Bhaskar, Ashish, Chung, Edward, & Dumont, Andre-Gilles
(2009)
Estimation of travel time on urban networks with midlink sources and sinks.
Transportation Research Record, *2121*, pp. 41-54.

https://doi.org/10.3141/2121-05

PAPER SUBMITTED FOR TRANSPORTATION RESEARCH BOARD 88TH ANNUAL MEETING ,
JANUARY 13-17, 2008

Travel time estimation on urban networks with mid-link sources and sinks

Ashish Bhaskar*
EPFL-ENAC-ICARE-LAVOC
Station 18, Lausanne, Switzerland
Ph: +41 21 693 2341;
Fax: +41 21 693 63 49
ashish.bhaskar@epfl.ch

Edward Chung
EPFL-ENAC-ICARE-LAVOC
Station 18, Lausanne, Switzerland
Ph: +41 21 693 2343
Fax: +41 21 693 63 49
edward.chung@epfl.ch

André-Gilles Dumont
EPFL-ENAC-ICARE-LAVOC
Station 18, Lausanne, Switzerland
Ph: +41 21 693 2345
Fax: +41 21 693 63 49
andre-gilles.dumont@epfl.ch

15/11/2008

*Corresponding Author

Number of Words in the Text	4747
Number of Figures in the Text	11(=2750 words)
Total Equivalent Words	7497

ABSTRACT

A model for average travel time estimation on signalized urban networks by integrating cumulative plots with probe vehicle information is presented. The integration is performed with the aim to reduce the relative deviations in the cumulative plots occurred due to mid-link sources/sinks. During under-saturated traffic conditions, concept of virtual probe is introduced and the accurate travel time can be obtained even with unavailability of real probe. For saturated and over-saturated traffic conditions and travel time estimation interval of 360 seconds: only one probe per estimation interval or 3% of vehicles traversing the link as probes has the potential to provide accurate travel time.

Keywords: *Urban travel time, Cumulative plots, Probe vehicles*

INTRODUCTION

Reducing congestion maximizes the efficiency and capacity of the network. Travel time information has the potential to reduce congestion on both temporal and spatial scale. For the efficient management of the whole network, monitoring traffic conditions of the urban routes in addition to the freeways is inevitable for *ITS*.

Travel time for a vehicle is the time needed to travel from point upstream (*u/s*) to point downstream (*d/s*) on the network. Here, we are interested in estimating average travel time for all the vehicles that departs downstream during certain travel time estimation interval (T_{EI}).

Literature Review

Different traffic data retrieval systems used for travel time estimation can be broadly categorized into fixed sensors and mobile sensors.

Fixed sensors such as inductive loop detectors provide temporal traffic state information, though only point-based. Loop detectors are the oldest and most widely used traffic data sources and hence, majority of travel time estimation models are based on detector data. Researchers have proposed a number of models with various degrees of complexities ranging from simple regression based [1-9], traffic flow theory based [10, 11], pattern recognition [12-21], to advance neural network based [22-25].

Regression based, pattern recognition based and neural network based models are data driven and are “non transferable”.

Regression based model defines its parameters by “best fitting” the observed data. Such models are unable to predict travel time for traffic conditions that are different from those assumed in the models formulation. Nevertheless, they are simple and fast to compute and perhaps are favorable for transport planning and policy applications. Generally, they are not suitable for *ITS* applications where more accurate and reliable travel time in real-time is required.

Neural network based models are more robust than regression based as they utilize the data to build the model structure as well as its parameters. These require large number of observations and hence are computationally intensive. Moreover, they can be like a black box; and care should be taken to verify the reliability of the output and that the model is applied well within the limits for which it is trained. Nevertheless, neural network is applied in different engineering discipline and many promising results have been reported in literature.

Pattern recognition models, such as k-NN technique, match the current traffic pattern with historical database and can fail to predict travel time for traffic conditions absent in database.

Mobile sensors such as probe vehicle provide accurate travel time for the vehicle. They represent the random sample of the population of the vehicles traversing the link. Therefore, average travel time estimation for all the vehicles traversing the link can be estimated by statistical sampling techniques [26]. Researchers [27] have shown interest to determine minimum number of probes during each estimation interval required for statistically significant travel time estimation.

Researchers have also applied data fusion techniques [28-33] to fuse data from different sources, specifically detector and probe vehicles, with the aim to improve the accuracy and reliability of the estimates.

Majority of the above research is limited to freeways, and cannot be applied to urban networks, where problem is rather more challenging due to number of reasons for instance: vehicle interaction with external control such as signals; significant proportions of flow from/to mid-link sources/sinks etc. There are avenues for improvement in travel time prediction models especially in terms of transferability and robustness with respect to urban complexities.

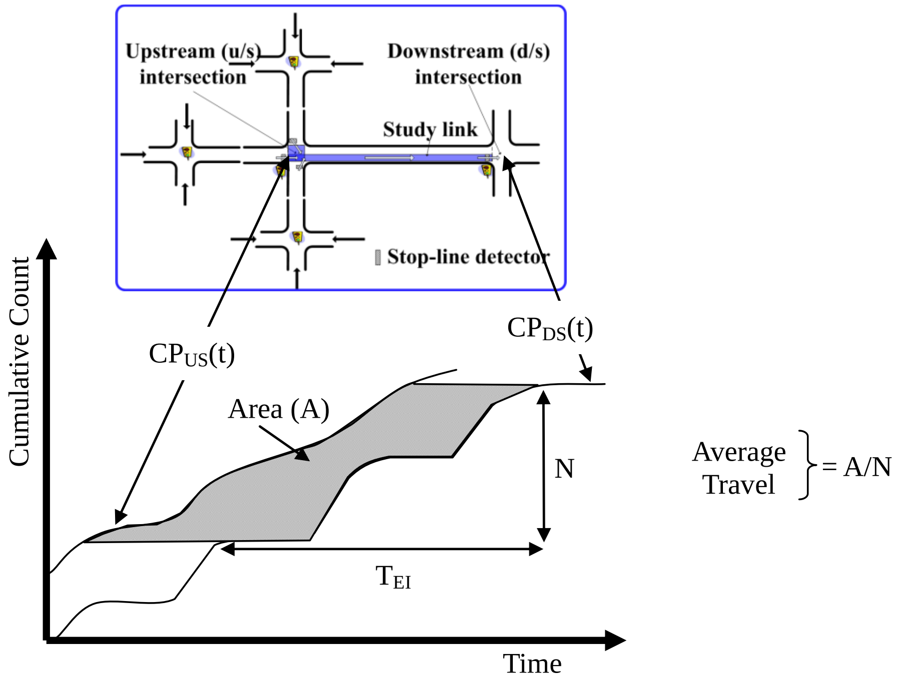
This paper is structured as follows: first the effect of mid-link sources/sinks on classical analytical methodology, based on cumulative plots, for travel time estimation is discussed, followed by the relation between the cumulative plots and probe vehicle. Thereafter, algorithm steps are discussed. Finally, the results for model testing controlled environment are presented.

CUPRITE MODEL

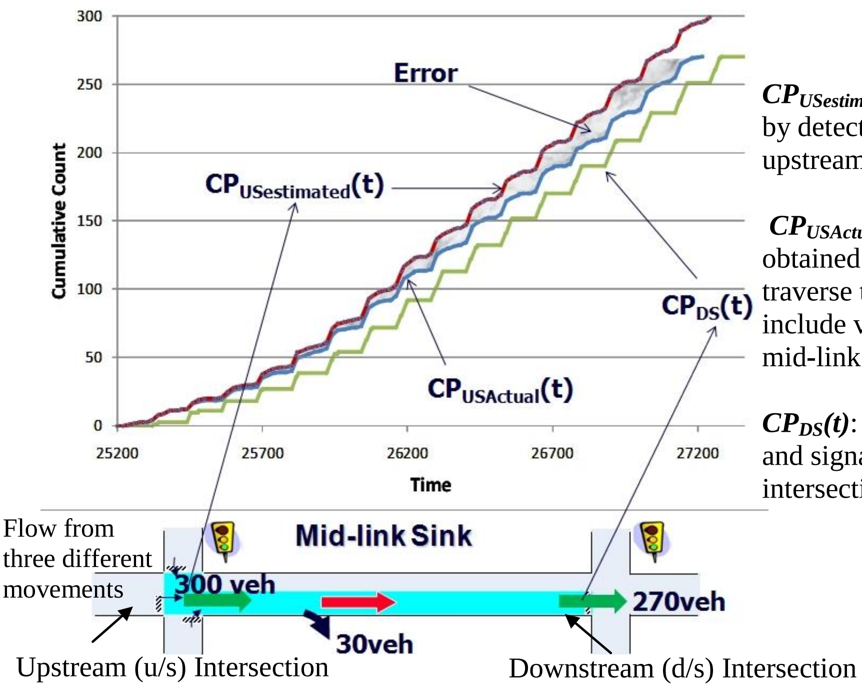
The classical analytical principle for travel time estimation defines, total travel time (*FIGURE 1* a) for all the *N* vehicles, that departs during T_{EI} at the downstream, as area *(A)* between cumulative plots at upstream $(CP_{US}(t))$ and at downstream $(CP_{DS}(t))$. Average travel time per vehicle is *A/N*.

If data from stop-line loop detector (counts) and signal controller (signal green time and red time) is available then accurate cumulative plots can be obtained [34] by integrating detector data with signal controller data. However, for real application there are certain issues to be addressed such as, a) relative deviation in cumulative plots due to mid-link sources and sinks (e.g. side street, parking etc.), and detector counting error; and b) unknown real turning proportions. Real turning proportions are required to estimate i) proportion of the flow to the study link from shared-use lane at upstream intersection and ii) cumulative plot for each movement of the multi-lane link.

