**Proceedings of the 2007 IEEE
Intelligent Transportation Systems Conference
Seattle, WA, USA, Sept. 30 - Oct. 3, 2007**

Analysis of Time- and Space-domain Sampling for Probe Vehicle-based Traffic Information System

Jun Hong, Xuedan Zhang, Jianshu Chen, Zhongya Wei, Jiannong Cao and Yong Ren

Abstract— **Using vehicles as probes is a flexible and low-cost way to obtain real-time traffic information. A key problem of using probe vehicles is to determine vehicle's sampling period and probe sample size. This paper addresses the sampling issue of using probe vehicles for detecting traffic information. An extended Nyquist sampling theorem based on signal processing theory is proposed to derive bounds on the sampling period and probe sample size. 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. We demonstrate through simulations that the traffic situations can be reconstructed from the samples without distortion according to the designed sampling period and sample size. The methods proposed in this paper can be a valuable tool for designers on planning parameters in the probe vehicle-based traffic information system.**

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, traffic accident detection and so on. One possible approach to obtain traffic information is to use vehicles running on the road as probes. The key idea of using probe vehicles to collect traffic information is that a vehicle traveling 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]. The probe vehicles can report data on their positions, speeds, space between vehicles and so on. 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. For probe vehicles to be the basis of a real-time traffic information system, it is necessary that the traffic situations sampled by them are reliable and adequate. Hence, a key problem of using probe vehicles as a source of data is to determine the probe sample size.

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`
Jianshu Chen is with the Department of Electrical Engineering, Tsinghua University, Beijing, 100084, P.R.China `jsc06@mails.thu.edu.cn`
Zhongya Wei is with the Department of Electrical Engineering, Tsinghua University, Beijing, 100084, P.R.China `hze@thu.edu.cn`
Jiannong Cao is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, P.R.China `csjcao@comp.polyu.edu.hk`
Yong Ren is with the Department of Electrical Engineering, Tsinghua University, Beijing, 100084, P.R.China `reny@thu.edu.cn`

Over the past decades, numerous studies discussing probe sample size have been carried out mainly based on two theories, quality-control theory [2]–[7] and large sampling theory [8]–[11]. Most of research focuses on estimating the sample size 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. Moreover, there have been few systematic efforts using theoretical analysis to solve the probe sampling problems. This paper deals with the issues of estimating the sampling period and the sample size for traffic information system using probe vehicles. The approach uses concepts from signal processing theory and an analytical model based on the Nyquist sampling theorem is employed to derive the bounds of vehicle's sampling period and probe sample size. To our best knowledge, this is the first paper introducing spatio-temporal characteristics and approach in signal processing 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 and the number of probe vehicles that are desirable in a traffic network in order to achieve certain accuracy in traffic situation estimation.

The rest of the paper is organized as follows. Section II formally defines the problems and presents some theorems which will be used in the following sections. Section III demonstrates our methodology for deriving vehicle's sampling period and sample size. Section IV estimates some important parameters which are relative to sampling period and sample size. Section V evaluates the performance of the designed parameters through computer simulations. Section VI concludes the paper.

II. FORMULATION

Since traffic situations are continuous signals, it is desirable that samples taken at discrete points can be used to represent them. The typical method of obtaining a discrete representation of a continuous signal is through periodic sampling. We use approaches introduced in signal processing to deal with the sampling problems of using probe vehicles for detecting traffic information.

Traffic is a spatio-temporal process which consists of information in terms of both time and space, so the traffic sampling issue can be divided into two sub-problems: time-domain sampling and space-domain sampling.

1-4244-1396-6/07/$25.00 ©2007 IEEE.

A. Time-domain Sampling of Traffic Situations

Traffic situation at a given point on the road network, denoted by $x_a(t)$ is a continuous-time signal, and under reasonable constraints it can be quite accurately represented by samples taken at discrete points in time.

