Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

Spatial and Temporal Analysis of Probe Vehicle-based Sampling for Real-time Traffic Information System

Jun Hong, Xuedan Zhang, Zhongya Wei, Li Li and Yong Ren

***Abstract*— Using vehicles as probes is a flexible and low-cost way to obtain real-time traffic information. This paper addresses the sampling issues of using probe vehicles for detecting traffic information in a road network. A spatial and temporal analysis model based on signal processing theory is established and used to derive bounds on the sampling period, transmitting period and sample sizes of probe vehicles. We also develop a Traffic & Information-collecting Simulation Platform (TISP), to simulate the traffic flows in a road network and generate probe vehicle data for analysis. The simulation results find that traffic flow has strong correlation in terms of time and space, which is critical to the sampling problem, and the system requires 2% probe penetration to guarantee the information integrity.**

I. Introduction

Real-time traffic information is an essential and important factor in such Intelligent Transportation System (ITS) as congestion management, traffic control, route guidance and so on. Using vehicles as probes is a flexible and low-cost way to obtain real-time traffic information. The key idea of using probe vehicles to collect traffic information is that a vehicle running on the road which is a part of the traffic is reasonable representative of the behavior of the traffic. The trajectory followed by a vehicle is an integral part of the highway travel experience and hence important for a traffic management system [1]. Compared to the conventional stationary detectors installed on the road, probe vehicles may provide benefits such as an easier implementation, more precise information and lowered costs for constructing and maintaining the information system. A key problem of using probe vehicles is to determine the probe sample sizes.

In recent years, there has been much research on the sampling problem of probe vehicles [1]–[11], with most of research focusing on estimating the sample sizes on a certain space during some period. However, the size of the sampling space and the length of the period, which are two essential factors determining the sample set, were not considered carefully. This paper deals with the issues of estimating the sampling period, data transmitting period and the sample sizes for traffic information system using probe vehicles. The approach uses concepts from signal processing theory and the correlation coefficient is used to divide the sample set in time and space. To our best knowledge, this is the first paper introducing both spatial and temporal characteristics to the domain of probe vehicles sampling. In this paper, the analysis may provide an insight for probe vehicle-based traffic information system designers on the sampling period, the transmitting period and the number of vehicles that are desirable in a traffic network in order to achieve certain coverage and accuracy in traffic situation estimation.

The rest of the paper is organized as follows. Section II presents the overview of related works. Section III describes the sampling model and formally defines the problems. Section IV presents our estimating algorithm. Section V evaluates the performance of the proposed algorithm through computer simulations. Section VI concludes the paper.

Jun Hong is with the Department of Electrical Engineering, Tsinghua University, Beijing, 100084, P.R.China fionehong@gmail.com
Xuedan Zhang is with the Department of Electrical Engineering, Tsinghua University, Beijing, 100084, P.R.China xuedanzhang@gmail.com
Zhongya Wei is with the Department of Electrical Engineering, Tsinghua University, Beijing, 100084, P.R.China hze@tsinghua.edu.cn
Li Li is with the Department of Automation, Tsinghua University, Beijing, 100084, P.R.China li-li@tsinghua.edu.cn
Yong Ren is with the Department of Electrical Engineering, Tsinghua University, Beijing, 100084, P.R.China reny@tsinghua.edu.cn

II. Related Work

Over the past decades, numerous probe sample size studies have been carried out mainly based on two theories, quality-control theory [2]–[7] and large sampling theory [8]–[11].

A. Quality-control Theory

Institute of Transportation Engineers (ITE) first provided a method using concepts from quality-control theory to determine the minimal probe sample sizes given by [2]:

$$N = \left(\sigma \frac{K}{e}\right)^2$$

where N is the minimum number of sample size, σ is the estimated sample standard deviation, K is the constant corresponding to the desired confidence level and e is the permitted error in the average speed estimate.