In this paper, it is assumed that stop-line loop detector and signal controller data is available and cumulative plots are generated. The gain/loss of vehicles from mid-link source/sink is unknown. The model, CUmulative plots & PRobe Integration for Travel timE estimation (*CUPRITE*), developed in this paper addresses the issue of relative deviation in cumulative plots due to mid-link sources and sinks by integrating the cumulative plots with probe data. For under-saturated situation virtual probes are defined and for saturated and over-saturated situations real probe data is used.



(a) Classical Analytical Procedure for travel time estimation



CP*$_{USestimated}$*(t) : *CP*$_{US}$*(t)* estimated by detector and signal controller at upstream intersection.

CP*$_{USActual}$*(t): Actual *CP*$_{US}$*(t)* obtained from vehicles which traverse the whole link. It does not include vehicles which are lost in mid-link sink.

CP*$_{DS}$*(t): Obtained from detector and signal controller at downstream intersection.

(b) Effect of mid-link sink on the classical analytical procedure

FIGURE 1a) Illustration of classical analytical procedure b) Effect of mid-link sink on the classical analytical procedure.

Effect of Mid-link Sources/Sinks on Cumulative Plots

An urban link can have different mid-link infrastructures such as, a side street, parking etc. Depending on the time of the day or day of the week, these mid-link infrastructures can act as sink, source or both. A significant proportion of the flow can be from/to the mid-link source/sink. This proportion is a dynamic entity and varies with time and one can easily observe on average around *10%* loss (or gain) to a side street.

FIGURE 1 b), presents an example where *300* vehicles are observed at upstream and *10%* of the vehicles are lost in the mid-link sink (one-way side street) resulting in only *270* vehicles observed at downstream. In this example $CP_{US}(t)$ is overestimated resulting in relative deviation from that of $CP_{DS}(t)$. The shaded area in the figure represents the error in travel time estimation and if left unchecked, can exponentially grow with time.

For a mid-link source, there will be more counts at downstream than that at upstream, i.e. $CP_{US}(t) < CP_{DS}(t)$. In such situations area between the plots is negative and hence travel time cannot be obtained.

Probe vehicle data and Cumulative plots

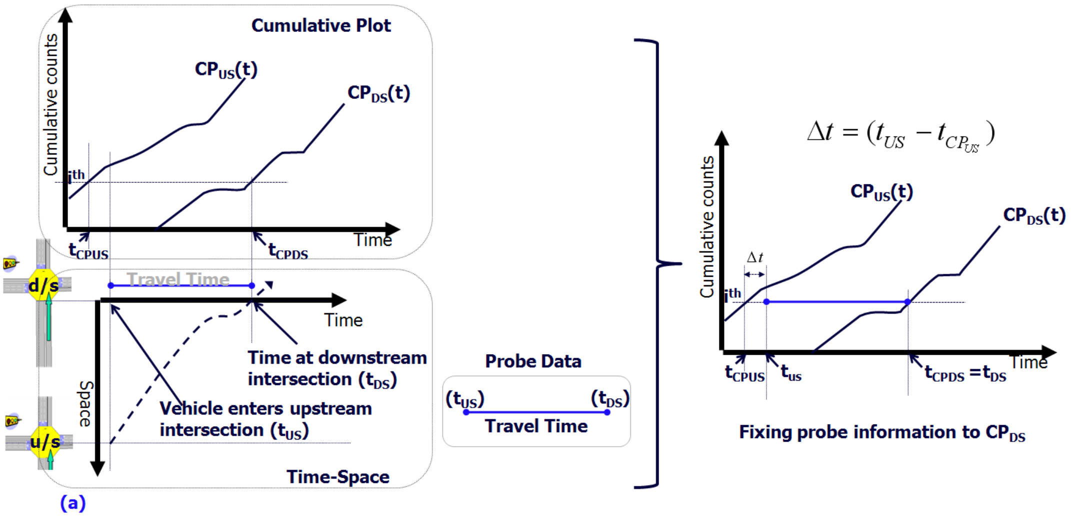
Here probe vehicles are the vehicles which can provide time information when at the intersection (position where cumulative plots are generated). Generally probe vehicles, such as taxi fleets, are equipped with *GPS* and can provide data for their position and time. To address the issues with probe vehicles such as frequency of data, map-matching of data, urban cannon etc. is beyond the scope of this paper. We assume that the time when probe is at upstream (t_{US}) and downstream (t_{DS}) intersection can be accurately obtained.

Under First-In-First-Out (*FIFO*) traffic condition the horizontal distance between the plots provides travel time for the i^{th} vehicle and the time when it is at upstream (t_{CPUS}) and downstream (t_{CPDS}). If we fix the probe information to the downstream cumulative plot (*FIGURE 2*a) i.e., $t_{DS} = t_{CPDS}$, then the probe vehicle is the i^{th} vehicle in the cumulative plots and we define $\Delta t = t_{US} - t_{CPUS}$.

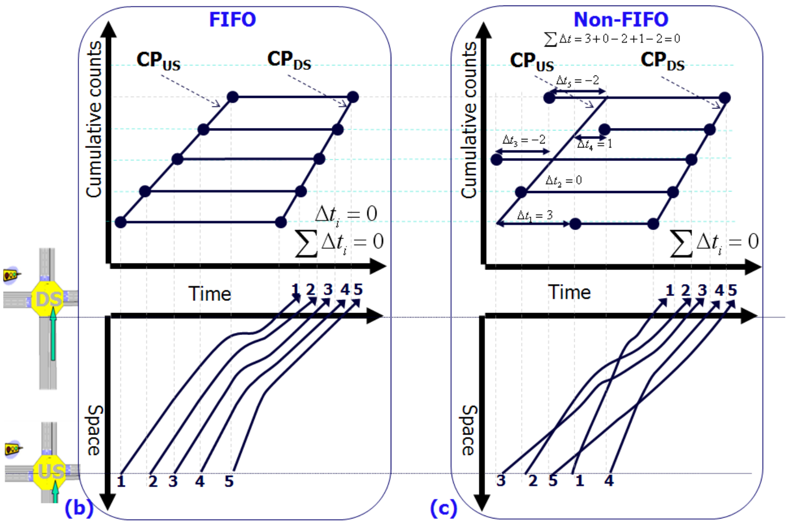
If there is no relative deviation in the cumulative plots then for *FIFO* each Δt should be zero (FIGURE 2b) and for *non-FIFO* condition (FIGURE 2c) Δt may or may not be zero. However, sum of Δt for all the vehicles in the cumulative plots should be zero ($\sum\Delta t=0$). Due to this property the area between the plots represents total travel time, as long as all the vehicles represented in $CP_{US}(t)$ are also represented at $CP_{DS}(t)$.

The above property is when the summation is performed for all the vehicles (populations). However, probe vehicles are only a random sample from the population. We make a hypothesis that relative deviation in the cumulative plots can be reduced by fixing the probe information to $CP_{DS}(t)$ (or $CP_{US}(t)$) and redefine $CP_{US}(t)$ (or $CP_{DS}(t)$) such that property of $\sum\Delta t = 0$ is satisfied.

In this paper we fix the probe information to $CP_{DS}(t)$ and define the set of points through which $CP_{US}(t)$ should pass.



Fixing probe information to downstream intersection cumulative plots



Probe vehicle and cumulative plots for *FIFO* and *non-FIFO* situation

FIGURE 2 Probe vehicle and cumulative plots.

How to redefine $CP_{US}(t)$?

The information from the probe is utilized to define points through which $CP_{US}(t)$ should pass. Say, we have: a) reference point $(t_{Ref}, CP(t_{Ref}))$, i.e., the point in which we have confidence that it is a correct point on the plot; and b) point (t_{topass}, Y_{topass}) through which $CP_{US}(t)$ should pass. Then, (eq (1) and *FIGURE 3*, left diagram) we redefine $CP_{US}(t)$ by applying correction on it such that all points on the plot:

i. before time t_{Ref} have zero correction;
ii. between t_{Ref} to t_{topas} are scaled vertically; and
iii. beyond t_{topass} are shifted vertically so that the redefined curve is parallel to $CP_{US}(t)$ and is continuous.

$$CP(t) = CP(t) + Correction$$

$$Correction = \begin{cases} 0 & \forall\ t \le t_{Ref} \\ (scale - 1) * (CP(t) - CP(t_{\mathrm{Ref}})) & \forall\ t_{Ref} < t < t_{topass} \\ (scale - 1) * (CP(t_{topass}) - CP(t_{\mathrm{Ref}})) & \forall\ t \ge t_{topass} \end{cases} \quad (1)$$

$$scale = \begin{cases} \dfrac{Y_{topass} - CP(t_{\mathrm{Ref}})}{CP(t_{topass}) - CP(t_{\mathrm{Ref}})} & \text{if } CP(t_{topass}) \ne CP(t_{\mathrm{Ref}}) \\ 1 & \text{if } CP(t_{topass}) = CP(t_{\mathrm{Ref}}) \end{cases}$$

