by the wireless carrier with the resulting data consisting of a randomly assigned probe vehicle identification number, time stamp and position. There are a number of techniques in the literature to estimate position of a cell phone namely, time difference of arrival, angle of arrival, and timing advance (Izadpanah and Hellinga, 2007; Lovell, 2001; Drane and Rizos, 1998). The data are then transmitted to a processing center for deriving information on traffic conditions such as link travel times and speed, incidents, and queues.

* Corresponding author. Tel.: +1 519 888 4567; fax: +1 519 888 4349.
E-mail addresses: bhellinga@uwaterloo.ca (B. Hellinga), pizadpan@uwaterloo.ca (P. Izadpanah), htakada@engmail.uwaterloo.ca (H. Takada), lfu@uwaterloo.ca (L. Fu).

0968-090X/$ - see front matter
doi:10.1016/j.trc.2008.04.002

Inferring traffic conditions from position data requires five steps as follows:

1. Map matching.
2. Path identification.
3. Probe filtering.
4. Travel time allocation.
5. Travel time aggregation.

The first step of this inference procedure is to address the issue that position estimates of a vehicle reported by the cellular phone location referencing system usually contain errors (due to several sources including non-line-of-sight and multi-path propagation (Takada, 2006)) and therefore may not correspond to the actual position of the vehicle on the road network. Consequently, it is necessary to determine the most likely position of the vehicle on the road network given the reported location. This process is termed *map matching*.

When positions are obtained relatively infrequently there may be more than one possible path on the road network between two consecutively matched positions. Thus it is necessary to identify the most likely path taken by the vehicle to travel between two consecutive positions. This process, called *path identification*, represents the second step.

In anonymous tracking systems, such as cellular phone based systems, it is not known *a priori* that the unit (e.g. cell phone) being sampled (i.e. for which location estimates have been obtained) is actually a vehicle. The cell phone may be stationary in a building, or in the possession of a pedestrian on the sidewalk, a person on a bus, on a bike, etc. Consequently, it is necessary to filter the sampled units to use only data from units in vehicles. This process is called *probe filtering*.

The travel time along a path between two consecutively reported locations is simply equal to the difference between the two consecutive reported times associated with the two location reports. However, the identified vehicle path may cover a partial link and/or several links. Consequently, if the goal is to derive travel times of individual links there is a need to allocate the path travel time to the individual links and/or partial links traversed by the vehicle. This fourth step of the inference procedure is called *travel time allocation*.

The last step in the process, called travel time aggregation, is to combine link travel times from individual probe vehicles into aggregate estimates of the current (or more accurately the recent past) average link travel time for all vehicles.

This paper focuses only on step 4 – travel time allocation and compares performance of two travel time allocation schemes. Consequently, it does not consider the magnitude and distribution of errors associated with steps 1, 2 or 3. Following a brief summary of previous research, the problem is formally posed, a solution algorithm is proposed, and a set of performance measures is introduced. Finally, the performance of the proposed algorithm is compared with the performance of a benchmark method using simulated data.

2. Previous research

The majority of previous relevant research has focused on the steps of map matching, path identification and travel time aggregation (Takada, 2006; Fountain and Smith, 2004; Cayford, 2003; Pyo et al., 2001; Bernstein and Kornhauser, 1996). The steps of filtering and travel time allocation have received little attention in the literature. One reason for the lack of research on these two problems is that they do not arise with dedicated probe systems which typically use on-board GPS with high measurement frequency (e.g. on the order of one or more position measurements per second) and high position accuracy. Furthermore, these systems often have on-board digital road map databases and computational resources for processing the data on-board the vehicle. Consequently, a dedicated probe is able to track its progress along a link with a temporal resolution as low as 1 s or less and directly determine the time it entered and exited each link and thus determine link travel times directly. In contrast, anonymous tracking systems may have lower positioning frequency (on the order of one reading per minute), larger location errors and no on-board processing. As a result of these differences, anonymous tracking systems are able to provide only the reported position with accompanying time stamp to a central data processing center (CDPC). The CDPC has a digital map database and uses appropriate map matching, path identification, and filtering techniques to determine whether or not a probe is a vehicle. If the probe is determined to be a vehicle then the CDPC estimates the vehicle's most likely traversed path between two consecutive matched positions.

Several wireless area-wide road conditions monitoring systems have been developed into commercial products and are now being deployed in North America and elsewhere (Izadpanah and Hellinga, 2007; AirSage, 2006; iTIS Holdings, 2006; Applied Generics, 2004; Cell-Loc, 2002). Unfortunately, due to the proprietary nature of these commercial systems, there is little or no detailed information publicly available regarding the specific models and algorithms used within these systems or how well they perform. Consequently, it is not possible to compare the performance of the travel time allocation model proposed in this paper with the performance of existing commercial systems.

Conventionally, travel time associated with any probe vehicle trajectory can be allocated to the partial links and/or links which constitute the trajectory proportionally to distance, free flow speed, or free flow travel time of the segment. If the route travel time is allocated on the basis of the free flow travel time of individual links and partial links, then both distance and free flow speed are simultaneously considered. In this study, allocation of travel time proportional to free flow travel time is used as the benchmark method against which the performance of the proposed method is compared.

3. Travel time allocation – problem description and solution algorithm

3.1. Network model

Consider a road network consisting of a set of n nodes $\mathbf{N}$ ($\mathbf{N} = \{n_a\}$, $a = 1, 2, \ldots, n$) and a set of m links $\mathbf{L}$ ($\mathbf{L} = \{l(n_a, n_b)|n_a, n_b \in \mathbf{N}\}$). Each node is a geographical location on the road network representing a network feature such as signalized or unsignalized intersections, shape points, dead ends of a road segment, crosswalks, or locations of a change in the road attribute. Each node (n_a) can be defined by its two dimensional coordinates, that is, $n_a = (x_a, y_a)$. Other features associated with the node may be available as part of the map database (e.g. traffic control device such as traffic signal or stop sign, turning movement restrictions, etc.), but are not assumed in the proposed model.

A link is the representation of a road segment connecting two nodes. Each link is assumed to be a directed segment of a straight line in the map database. This assumption ensures the feasibility of inferring the complete link on the basis of the location of its end nodes. It is also assumed that at most one link exists from one node to another and vehicles can traverse each link in only one direction. The link from node n_a to node n_b can be defined as a continuous set of locations, denoted by $l(n_a, n_b)$, that are located on the line between n_a to n_b, that is $l(n_a,n_b) = \{(1-\lambda)n_a + \lambda n_b | 0 \leqslant \lambda \leqslant 1\}$, where λ is a location parameter that is used to identify any location on the link. This representation is convenient because increasing values of λ correspond to the forward movement of vehicles on the link. It is assumed that two attributes are associated with each link, namely, free flow speed and link length. The length of the link may be calculated by taking the Euclidean distance between n_a and n_b, or may be taken directly from the map database. With free flow speed and link length, the free flow travel time of this link can be calculated as the ratio of link length to free flow speed. Other attributes such as number of lanes, vertical gradient, lane width, etc. may be available but are not required by the proposed model.

A sampled mobile probe k ($k = 1, 2, \ldots, K$) periodically reports its locations (Fig. 1). The location reported by mobile probe k at time $t_{k,i}$ ($i = 0,1,2,\ldots$) is denoted as $\tilde{m}_k(t_{k,i})$ and is defined as $\tilde{m}_k(t_{k,i}) = (\tilde{x}(t_{k,i}), \tilde{y}(t_{k,i}))$.

