Study on the Reliability of Travel Time Estimation by Probe Vehicle System

Tetsuhiro ISHIZAKA a, Atsushi FUKUDA b

a College of Science and Technology, Nihon University, 7-24-1 Narashinodai, Funabashi, Chiba, 274-8501, Japan, E-mail: ishizaka.tetsuhiro@nihon-u.ac.jp

b E-mail: fukuda.atsushi@nihon-u.ac.jp

Abstract: In order to provide reliable travel time information by probe vehicle system, a huge amount of travel time data should be collected. To minimize cost of data collection from probe vehicle with certain reliability level, the number of required probe vehicles and its optimal regional allocation should be examined. Therefore, this study proposes a new methodology to estimate the number of probe vehicles required to ensure the reliability of travel time estimation. It aims to verify the feasibility of reducing the number of probe vehicles from that required by the conventional methodology. Also, the feasibility of reducing probe vehicles by estimating travel time with existing traffic detectors is verified. The study concluded that the required number of probe vehicles estimated by the new methodology is less than that estimated by the conventional methodology, and that the use of probe vehicle systems leads to quick development of traffic information collection system.

Keywords: Probe vehicle system, Number of probe vehicles, Reliability of average travel time

1. INTRODUCTION

Probe vehicle system is expected to become an efficient traffic information collection system with wide applications. Average travel time is one of the traffic information estimated by the travel time data from probe vehicles. The average travel time estimated by a probe vehicle system is expected to provide more accurate information in comparison with travel time data estimated by on-road detectors. That is why accurate link travel time data from probe vehicles are used directly for estimation. Probe vehicle systems have the capability to estimate the average travel time throughout an entire city area since probe vehicles can travel everywhere.

However, probe vehicle systems have some critical problems to be solved before achieving real time travel time observation. One critical problem is the reliability of travel time estimation, which is defined in this study as the accuracy of estimated average travel time, as well as the coverage area over which probe vehicle systems can estimate reliable travel time at any time. The reliability of travel time estimation depends on how many probe vehicles are running throughout the whole city area. Therefore, many studies have tried to estimate the number of probe vehicles required to ensure the reliability of travel time estimation and concluding that an impractically high number of probe vehicles would be required. The author also estimated the required number of probe vehicles using a micro simulation model for real time travel time estimation in Bangkok road network. It was also concluded that the required number of probe vehicles would be enormous. Therefore, it is necessary to develop a new methodology that is capable of reducing the required number of probe vehicles.

This study proposes a new methodology that can reduce the required number of probe vehicles thereby bringing the required number under the actual number, and apply it to Bangkok road network. The use of probe vehicle system as a traffic data collection system with less investment cost would be enhanced if the new methodology could reduce the
number of probe vehicles. Furthermore, by comparing with fixed location data collection system using on-road detectors, the probe vehicle system may require less infrastructure investment. Finally, the most effective data collection system for all links in whole city in view of data reliability, data collection cost and investment cost (see Figure 1) can be achieved.

2. LITERATURE REVIEW

Previous studies have attempted to estimate the number of probe vehicles required to estimate average travel time with high reliability. Boyce, Hicks, and Sen (1991) estimated that the number of required probe vehicles was 4700 in order to estimate reliable travel time for at least 50 percent of all road sections every 5 min for the North Shore network in Chicago. Srinivasan and Jovanis (1996) have estimated that 3500 vehicles would be required to collect traffic information at 80 percent of arterial and freeway links every 10 minutes for the road network (432 km²) in Sacramento, CA, U.S.A. These studies used a traffic assignment method such as stochastic or dynamic assignment. Boyce, Hick, and Sen (1991) assumed that the traffic information of a link has been collected if probe vehicles passed through the link at least once. Since one time per link is not sufficient to reliably estimate average travel time, the results based on this assumption cannot reliably estimate travel time.

On the other hand, a micro simulation model has been accepted instead of the assignment methods because micro simulation can consider the dynamic travel pattern of each vehicle. Chen and Chien (2000) used the micro simulation model to estimate the number of probe vehicles. However, the dynamic distribution of probe vehicles was not considered since the network in their study consists of only one freeway and five intersections. Cheu, Xie, and Lee (2002) also used the micro simulation model for the road network of the Clemet town area in Singapore. All previous studies aimed at estimating the minimum number of probe vehicles did not consider the difference in cost of each data transmission method. Overall, these studies find the minimum sampling size to be from 10% to 20%.