Since the tables of ITE's Manual of Transportation Engineering Studies contain systematic numerical errors, Quiroga and Bullock [3] provided new tables to correct them. Besides, they presented a hybrid method, which combines sample range R and t distribution statistic together, to estimate the required sample sizes:

$$N = \left(R \frac{t_\alpha}{e}\right)^2$$

where t_α is the t distribution statistic for a confidence level of $1 - \alpha$.

Li [4] confirmed that ITE's Manual of Transportation Engineering Studies usually underestimates the sample sizes while the hybrid method developed by Quiroga overestimates them. As a result, a sample size adjustment is added to help estimate the minimum sample sizes:

$$N = \left(\sigma \frac{Z_\alpha}{e}\right)^2 + \varepsilon_N$$

where Z_α is the normal distribution statistic for a confidence level of $1 - \alpha$ and ε_N is the sample sizes adjustment. However, the author did not describe how to get ε_N.

For all the above methods based on quality-control theory, there are some requirements for representative sample [2]:

1) the sample must be selected without bias,
2) the components of the sample must be completely independent of one another,
3) there should be no underlying differences between areas from which the data are selected and
4) conditions must be the same for all items constituting the sample.

Obviously, in real road networks, it is difficult to satisfy these requirements and calculate the size of the selected area which may impact the results greatly.

B. Large Sampling Theory

A simple approach based on the large sampling theory is introduced by Chul Gyu Park [8]. The whole network are partitioned into groups of highly correlated links, and then the total number of probe vehicles may be chosen so that all the link groups maintain predetermined number of probe cars with desired level of certainty, i.e. a large enough N is chosen, so that $\Pr\{x_i \geq \alpha_i\} \geq \beta$, where x_i is the number of probe samples on the i-th link (link group), α_i is the predetermined number of probe cars and β is the desired level of certainty. However, they did not mention the method of partitioning links, and how to determine α_i is unknown. Xiaowen Dai [9] used the simulation approach to estimate the number of probe vehicles required for various types of road networks. The main performance metric is the network coverage C which is defined as $C = P\{x_i \geq \alpha_i\}$. But there are few details about the based theory. Wang Li [10] established an analytical model based on large sampling theory to study the sample sizes of probe vehicles. In their sample size model, the network coverage is defined as the ratio of the number of the links where the probe samples are more than one, i.e. $x_i \geq \alpha_i(\alpha_i = 1)$, to the number of total links. The model is an ideal case of Chul Gyu Park's. Guiyan Jiang [11] studied some cases performed with simulation to analyze the impact of probe vehicles sample size on accuracy of average link travel time estimation. The proposed method is based on large sampling method. However, they did not provide much description of the analytical model except the definition of average error.

III. Formulation

Suppose the length of the road is S, the observation time is T, the number of probe vehicles is M and the speed of vehicles is denoted by v. The sampling period is τ, i.e. the probe vehicles collect traffic information at τ intervals. The information updating period is ς, that is, the probe vehicles transmit traffic data at ς intervals. Table I lists the notations used in this paper.

A. Sampling Period

Sampling period, which has great impacts on information integrality, data processing workload and communication cost, is critical for the design of traffic information system. If the sampling period is too long, some essential information may be missed. If it is too short, tremendous amounts of resources will be required to process and transfer data.

Definition 1: traffic space correlation, denoted by $\rho_s(P,Q)$, is the traffic correlation degree between traffic situations of two spots, P and Q.

If $\rho_s(P,Q) \geq \rho_{th}^s$ and $\rho_s(P,R) \geq \rho_{th}^s$, the spots between P and Q have strong correlation, so do the spots between P and R, that is, the traffic information of P is sampled, then the traffic situations between R and Q can be estimated, where $\overrightarrow{PR}$ and $\overrightarrow{PQ}$ are opposite. If Q and R are also satisfied that $\rho_s(P, Q + l_v) < \rho_{th}^s$ and $\rho_s(P, R - l_v) < \rho_{th}^s$, where l_v is the average length of vehicles, Q and R are called critical correlation spot of P and the distance between them is called correlation distance of P, denoted by $L_c(P)$. Note that ρ_{th}^s is the predetermined correlation degree and it is relative to the allowable error and the confidence level.