For each reported location $\tilde{m}_k(t_{k,i})$, the map matching process provides an estimate of the true position of the mobile probe. The estimated location for mobile probe k at time $t_{k,i}$ is defined as $\hat{m}_k(t_{k,i}) = (\hat{x}_k(t_{k,i}), \hat{y}_k(t_{k,i}))$. The true location of the mobile probe is defined as $m_k(t_{k,i}) = (x_k(t_{k,i}),\ y_k(t_{k,i}))$. In practice, the map matching process introduces errors and $\hat{m}_k(t_{k,i})$ may not be equal to $m_k(t_{k,i})$. However, the focus of this paper is strictly on the performance of travel time allocation methods and therefore the impact of map matching errors is not considered (i.e. it is assumed that $\hat{m}_k(t_{k,i}) = m_k(t_{k,i})$). Consequently, in the following model developments, $m_k(t_{k,i})$ represents the matched location of mobile probe k at time $t_{k,i}$.

If the mobile probe being tracked is a traveling vehicle with its movement constrained by the road network, then the path identification process estimates the route traveled by the mobile probe between two consecutive locations, $m_k(t_{k,i})$ and $m_k(t_{k,i+1})$, denoted by $r_k(t_{k,i},t_{k,i+1})$, and defined as a sequence of links on the road network

$$r_k(t_{k,i}, t_{k,i+1}) = \left\{ l(m_k(t_{k,i}), n_{a_1}), l(n_{a_1}, n_{a_2}), \ldots, l(n_{a_{J-1}}, n_{a_J}), l(n_{a_J}, m_k(t_{k,i+1})) \right\} \tag{1}$$

Note that the first link and the last link in the route may represent only a portion of a link, depending on the starting and ending location of the probe; these links are therefore called partial links. To simplify our subsequent discussion, the notation of the path is redefined as follows:

$$r_k(t_{k,i}, t_{k,i+1}) = \left\{ l_{k,i,0}, l_{k,i,1}, \ldots, l_{k,i,J(k,i)-1}, l_{k,i,J(k,i)} \right\} = \left\{ l_{k,i,j} | 0 \leqslant j \leqslant J(k,i) \right\} \tag{2}$$

where $J(k,i)$ is a more concise representation of partial link $l(n_{a_J}, m_k(t_{k,i}+1))$.

For the purpose of this research, it is assumed that the traversed path $r_k(t_{k,i},t_{k,i+1})$ has been identified by the preceding steps of the traffic monitoring system and the focus of this research is therefore on the problem of allocating the traversal time, i.e., $(t_{k,i+1} - t_{k,i})$, to the individual links (i.e. $l_{k,i,j}$) on the route $r_k(t_{k,i},t_{k,i+1})$.

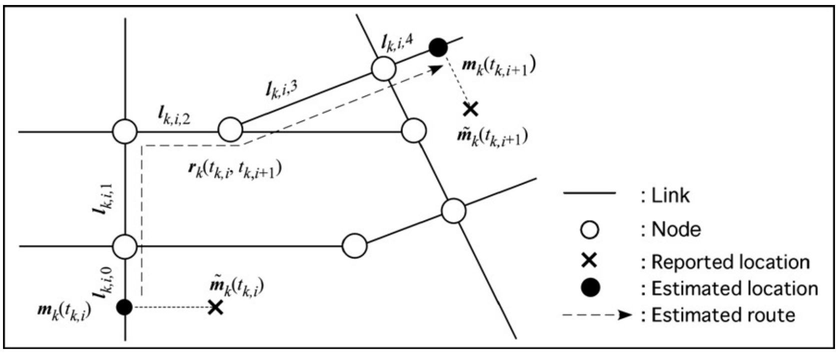

Fig. 1. Definitions for vehicle trajectory.

As illustrated in Fig. 2, the time interval between two consecutive reported locations can be decomposed into four constituent parts as follows:

1. Minimum travel time, or free flow travel time, for the estimated route, which includes free flow travel time plus the minimum transition times (time required when the vehicle is moving from one link to another adjacent link e.g. left turn).
2. Stopping time.
3. Deceleration and acceleration time.
4. Delay due to traffic congestion.

The free flow travel time of a link is calculated as the link length divided by the free flow speed which is assumed to be available from the digital road map database.

$$\tau_f(l(n_a,n_b)) = \frac{|l(n_a,n_b)|}{S_f(n_a,n_b)} \tag{3}$$

where $\tau_f(l(n_a, n_b))$ is the free flow travel time for complete or partial link $l(n_a, n_b)$; $S_f(n_a, n_b)$ is the free flow speed for complete or partial link $l(n_a, n_b)$; $|l(n_a,n_b)|$ is the length of complete or partial link $l(n_a, n_b)$.

The length of link $l(n_a, n_b)$, $|l(n_a,n_b)|$, may be calculated by computing the Euclidean distance between n_a and n_b.

The stopping time, denoted as $\tau_s(l_{k,i,j})$ for complete or partial link $l_{k,i,j}$, reflects the stopped delay caused to the mobile probe by traffic control devices. Note, it is not assumed that detailed information regarding the location, type, and operating characteristics (e.g. signal timing plans) of traffic control devices is known and therefore stopped time cannot be directly estimated using conventional intersection delay estimation methods.

Acceleration and deceleration time is assumed to be included within $\tau_s(l_{k,i,j})$ when these delays are caused by traffic controls and within $\tau_f(l_{k,i,j})$ when caused by geometry. Consequently, acceleration and deceleration time is not separately computed.

The final component of the travel time is the time associated with congestion, denoted as $\tau_c(l_{k,i,j})$. Congestion time results when the mobile probe travels at a speed less than the free speed due to the impedance of other vehicles.

Based on the above definitions, route travel time between two reported locations can be expressed as:

$$t_{k,i+1} - t_{k,i} = \sum_{j=0}^{J(k,i)} \{\tau_f(l_{k,i,j}) + \tau_s(l_{k,i,j}) + \tau_c(l_{k,i,j})\} \tag{4}$$

Note from Eq. (4) that both $\tau_s(l_{k,i,j})$ and $\tau_c(l_{k,i,j})$ are unknown and need to be determined separately before the total travel time can be allocated to individual links. The following section describes the benchmark and proposed methods for calculating these two components.

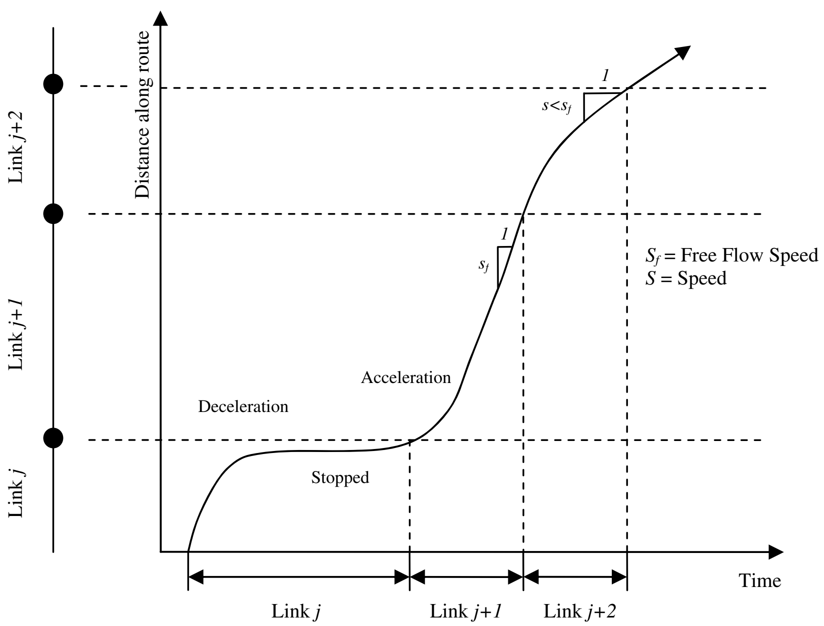

Fig. 2. Travel time components.