In order to reduce the number of probe vehicles, Toi (1990) proposed the theoretical disposition of probe vehicles under the condition of covering traffic volume of OD pair. Horiguchi (2002) proposed the strategic disposition of taxi probe vehicles and showed the practical use of commercial vehicles as probe vehicles. The mix of commercial vehicles having different characteristics regarding running time period and area was important for effective data collection. Fushiki (2006) showed the methodology to calculate the number of
probe vehicles considering traffic information update frequency.

On the other hand, the combination of probe vehicles and traffic detectors was verified by Horiguchi (2000). The optimal distribution of traffic detectors to collect travel time data of other links from probe vehicles was examined. Nakamura (2003) examines the travel time estimation by using probe vehicles and traffic detectors to increase reliability of estimated travel time.

In the above studies using the micro simulation model, the percentage of the probe vehicle OD (Origin and destination) flow of the total OD flow was used to calculate the required number of probe vehicles. Previous studies set this percentage at levels such as 10%, 20%, 40%, 60%, 80% and 100% in the simulation model. Initially, a 10% probe vehicle OD flow is set in the model. The flowchart of this conventional methodology is shown in Figure 2. If the simulation run results satisfied the reliability condition in all links as described in Section 2.3, then these studies found the minimum percentage of probe vehicles to be 10%. However, if the reliability condition was not met for even one link, the next percentage was used and the simulation was run again. This process continued until all links satisfied the reliability condition. However, since percentage is an exogenous value, these studies cannot estimate the minimum number of probe vehicles. Additionally, differences in area and road length as well as differences in the traffic volume of each OD pair caused these studies to overestimate the required percentage of probe vehicle OD flow.

3. METHODOLOGY

3.1 Proposed Methodology for Estimating the Required Number of Probe Vehicles

A methodology is needed that can reduce the required number of probe vehicles. This study proposes a methodology for estimating the required number of probe vehicles relative to each OD pair flow. In the case of low number of probe vehicles, an unfavorable geographical distribution of the probe vehicles causes links where the travel time cannot be estimated. To prevent this situation from occurring, the probe vehicles have to be optimally deployed by estimating the required number of probe vehicles relative to each OD pair flow. The required number of probe vehicles relative to each OD pair flow that shows the desired spatial distribution of deployed probe vehicles is estimated using the following equation.

\[
\begin{align*}
\text{Min} & \quad \sum_{\tau} \sum_{\nu} D_{ij}^{\tau} P_{ij}^{\tau} \\
\text{Subject to} & \quad n_a^{\tau} \leq x_a^{\tau}
\end{align*}
\]

where,

- \( D_{ij}^{\tau} \): OD traffic flow of OD pair \( ij \) in time interval \( T \),
- \( P_{ij}^{\tau} \): percentage of probe vehicle OD pair \( ij \) in time interval \( T \),
- \( n_a^{\tau} \): number of probe vehicles required in link \( a \) in time interval \( T \), and
- \( x_a^{\tau} \): link probe vehicles flow calculated by using simulation in time interval \( T \).

In this equation, the total required number of probe vehicles \( \sum_{\tau} \sum_{\nu} D_{ij}^{\tau} P_{ij}^{\tau} \) is minimized subject to maintaining the reliability of travel time estimation. The link probe vehicle flow \( x_a^{\tau} \) has to exceed the required number of probe vehicles \( n_a^{\tau} \) on all links. The methodology of
calculation of the required number of probe vehicles $n_a^T$ will be explained in the next section. $P_{ij}^T$ is calculated by averaging $P_{ija}$ in the following equation.

$$P_{ij}^T = \frac{\sum_{a \in A} P_{ija}^T \beta_{ija}^T}{N_A^T}$$

where,

- $N_A^T$: number of links ($a$) in time interval $T$,
- $\beta_{ija}^T$: equal to 1 when traffic flow on link $a$ of $P_{ij}^T$ is observed; 0 otherwise in time interval $T$,
- $P_{ija}^T$: percentage of required number of probe vehicles on link $a$ in time interval $T$.

To make this equation easier to understand, the steps involved are described in Figure 2. In Step 1, the probe vehicle OD matrix is calculated by multiplying the percentage of probe vehicles ($P_{ij}$) and Total OD. A micro simulation is conducted using probe vehicle OD and total OD in Step 2. In Step 3, the simulation results are used to check the reliability condition, which is described earlier in Chapter 2.2. If the reliability condition of all links is satisfied, $P_{ij}$ is recognized as the required number of probe vehicles. If not, $P_{ij}$ is increased in Step 4 and the process is repeated from Step 1.

To compare with the conventional method (see Figure 2), the percentage of probe vehicles is calculated by each OD pair on the new proposed methodology. Therefore, the proposed method would reduce the total number of probe vehicles required for a highly reliable traffic data collection in the whole city because this method can estimate less required number of probe vehicles by each OD pair. This method can determine the optimal number of probe vehicles by considering regional traffic situation (OD pair).