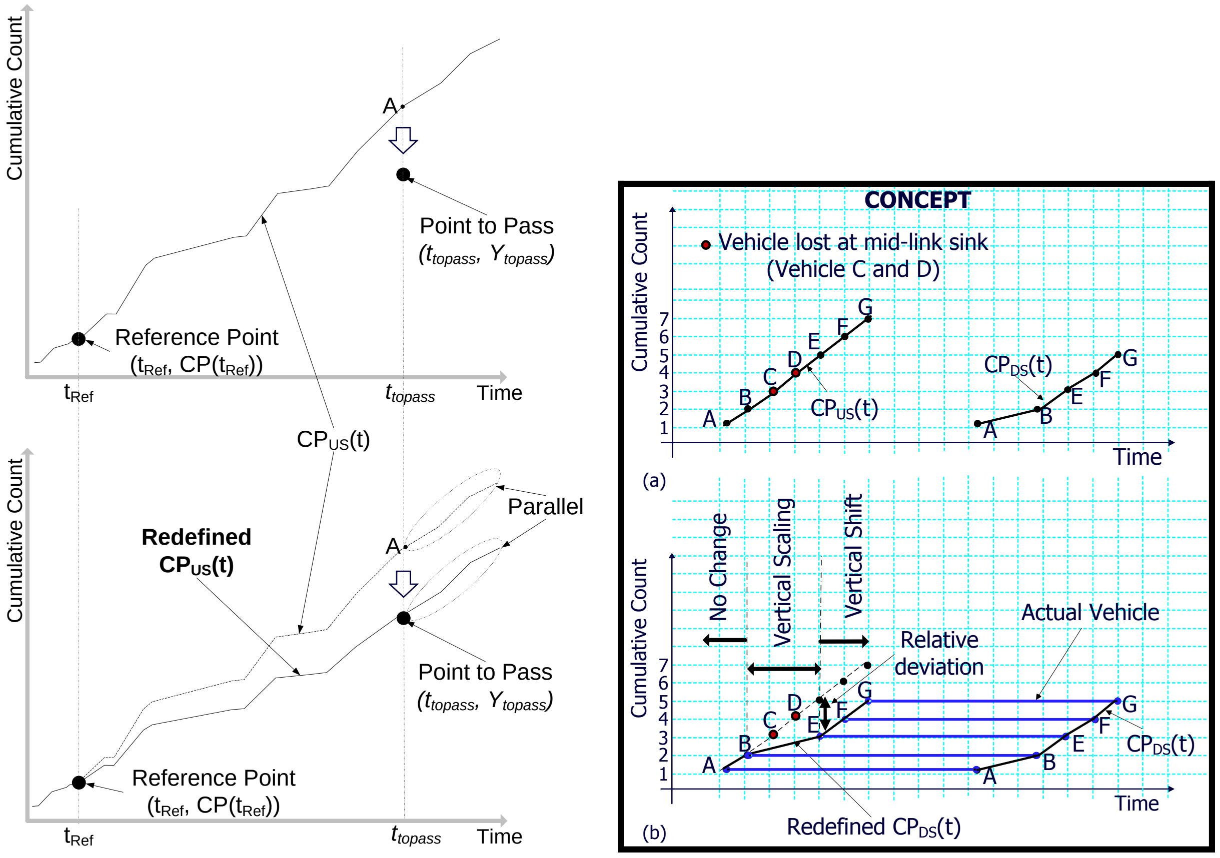


FIGURE 3 Redefining $CP_{US}(t)$ based on vertical scaling and shifting.

Let us consider an example. *FIGURE 3a*, has seven vehicles ($CP_{US}(t)$) detected at upstream (*A* to *G*) and two of them (*C* and *D*) are for mid-link sink therefore, at downstream ($CP_{DS}(t)$) only five vehicles are detected. For simplicity assuming *FIFO* discipline. The rank of vehicles *E*, *F* and *G* are *5*, *6* and 7 at $CP_{US}(t)$ and the *3*, *4* and *5* at $CP_{DS}(t)$, respectively. The presence of mid-link source/sink only affects the rank of the vehicle in the plots which results in relative deviation between the plots. In *FIGURE 3b*, the information for departing vehicle is fixed to $CP_{DS}(t)$ and thereafter $CP_{US}(t)$ is redefined. Before point *B* there is no change; between *B* and *E* it is scaled vertically; and after *E* it is shifted vertically. The vertical distance defines the relative deviation; hence the correction is applied only on the vertical axis and not on the horizontal axis.

To define the points from where $CP_{US}(t)$ should pass, *CUPRITE* considers probe. Also if following "*conditions for virtual probe*" are satisfied then the information is incorporated.

Virtual Probe:
Virtual probe is defined as virtual vehicle, that during under-saturated traffic flow, departs from the downstream at the end of signal green interval and its travel time is free-flow travel time of the link. The probe is not real and is defined with the aim of reducing the relative deviation amongst the cumulative plots.

Here, traffic signal control cycle is from start of effective red interval (t_{RS}) to end of effective green interval (t_{GE}). For under-saturated traffic conditions vehicle queue should vanish at the end of each signal cycle and travel time for the vehicle entering the intersection during the end of signal cycle should be close to free-flow travel time (t_{ff}) of the link. Therefore, during under-saturated traffic conditions we can define virtual probe such that it is observed at upstream and downstream at time $t_{GE} - t_{ff}$ and t_{GE}, respectively (i.e. for virtual probe $t_{US} = t_{GE} - t_{ff}$ and $t_{DS} = t_{GE}$.) (*FIGURE 4*). Note: virtual probe is only defined if the following *"conditions for virtual probe "* are met:

1. ***Absence of source for significant mid-link delay***: Travel time of the virtual probe is free-flow travel time of the link, therefore on the study link the following sources for significant mid-link delay should be absent:
   a. *Mid-link intersection:* $CP_{US}(t)$ and $CP_{DS}(t)$ should be for stop-line location of two consecutive intersections. Else, unknown delay at mid-link intersection(s) results in non free-flow conditions.
   b. *Mid-link on-street bus stop:* On-street bus stop blocks the flow of vehicles following the bus resulting in non free-flow conditions.
2. ***No-Leftover-Queue:*** Virtual probe is defined only for under-saturated condition with logic of zero queue length at the end of signal green interval. Traffic condition is defined as under-saturated if counts during the signal cycle (or more specifically during signal green time) is less than the corresponding capacity i.e., $\mathrm{CP}(\mathrm{t_{GE}}) - \mathrm{CP}(\mathrm{t_{RS}}) < s * g$; where: $CP(t)$ is the cumulative count at time t; s, g and $s*g$ are saturation flow rate, effective signal green time and capacity, respectively[1]. To define the above equation it is assumed that there is no spill-over from downstream link. Else, vehicles are restricted to flow resulting in low counts at stop-line detector. Capacity is generally not corrected to account for the spill-over from downstream link due to which the above equation is satisfied and system indicates under-saturated situation. Though, for spill-over cases, queue may or may not vanish and hence virtual probe should not be defined.
3. ***Presence* of relative deviation in cumulative plots:** If all the above conditions are met, then theoretically relative deviations amongst the cumulative plot exist if $t_{GE} - CP^{-1}_{US}(CP_{DS}(t_{GE})) \neq t_{ff}$. t_{ff} is a statistical estimator and its actual value can vary from driver to driver. Moreover, practically there can be presence of minor mid-link delays such as interaction with the vehicles from the mid-link source/sinks/pedestrian, etc. Therefore, certain confidence should be taken into account to define if there is a presence of relative deviation in cumulative plots. Hence, to define virtual probe the following equation should be satisfied: $CP^{-1}_{US}(CP_{DS}(t_{GE})) \notin [t_{ff} - \delta, t_{ff} + \delta]$; where δ

[1] To take into account the error in estimation of capacity we can re-write the equation as: $s * g - \left[\mathrm{CP}(\mathrm{t_{GE}}) - \mathrm{CP}(\mathrm{t_{RS}})\right] > \Delta$; where: Δ is a calibration parameter.

is a calibration parameter taking into account the variation in the estimation of t_{ff}. It can be considered equal to the standard deviation of the estimate of t_{ff}.

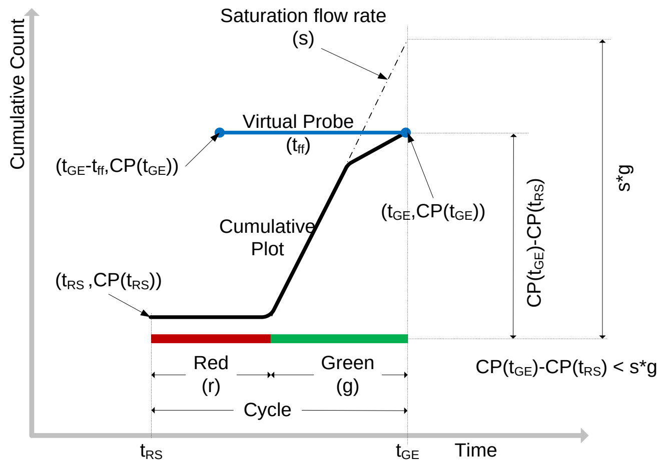


FIGURE 4 Virtual Probe during under-saturated situation.

How to define the points from where the plot should pass?

Say, we have *n* probe vehicles and the database for the probe is defined as list of $[t_{US}]$ and list of $[t_{DS}]$ where the size of each list is *n*. The value of j^{th} element in the list represents the data from the j^{th} probe.

The list $[t_{US}]$ and $[t_{DS}]$ can be appended with additional elements satisfying the *"conditions for virtual probe"*. For this, if the conditions are satisfied then for each under-saturated signal cycle: time corresponding to the end of the green time (t_{GE}) is appended to the list $[t_{DS}]$; and (t_{GE}- t_{ff}) is appended to the list $[t_{US}]$.

Following steps defines the points from where $CP_{US}(t)$ should pass:

Step1: Sort list $[t_{DS}]$ in ascending order of its values. This is required as the probe information is fixed with respect to $CP_{DS}(t)$.

Step2: Sort list $[t_{US}]$ in ascending order of its values. This is required to make sure that the redefined CP_{US} : a) is monotonically increasing; and b) satisfies $\sum \Delta t = 0$.

Step3: The required points through which $CP_{US}(t)$ should pass are (t_{USj}, $CP_{DS}(t_{DSj})$); where t_{USj} and t_{DSj} are j^{th} value in the sorted list of $[t_{US}]$ and $[t_{DS}]$, respectively.

For better understanding an example is presented in FIGURE 5 where we have four probes and corresponding list of $[t_{US}]$ and $[t_{DS}]$. The example is for *non-FIFO* with $t_{DS1}<t_{DS2}<t_{DS3}<t_{DS4}$ and $t_{US2} <t_{US1}<t_{US4}<t_{US3}$.