Periodic sampling is used to obtain a discrete-time representation of $x_a(t)$, wherein a sequence of samples, $x(n)$, is obtained from the continuous-time signal $x_a(t)$ according to the relation:

$$x(n) = x_a(nT)$$

where T is the sampling period, and its reciprocal, f_s, is the sampling frequency, in samples per second.

Since many continuous-time signals can produce the same output sequence of samples, the sampling operation is generally not invertible. As a result, the sampling period T should be restricted to ensure that the signal can be recovered without distortion. Nyquist sampling theorem provides an insight on determining the sampling period.

Theorem 1 (Nyquist Sampling Theorem): Let $x_c(t)$ be a bandlimited signal whose maximum frequency is f_{x_c}, i.e.,

$$X_c(j2\pi f) = 0, for\, |f| \geq f_{x_c}$$

where X_c is the Fourier transform of $x_c(t)$. Then $x_c(t)$ is uniquely determined by its samples $x(n) = x_c(nT)$, $n = 0, \pm 1, \pm 2 \cdots$, if the sampling frequency f_s is satisfied:

$$f_s = \frac{1}{T} \geq 2 f_{x_c}.$$

The frequency $2f_{x_c}$ that must be exceeded by the sampling frequency is called the Nyquist frequency. The proofs of Theorem 1 can be found in [12].

Recurring to Nyquist sampling theorem, we derive the upper bound on the sampling period T of a given point's traffic situation $x_a(t)$, i.e.,

$$T \leq \frac{1}{2f_x}.$$

The frequency f_x, called the cut-off frequency, is the maximum frequency contained in the signal $x_a(t)$.

B. Space-domain Sampling of Traffic Situations

Since traffic situation at a given time along the road is a continuous-space signal, the sampling process used to derive a discrete-space representation of the signal can be obtained by mapping space to time, and thus the space-domain sampling is equivalent to the time-domain sampling.

The traffic situation along the road, denoted as $y_a(s)$, is sampled uniformly every S meters to produce the discrete-space signal

$$y(n) = y_a(nS)$$

where S is the sampling distance and its reciprocal, h_s, called sampling space-rate, is the number of samples per meter.

We propose a space sampling theorem, which is the extension of Nyquist sampling theorem, to solve the sampling problem for a continuous-space signal.

Theorem 2 (Extended Nyquist Sampling Theorem): Let $y_c(s)$ be a bandlimited signal whose maximum space-rate is h_{x_c}, i.e.,

$$Y_c(j2\pi h) = 0, for\, |h| \geq h_{y_c}$$

where Y_c is the Fourier transform of $y_c(s)$. Then $y_c(s)$ is uniquely determined by its samples $y(n) = y_c(nS)$, $n = 0, \pm 1, \pm 2 \cdots$, if the sampling space-rate h_s is satisfied:

$$h_s = \frac{1}{S} \geq 2h_{y_c}.$$

The frequency $2h_{y_c}$ that must be exceeded by the sampling space-rate is called the Nyquist space-rate.

From Theorem 2, we know that the sampling distance S of traffic situation $y_a(s)$ should be satisfied,

$$S \leq \frac{1}{2h_y}$$

where h_y is the maximum space-rate contained in the signal $y_a(s)$. We call h_y the cut-off space-rate.

The critical parameters of the probe vehicle-based traffic information system, including vehicle's sampling period and probe sample size, will be determined so that the two sampling theorems' requirements can be satisfied.

III. Methodology

A. Vehicle's Sampling Period

Probe vehicles traveling on the road sample the traffic situations periodically, and the vehicle's sampling period is denoted as τ. That is, the probe vehicles collect traffic information at every τ seconds.

Vehicle's sampling period has great impacts on information integrality, data processing workload and communication cost, and it is critical for the design of traffic information system. If vehicle's sampling period is too long, some essential information may be missed. If it is too short, numerous resources will be required to process and transfer data. Note that the "vehicle's sampling period" is different from the "sampling period" mentioned in Section II-A.