Since the traffic situations of P can just represent the traffic situations between Q and R, there should be at least one sampling within $L_c(P)$. Hence, when the speed of probe vehicle is v, the sampling period should be satisfied:

$$v\,\tau \leq L_c(P)$$

TABLE I

Index of symbols used in the paper

S	length of road	T	total observation time
v	velocity of vehicle	M	number of probe vehicles
τ	sampling period	$\rho_s(P,Q)$	traffic correlation degree between P and Q
ς	transmitting period	s_i	detected road length by the i-th probe vehicle
C	coverage	t_i	detected time period by the i-th probe vehicle
l_v	average length of vehicles	$\rho_t(p,\tilde{t})$	traffic correlation degree between t_1 and $t_1 + \tilde{t}$
L_c	correlation distance	$r(v_P, v_Q)$	the correlation coefficient between v_P and v_Q
ε	mean squared error	k	number of data in a data packet
t_c	correlation time		

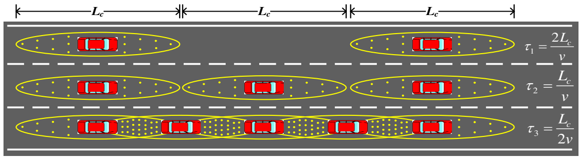


Fig. 1. Results of Different Sampling Periods

That is,

$$\tau \leq \frac{L_c\left(P\right)}{v} \tag{1}$$

According to (1), we acquire the upper bound of sampling period. Note that the probes equipped on vehicles can detect real-time speed , so it can adjust its sampling period dynamically according to the detected speed. The expression of function ρ_s and the estimation of ρ_{th}^s and $L_c\left(P\right)$ will be presented in Section IV.

Figure 1 shows the different results brought by different sampling periods. The field with yellow dots represents correlation distance of the vehicle's position, i.e., the traffic situations of the field can be estimated by the sampled information of the vehicle. As shown in the figure, it is obvious that when the sampling period is longer than L_c/v, there are some places whose traffic situations remain unknown.

B. Transmitting Period

The size of probe vehicle's sample data, which just consist of position and speed information, is very small. If probe vehicles send traffic data immediately after sampling, the communication network will be accessed frequently with numerous short data packets. As a result, the communication network is busy while its efficiency is quite low. It is an efficient way to pack several samples into a data packet.

Definition 2: traffic time correlation, denoted by $\rho_t\left(P,\tilde{t}\right)$ is the traffic correlation degree between P's traffic situations of two time points whose interval is $\tilde{t}$.

If $\rho_t\left(P,\tilde{t}\right) \geq \rho_{th}^t$, the traffic situations within two time points, t_1 and t_2, have strong correlation, where $t_2 = t_1 + \tilde{t}$. That is, the traffic information of t_1 is sampled, then we can estimate the traffic situations between t_1 and t_2. The span between t_1 and t_2 is called the correlation time, denoted by t_c. Suppose that the probe vehicles transmit data at time points $0, \varsigma, 2\varsigma\ \cdots$, which are called updating time. It is required that the information transmitted by probe vehicles is all valid and the traffic situations at any time can be estimated. At time $n\varsigma\,(n = 0, 1\cdots)$, the probe vehicle transmits the information collected between $(n-1)\,\varsigma$ and $n\varsigma$, and predict the traffic situations between $n\varsigma$ and $(n+1)\,\varsigma$.