This methodology is not a theoretical mechanism to estimate a global optimum solution for the required number of probe vehicles. Finding concrete value as a global optimum solution is difficult and meaningless because the traffic situation dynamically changes every day. However, it is practically important to verify the feasibility of reducing the number of probe vehicles.

### 3.2 Reliability of Estimated Average Travel Time

This section describes the procedure used to check the reliability of the estimated average travel time (step 3). This same methodology is used in both the proposed methodology and the existing methodology.

Recent studies have checked reliability using statistical sampling theory which can estimate the minimum sample size of probe vehicles required for estimating the average population. If the number of probe vehicles exceeds the minimum sample size of each link, traffic information with high reliability can be collected by probe vehicles. The minimum sample size of probe vehicles is given by:

$$n = \left( \frac{t \times s}{\varepsilon} \right)^2 = \left( \frac{t \times c.v.}{\varepsilon} \right)^2$$

(3)
where \( t \) is the \( t \)-statistic from the Student’s \( t \) distribution for a specified confidence level, \( s \) is the standard deviation of travel time, and \( \varepsilon \) is the maximum specified allowable error. The coefficient of variation \( c.v. \) and relative error \( e \) are given by:

\[
c.v. = \frac{s}{\bar{x}},
\]

\[
e = \frac{\varepsilon}{\bar{x}},
\]

If the sample size exceeds 30, the normal distribution can be used instead of Equation (3) by using a \( z \) value. This sample size is given by:

\[
n = \left( \frac{z \times s}{\varepsilon} \right)^2.
\]

### 3.3 Development of a Micro Traffic Simulation Model

This study used this basic statistical sampling theory to estimate the number of probe vehicles required for accurate data collection for the Bangkok road network. Figure 3 shows the entire Bangkok road network. The dotted line defines the study area containing 1896 links and 1310 nodes. The OD demand for probe vehicles was configured as part of a taxi OD matrix that was developed through the field survey data. The field survey conducted by the authors in 2002 showed the running pattern of several taxis such as origin, destination, coverage area, travel time, trip length and frequency of trip. In the field survey, the data collection of taxis was conducted for four months by using GPS and data logging system. The details of the field survey were described by Tetsuhiro Ishizaka et al. (2004). The other OD matrix was
configured using data from UTDM (Urban Transport Database and Model Development Project), provided by the OTP (Office of Traffic and Transport Policy and Planning). This OD matrix consists of passenger cars, taxis, pickup trucks, buses, etc. The total passenger car OD demand is approximately 2.3 million trips per day and the total taxi OD demand is 0.7 million trips per day. PARAMICS version 5 was employed with an original API sampling theory simulation program developed by the authors.

The simulation model was calibrated with travel time data and traffic volume. The travel time data was collected by probe vehicles during the field survey conducted by the authors in Bangkok. The several links on which characteristics, such as average travel time by period and area, have been shown by the field survey were selected for the validation of traffic flow in the simulation model. The detailed traffic situation was also described by Tetsuhiro Ishizaka et al. (2004). The traffic volume at intersections collected by the Bangkok Metropolitan Administration (BMA) was used as data for validation. This traffic volume data was counted for seven vehicles types from 7 a.m. to 7 p.m. at 250 intersections in 2002. Its interval of data aggregation is one hour. Unfortunately, the vehicle types of OD matrix and traffic volume data do not match. There was especially no data classification about taxi on the traffic volume data as it was included in the passenger car class. Therefore, the traffic volume of passenger cars including taxis was used when the validation of the developed micro simulation model was conducted.

In this study, taxis were used as probe vehicles in the low percentage of probe vehicle case because their daily running time is over twenty hours and they generate more data per vehicle than other vehicles.

The developed simulation model was validated by comparing observed link traffic volume and estimated link traffic. The link traffic volume was obtained by aggregating intersection traffic volume collected by BMA (2002). The result of validation is shown in Figure 4. It is concluded that the developed simulation model can simulate actual traffic situation in Bangkok.

Figure 3. Bangkok road network and the study area
4. REQUIRED NUMBER OF PROBE VEHICLES

4.1 Required Number of Probe Vehicles Estimated by Conventional Methodology

The required number of probe vehicles was estimated using the conventional methodology in Figures 5 and 6 for comparison with the results of the proposed methodology. The required number of probe vehicles is shown on the horizontal axis as the percentage of probe vehicles relative to total OD flow. Three links with the required number of probe vehicles are illustrated to indicate the required number of probe vehicles, which is influenced by the time interval for travel time estimation. The link coverage means the number of links that satisfied the reliability condition of step 3 in Chapter 2.