3.2. Benchmark travel time decomposition method

The benchmark travel time decomposition method allocates $\tau_s(l_{k,i,j})$ and $\tau_c(l_{k,i,j})$ (i.e. travel time in excess of the free flow travel time) in proportion to the free flow travel time (Eq. (5)).

$$t(l_{k,i,j}) = \frac{\tau_f(l_{k,i,j})}{\sum_{q=0}^{J(k,i)} \tau_f(l_{k,i,q})}(t_{k,i+1} - t_{k,i}) \tag{5}$$

where in Eq. (5) $t(l_{k,i,j}) = \tau_s(l_{k,i,j}) + \tau_c(l_{k,i,j})$. Note that in this method there is no need to explicitly compute $\tau_s(l_{k,i,j})$ and $\tau_c(l_{k,i,j})$.

3.3. Proposed travel time decomposition method

The proposed travel time decomposition method attempts to provide a more accurate allocation of travel time by recognizing that vehicles are more likely to incur stopping delay (i.e. $\tau_s(l_{k,i,j})$) at the downstream rather than upstream end of a link, especially when the link is influenced by a traffic control device. However, we assume that detailed information regarding the location and operating characteristics of traffic control devices (e.g. traffic signals) is not known and therefore it is not possible to directly determine:

(1) the fraction of the total route travel time associated with stopped delay; and
(2) where along the route (i.e. which links) the probe vehicle experiences this stopped delay.

In the next two sections we present a method for over-coming both of these challenges.

3.3.1. Computing congestion time

The proposed approach to determining link congestion time is based on the assumption that the degree of congestion on each of the links on the route is nearly equal. This assumption is considered to be reasonable when the number of links on $r_k(t_{k,i},t_{k,i+1})$ is relatively small, which is expected to be the case when cellular phones are polled at an interval of 1 min or less. When traffic demand is low, delays due to congestion should be relatively small and therefore similar. On the other hand, when traffic demand is high, all links within close proximity are expected to experience relatively similar degrees of congestion. Congestion due to unexpected events tends to spread quickly over a number of links as drivers seek alternate routes and queues grow. Furthermore, probes will traverse fewer links (or partial links) within a polling interval as congestion increases.

We define a congestion index, w, as the ratio of the congestion time on the route to the sum of the congestion time and free speed time on the route (Eq. (6))

$$w = \frac{\sum_{j=0}^{J(k,i)} \{\tau_c(l_{k,i,j})\}}{\sum_{j=0}^{J(k,i)} \{\tau_c(l_{k,i,j}) + \tau_f(l_{k,i,j})\}} \tag{6}$$

Using this definition, the minimum value of w is zero and occurs when traffic demand is very low and the probe travels at the free speed. The maximum value of w is always less than 1.

We re-write Eq. (6) by expressing the time associated with congestion along the route $r_k(t_{k,i}, t_{k,i+1})$, which is denoted by τ_c (i.e. $\tau_c = \sum_{j=0}^{J(k,i)} \{\tau_c(l_{k,i,j})\}$), as a function of the unknown congestion index and the known free flow travel (Eq. (7)).

$$\tau_c = \frac{w}{1-w} \sum_{j=0}^{J(k,i)} \{\tau_f(l_{k,i,j})\} \tag{7}$$

The minimum value of τ_c is 0 and occurs when $w = 0$. The maximum value of τ_c occurs for the maximum value of w, which occurs when vehicles travel at a speed less than free flow speed due to traffic congestion and experience no delay caused by traffic control devices. This maximum value is obtained by substituting $\tau_s(l_{k,i,j}) = 0$ into Eq. (5), and defined as follows:

$$\tau_{c,\max} = T_c(k,i) = t_{k,i+1} - t_{k,i} - \sum_{j=0}^{J(k,i)} \{\tau_f(l_{k,i,j})\} \tag{8}$$

Using Eqs. (7) and (8), the maximum value of w can be determined by

$$w_{\max} = \frac{\tau_{c,\max}}{t_{k,i+1} - t_{k,i}} \tag{9}$$

It is not possible to resolve Eq. (7) at this point because τ_c is a function of w which is unknown. To resolve this issue, it is further assumed that the degree of congestion on the route traversed by the mobile probe during the most recent polling interval is not substantially different from the degree of congestion experienced on the route traversed by this same mobile probe during the previous polling interval. Based on this assumption, we introduce a model to capture the likelihood that a certain degree of congestion is experienced by a mobile probe when traversing a given link k, denoted by $P_w(k,i,w)$.

$$P_w(k,i,w) = \min\left(1, \frac{T_c(k, I_p(i)) + T_c(k,i)}{t_{k,I_p(i)+1} - t_{k,I_p(i)} + t_{k,i+1} - t_{k,i}} \frac{1}{w}\right) \quad (10)$$

where $I_p(i)$ is the largest integer less than i for which $T_c(k,I_p(i))$ is less than $t_{k,I_p(i)+1} - t_{k,I_p(i)}$. This requirement ensures that for the previous route being considered, the mobile probe has not remained stationary for the entire polling interval.

Eq. (10) is structured to reflect two assumptions:

First, it is assumed that when all other attributes are held constant, the likelihood of a particular level of congestion occurring increases as the maximum delay due to congestion (i.e. $T_c(k,i)$) increases.

Second, it is assumed that for a given maximum delay due to congestion, very high levels of congestion are less likely than lower levels of congestion.

These two assumptions and the relationship defined by Eq. (10) are illustrated in Fig. 3 for two sample cases. For each case, the polling interval (i.e. $t_{k,i+1} - t_{k,i}$) is assumed to be 30 s and the maximum delay due to congestion for the probe's route during the previous interval (i.e. $T_c(k,I_p(i))$) is assumed to be 5 s.

The impact of assumption 1 is observed by comparing the values of $P_w(k, i, w)$ for case 1 and 2 for a given value of w (say $w = 0.6$). For case 1, the maximum delay due to congestion is 5 s and for case 2 is 15 s. Given that all other attributes between the two cases are the same, it is expected that the link in case 2 is more heavily congested. This is reflected by the higher likelihoods for all levels of congestion for case 2 compared with case 1.

The impact of assumption 2 is reflected in both curves by the decrease in $P_w(k, i, w)$ for increasing level of congestion.

3.3.2. Computing stopping time

Stopping time is associated with the delay experienced as a result of stopping for a traffic control device. However, it is not possible to determine directly if the mobile probe has stopped along the route, and if it has stopped, where the probe stopped and for how long. Furthermore, it is assumed that specific characteristics of the road network, such as the location of traffic signals, stop signs, etc, are not known and therefore it is not possible to develop models that rely on signal timing information, etc. Nevertheless, it can be assumed that if a traffic control device exists on a link, it is located at the downstream end of the link, and therefore, if a vehicle stops on a link, it is more likely to do so near the downstream end of the link than at the upstream end.

It is also expected that the queues created by traffic control devices are likely to become larger as the level of congestion increases, and therefore, the likelihood of stopping at a location near the upstream end of the link increases as the level of congestion increases.

The probability of stopping is defined on the basis of the stopping likelihood function provided in Eq. (11).

$$h(\lambda, w) = (1-w)e^{p(\lambda-1)} + C_2 w \quad (11)$$

where $p = \frac{C_1}{w}$ and C_1 and C_2 are model parameters that are used to reflect the stopping likelihood pattern of a link.