In Figure 2, the blue solid line is the route of a probe vehicle, the blue dashed is the upper bound of the area where the traffic situations can be estimated, the yellow field is the possible area where the traffic situations can be estimated if the sampled data are transmitted immediately, and the diagonal field is the area where the traffic situations can be estimated by the sampled data transmitted at a ς interval. When $\varsigma > t_c$, as shown in Figure 2(a), take a point (s_1, t_1) for example, the information collected at (s_1, t_1) can estimate the traffic situations between t_1 and $t_1 + t_c$, but it is sent at ς $(\varsigma > t_1 + t_c)$, so at that time the traffic situations duringt_1 to $t_1 + t_c$ is overdue. Hence the data sampled between $n\varsigma$ and $(n+1)\,\varsigma - t_c$, circled by green solid curve, are useless. Besides, the traffic situations between $n\varsigma + t_c$ and $(n+1)\,\varsigma$ cannot be estimated and remain unknown. Thus ς should be no more than t_c, i.e. $\varsigma \leq t_c$. Actually, t_c is not a constant under various traffic situations, so ς is always set to be the minimum value of t_c.

Consequently, it is acceptable that the probe vehicles transmit the data at t_c intervals, and the efficiency of the communication network can be improved greatly. Probe vehicles collect information at τ intervals, and at the updating time there will be k spots' information, where $k = \lceil \varsigma/\tau \rceil$.

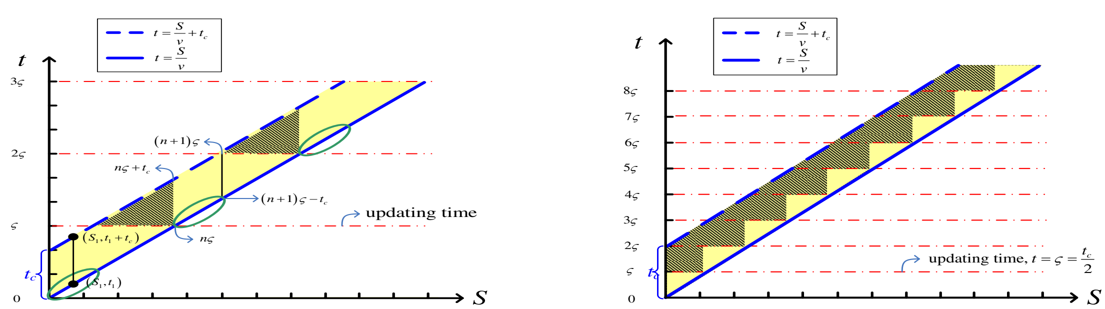
C. Sample Size

We use a two-dimensional space to describe the road network coverage, which is an important metric used to measure the performance of any probe-based traffic system.

Definition 3: coverage area is the product of covered road length and covered time of which the traffic information is known. Investigation area is the product of total road length and total investigated time. Coverage is the ratio of coverage area to investigation area.

Coverage:

$$C = \frac{\bigcup\limits_{i=1}^{M} s_i t_i}{ST}$$



(a) $\varsigma = 1.5t_c$ (b) $\varsigma = 0.5t_c$

Fig. 2. Traffic Estimation Area of Different Transmitting Period

where s_i is the length of the road whose traffic information is detected by the i-th probe vehicle, t_i is the time period, during which the traffic information is detected by the i-th probe vehicle. If the delay of transmitting data is ignored, the coverage area of each datum is $L_c t_c$. Since probe vehicles collect information at τ intervals, during the period T, the number of data collected by each probe vehicle is $\lceil T/\tau \rceil$. If $\tau \leq L_c/v$, the corresponding coverage area is $L_c t_c \lceil T/\tau \rceil$. Otherwise, it is $t_c v T$. Suppose that the coverage area of any two probe vehicles is orthogonal, then the coverage of M probe vehicles is given by:

$$C = \begin{cases} \min\left\{\dfrac{M t_c v}{S}, 100\%\right\} & \tau \leq \frac{L_c}{v} \\ \min\left\{\dfrac{M L_c t_c \lceil \frac{T}{\tau} \rceil}{ST}, 100\%\right\} & else \end{cases}$$