How to define the reference points?

$CP_{US}(t)$ and $CP_{DS}(t)$ are initially two independent cumulative plots. When the traffic condition is free-flow (for instance during night) then counts for cumulative plots can be initialized to zero. This is the initial reference point (P_0). Say $[P_1, P_2, P_3, ..., P_n]$ is the list of *n* points from where $CP_{US}(t)$ should pass then for redefining $CP_{US}(t)$ for point P_i, the reference point is P_{i-1}.

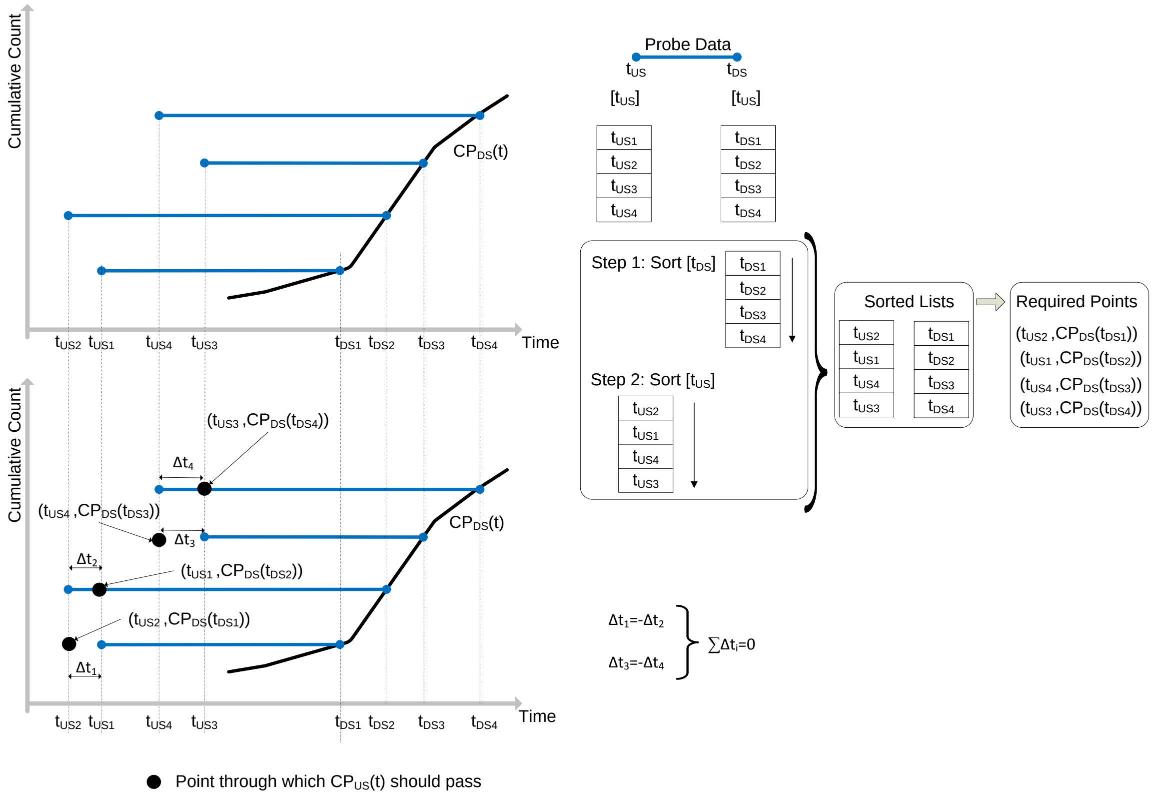


FIGURE 5 Points from where the plot should pass.

Summary of the Algorithm

The summary of the algorithm is as follows (*FIGURE 6*):

1. Cumulative plots are defined by integrating signal controller data with detector data.
2. Probe data (list of $[t_{US}]$ and $[t_{DS}]$) is defined by fixing real probe data with $CP_{DS}(t)$.
3. The above list is appended by virtual probe data if conditions for virtual probe are satisfied.
4. Points through which $CP_{US}(t)$ should pass are defined.
5. $CP_{US}(t)$ is redefined by vertical scaling and shifting the plots so that it passes through the points defined in Setp 4.
6. Finally, average travel time is defined as the ratio of the area between the plots and number of vehicles departing.

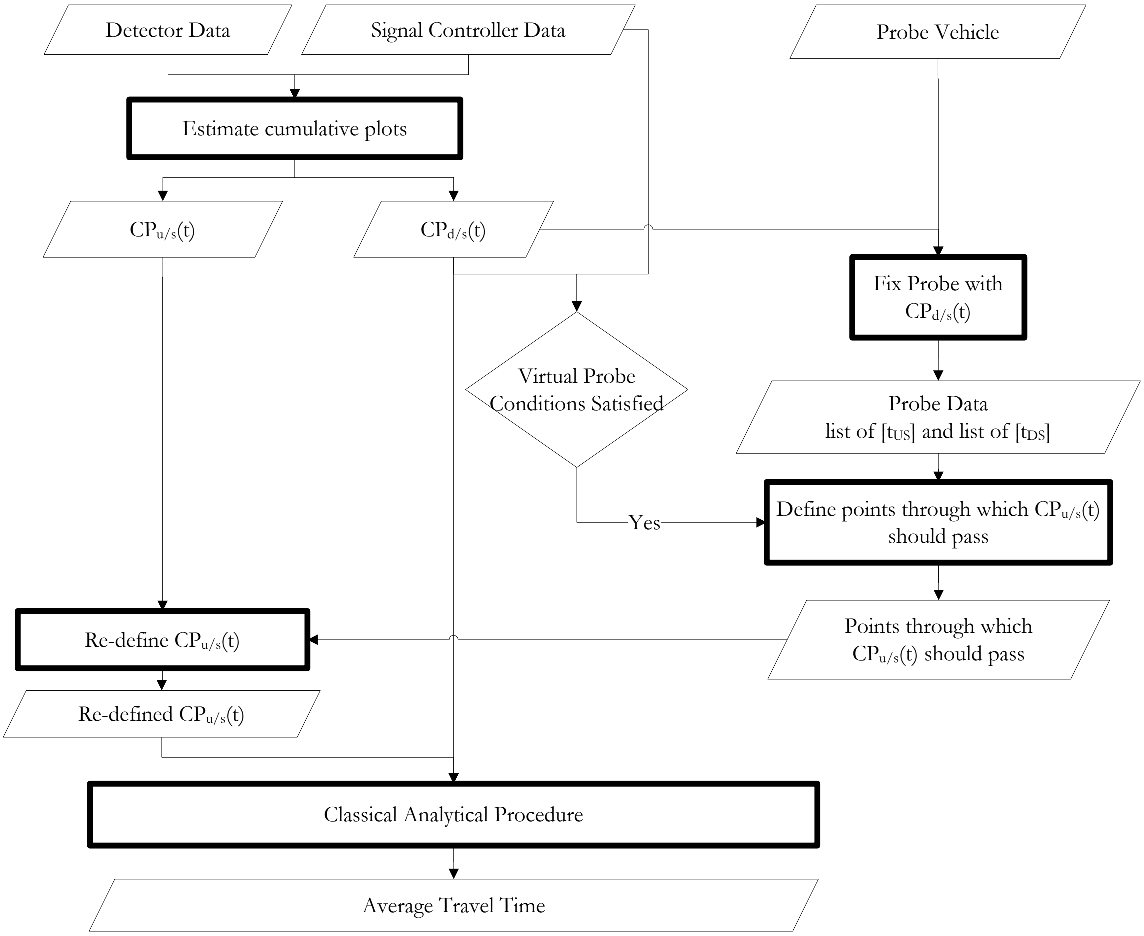


FIGURE 6 ***CUPRITE architecture*.**

MODEL TESTING

The model is tested under controlled environment using, AIMSUN [35] on a) one lane link (*FIFO* condition) and b) two lane link (*non-FIFO* condition) between two consecutive signalized intersections. On the study link the flow is from three different directions at upstream intersection. Only through movement at downstream intersection is considered. Bottleneck is at downstream intersection. Vehicles for the mid-link sink (or source) are randomly selected from the vehicles traversing the link.

The performance of the model, defined in terms of accuracy (%) (eqn (2)) is evaluated as:

$$Error_i = \left(\frac{|actual_i - estimated_i|}{actual_i}\right); \quad Accuracy_i = (1 - Error_i) * 100$$

$$MAPE = \frac{\sum_{i=1 \, to \, N} Error_i}{N} \qquad (2)$$

$$Accuracy(\%) = (1 - MAPE) * 100$$

Where, N is the total number of estimation intervals. $Actual_i$, $estimated_i$, $Error_i$, and $Accuracy_i$ are the average actual travel time, average estimated travel time, absolute relative error and accuracy for i^{th} estimation interval, respectively. *MAPE* is the mean absolute percentage error.

The simulation parameters are:

a) Signal cycle time = 120 seconds
b) Scenarios for different flow to capacity ratio in the range of 0.5 to 1.2 at downstream intersection.
c) T_{EI} = 360 seconds.