The likelihood is a function of both the position on the link (λ) and the level of congestion (w) and reflects the expectation that when a link experiences relatively low levels of congestion, queues formed upstream of traffic control devices are relatively short and vehicles that are required to stop as a result of the queue tend to do so near the downstream end of the link. However, as the link becomes increasingly more congestion, queues become longer, and the likelihood that a vehicle is required to stop farther upstream also increases. As illustrated in Fig. 4, the proposed likelihood function reflects the impact of location on the link and level of congestion in a way that is consistent with traffic engineering expectation. Consequently, any other function that behaves in the same manner can be chosen as long as the range of the likelihood function is

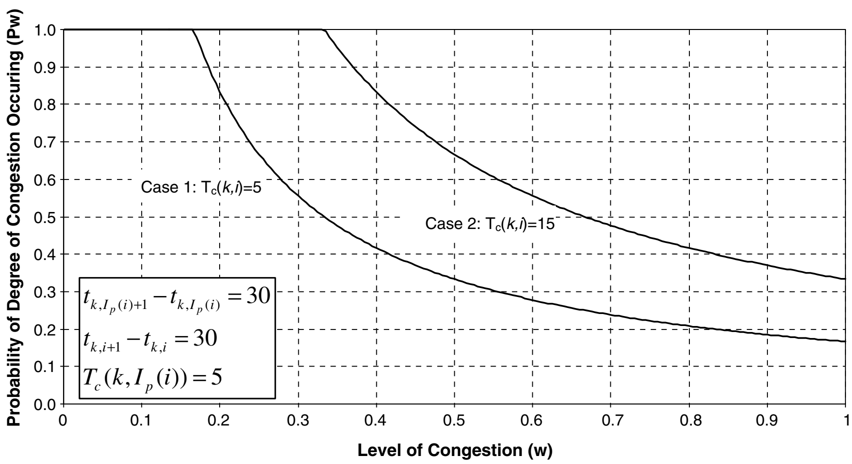

Fig. 3. Probability of congestion as a function of level of congestion.

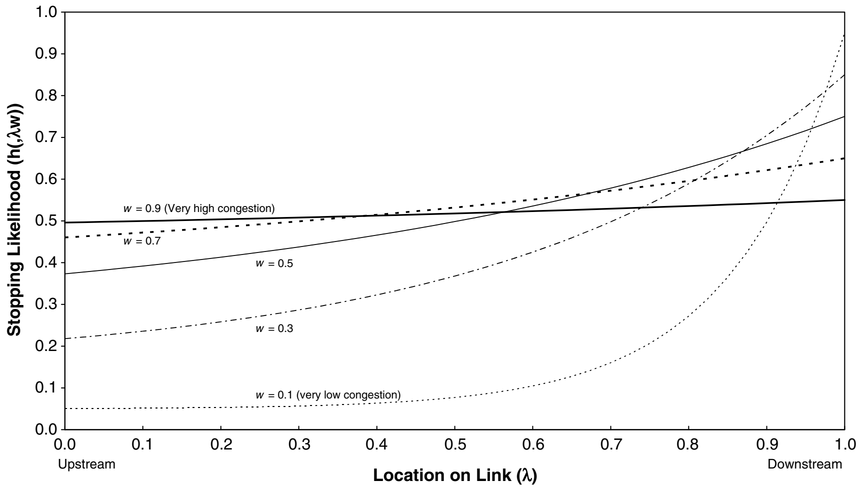

Fig. 4. Stopping likelihood as a function of level of congestion and location on links.

constrained between 0 and 1. The parameters C_1 and C_2 are chosen in a way to ensure that the range of the function is between 0 and 1. A sensitivity analysis is presented at the end of the paper to clarify the importance of these two parameters on the accuracy of the proposed travel time allocation method.

The probability of stopping on a link $l_{k,i,j}$ which is on the route $r_k(t_{k,i},t_{k,i+1})$ can be determined by integrating the likelihood function along the length of the link:

$$\begin{aligned} H_s(l_{k,i,j},w) &= \frac{1}{\lambda_2-\lambda_1}\int_{\lambda_1}^{\lambda_2} h(\lambda,w)\mathrm{d}\lambda \\ &= \frac{1}{\lambda_2-\lambda_1}\int_{\lambda_1}^{\lambda_2}\{(1-w)\mathrm{e}^{p(\lambda-1)}+C_2w\}\mathrm{d}\lambda \\ &= \frac{1}{\lambda_2-\lambda_1}\left[\frac{1-w}{p}\mathrm{e}^{p(\lambda-1)}+C_2w\lambda\right]_{\lambda_1}^{\lambda_2} \\ &= \frac{1-w}{p(\lambda_2-\lambda_1)}\left(\mathrm{e}^{p(\lambda_2-1)}-\mathrm{e}^{p(\lambda_1-1)}\right)+C_2w \end{aligned} \tag{12}$$

If it is assumed that a mobile probe stops at most once on the route $r_k(t_{k,i}, t_{k,i+1})$ then the probability of stopping on link $l_{k,i\,J}$ is given by

$$P_s(l_{k,i,J},w) = \begin{cases} H_s(l_{k,i,0},w) & \text{if } J(k,i)=0 \\ H_s(l_{k,i,J,w})\prod_{j\neq J} 1-H_s(l_{k,i,j},w) & \text{otherwise} \end{cases} \tag{13}$$

Finally, the estimated stopping time can be obtained by integrating over the whole range of possible levels of congestion (w) on the link.

$$\tau_s(l_{k,i,J}) = \int_0^{w_{\max}} \tau_s \frac{P_w(k,i,w)P_s(l_{k,i,J},w)}{Q_s(k,i)}\mathrm{d}w \tag{14}$$

where

$$\tau_s = t_{k,i+1}-t_{k,i}-\sum_{j=0}^{J(k,i)}\{\tau_f(l_{k,i,j})\}-\tau_c \tag{14a}$$

$$Q_s(k,i) = \int_0^{w_{\max}} P_w(k,i,w)\sum_{j=0}^{J(k,i)} P_s(l_{k,i,j},w)\mathrm{d}w \tag{14b}$$

In Eq. (14a), τ_s denotes stopping time for the router$_k(t_{k,i},t_{k,i+1})$. Then the estimated congestion time can be obtained by integrating over the whole range of w,

$$\tau_c(l_{k,i,J}) = \int_0^{w_{\max}} \delta_{k,i,J}\tau_c \frac{\sum_{j=0}^{J(k,i)} P_w(k,i,w)P_s(l_{k,i,J},w)}{Q_s(k,i)}\mathrm{d}w \tag{15}$$

where:

$$\delta_{k,i,J} = \frac{\tau_f(l_{k,i,J})}{\sum_{j=0}^{J(k,i)}(\tau_f(l_{k,i,j}))} \tag{15a}$$

Eq. (15) implies that the time associated with traffic congestion is assigned to each link according to the proportion of the minimum travel time of the link to the minimum travel time of the route. Finally,

$$t(l_{k,i,j}) = \tau_f(l_{k,i,j}) + \tau_s(l_{k,i,j}) + \tau_c(l_{k,i,j}), j = 0, 1, 2, \ldots, J(k,i) \tag{16}$$

where $t(l_{k,i,j})$ denotes travel time of link or partial link j in time interval ($t_{k,i}$,$t_{k,i+1}$) when mobile probe k is tracked.

4. Illustrative example

Consider a mobile probe that traverses a portion of a road network. The route traversed during time interval ($t_{k,i}$, $t_{k,i+1}$) consists of two partial links and one complete link (Fig. 5). Assume $t_{k,i+1} - t_{k,i} = 60$ s. The free flow travel time (i.e.$\tau_f(l_{k,i,j})$) of each complete link is equal to 15 s. However, the mobile probe traverses only 2/3 of the first link and 1/3 of the last link on the route and therefore the free speed travel times can be determined as 10, 15, and 5 s, respectively.

On the basis of Eq. (8), $T_c(k,i) = 60 - (10 + 15 + 5) = 30$ s. Assume that the minimum travel time for the previous route traversed by the mobile probe was greater than zero and therefore, $I_p(i) = i-1$. Assume $T_c(k,i-1) = 5$ s and $t_{k,i} - t_{k,i-1} = 90$ s. Consequently, $P_w(k,i,w)$ can be computed using Eq. (10) for each value of w (e.g. for $w = 0.3$, $P_w(k,i,w) = 0.78$).