According to the extended Nyquist sampling theorem, there should be at least one sample within a distance of $\frac{1}{2h_y}$ meters, so that the traffic situation $y_a(s)$ can be represented accurately. Suppose the speed of probe vehicle is v, the traffic situations along the road can be obtained if the following inequality holds,

$$v\,\tau = S \leq \frac{1}{2h_y}$$

that is,

$$\tau \leq \frac{1}{2vh_y}. \quad (1)$$

According to (1), we acquire the upper bound of the vehicle's sampling period.

In the real situation, the vehicle speed is variable, which means that the vehicle's sampling period is not a constant and how to determine its value remains a problem. Fortunately, the probe equipped on the vehicle can detect real-time vehicle speed, so it can adjust the sampling period dynamically according to the detected speed.

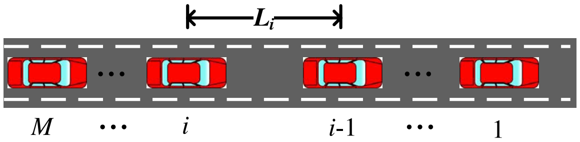


Fig. 1. Road with Probe Vehicles

B. Sample Size

As shown in Figure 1, the length of the road is L, the number of probe vehicles running on the road is M, the speed of the i-th probe vehicle is v_i, the distance between the (i-1)- and i-th probe vehicles is L_i, and the position of the i-th probe vehicle is denoted as s_i.

According to Theorem 1, the sampling period for a given point's traffic situation $x_a(t)$ should be less than $\frac{1}{2f_x}$ seconds, which implies that the time interval between two adjacent vehicles' arrivals at a certain point should be less than $\frac{1}{2f_x}$ seconds. That is, the i-th probe vehicle should reach s_{i-1} within $\frac{1}{2f_x}$ seconds, i.e.,

$$\frac{L_i}{v_i} \leq \frac{1}{2f_x}, i = 1, 2 \cdots M.$$

By adding the M inequalities together, the following relation can be obtained:

$$\sum_{i=1}^{M} L_i \leq \frac{1}{2f_x} \sum_{i=1}^{M} v_i.$$

Note that the the sum as i runs from one to M of the L_i is L, then

$$L \leq M \frac{\bar{v}}{2f_x}$$

where $\bar{v}$ is the average speed of the probe vehicles. As a result, the lower bound of the sample size is derived:

$$M \geq \frac{2Lf_x}{\bar{v}}.$$

C. Reconstruction of Traffic Situations

Discrete samples are used to represent the traffic situations, which are continuous signals, so the remained problem is to reconstruct the original traffic situations from the sampled signals. In time-domain sampling, when the sampling is performed at the required frequency, which is greater than the Nyquist frequency, the traffic situations can be recovered from the samples without distortion with the formula,

$$x_a(t) = \sum_{n=-\infty}^{\infty} x(nT) \frac{\sin \pi (t/T - n)}{\pi (t - nT)/T}.$$

Similarly, we can get the formula for reconstructing the space-domain signals,

$$y_a(s) = \sum_{n=-\infty}^{\infty} y(nS) \frac{\sin \pi (s/S - n)}{\pi (s - nS)/S}.$$

Consequently, the traffic situations in terms of time and space can be obtained from the samples collected by probe vehicles.

IV. Parameter 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. The simulator TISP is validated with real data obtained from Wudaokou road, which is in front of the east gate of Tsinghua University.

With TISP, we can simulate the traffic behavior and collect data to test the performance of traffic information system under various conditions.

B. Nyquist Frequency Estimation

In order to determine the Nyquist frequency of the traffic situations at a given point, we conduct a survey on the "duration of changed state". 37 drivers were invited to give a value of t_{th}. If the traffic situations change from one state to another, whose duration only lasts less than t_{th} seconds, and then restore, the driver thinks that there is no need to know the changed state and it can be ignored. That is, if a sharp change in traffic situations, for example, an accident, takes place at a given point, and the traffic situation recovers soon, probe vehicles do not have to sample and report the accident. The data collected in the survey suggests that the average value of t_{th} is 292 seconds, i.e. changed state whose duration is less than 292 seconds can be ignored. Note that the value of t_{th} may be various under different environments.