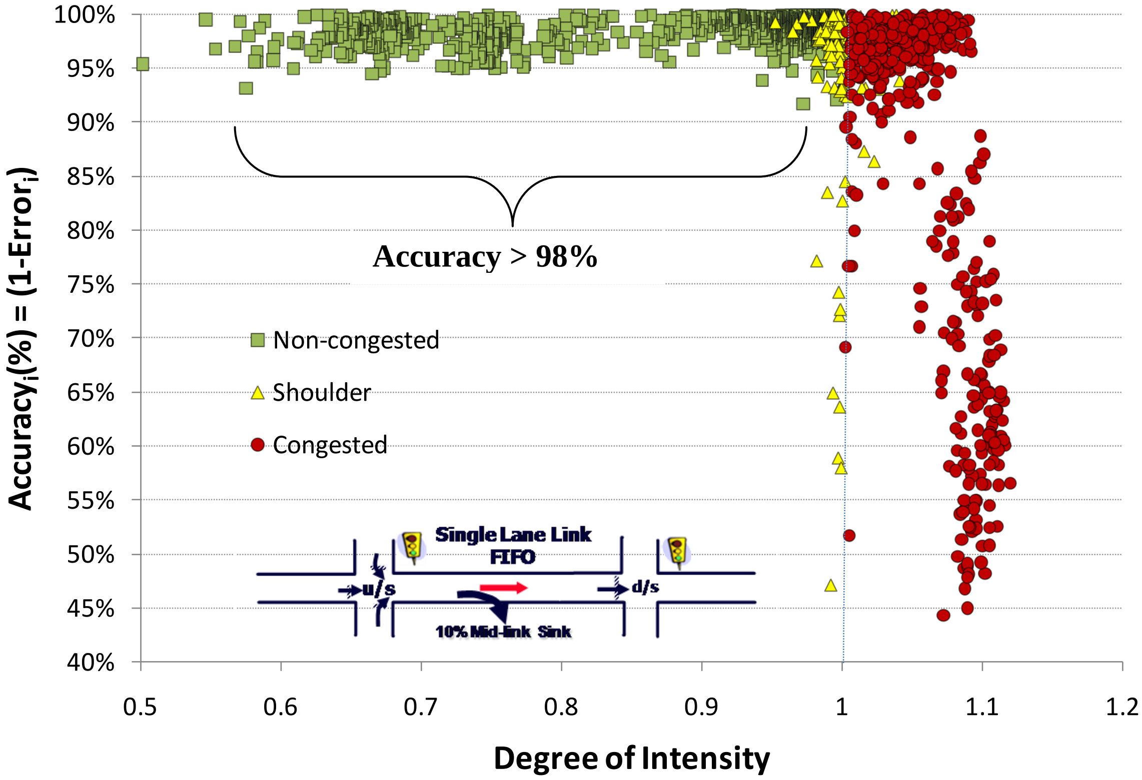
Sink percentage is defined as the ratio of vehicles lost in the sink to the vehicle observed at upstream. Source percentage is defined as the ratio of vehicles gained from the source to the vehicles departing from downstream. In the present analysis 5%, 10%, 15% and 20% of sink and source are considered.

Under-saturated Condition: No real probe consideration

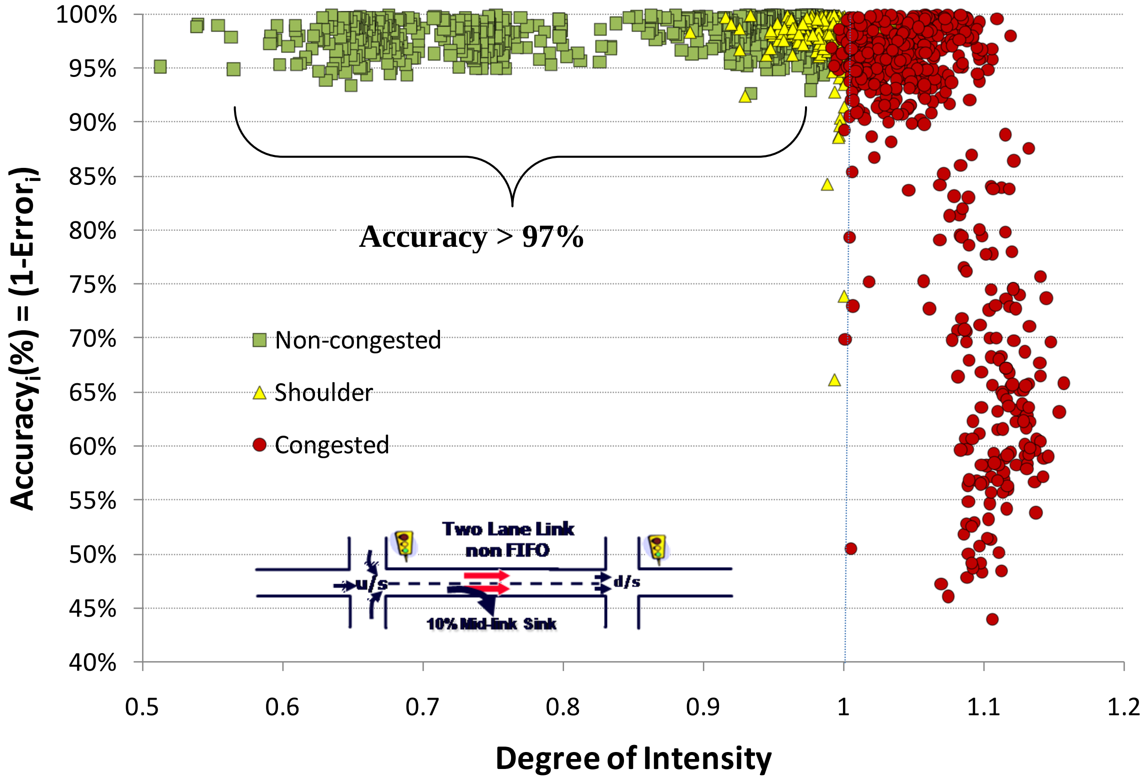
FIGURE 7 represents the data from simulation with no real probe consideration. In the present analysis, virtual probes can be defined for under-saturated traffic conditions resulting in consistent accuracy of more than 98% and 97% for *FIFO* and *non-FIFO* network, respectively. It indicates that if conditions for virtual probe are met then travel time can be accurately estimated even if real probe is unavailable.

Virtual probe, checks that the relative deviations amongst the cumulative plots are corrected during under-saturated (non-congested) traffic conditions. During congestion build-up and dissipation (shoulder) process, traffic is from under-saturated to saturated situation and vice versa, respectively. The use of virtual probe makes sure that in absence of real probes the errors during shoulder conditions is low compared to the case when no correction is applied during the under-saturated situation.

For saturated and over-saturated conditions virtual probe does not exist, resulting in significant decrease in accuracy, and accuracy can be significantly enhanced by considering real probes.



(a)



(b)

FIGURE 7 Simulation for different traffic flow conditions with no real probes.

Saturated and Over-saturated Condition
Probe information is only available after its departure from downstream. Hence, here we present the results for *online* and *offline* applications of travel time estimation.

Online and offline applications
Say the current time is t_c and we are interested in estimating travel time for all the vehicles which have departed during the last T_{EI} seconds. For *online* application, information available until the current time is utilized and any change in the $CP_{US}(t)$, due to availability of future probe data after time t_c, is not considered for travel time estimation in the interval from t_c -T_{EI} to t_c.

For *offline* application we assume that probe vehicle and cumulative plots for all the estimation time intervals are available and $CP_{US}(t)$ is redefined with the complete information. Thereafter, travel time for each estimation interval is estimated.

FIGURE 8 represents the cumulative plots ($CP_{US}(t)$ and $CP_{DS}(t)$) estimated till the current time indicated in the figure. The plots are for a *10%* sink case. Actual cumulative plots ($CP_{USactual}(t)$ and $CP_{DSactual}(t)$) are the accurate cumulative plots to be used for travel time estimation. They are obtained from individual simulated vehicles traversing the complete link. *FIGURE 8a, b* and *c* are for *online* application and *d* is for *offline*.

a. *FIGURE 8a:* t_c = *7:18:00*. Traffic condition is under-saturated therefore, virtual probes are used and we can see that redefined $CP_{US}(t)$ is close to $CP_{USActual}(t)$.
b. *FIGURE 8b:* t_c = *7:24:00*. Saturated traffic condition with no probe information. As saturated traffic condition so, virtual probe cannot be defined. Due to this there is a deviation in redefined $CP_{US}(t)$ from that of $CP_{USactual}(t)$ (Refer zoomed portion of the figure).
c. *FIGURE 8c,* t_c =*7:30:00*. Oversaturated traffic condition with probe information. Here, there are actually two probes observed at upstream, but only one of them has departed from the downstream. Therefore, for the current period only the first probe is considered to redefine $CP_{US}(t)$. Note: as $CP_{US}(t)$ is redefined, the error in the previous estimation interval (*7:18:00* to *7:24:00*) is also corrected (Refer zoomed portion of the figure). For online application to estimate travel time for estimation interval from *7:18:00* to *7:24:00*, the plots represented in *FIGURE 8b* are considered. However, if time series modeling is to be performed and one is interested in time series of travel time then the errors performed in the previous intervals can be corrected.
d. *FIGURE 8d*, is an example for *offline* estimation. $CP_{US}(t)$ is redefined with all the probes and travel time for each estimation intervals are estimated. It can be seen that redefined $CP_{US}(t)$ is close to $CP_{USActual}(t)$ hence, *offline* estimation should have better accuracy than that of *online*.