The probability of the mobile probe stopping on each of the three links on the route can be computed using Eq. (12). For example, for $w = 0.3$, $p = 2.33$, $C_1 = 0.7$, $C_2 = 0.5$, and for the first link on the route $\lambda_1 = 1/3$, $\lambda_2 = 1.0$ and thus

$$H_s(l_{k,i,0}, 0.3) = \frac{1-0.3}{2.33(1.0-0.333)}(e^{2.33(1-1)} - e^{2.33(0.333-1)}) + 0.5(0.3) = 0.505$$

Similarly, $H_s(l_{k,i,1},0.3) = 0.421$ and $H_s(l_{k,i,2},0.3) = 0.253$. Eq. (13) is used to determine the probability of stopping on each link.

$$\begin{aligned} P_s(l_{k,i,0}, 0.3) &= H_s(l_{k,i,0}, 0.3)1 - H_s(l_{k,i,1}, 0.3)1 - H_s(l_{k,i,2}, 0.3) \\ &= (0.505)(1-0.421)(1-0.253) \\ &= 0.219 \end{aligned}$$

Similarly, $P_s(l_{k,i,1}, 0.3) = 0.156$ and $P_s(l_{k,i,2}, 0.3) = 0.072$. The sum of these probabilities = 0.447. From Eq. (7), for $w = 0.3$, $\tau_c = 12.9$ s. From Eq. (14b), $Q_s(k,i) = 0.154$ and the term $\sum_{j=0}^{J(k,i)} P_w(k,i,w)P_s(l_{k,i,J}, w)$ in Eq. (15) is equal to 0.347. For the first link, from Eq. (15a) $\delta_{k,i,0} = (10/30)$. From Eq. (8), when $\tau_c = T_c(k,i) = 30$ s, $w_{\max} = (30/60) = 0.5$.

The calculation of $\tau_c(l_{k,i,0})$ is completed by integrating over all levels of w between 0 and 0.5 (as per Eq. (15)) resulting in $\tau_c(l_{k,i,0}) = 3.63$ s. Similarly, $\tau_c(l_{k,i,1}) = 5.44$ and $\tau_c(l_{k,i,2}) = 1.81$ s.

From Eq. (14a), for $w = 0.3$, $\tau_s = 17.1$ s and from Eq (14), $\tau_s(l_{k,i,0}) = 9.81$ s. Similarly, $\tau_s(l_{k,i,1}) = 6.84$, and $\tau_s(l_{k,i,2}) = 2.47$ s. Table 1 tabulates the calculated values for each component of travel time associated with each link.

The calculations associated with the benchmark method are straightforward. For example, the travel time associated with the first link ($j = 0$) is equal to the polling interval duration ($t_{k,i} - t_{k,i-1}$) multiplied by the free flow travel time for the link divided by the free flow travel time for the route ($90 \times 10/30 = 20$ s).

Table 1
Sample calculations

		Link number			Sum
		$j = 0$	1	2	
Proposed method	τ_f	10.00	15.00	5.00	30.00
	τ_s	9.81	6.84	2.47	19.12
	τ_c	3.63	5.44	1.81	10.88
	Total	23.44	27.28	9.28	60.00
Benchmark method	Total	20.00	30.00	10.00	60.00

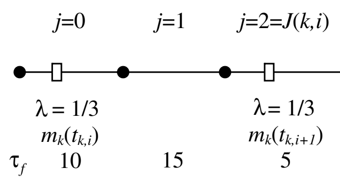

Fig. 5. Example path of mobile probe.

5. Performance evaluation

5.1. Evaluation network

The performance of the proposed method was evaluated using a hypothetical arterial network shown in Fig. 6. Traffic flows on the network were modeled using the INTEGRATION microscopic traffic simulation model (Van Aerde, 2002a,b). The test network is composed of 32 links, 18 nodes, 3 signalized intersections and 1 unsignalized intersection. All links were assigned a free speed of 60 km/h and a saturation flow rate of 1900 passenger car per hour. The network was simulated for a period of 25 min. Data from the first 5 min of simulation were considered to be part of the warm-up period and were not used in the analysis.

Fig. 7 illustrates the link traversal times experienced by individual vehicles as a function of simulation time for two sample links – link 13 which is controlled by a traffic signal and link 3 which in not controlled by any type of traffic control device (Fig. 6). The results in the figure for the link controlled by the traffic signal clearly reflect the significant influence that the traffic signal has on the link traversal times. Vehicles that arrive at the intersection just as the signal turns red may experience link traversal times that are approximately eight times as large as the traversal times of vehicles that incur no signal delay.

In this study all generated vehicles are treated as probe vehicles and positions of each probe vehicle are reported at a pre-defined fixed frequency. It should be noted that parameters C_1 and C_2 that are used in Eq. (11) were assumed to be 0.7 and 0.5, respectively.

5.2. Performance measures

An observation is considered to be two consecutive location references for an individual probe vehicle. Each observation may be categorized into one of three types as illustrated in Fig. 8:

1. Both of the reported positions lie on the same link (Fig. 8a).
2. The first and second reported positions are located on adjacent links (Fig. 8b).
3. At least one full link exists between the two reported positions (Fig. 8c).

In the first case, the vehicle path consists of only a partial link. In the second case, the vehicle path consists of two partial links. In the third case, the route between the two consecutive reported locations is composed of a partial link at both ends of the route and at least one full link in between.

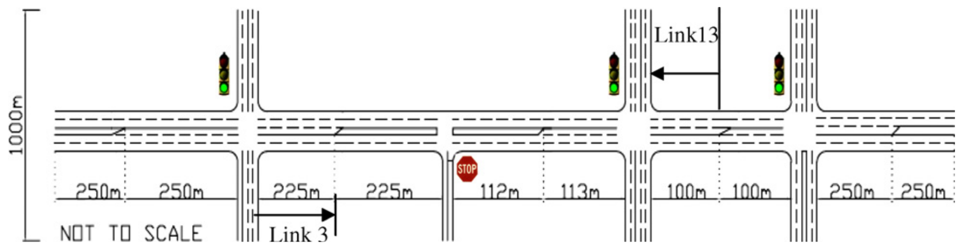

Fig. 6. A hypothetical network to evaluate performance of the proposed method.

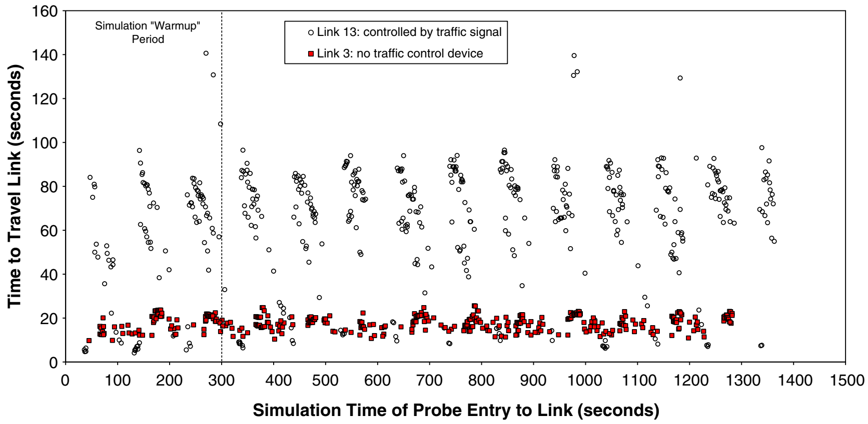

Fig. 7. Variation of travel time for two sample links.

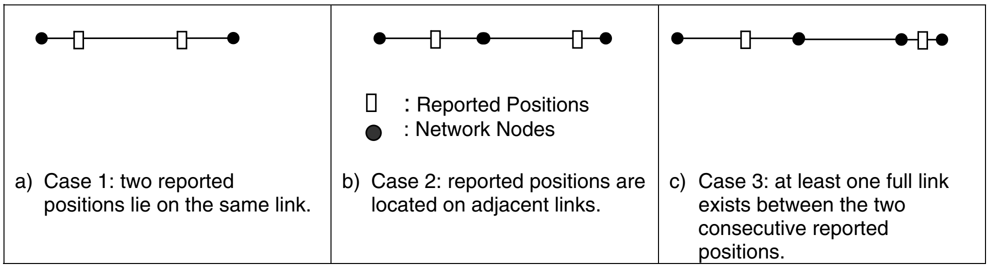

Fig. 8. Three different travel time decomposition cases that can occur.