Figure 5. Percentage of links satisfying the reliability condition when using taxis with the conventional method
As illustrated in the figure, the link cover percentage had the same trend regardless of the time interval used to aggregate travel time. When all the taxis are used as probe vehicles, around 40% of the links in the network can be covered. In other words, it is difficult to collect travel time information from the entire network using a probe vehicle system if only taxis are used. The estimation exceeds the results of previous studies. This overestimation could be attributed to the unique network conditions: heavy traffic congestion, long link lengths, and the long red periods traffic signal system. As a result of these unique network conditions, many of the probe vehicles cannot pass end of a link within the allotted time.

Using only taxis as probe vehicles was not possible to reach 100 percent link coverage. Therefore, passenger cars were also used as probe vehicles, as shown in Figure 6. In this case, over 70% of the links could be covered by the probe vehicle system using 30% of total OD.

4.2 Required Number of Probe Vehicles Estimated by the Proposed Methodology

The minimum percentage of the OD pair traffic flow is estimated using the proposed methodology in Figures 7 and 8. Compared with Figures 5 and 6, the required number of probe vehicles estimated using the proposed methodology is less than 10%. This result shows that the proposed methodology can reduce the required number of probe vehicles and can contribute to the efficient collection of data using probe vehicles.

The minimum percentage result is compared with the zone traffic congestion level, total traffic flow, and average trip length. Figure 9 illustrates the relationship between the traffic congestion level and the optimal percentage. The optimal percentage decreases as the traffic congestion level rises to one. Beyond this peak, the optimal percentage gradually decreases.
Figure 7. Percentage of links satisfying the reliability condition using taxis with the proposed method.

Figure 8. The percentage of links satisfying the reliability condition by total OD using the proposed method.

Figure 9. Optimal percentage and congestion level.
5. REQUIRED NUMBER OF PROBE VEHICLES UNDER THE COMBINATION OF TRAFFIC DETECTOR AND PROBE VEHICLE SYSTEM

When introducing probe vehicle systems, the examination of the strategy that makes the best use of existing traffic data collection system is important. Especially, in the developing cities where traffic data collection system is yet to be developed and applied, it is significant how soon the collection of traffic information can begin.

The proposed methodology of estimating the number of probe vehicles was described in Chapter 3. The link from which travel time data is being collected in Japan is about 1/3 within all links. Therefore, 1/3 of the 1876 links in the study area was assumed to be the travel time data collected by traffic detectors. It was assumed that traffic detectors were installed on arterial roads in the central area of Bangkok. Probe vehicle system plays a role to collect the traffic data on sub-urban area and minor road in central area. However, the travel time estimated by the probe information system was used if the probe vehicle system obtained the prescribed accuracy on travel time even at the link where traffic detector was installed. The simulation calibration and reliability condition of travel time estimation are the same as in Chapter 4.

Simulating the condition with existing traffic detectors, the required number of probe vehicles was estimated as shown in Figure 10. As a result, the simulation result showed that the percentage of probe vehicles was reduced by about 8% with the same number of links of 80% in comparison with Figure 8. In the case of this simulation result, Figure 11 shows whether probe vehicle system or traffic detector on link can collect reliable traffic data. The accuracy of the travel time collected by probe vehicle system was confirmed even at the link where the vehicle detector was installed. These links where probe vehicle system can estimate reliable travel time is over two thirds of the total number of links. It is concluded that the combination of probe vehicle system and exiting traffic detectors can contribute to the efficiency and ease of traffic data collection.

Figure 10. The percentage of links satisfying the reliability condition by total OD using the proposed method
6. CONCLUSION

This study clarified the feasibility of travel time information collection by a probe vehicle system by calculating the number of probe vehicles required for reliable travel time estimation using the proposed new methodology. It is concluded that a probe vehicle system could be used as a traffic information collection system, and would be particularly appropriate for collection from a wide area since 80 percent of the links can be covered by approximately 28% of total OD flow. This study tried to reduce the required number of probe vehicles by optimizing the percentage of each OD pair. It concluded that the number of probe vehicles estimated by a new methodology is less than the conventional methodology, and that the use of probe vehicle systems leads to quick development of traffic information collection system. The optimal percentage consequently has a linear relationship with the zone congestion level. Further studies will examine whether the number of probe vehicles can be decreased by demonstrating the relationship between the optimal percentage and other zone characteristics.

REFERENCES


Chen, M., Chien, S.I.J. (2000) Determining the number of probe vehicles for freeway travel time estimation using microscopic simulation. *Transportation Research Record*, 1719,
61-68.