When $\tau \leq L_c/v$, to ensure the coverage is more than a predetermined value C_{th}, the following relation must be satisfied:

$$M \geq \frac{S C_{th}}{v t_c}$$

Especially, if $C_{th} = 1$, M should satisfy:

$$M \geq \frac{S}{v t_c}$$

IV. Model and Estimation

A. Traffic Simulator

With an object oriented programming language, we develop the Traffic & Information-collecting Simulation Platform (TISP) simulator based on car following decision model and lane changing model [2]. The traffic simulator is an application that simulates on real time the behavior of the vehicles in a bidirectional four-lane road and the process of traffic information collection based on probe vehicles. Each vehicle on the road has a random destination and travels following the set rules. The simulator creates some random accidents occasionally. The results of the collected traffic information are observed online and output to a file when the simulation ends. With TISP, we can simulate the traffic behavior and collect data to test the performance of traffic information system under various situations.

B. Correlation Distance

In our simulation, the average velocity is chosen to represent traffic situations and the least-square method is used to predict the traffic situations which are not sampled directly. Hence, $\rho_s(P,Q)$ is defined as the correlation coefficient between the average velocity of P and Q, i.e.

$$\rho_s(P,Q) = r(v_P, v_Q) = \frac{COV(v_P, v_Q)}{\sigma_{v_P} \sigma_{v_Q}}$$

where r is the correlation coefficient, v_P and v_Q is the average velocity of P and Q respectively, σ_{v_P} and σ_{v_Q} is the standard deviation of v_P and v_Q respectively , and $COV(v_P, v_Q)$ is the covariance of v_P and v_Q.

Suppose v_P is observed and a good estimate of v_Q, say $\hat{v}_Q = \hat{v}_Q(v_P)$, is desired. The quality of the estimator can be measured by the resulting mean squared error defined as

$$\varepsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(v_Q(i) - \hat{v}_Q(i)\right)^2}$$

where n is the number of observation data. Based on least-square optimization, the designed estimator is a linear function of the observation, i.e. $\hat{v}_Q = a v_P + b$, where a and b are fixed constants chosen to minimize the mean squared error, and the minimum error can be achieved, i.e.

$$\varepsilon = \sqrt{\frac{1}{n}\left(1 - r^2(v_P, v_Q)\right)\sigma_{v_Q}^2}$$

Based on the result, we can get the threshold of the correlation coefficient ρ_{th}^s, according to a predetermined mean squared error ε_{th}^s. It is required that

$$\varepsilon = \sqrt{\frac{1}{n}\left(1 - r^2(v_P, v_Q)\right)\sigma_{v_Q}^2} \leq \varepsilon_{th}^s$$

i.e.

$$r(v_P, v_Q) \geq \rho_{th}^s = \sqrt{1 - \frac{n\left(\varepsilon_{th}^s\right)^2}{\sigma_{v_Q}^2}}$$

Algorithm 1 Correlation Distance Search

Initialization: the search step size is set to be the average length of the vehicle l_v
$X = p(P) + l_v$
$Y = p(P) - l_v$
$Q = P$
while $r(v_P, v_X) \geq \rho_{th}^s$ **do**
  $Q = X$
  $X = p(X) + l_v$
end while
$R = P$
while $r(v_P, v_Y) \geq \rho_{th}^s$ **do**
  $R = Y$
  $Y = p(Y) - l_v$
end while
$L_c(P) = p(Q) - p(R)$

Once the threshold of the correlation coefficient is computed, Algorithm 1 is used to calculate the correlation distance. The algorithm starts by choosing two starting spots, X and Y, whose positions are set to be $p(P) + l_v$ and $p(P) - l_v$ respectively, where the output of the function p is the position of the spot. The algorithm then search the critical correlation spot Q and the relation $r(v_P, v_Y) \geq \rho_{th}^s$ is checked. If it is satisfied, Q is set to be X, and then X will be updated by the spot whose position is $p(X) + l_v$. This step will be repeated until $r(v_P, v_X) < \rho_{th}^s$. Next, the algorithm searches the critical correlation spot R in the same way. The correlation distance of P is obtained by $L_c(P) = p(Q) - p(R)$.