According to signal processing theory, we know that if the duration of a impulse is t, its frequency f approximates $\frac{1}{t}$. The frequency of the signals with a less than 292 second duration is greater than $\frac{1}{292}$. Based on the survey result, when estimating the traffic situations, we can filter out the contained impulses whose frequency is greater than $\frac{1}{292}$. So the cut-off frequency of the filtered traffic situations is $\frac{1}{292}$, and the lower bound of the sample size is

$$M \geq \frac{2L}{t_{th}\bar{v}} = \frac{L}{146\bar{v}}. \quad (2)$$

By (2), designer of probe vehicle-based traffic information system can predetermine the number of probe vehicles according to the size of the road network.

C. Nyquist Space-rate Estimation

In our simulation, the time-mean speed v_t is chosen to represent traffic situations [13]. The precise definition of v_t, at a given point p on the road and a given time t, is included in the following instruction for measuring it. Draw two lines across the road, a shot distant dp apart, to form a slice of road with the point p in the middle. Take averages over a time interval of length ς, with the time t in the middle. The number of vehicles crossing the slice dp in time ς is n. Then time-mean speed is the average speed of vehicles crossing the slice, namely

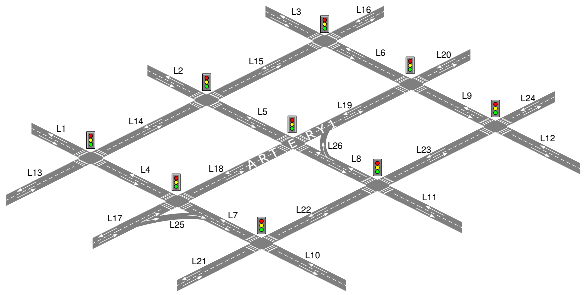


Fig. 2. Diagram of Simulated Traffic Network

$$v_t = \frac{1}{n} \sum \frac{dp}{dt}$$

Note that if there is no vehicle passing the slice, i.e. $n = 0$, the time-mean speed is set to be the average free speed.

With the simulator, a series of traffic data are produced, then the Fourier transform of $y_a(s)$ is obtained by the following formula,

$$Y(h) = \int_{-\infty}^{+\infty} y_a(s) e^{-j2\pi hs} ds.$$

Hence, the Nyquist space-rate of $y_a(s)$ can be observed from $Y(h)$.

V. Simulation

A. The Simulation Traffic Network

We employ the simulator introduced in Section IV to simulate an urban traffic network shown in Figure 2, which consists of nine intersections and 26 road segments. Speed data on four 0.75-kilometer links ($L17$, $L18$, $L19$ and $L20$), which all have four lanes in each direction, were selected to analyze. For simplicity, the artery comprising the four links $L17$, $L18$, $L19$ and $L20$ is called Artery 1.

B. Simulation Scenarios

The source of traffic is taken to be at each end of the links, and the time-mean speed v_t is observed. We consider the case when the traffic flow rises to a peak gradually, producing traffic congestion. The rise of flow can be easily realized by increasing arrival rates of vehicles, which result in different road network occupancies. The following three traffic conditions (with incidents) were adopted to simulate the traffic situations under peak and off-peak hours,

Scenario 1 (from off-peak to peak) : the average network occupancy is increasing from approximately 8% to 50%.

Scenario 2 (off-peak) : the average network occupancy is around 8%.

Scenario 3 (peak) : the average network occupancy is around 50%.

The precise definition of occupancy is the ratio of the total length of the vehicles to the total length of the road. Each simulation period is 4500 seconds, and the beginning part of which is a warm up period of 1500 seconds.

C. Traffic Situation at a Given Point

Since the probe sample size is predetermined and it is inconvenient to adjust the number of probe vehicles to different traffic situations dynamically, the predefined sample size should be able to achieve reliable traffic information under various traffic conditions. Scenario 1 that the traffic flow rises gradually from off-peak to peak state is simulated to evaluate the performance of the designed sample size under a variety of traffic conditions. .