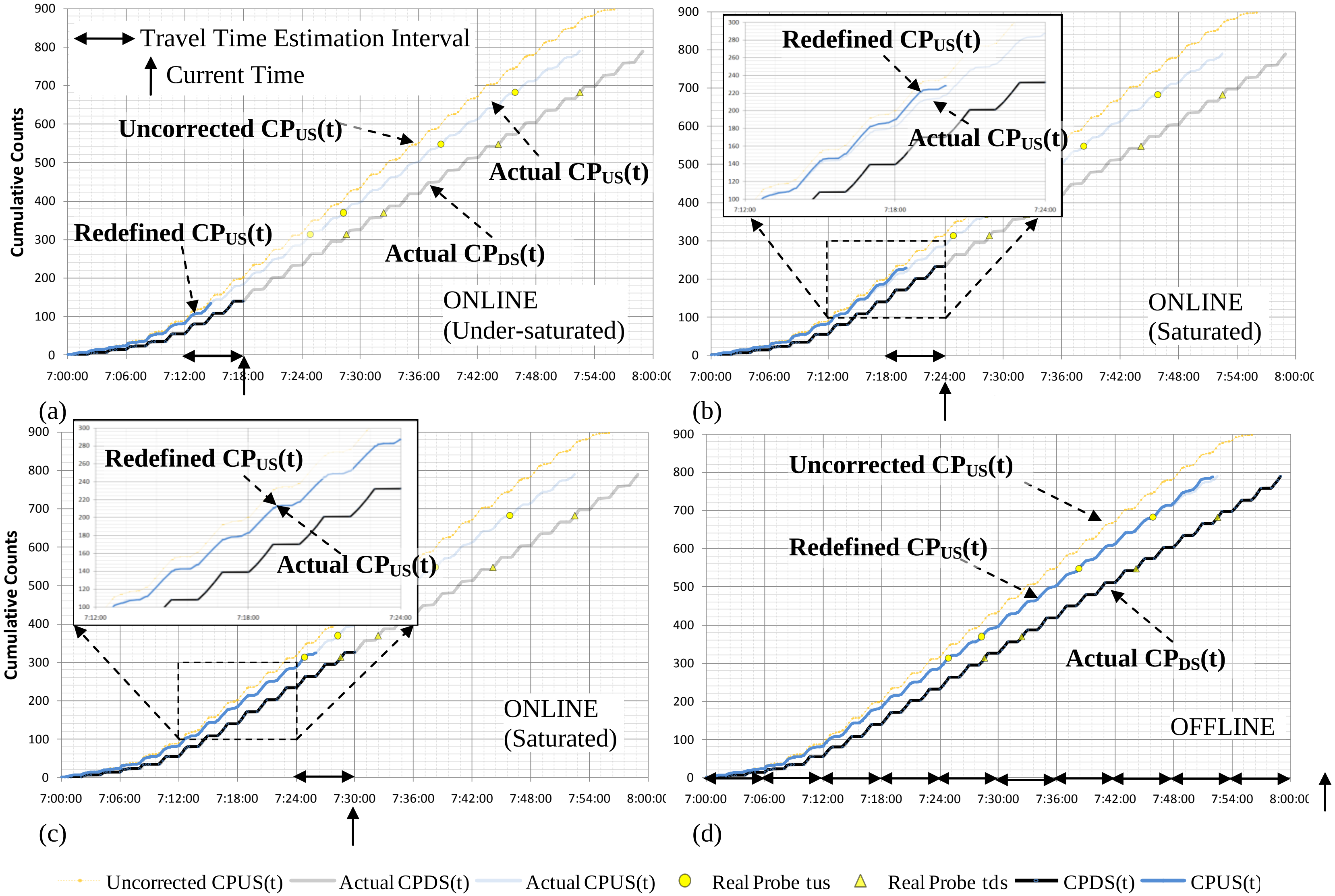


FIGURE 8 Example for *online* and *offline* applications.

Minimum probes required

It is interesting to know the minimum probes required for statistically accurate travel time estimation. The performance of the model is evaluated for:

a) Fixed number *n* during each estimation interval.
b) Percentage *p* of all vehicles traversing the link. This percentage is an indirect representation of the market penetration of probes in vehicles traversing the link during certain time periods.

While presenting the results, comparison of *CUPRITE* with model solely based on probe data here referred as "*Probe-Only"* is also provided. "*Probe-Only*" (eqn (3)) assumes that probe represents a random sample from all the vehicles (population) and average of the travel times from the probes *(t_i)* is the representative of the population, given that the sample size *(n_p)* is at least a minimum value.

$$Average\,Travel\,Time = \frac{\sum_{i=1}^{n_p} t_i}{n_p}; \quad n_p \geq 1 \tag{3}$$

Note: Comparison is only possible when at least one probe per estimation interval is present.

Fixed number of probes

10% Sink Case:

FIGURE 9*a* and FIGURE 9*b* represent the accuracy of estimation versus fixed number of probes per estimation interval for *FIFO* and *non-FIFO* network, respectively. It is observed that:

a) Accuracy from *CUPRITE* is close to 98% and 95% for *FIFO* and *non-FIFO* networks with at least one probe per estimation interval.
b) If we have only a few probes per estimation interval (less than 5 probes) then there is significant benefit of integrating probes with cumulative plot. If the number of probes per estimation interval is large (more than 10) then the probes are good representative of the population of the vehicles and there is little benefit of integrating probes with cumulative plots.
c) As expected, o*ffline* application performs better than *online* application and accuracy increases with increase in number of probes for both *CUPRITE* and *Probe-Only*.

Comparative Case:

FIGURE 10 provides comparative results of the accuracy versus fixed number of probes per estimation intervals for *non-FIFO* network under different sink and source percentages. The values of the accuracy are rather stable if at least one probe per estimation interval is available. It can be concluded that: if probe data is available then the *CUPRITE* model is not sensitive to the percentage of mid-link sink/source. The results also indicate that only one probe per estimation interval has the potential to provide accurate estimates.

As expected, if no probe data is available then the accuracy for *CUPRITE* model decreases with increase in the percentage of the sink or source.

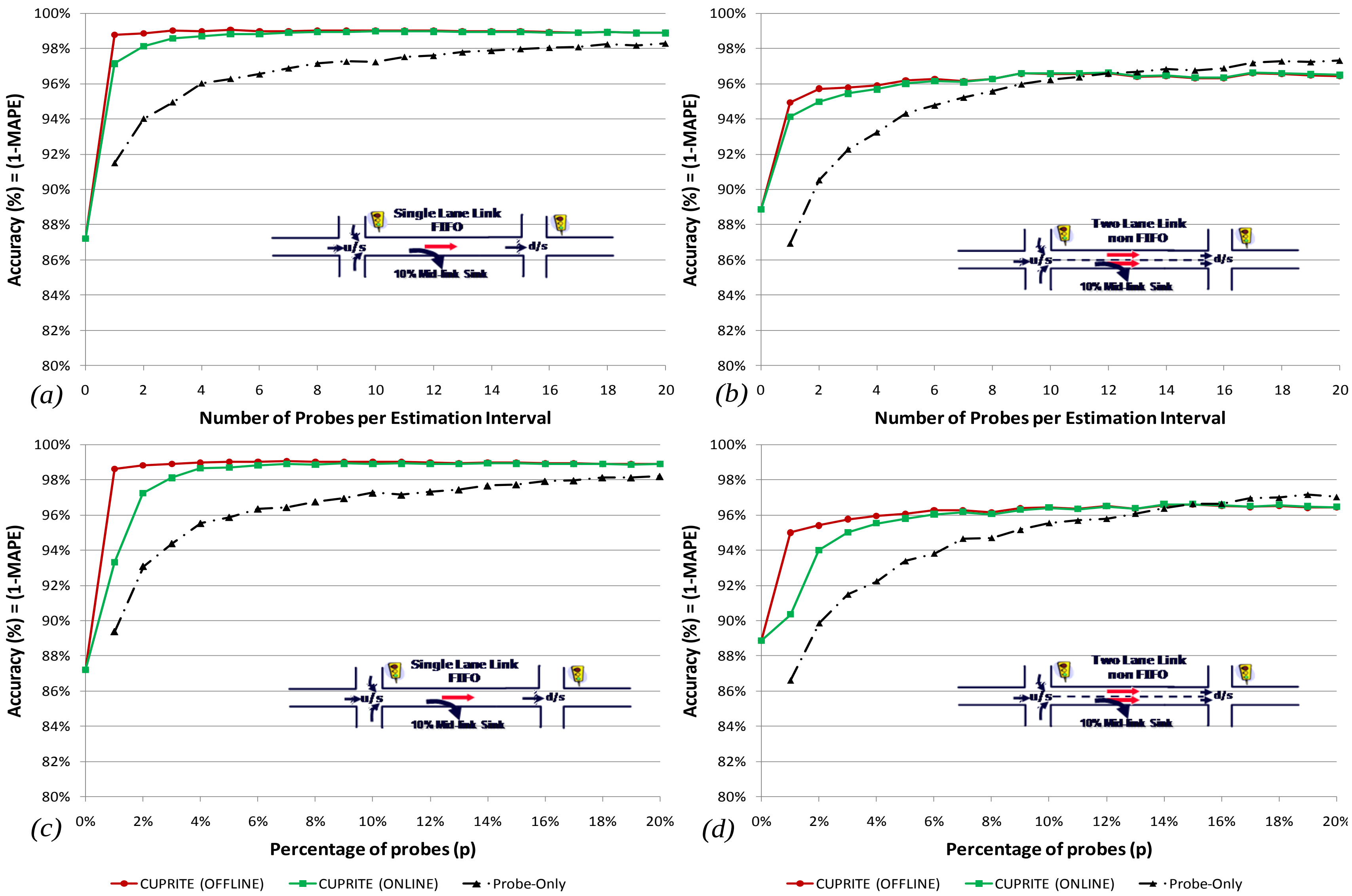


FIGURE 9 10% sink results: (a,b) for fixed number of probes per estimation interval (c,d) for p% of vehicles as probes.