C. Correlation Time

Similarly, the correlation time can be obtained by the method introduced in Section IV-B. The relationship between

Algorithm 2 Correlation Time Search

Initialization: the search step size is set to be Δt
$Z = t_1 + \Delta t$
$t_2 = t_1$
while $r\left(v_{t_1}, v_Z\right) \geq \rho_{th}^t$ **do**
  $t_2 = Z$
  $Z = Z + \Delta t$
end while
$t_c = t_2 - t_1$

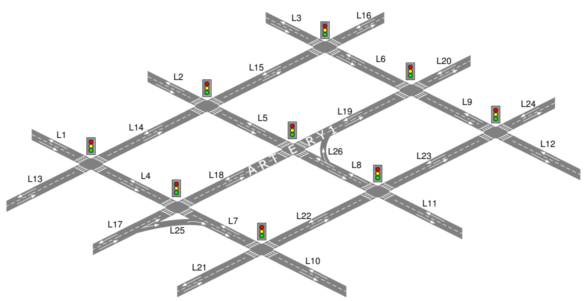


Fig. 3. Diagram of Simulated Traffic Network

the predetermined mean squared error ε_{th}^t and the threshold of the correlation coefficient ρ_{th}^t is

$$\rho_{th}^t = \sqrt{1 - \frac{n\left(\varepsilon_{th}^t\right)^2}{\sigma_{v_{t_2}}^2}}$$

The correlation time can be obtained by Algorithm 2.

V. Simulation

A. The Simulation Traffic Network

We employ the simulator introduced in Section IV to simulate an urban traffic network shown in Figure 3, which consists of 9 intersections and 26 road segments. Speed data on four 0.75-kilometer links ($L17$, $L18$, $L19$ and $L20$), which all have 4 lanes in each direction, were selected to analyze. For simplicity, the artery composed by the four links $L17$, $L18$, $L19$ and $L20$ is called Artery 1.

B. Simulation Scenarios

Different arrival rates of vehicles, which result in different road network occupancies, were adopted to simulate the traffic situations under peak and off-peak hours. The following three scenarios were simulated,

Scenario 1 : the occupancy is around 4%.

Scenario 2 : the occupancy is around 20%.

Scenario 3 : the occupancy is around 50%. where the occupancy is defined as the ratio of the total length of the vehicles to the total length of the road, i.e.

$$\frac{number\ of\ vehicles \times length\ of\ vehicle}{number\ of\ lanes \times length\ of\ road} \tag{2}$$

The occupancies of the links are diverse, because of different road qualities, different densities of the surrounded residential areas, different densities of the bus stations, and so on. Each simulation period is 6000 sec, and the beginning part of which is a warm up period of 1000 sec.

C. Evaluation of Correlation Distance

Figure 4 plots the correlation distances of Artery 1, at three different occupancies, where ρ_{th}^s is set to be 0.9 and the occupancy (OCC) of each link is listed in the figure. It is obvious that in all the scenarios the correlation distances within a link are almost the same, except some special points which are very close to the intersections. The correlation distance decreases slightly with increasing road occupancy and is highly relative to the length of the link.

When traffic is light, the vehicles can travel with a free speed at any place of the highway. So the correlation distance approximates the length of the highway. As traffic increasing, the movement of the vehicles are influenced by traffic light and other vehicles, so the correlation distance decreases.

The traffic situations of the points near the intersections are highly impacted by the traffic light. Therefore the correlation distance of the special points close to the intersections, is always near 0. Since the number of the "special points" is very small and in a real system, the traffic situations near the traffic lights can be predicted easily according to the control of the traffic lights, the points can be ignored.