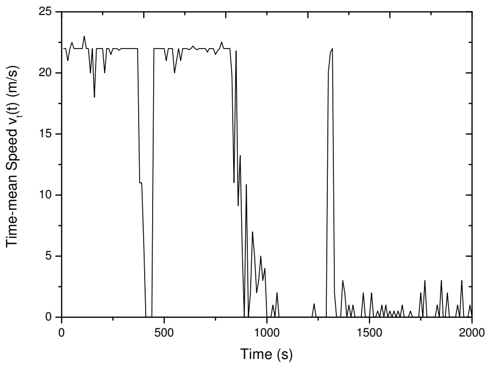
Figure 3 illustrates the time-domain sampling and recovering of traffic situations under Scenario 1. Figure 3(a) shows the fluctuation in traffic situations $v_t(t)$ at a given point over a period of 2000 seconds. The corresponding spectrum of $v_t(t)$ is shown in Figure 3(b). Based on the result derived in Section IV-B, we have $M \geq \frac{3000}{146\bar{v}}$. According to the real data collected from Wudaokou road, the probability that the average speed of vehicles exceeds 4 m/s is greater than 90%. So $\bar{v}$ is set to be 4 m/s here, and hence five probe vehicles are running on Artery 1 to sample traffic information. The samples collected by the probe vehicles are shown in Figure 3(c). Note that the sampling period is not uniform, because the arrival time of the probe vehicles is not fixed. This phenomena influences the sampling result little.

Figure 3(d) shows the recovery of the continuous-time signal from its samples. At some time points, the traffic situations are not recovered correctly. This is because the duration of the states at these time points is too short, less than 292 seconds. And according to the results in SectionIV-B, the "impulses" are ignored. However, the distortion is acceptable since the lost information is trivial and the drivers care little about it.

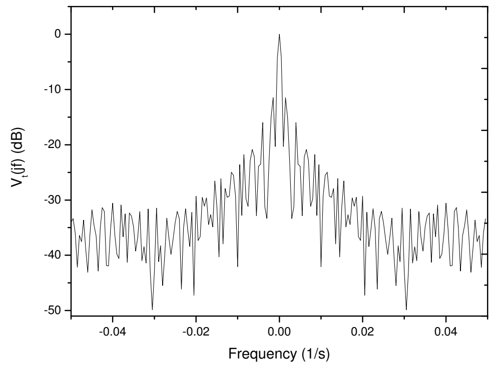
D. Traffic Situation at a Given Time

Figure 4 and 5 illustrate the space-domain sampling and recovering of traffic situations. The curves in Figure 4(a) and 5(a) are fluctuation in traffic situations $v_t(s)$ along Artery 1 for two different network occupancies respectively. The corresponding spectrums of the two traffic situation signals are shown in Figure 4(b) and 5(b). It can be seen from the spectrums that the cut-off space-rate of the two signals approximates 0.013 and 0.025 respectively. As a result, 150 meters is taken as the sampling distance for the traffic situations under off-peak hours, and under peak hours when the cut-off space-rate is higher, 75 meters is chosen to be the sampling distance. The samples collected by probe vehicles at the determined sampling distance, are shown in Figure 4(c) and 5(c).

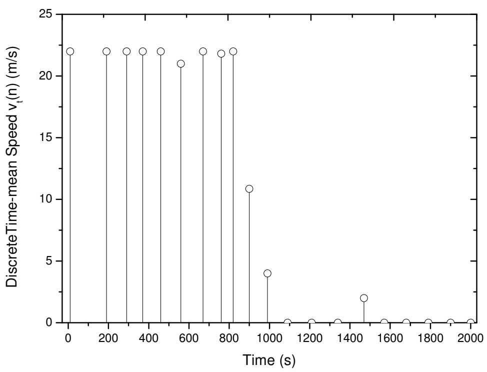
Figure 4(d) and 5(d) show the recovery of the continuous-space signals from their samples. It is satisfying that the original traffic situations can be almost reconstructed from the samples without distortion. Note that the cut-off space rate grows up with increasing road occupancy. When the traffic is light, there is no obstacle on the road and the vehicles can travel with a free speed at any place of the highway. So the traffic situations along the road change slowly. As the traffic increasing, the movement of the vehicles is influenced by traffic light and other vehicles, so the traffic