For the first case, the travel time of the partial link is equal to the polling interval of the location regardless of the method used. However, this is not true for cases 2 and 3 and therefore, for these cases, the travel allocation method used does have an impact on the accuracy.

In this study, the link travel time estimation errors associated with each partial or full link for each probe vehicle route is determined by comparing the link (or partial link) travel time estimated by the proposed method and the benchmark method with the corresponding true travel time as extracted directly from the simulation model.

The performance of the proposed method and the benchmark is quantified on the basis of two measures of accuracy. The first measure, $E_{l(n_a,n_b)}$, defined in Eq. (17), can be used to compare the performance of the two travel time estimation methods at the individual link level.

$$E_{l(n_a,n_b)} = \frac{1}{\mathrm{ATT}_{l(n_a,n_b)}} \sqrt{\frac{\sum_{r=1}^{N}(TA_r - TT_r)^2}{N}} \tag{17}$$

where $E_{l(n_a,n_b)}$ is the average normalized error associated with travel time allocation for link $l(n_a, n_b)$; $R_{l(n_a,n_b)}$ is the observation set for link $l(n_a, n_b)$; r is the index denotes any individual observations in $R_{l(n_a,n_b)}$; N is the number of observations in $R_{l(n_a,n_b)}$; TA_r is the allocated travel time of observation r, which is calculated based on either the proposed or the benchmark travel time allocation method; TT_r is the true travel time of observation r; $\mathrm{ATT}_{l(n_a,n_b)}$ is the average travel time of all observation in $R_{l(n_a,n_b)}$, i.e., $\mathrm{ATT}_{l(n_a,n_b)} = \frac{1}{N}\sum_{r=1}^{N} TT_r$.

The second measure, $\overline{E}$, is an aggregate measure of performance and is obtained by averaging the link level error, $E_{l(n_a,n_b)}$, estimated using Eq. (17), over all links in the network.

$$\overline{E} = \frac{1}{|\mathbf{L}|} \sum_{j \in \mathbf{L}} E_j \tag{18}$$

where $\overline{E}$ is the average error for the network; $\mathbf{L}$ is the set of all links in the network; $|\mathbf{L}|$ is the dimension of the set of all links, $\mathbf{L}$; E_j is the the error obtained for link j using Eq. (17).

5.3. Results

Fig. 9 depicts the relationship between aggregate travel time estimation error ($\overline{E}$) and polling interval duration for both the proposed and benchmark methods. As can be seen in these results, the proposed method is superior to the benchmark method for all polling interval durations examined. Furthermore, the results show that estimation error is smallest for very short polling interval duration, but increases rapidly as the polling interval duration increases until a maximum error plateau is reached ($\overline{E} \approx 0.75$ for the benchmark method at a polling interval duration of 60 s; $\overline{E} \approx 0.65$ for the proposed method at a polling interval duration of 100 s).

The relative improvement in estimation accuracy provided by the proposed method can be computed as $(\overline{E}_B - \overline{E}_P)/\overline{E}_B$ where, $\overline{E}_P$ and $\overline{E}_B$ are the overall error for the network obtained using the proposed and the benchmark methods, respectively. The results in Fig. 9 suggest that the proposed method provides a reduction in overall estimation error of approximately 40% for polling interval durations of 35 and 60 s. The improvements are smaller for other polling interval durations (25% for a polling interval duration of 15 s; 14% for 90 s; and 9% for 100 s).

The results depicted in Fig. 9 can be explained on the basis of the proportion of case 1, 2, and 3 observations (as per Fig. 8) for each of the polling interval durations. Fig. 10 provides the fraction of observations of each type for each different polling interval duration. As expected, the fraction of case 3 observations increases as the polling interval becomes larger. Recall, that there is no error associated with the travel time allocation for case 1 observations and also it is expected that the travel time allocation error for case 3 is larger than for case 2. Consequently, overall estimation error is strongly correlated with the proportion of case 2, and case 3 observations and strongly negatively correlated with the proportion of case 1 observations.

It can be observed in Fig. 9 that for each estimation method, the estimation error is approximately equal for polling interval durations of 100 s and 90 s. In Fig. 10, it can be observed that for these two polling interval durations, the proportions of case 2 and 3 observations are very similar.

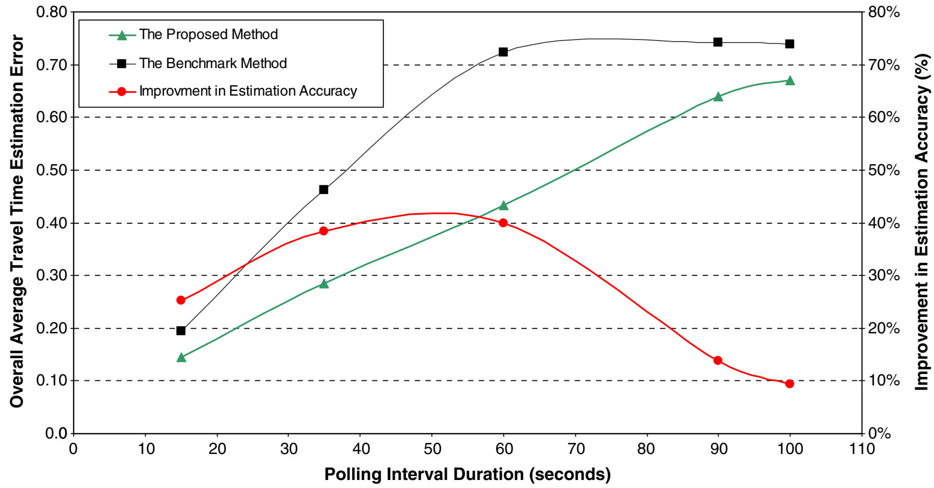

Fig. 9. Overall average travel time estimation error ($\overline{E}$).

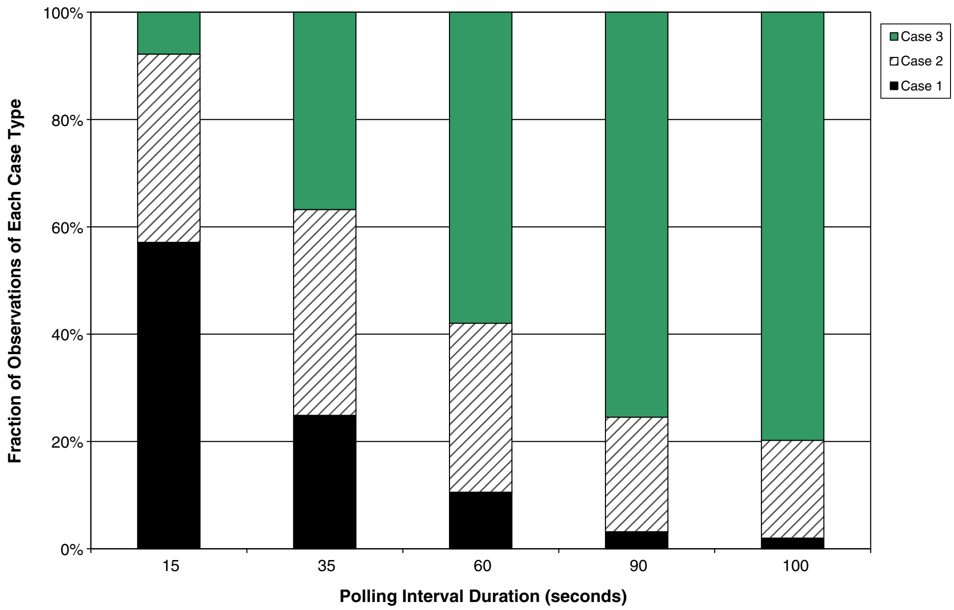

Fig. 10. Fraction of observations of each case type as a function of polling interval duration.