Consequently, based on the method introduced in Section III, the sampling period can be obtained. Take the link $L19$ for example, when its occupancy is less than 30%, the correlation distance is about 520m, so a probe vehicle traveling on the link with a speed $20m/s$ may take $26s$ as its sampling period.

D. Evaluation of Correlation Time

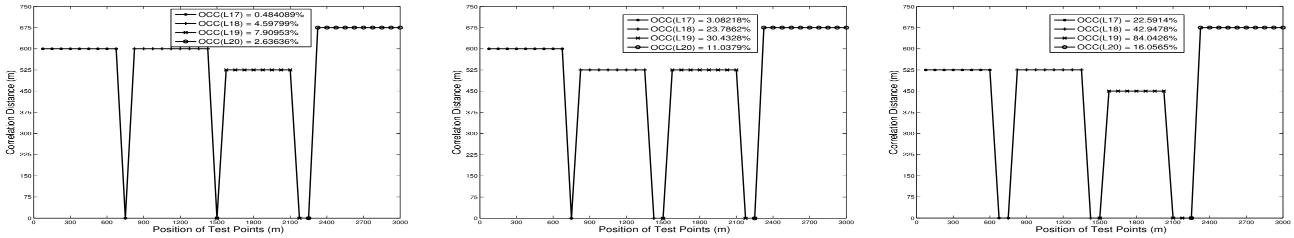
Figure 5 shows the correlation time of Artery 1, at three different network occupancies, where ρ_{th}^t is set to be 0.9 and the occupancy (OCC) of each link is listed in the figure. The figures imply that the correlation time depends on the distance from the traffic light and the road occupancy.

To the points far from the traffic lights, there is no restriction and the vehicles can travel with a free speed, so the correlation time on those places is long. When the traffic is light, vehicles on most part of the highway can travel with a free speed, so the correlation time can be very long if the traffic does not turn to heavy. As the traffic increasing, on the places suffered the impact of the traffic light, the traffic situations vary frequently because of the inter-action among vehicles and the control of the traffic lights.

According to the method introduced in Section III, the transmitting period and the minimum number of probe vehicles can be acquired. Take Artery 1 for example, when the occupancy is 50% and the average speed of the vehicles is 5m/s, since most of the correlation time is larger than 100s, the probe vehicle traveling on the link may take $100s$ as its transmitting period and the number of probe vehicles on Artery 1 should be more than 5 to guarantee $C_{th} = 0.9$.

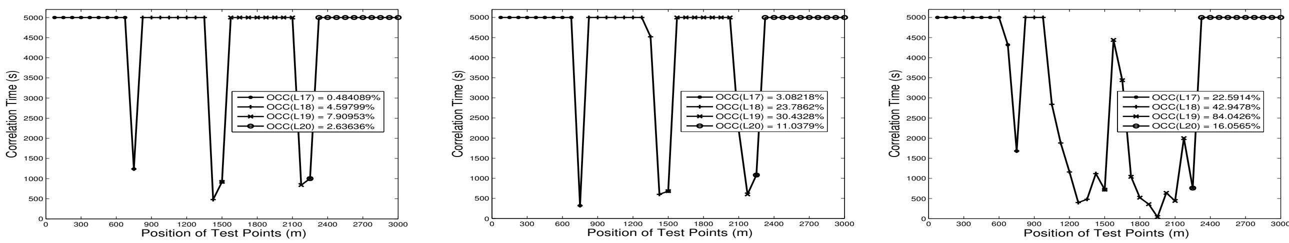
E. Evaluation of Coverage

Figure 6 shows the average coverage of Artery 1, versus its occupancy ranging from 0 to 60%, with different sample size. The coverage increases with the sample size, and reduces with the road occupancy. It is because when the traffic gets



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Fig. 4. Correlation Distance of Artery 1



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Fig. 5. Correlation Time of Artery 1