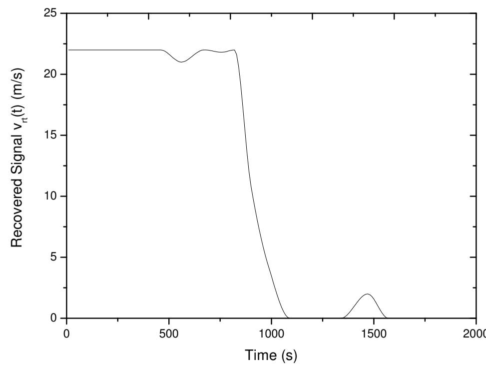
(a) Traffic Conditions at a Given Point



(b) Spectrum $V_t(jf)$ of Signal $v_t(t)$

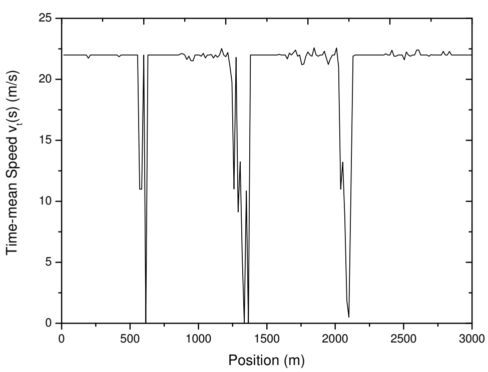


(c) Samples Taken by Probe Vehicles

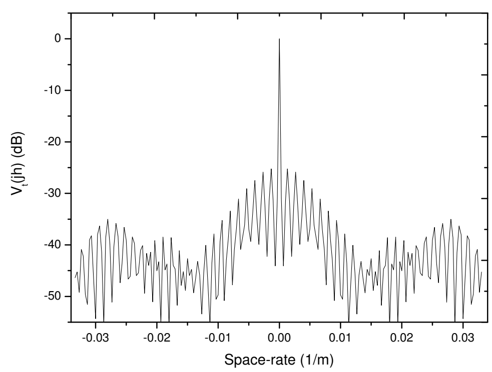


(d) Recovery of Signal $v_t(t)$

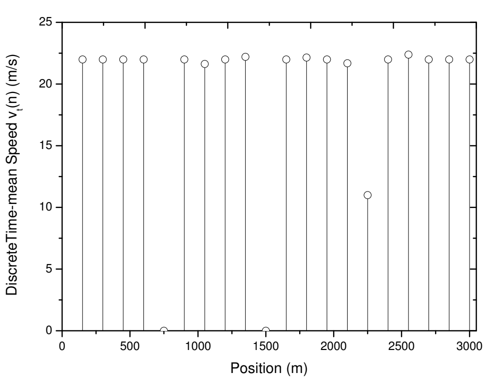
Fig. 3. Sampling of Time-continuous Traffic situations