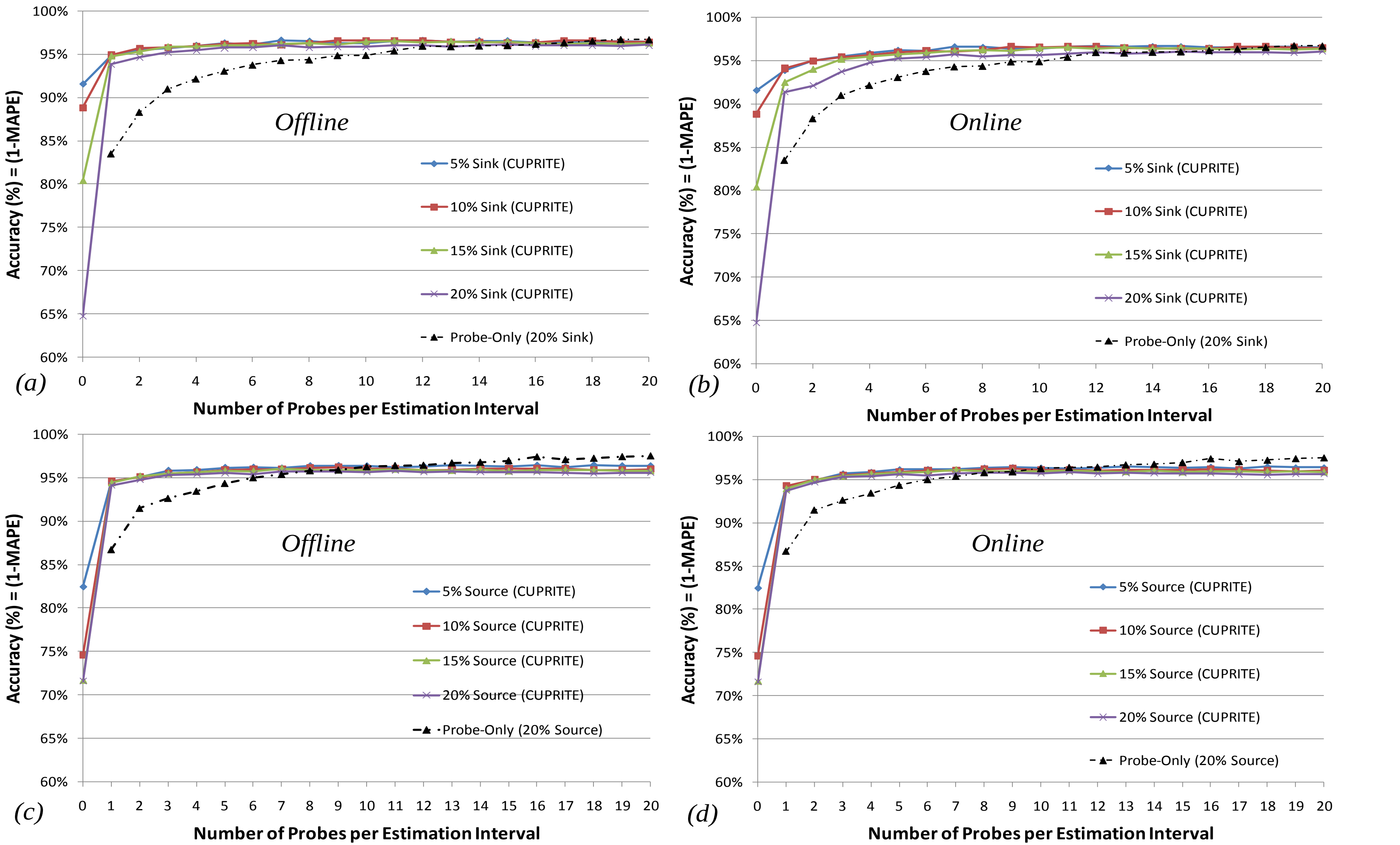


FIGURE 10 Accuracy versus number of probes for different percentage of mid-link sink (a and b) and source (c and d) from non-FIFO network.

Reliability of the estimates:
FIGURE 11a represents standard deviations (σ) of accuracies versus number of probes per estimation interval for *non-FIFO* network. It can be said that higher the σ, lower is the reliability of the accuracy estimate and vice versa. It is observed that the reliability of the *CUPRITE* model increases (σ decreases) with increase in number of probes and *CUPRITE* is more reliable than *Probe-Only*. Hence, integration of cumulative plots and probe not only increases the robustness of the travel time estimates using cumulative plots but also overcomes the issue of uncertainty in the estimates from the use of probe vehicles.

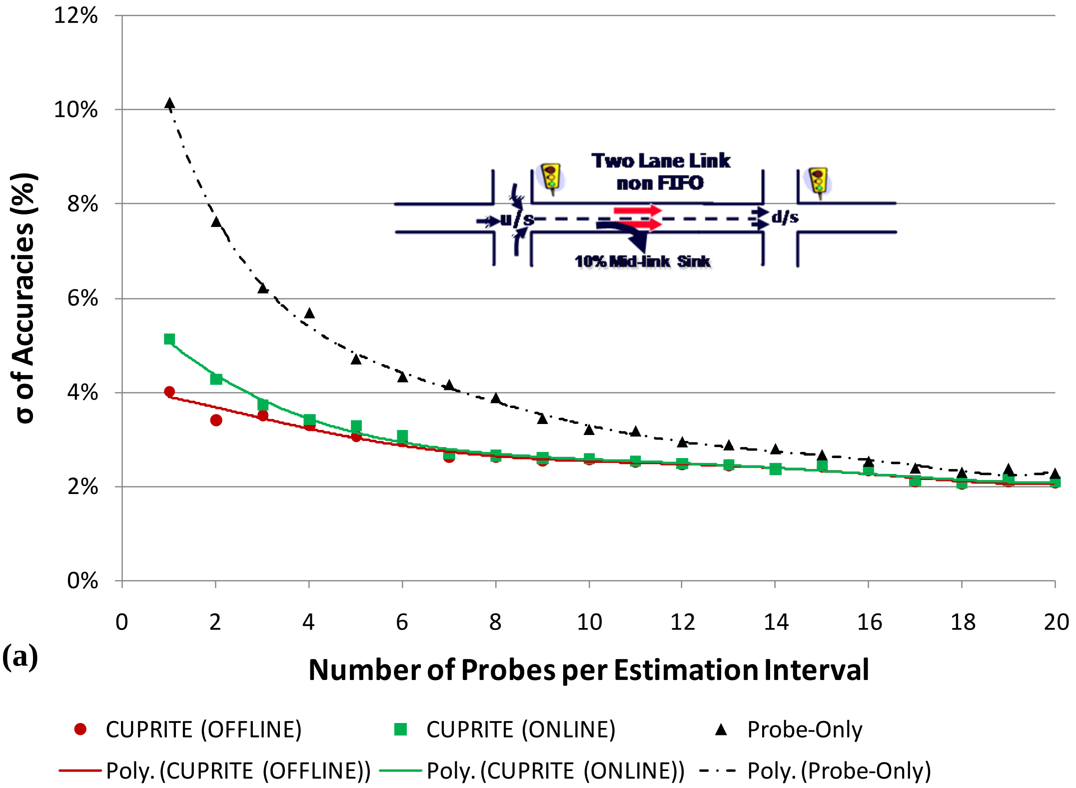
Percentage p of all vehicles traversing the link
Here, *p%* of the vehicles is randomly selected from all the vehicles traversing the link during one hour of simulation. There may be certain estimation intervals with no probe information. *FIGURE 11*b represents the distribution of the number of probes per estimation intervals for different probe percentage from 10% sink and non-FIFO network. It is observed that around 5% of probes can cover all the estimation intervals with at least one probe per estimation interval, whereas, for less than *3%* of probes there can be significantly number of estimation intervals with no probe. For instance, *1%* probe can have more than *30%* of estimation intervals with no probe.

For *Probe-Only* (eqn (3)) travel time cannot be estimated if there is no probe. Therefore, travel time for estimation interval with no probe is assumed to be equal to the travel time of the previous estimation interval with at least one probe.

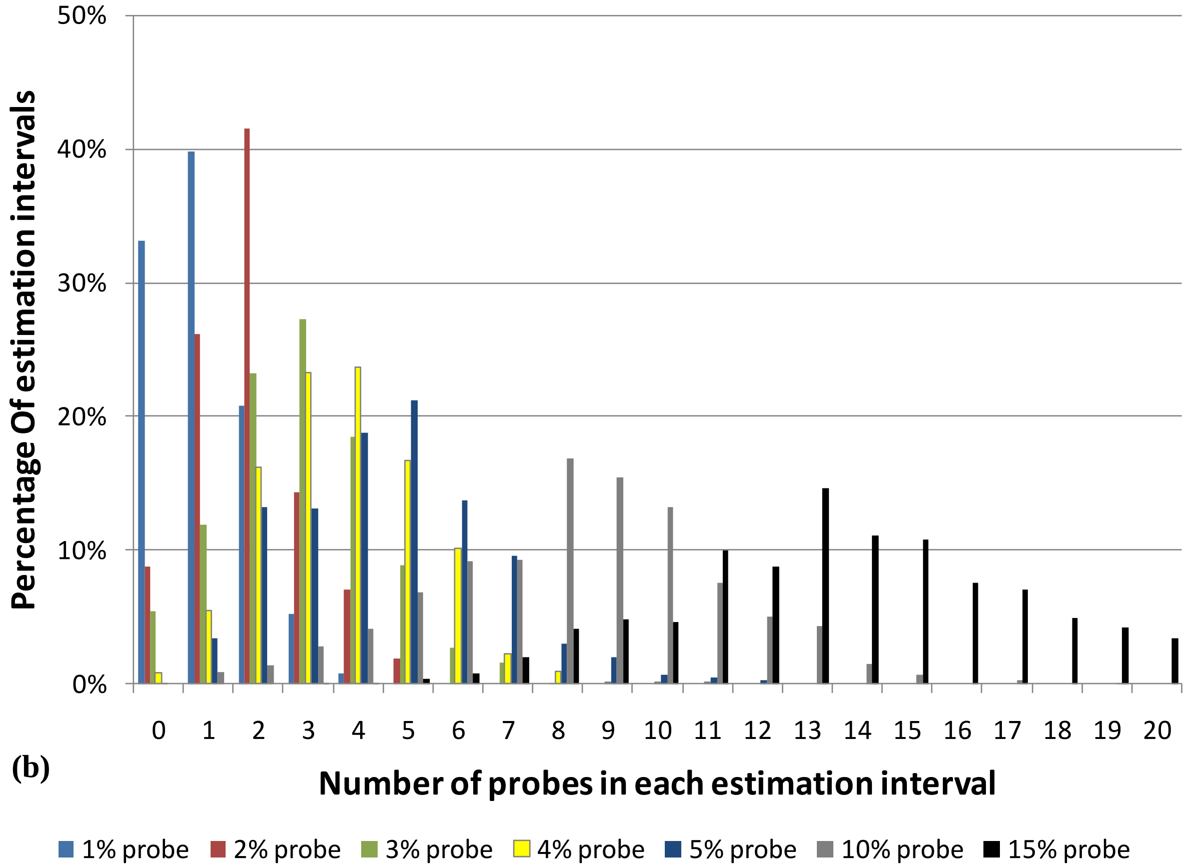
FIGURE 9c and FIGURE 9d represent accuracy versus probe percentage for saturated and over-saturated conditions. Each of the estimation interval considered may or may not have a probe. It is observed that:

a) For *offline* Application: The performance is consistent and accuracy is more than *98%* and *96%* for *FIFO* and *non-FIFO* network, respectively. As expected *offline* performs better than *online*.
b) For *online* application: there is increase in accuracy from 9*0%* to *98%* (*FIFO*) and *95%* (*non-FIFO*) for increase in probe from *0%* to 3*%*. This indicates that 3% of the vehicles traversing the link as probes have the potential to provide accurate travel time.
c) Accuracies for *Probe-Only* increases with increase in probe percentage. For low probe percentage (less than 5%) significant large number of estimation intervals is with no probe or a few number of probes which accounts for low accuracy. Integration of probe with cumulative plots for low probe percentage significantly enhances the accuracy. As percentage of probe increases, the number of probes per estimation interval also increases resulting in better accuracy. For instance, 15% probes generally provide 10 probes per estimation interval. For such cases, probes are good representative of the population of the vehicles and there is little benefit of integrating probes with cumulative plots.

The above model testing is for two independent conditions of source and sink. A mid-link infrastructure, such as parking, can simultaneously act as both source and sink. If the net loss of vehicles to sink and gain of vehicles from source is zero then the issue of relative deviation should not exist. In practice, source and sink percentage are dynamic in nature and for a larger time period such as one hour or so they may balance each other. Nevertheless, for each travel time estimation interval the effect of relative deviation exits and integration of probe vehicles with cumulative plots have potential to improve the accuracy.



(a)



(b)

FIGURE 11 a) Standard deviation of accuracy versus number of probes per estimation interval b) Percentage of estimation intervals with different number of probes per interval.

CONCLUSION

The model (*CUPRITE)* developed in this paper provides encouraging results for travel time estimation by integrating data from different sources: cumulative plots and probe vehicles. The integration provides better performance than method based on single data source only. It overcomes the issue of relative deviation in cumulative plots and uncertainty of travel times estimates from a few number of probes (sample size of one or two vehicles). It can provide accurate travel time for successive estimation intervals for both *offline* and *online* applications.

REFERENCES

[1] J. G. Wardrop, "Journey speed and flow in central urban areas," *Traffic Engineering and Control*, vol. 9, pp. 528-532, 1968.

[2] H. E. Gault, "An on-line measure of delay in road traffic computer controlled systems," *Traffic Engineering and ControL*, vol. 22 pp. 384-389, 1981.

[3] C. P. Young, "A relationship between vehicle detector occupancy and delay at signal-controlled junctions," *Traffic Engineering and Control*, vol. 29, pp. 131-134, 1988.

[4] V. P. Sisiopiku and N. M. Rouphail, "Toward the use of detector output for arterial link travel time estimation: a literature review," *Transportation Research Record*, pp. 158-165, 1994.

[5] V. P. Sisiopiku, N. M. Rouphail, and A. Santiago, "Analysis of correlation between arterial travel time and detector data from simulation and field studies," *Transportation Research Record*, pp. 166-173, 1994.

[6] C. Xie, R. L. Cheu, and D. H. Lee, "Calibration-free arterial link speed estimation model using loop data," *Journal of Transportation Engineering*, vol. 127, pp. 507-514, 2001.

[7] K. F. Petty, P. Bickel, M. Ostland, J. Rice, F. Schoenberg, J. Jiang, and Y. a. Ritov, "Accurate estimation of travel times from single-loop detectors," *Transportation Research Part A: Policy and Practice*, vol. 32, pp. 1-17, 1998.

[8] J. Rice and E. Van Zwet, "A simple and effective method for predicting travel times on freeways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, pp. 200-207, 2004.

[9] X. Zhang and J. A. Rice, "Short-term travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 11, pp. 187-210, 2003.

[10] J. S. Oh, R. Jayakrishnan, and W. Recker, "Section travel time estimation from point detection data," in *In proceedings of the 82th Annual Meeting of Transportation Research Board*, Washington, D.C., U.S.A., 2003.

[11] D. H. Nam and D. R. Drew, "Automatic measurement of traffic variables for intelligent transportation systems applications," *Transportation Research Part B: Methodological*, vol. 33, pp. 437-457, 1999.

[12] J. You and T. J. Kim, "Development and evaluation of a hybrid travel time forecasting model," *Transportation Research Part C: Emerging Technologies*, vol. 8, pp. 231-256, 2000.

[13] S. I. Bajwa, E. Chung, and M. Kuwahara, "A travel time prediction method based on pattern matching technique " in *21st ARRB and 11th REAAA Conference* Cairns, Australia,, 2003.

[14] S. Robinson and J. W. Polak, "Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method," *Transportation Research Record*, vol. 1935, pp. 47-56, 2005.

[15] D. J. Dailey, "Travel-time estimation using cross-correlation techniques," *Transportation Research*, vol. 27 B, pp. 97-107, 1993.
[16] B. Coifman, "Vehicle reidentification and travel time measurement, Part II: Uncongested freeways and the onset of congestion," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Oakland, CA, 2001, pp. 144-149.
[17] B. Coifman and M. Cassidy, "Vehicle reidentification and travel time measurement, Part I: Congested freeways," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Oakland, CA, 2001, pp. 138-143.
[18] B. Coifman and M. Cassidy, "Vehicle reidentification and travel time measurement on congested freeways," *Transportation Research Part A: Policy and Practice*, vol. 36, pp. 899-917, 2002.
[19] B. Coifman and E. Ergueta, "Improved vehicle reidentification and travel time measurement on congested freeways," *Journal of Transportation Engineering*, vol. 129, pp. 475-483, 2003.
[20] B. Coifman and S. Krishnamurthy, "Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure," *Transportation Research Part C: Emerging Technologies*, vol. 15, pp. 135-153, 2007.
[21] T. M. Kwon, "Blind Deconvolution of Vehicle Inductance Signatures for Travel-Time Estimation " University of Minnesota Duluth, http://conservancy.umn.edu/bitstream/436/1/200606.pdf2006.
[22] D. Park and L. R. Rilett, "Forecasting multiple-period freeway link travel times using modular neural networks," in *Transportation Research Record*, 1998, pp. 163-170.
[23] H. Chen, S. Grant-Muller, L. Mussone, and F. Montgomery, "A study of hybrid neural network approaches and the effects of missing data on traffic forecasting," *Neural Computing and Applications*, vol. 10, pp. 277-286, 2001.
[24] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting," *European Journal of Operational Research*, vol. 131, pp. 253-261, 2001.
[25] H. Liu, H. Van Zuylen, H. Van Lint, and M. Salomons, "Predicting urban arterial travel time with state-space neural networks and Kalman filters," in *Transportation Research Record*, 2006, pp. 99-108.
[26] B. R. Hellinga and L. Fu, "Reducing bias in probe-based arterial link travel time estimates," *Transportation Research Part C: Emerging Technologies*, vol. 10, pp. 257-273, 2002.
[27] K. K. Srinivasan and P. P. Jovanis, "Determination of number of probe vehicles required for reliable travel time measurement in urban network," *Transportation Research Record*, pp. 15-22, 1996.
[28] S. Berka, A. Tarko, N. M. Rouphail, V. P. Sisiopiku, and D.-H. Lee, "Data fusion algorithm for ADVANCE Release 2.0," *Advance Working Paper Series, No. 48, University of Illinois at Chicago, Chicago, IL*, 1995.
[29] M. Westerman, R. Litjens, and J.-p. Linnartz, "Integration Of Probe Vehicle And Induction Loop Data: Estimation Of Travel Times And Automatic Incident Detection," 1996.
[30] K. Choi and Y. Chung, "A data fusion algorithm for estimating link travel time," *ITS Journal: Intelligent Transportation Systems Journal*, vol. 7, pp. 235-260, 2002.
[31] C. Xie, R. L. Cheu, and D. H. Lee, "Improving arterial link travel time estimation by data fusion," *83rd Annual Meeting of the Transportation Research Board*, 2004.
[32] N. E. El Faouzi, "Data-driven aggregative schemes for multisource estimation Fusion: A road travel time application," in *Proceedings of SPIE - The International Society for Optical Engineering*, Orlando, FL, 2004, pp. 351-359.

[33] N. E. El Faouzi, "Data fusion in road traffic engineering: An overview," in *Proceedings of SPIE - The International Society for Optical Engineering*, Orlando, FL, 2004, pp. 360-371.

[34] A. Bhaskar, E. Chung, M. de Olivier, and A. G. Dumont, "Analytical Modelling and Sensitivity Analysis for Travel Time Estimation on Signalised Urban Network," in *Transportation Research Board 87th Annual Meeting*, Washington, D.C., 2008.

[35] J. Barceló, E. Codina, J. Casas, J. L. Ferrer, and D. Garcia, "Microscopic traffic simulation: A tool for the design, analysis and evaluation of intelligent transport systems," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 41, pp. 173-203, 2005.