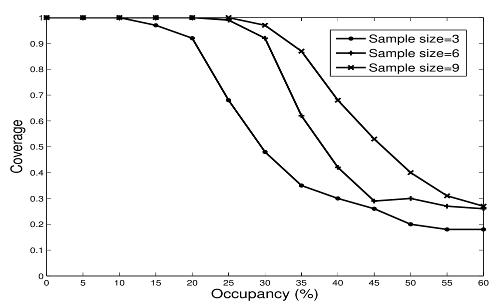


Fig. 6. Coverage of Artery 1

heavier, the correlation distance and correlation time become smaller, i.e. the vehicle collects a sample, the range which represents is smaller, and the vehicles are slowed down, so some of the places cannot be reached by probe vehicles within a certain period. It is found that the number of probe vehicles required for integral traffic information is about 2% of all the vehicles on the road. The percentage can be calculated easily by $M \times l_v/$(occupancy$\times$ total length of road)

VI. Conclusion

In this paper, we address some of the key issues involved in the design of such a traffic information system using vehicles as probes. The bounds of the sampling period, transmitting period and sample size of probe vehicles, which are three critical parameters that determine the performance of the system, are derived. Based on the results, system designers can choose parameters according to different performance requirements. A simulation platform is developed to study the properties of traffic flows. Based on the simulation, it is found that the number of probe vehicles required for good traffic information integrality is about 2% of all the vehicles on the road.

References

[1] K. K. Sanwal and J. Walrand, "Vehicles as probe," California PATH working paper, U.C. Berkeley-ITS-PWP-95-11, August 1995.

[2] H.D. Robertson, J.E. Hummer and D.C. Nelson, Manual of transportation engineering studies, Prentice Hall, Englewood Cliffs, NJ; 1994.

[3] C. A. Quiroga and D. Bullock, Determination of sample sizes for travel time studies, *Institute of Transportation Engineers Journal*, vol. 68(8), 1998, pp. 92-98.

[4] S. Li, K. Zhu, B.H.W. van Gelder, J. Nagle and C. Tuttle, "Reconsideration of Sample Size Requirements for Field Traffic Data Collection Using GPS Devices," in *Proceedings of the TRB 2002 Annual Meeting*, 2002.

[5] R. L. Cheu, C. Xie and D. Lee, Probe Vehicles population and Sample Size for arterial speed Estimation, *Computer-aided civil and infrastructure engineering*, vol. 17(1), 2002, pp. 53-60.

[6] S. Cohen, J.L. Bosseboeuf and N. Schwab, "Probe vehicle sample sizes for travel time estimation on equipped motorways," in *Proceedings of Eleventh International Conference on Road Transport Information and Control*, 2002, pp. 172-176.

[7] S.M. Turner, D.J. Holdener, "Probe Vehicle Sample Sizes for Real-Time Information: The Houston Experience", in *Proceedings of Vehicle Navigation and Information Systems (VNIS) Conference*, 1995, pp. 3-9.

[8] C. G. Park, J. Oh, and S. Kim, "Determination of optimal number of probe vehicles for real-time traffic flow information," in *the 5-th ITS Congress*, Seoul, Korea, 1998.

[9] X. Dai, M. A. Ferman, and R. P. Roesser, "A simulation evaluation of a real-time traffic information system using probe vehicles," in *Proceedings of IEEE Intelligent Transportation System Conference*, 2003, vol. 1, pp. 475-480.

[10] L. Wang, C. Wang, X. Shen, Y. Fan, "Probe vehicle sampling for real-time traffic data collection," in *Proceedings of the 2005 IEEE Intelligent Transportation System Conference*, 2005, pp. 222-224.

[11] G. Jiang, L. Gang, and Z. Cai, "Impact of Probe Vehicles Sample Size on Link Travel Time Estimation," in *Proceedings of IEEE Intelligent Transportation System Conference*, 2006, pp. 505-509.