Neither the proposed method nor the benchmark method assumes that information is available specifying the location and characteristics of traffic control devices. However, it is quite clear from the results provided in Fig. 8 that the link travel time characteristics of the links controlled by a traffic signal are significantly different from those links that are not controlled by a signal. Consequently, it is of interest to examine the accuracy of the proposed method and the benchmark method for each link class separately. Figs. 11 and 12 depict estimation error for the proposed and benchmark methods, respectively as a function of link class (i.e. controlled by a traffic signal or not) and the ratio of polling interval duration to the free flow travel time. This ratio is used for the *x*-axis as it simultaneously captures the impact of polling interval duration (and it's impact on the proportion of case 1, 2, and 3 observations) and link length (which also influences the proportion of each case type).

The following observations can be made from the results in Figs. 11 and 12:

1. Estimation error is generally larger for links not controlled by a traffic signal than for links that are controlled by a traffic signal. The larger error appears to result from a tendency for the proposed method to over estimate the time allocated to the unsignalized links. This observation seems to imply that overall estimation accuracy can be improved if the location of traffic control devices is known.

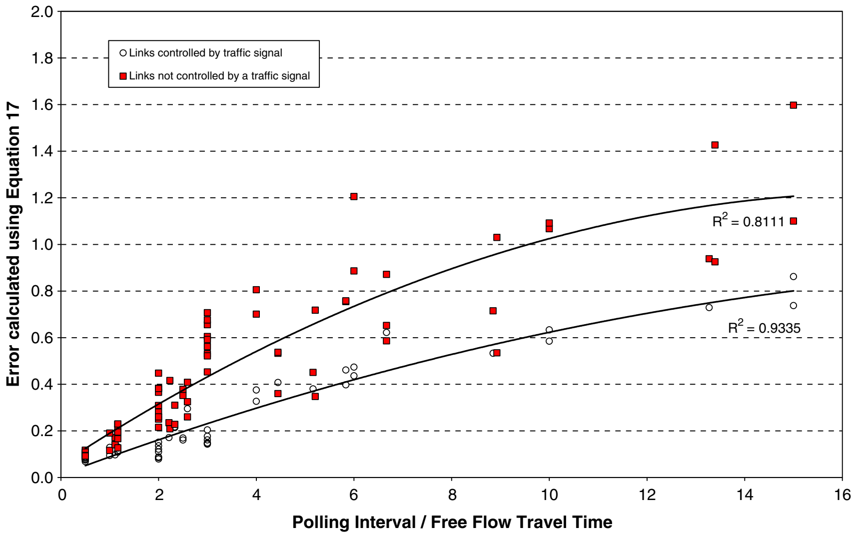

Fig. 11. The proposed method estimation error as a function of link class.

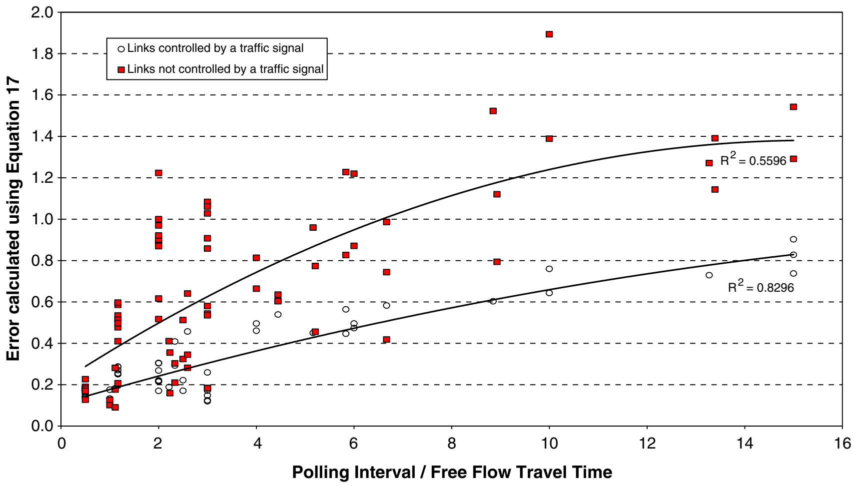

Fig. 12. The benchmark method estimation error as a function of link class.

2. As the polling interval to free flow travel time ratio increases the estimation error also increases but at a decreasing rate. A 2nd order polynomial regression, of the form

$$\widehat{E} = \beta_0 + \beta_1\alpha + \beta_2\alpha^2$$

was performed on the data from each link class separately. β_0, β_1, and β_2 are the regression coefficients; α is the ratio of polling interval duration to the free speed travel time of the link; and $\widehat{E}$ is the regression prediction of the estimation error.
The resulting regression relationships are illustrated in the figures, and with the exception of the model for the links not controlled by a traffic signal for the benchmark method, explain a large proportion of the variance exhibited in the data. The coefficients for each model are provided in Table 2.
3. The estimation errors for the proposed method are smaller than the errors from the benchmark method for links that are controlled by a traffic signal and links that are not controlled by a signal. The relative improvement in estimation accuracy is more clearly depicted in Fig. 13 which illustrates the percent reduction in error provided by the use of the proposed method for both link classes. The reduction in estimation error was computed as

$$\Delta(\alpha) = \frac{\widehat{E}_B(\alpha) - \widehat{E}_P(\alpha)}{\widehat{E}_B(\alpha)} \tag{19}$$

Table 2
Regression model coefficients

Coefficient	Regression model			
	Proposed method		Benchmark method	
	Signalized	Not signalized	Signalized	Not signalized
β_0	0.0093	0.0592	0.1093	0.2134
β_1	0.0782	0.1393	0.069	0.1523
β_2	−0.0018	−0.0042	−0.0014	−0.005

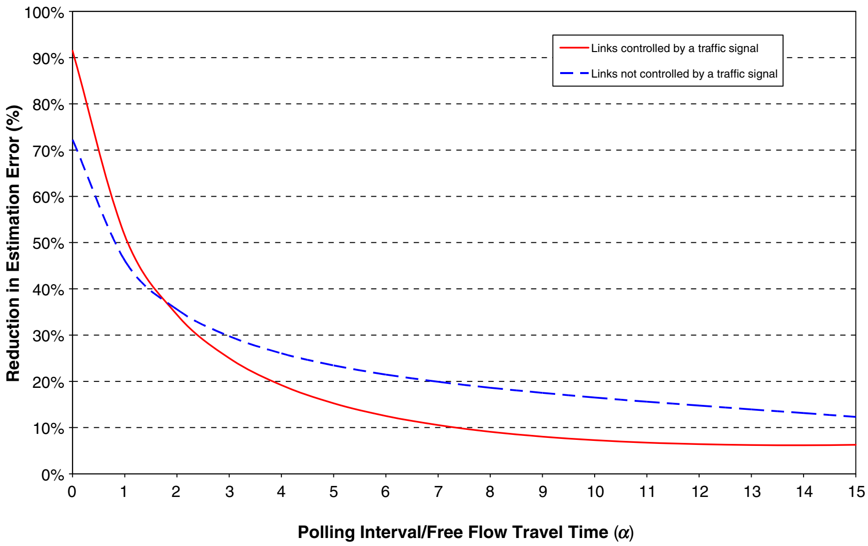

Fig. 13. Relative improvement in estimation accuracy provided by the proposed travel time allocation method.

where $\Delta(\alpha)$ is the fractional reduction in estimation error provided by the use of the proposed method for a value of α; $\widehat{E}_B(\alpha)$ is the estimation error as predicted by the appropriate regression model fitted to the data obtained from use of the benchmark method for a value of α; $\widehat{E}_P(\alpha)$ is the estimation error as predicted by the appropriate regression model fitted to the data obtained from use of the proposed method for a value of α; α is the ratio of the polling interval duration to the free speed travel time for the link.

The results from Fig. 13 indicate that for a short polling interval duration or for long links (i.e. α is small), the proposed method provides a reduction in estimation error of approximately 90% for links that are controlled by a traffic signal and approximately 70% for links not controlled by a traffic signal. The relative improvement provided by the proposed method decreases as the polling interval duration increases and/or the links become shorter; however, over the range of polling interval durations and link lengths considered, the proposed method is superior to the benchmark method.