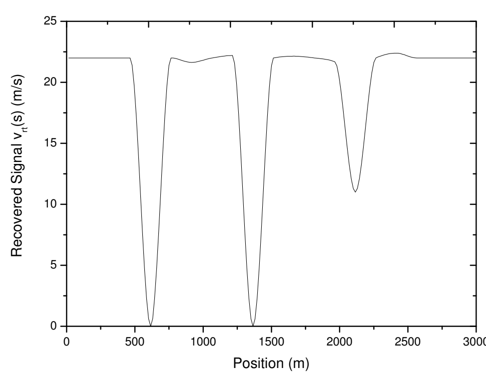
(a) Traffic Situations along Artery 1



(b) Spectrum $V_t(jh)$ of Signal $v_t(s)$

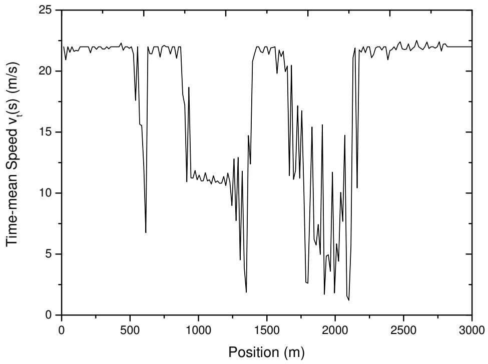


(c) Samples Taken by Probe Vehicles

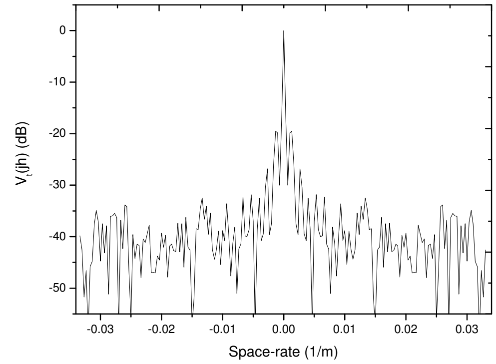


(d) Recovery of Signal $v_t(s)$

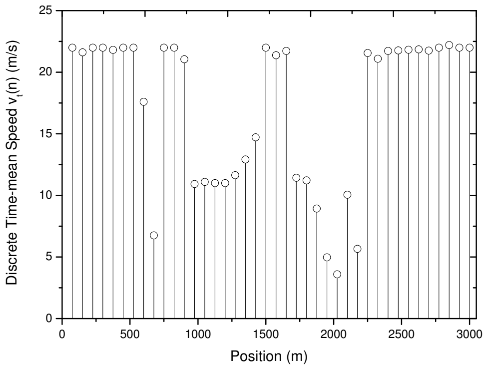
Fig. 4. Sampling of Space-continuous Traffic Situations under Scenario 2



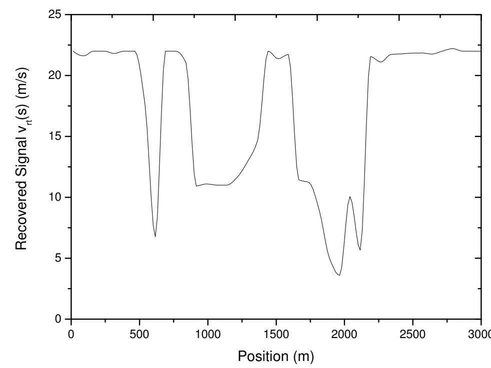
(a) Traffic Situations along Artery 1



(b) Spectrum $V_t(jh)$ of Signal $v_t(s)$



(c) Samples Taken by Probe Vehicles



(d) Recovery of Signal $v_t(s)$

Fig. 5. Sampling of Space-continuous Traffic Situations under Scenario 3

situations at different points vary a lot. Hence, with higher road occupancy, the sampling distance should be smaller so that the variation among different places can be obtained by the collected samples.

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 and sample size of probe vehicles, which are two critical parameters to determine the performance of the system, are derived. Based on the results, system designers can determine the probe sample size according to the size of road network in advance, and the vehicle's sampling period under off-peak and peak hours are also obtained in this paper. Furthermore, a simulation platform is developed to study the properties of traffic flows in a road network. We demonstrate through simulations that the traffic situations recovered from samples, which are obtained with designed parameters, are accurate and reliable.

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.

[12] A. V. Oppenheim, R. W. Schafer and J. R. Buch, Discrete-time Signal Processing, Prentice Hall, Englewood Cliffs, NJ; 1994.

[13] M. J. Lighthill, G. B. Whitham,"On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads," in *Proceedings of the Royal Society of London*, 1955, vol.229, pp. 317-345.