In Fig. 13, the improvement provided by the proposed method is larger for links controlled by a traffic signal until the ratio of polling interval duration to free flow travel time approaches a value of 2. When the ratio of polling interval duration to free flow travel time exceeds a value of approximately 2, the relative improvement provided by the proposed method is larger for links not controlled by a traffic signal. However, this result should be viewed with caution as it is dependent on the regression model results and the model for links not controlled by a traffic signal for the benchmark has a relatively low R^2 value.

6. Sensitivity analysis

Eq. (11) provides a model which describes stopping likelihood of vehicles along a link. According to this equation, the stopping likelihood is a function of both position on the link (λ) and the level of congestion (w). Two model parameters denoted by C_1 and C_2 were used to reflect the stopping likelihood pattern of the link. In the previous sections of this paper values of C_1 = 0.7 and C_2 = 0.5 were used. There exists the issue of how to select appropriate values and the sensitivity of the performance of the proposed travel time allocation method (in terms of the average error for the network ($\overline{E}$)) to these two model parameters.

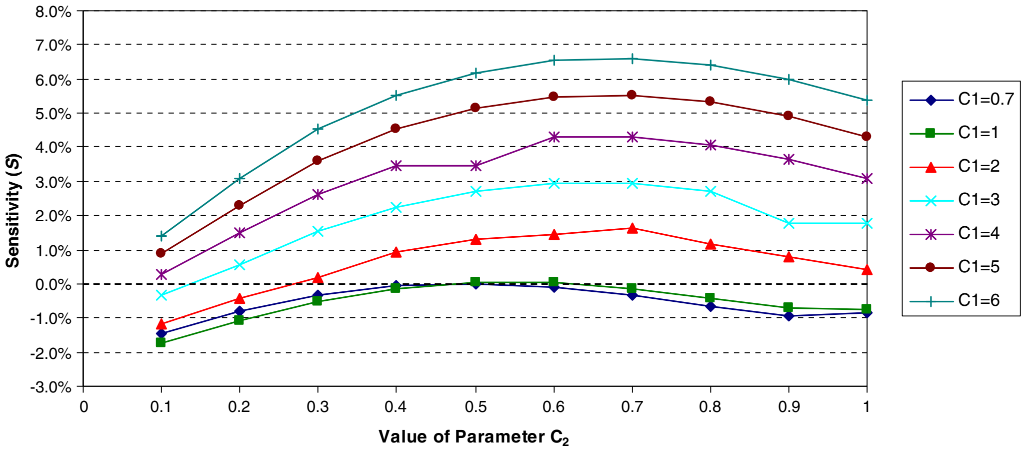

Fig. 14. Sensitivity of average error for the network, $\overline{E}$ to model parameters C_1 and C_2 for polling interval duration of 60 s.

Examination of Eq. (11) shows that the stopping likelihood increases for larger values of C_2. Furthermore, in order to ensure that the stopping likelihood function varies between 0 and 1, C_2 must be restricted to be between 0 and 1. Conversely, as C_1 increases, the stopping likelihood deceases (though remains positive) and consequently, C_1 may take on any positive value.

The sensitivity of the performance of the proposed travel time allocation method to the value of C_1 and C_2 is illustrated in Fig. 14. The y-axis represents the sensitivity (S) measured as

$$S = \frac{\overline{E}_{C_1,C_2} - \overline{E}_{0.7,0.5}}{\overline{E}_{0.7,0.5}} \tag{20}$$

where $\overline{E}_{C_1,C_2}$ is the average network error obtained from the proposed travel time allocation method using parameter values C_1 and C_2; $\overline{E}_{0.7,0.5}$ is the average network error obtained from the proposed travel time allocation method using parameter values $C_1 = 0.7$ and $C_2 = 0.5$.

The polling interval was held equal to 60 s. Every individual curve in the figure is associated with a constant value for C_1. The results demonstrate that the proposed travel time allocation method is relatively insensitive to the values of C_1 and C_2. Over the range of values considered in the sensitivity analysis, the overall network error only changes by between −1.7% and 6.6%. The proposed travel time estimation method performs better than the benchmark method for all parameter value (C_1 and C_2) combinations considered in the sensitivity analysis.

7. Conclusions

This research addresses the problem of travel time allocation which is one of the steps required to obtain average travel time for individual links in a road network on the basis of position data obtained from anonymously tracked probe vehicles. In this study a method for travel time allocation was proposed and the performance of the proposed method was evaluated against a benchmark using data from a simulated network. It was found that the frequency of the location referencing and the link free speed travel time are important factors influencing the accuracy of the travel time estimates.

It was observed that the proposed method improves the accuracy of the travel time allocation by an average of 40% for all links in the network for a polling interval duration of 60 s. However, improvements can be as large as 90% for long links that are controlled by a traffic signal.

It is observed that the proposed method tends to over estimate the travel time for links that are not controlled by a traffic signal. Consequently, if the location and type of traffic control device was assumed to be known (not an unreasonable assumption given that some electronic map databases already contain this information) then it is likely that this information could be incorporated within a modified form of the proposed method in order to improve the accuracy of the travel time estimates.

It is recommended that future research efforts address the following: (a) develop a travel time allocation method that can make use of information about the type and location of traffic control devices; (b) quantify the importance of travel time allocation accuracy on the accuracy of aggregate link travel time estimates; (c) explore methods by which travel times estimated for partial links can be used to estimate aggregate link travel times; and (d) evaluate the proposed method for an actual road network.

Acknowledgement

This research was financially supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

AirSage Company Website, 2006. Accessed from <http://www.airsage.com>.
Applied Generics Ltd., 2004. RoDIN24 road traffic monitor (GSM ed.) (accessed from <http://www.appliedgenerics.com>).
Bernstein, C., Kornhauser, A.L., 1996. An introduction to map matching for personal navigation assistants. New Jersey, TIDE Center.
Cayford, R., 2003. Method and system for electronically determining dynamic traffic information. United States Patent No. 6,560,532.
Cayford, R., Yim, Y., 2006. A field operation test using anonymous cell phone tracking for generating traffic information. In: The 85th Annual Meeting of the Transportation Research Board CD-ROM, Washington, DC, USA.
Cell-Loc Inc., 2002. Traffic monitoring application of cellular positioning technology: proof of concepts (accessed from <http://www.tc.gc.ca/tdc/summary/13900/13936e.htm>).
Drane, C, Rizos, C., 1998. Positioning systems in intelligent transportation systems. Artech House, INC., ISBN: 0-89006-536-5.
Fountain, M.D., Smith, B., 2004. Improving the effectiveness of traffic monitoring based on wireless location technology. Virginia Transportation Research Council, Final Report, VTRC 05-R17.
iTIS Holdings Company Website, 2006. (accessed from <http://www.itisholdings.com>).
Izadpanah, P., Hellinga, B., 2007. Wide-area wireless traffic conditions monitoring: reality or wishful thinking? In: Proceedings of Annual Conference of the Canadian Institute of Transportation Engineers, Toronto, Ontario.
Lovell, D.J., 2001. Accuracy of speed measurements from cellular phone vehicle location systems. Journal of Intelligent Transportation Systems 3, 303–325.
Pyo, J.S., Shin, D.H., Sung, T.K., 2001. Development of a map matching method using the multiple hypothesis technique. In: Proceedings of IEEE Intelligent Transportation Systems Conference, Oakland, CA, USA.
Takada, H., 2006. Road traffic condition acquisition via mobile phone location referencing. PhD Dissertation, University of Waterloo, Ontario, Canada.
Van Aerde, M., Assoc., Ltd., 2002a. Integration© Release 2.30 for Windows: User's Guide – vol. I: Fundamental Model Features.
Van Aerde, M., Assoc., Ltd., 2002b. Integration© Release 2.30 for Windows: User's Guide – vol. II: Advanced Model Features.