

## Research Article

# The Optimal Sampling Period of a Fingerprint Positioning Algorithm for Vehicle Speed Estimation

Ding-Yuan Cheng,<sup>1</sup> Chi-Hua Chen,<sup>2,3</sup> Chia-Hung Hsiang,<sup>2,3</sup> Chi-Chun Lo,<sup>2</sup>  
Hui-Fei Lin,<sup>4</sup> and Bon-Yeh Lin<sup>2,3</sup>

<sup>1</sup> Department of Information Management, Hwa Hsia Institute of Technology, No. 111 Gongzhuang Road, Zhonghe District, New Taipei 235, Taiwan

<sup>2</sup> Institute of Information Management, National Chiao Tung University, No. 1001 University Road, Hsinchu 300, Taiwan

<sup>3</sup> Telecommunication Laboratories, Chunghwa Telecom Co., Ltd, No. 12, Lane 551, Min-Tsu Road Sec. 5, Yang-Mei, Taoyuan 326, Taiwan

<sup>4</sup> Department of Communication and Technology, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan

Correspondence should be addressed to Ding-Yuan Cheng; [kewas.cheng@gmail.com](mailto:kewas.cheng@gmail.com)

Received 5 July 2013; Revised 19 September 2013; Accepted 19 October 2013

Academic Editor: Yi-Chung Hu

Copyright © 2013 Ding-Yuan Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Using cellular floating vehicle data is a crucial technique for measuring and forecasting real-time traffic information based on anonymously sampling mobile phone positions for intelligent transportation systems (ITSs). However, a high sampling frequency generates a substantial load for ITS servers, and traffic information cannot be provided instantly when the sampling period is long. In this paper, two analytical models are proposed to analyze the optimal sampling period based on communication behaviors, traffic conditions, and two consecutive fingerprint positioning locations from the same call and estimate vehicle speed. The experimental results show that the optimal sampling period is 41.589 seconds when the average call holding time was 60 s, and the average speed error rate was only 2.87%. ITSs can provide accurate and real-time speed information under lighter loads and within the optimal sampling period. Therefore, the optimal sampling period of a fingerprint positioning algorithm is suitable for estimating speed information immediately for ITSs.

## 1. Introduction

Intelligent transportation systems (ITSs) have become increasingly popular. For ITSs, establishing an effective real-time traffic information system is essential. Real-time traffic information includes information on the average vehicle speed, travel time, traffic flow, and traffic incidents.

Cellular floating vehicle data (CFVD) technology is used to measure and forecast real-time traffic information based on anonymously sampling mobile phone positions. CFVD is an immediate, cost-effective, and easily deployed solution. Nearly every person owns a cell phone. Thus, using mobile stations (MSs) as probes to collect traffic information is feasible because MSs can be located using location services (LCS), and the speed of mobile phones can be estimated according to two sets of positioning information and time difference information.

To estimate speed, handover signals were retrieved twice from cellular networks, and the time of occurrence was analyzed to infer traffic information. However, the location accuracy of some handover pairs was low. Consequently, a fingerprint positioning algorithm (FPA) is proposed to increase the location accuracy; however, the speed report rate of the FPA has not been investigated.

When analyzing location samples to estimate speed, excessively short sampling periods may generate substantial data computation costs and increase server loads. Conversely, excessively long sampling periods may result in limited speed report data. If speed report data are insufficient, then the accuracy of traffic information declines.

When there are more than two location data for the same call on the same road, the received signal strength (RSS) of measurement reports (MRs) from MSs can be analyzed.

Speed estimates are then computed and reported [1]. In this study, an FPA was used to increase the location and speed estimation accuracy, and a model for analyzing speed-reporting rates is proposed based on the communication behaviors, traffic conditions, and two consecutive fingerprint positioning locations for the same call to determine the optimal sampling period and estimate vehicle speed.

ITSs can provide accurate, real-time speed information under light loads and within the optimal sampling period. Therefore, the optimal sampling period of an FPA is suitable for estimating speed information for ITSs immediately.

The remainder of this thesis is structured as follows. In Section 2, the location determination and speed estimation methods are discussed and compared. The study provides speed estimates based on the FPA and proposes an analytic model for analyzing the speed-reporting rates in Section 3. The numerical and simulation results of speed estimation accuracy and sampling period efficiency are discussed in Section 4. Finally, Section 5 provides the research conclusion.

## 2. Literature Review

This chapter presents a discussion of the background information concerning ITS and research on estimating traffic information by using cellular networks.

*2.1. Intelligent Transportation System.* Because of the economic growth and technological evolution, human mobility has dramatically improved. The proliferation of privately owned transportation equipment and the phenomena of population aggregation caused by urbanization are arguably the major factors contributing to the unbalanced traffic-network usage problem. The traffic volume of some roads often exceeds the road capacity during peak hours, leading to problems such as traffic congestion, traffic accidents, environmental pollution, and energy waste.

Conventional solutions, such as building new road networks, enhancing traffic sign controls, and implementing reversible or offset lanes, have long been adopted. However, these solutions present their own problems. For example, building and expanding road networks require considerable cost and time, and doing so is particularly difficult in urban areas. Traffic sign control and optimal route selection may not be viable when timely and accurate traffic information is unavailable.

With the remarkable advances in information and communications technology, the government, industry, and academia have hoped that the use of modern technology could construct the so-called ITS. The main goals of an ITS are to alleviate growing traffic problems and to use transportation resources effectively.

The effective operation of ITS applications depends on accurate real-time traffic information. Necessary traffic information includes the average speed, traffic volume, density, travel time, traffic accidents, traffic congestion, and weather conditions of a specific segment of a roadway at a specific time. The information is generally collected using preinstalled equipment at certain locations and transmitted to users and traffic control centers to facilitate additional decisions.

The data collection equipment can be installed along the roadway and is known as stationary vehicle detectors (VDs).

Conventionally, VDs are deployed on major roadways to detect the average vehicle speed and traffic flow. However, installing and maintaining such devices is costly. For example, according to a survey [2] conducted by the Texas Transportation Institute in 2002, the cost of building each VD ranged between \$US1,000 and \$US1,500. In Taiwan, one VD for every kilometer was deployed along National Highway No. 1, which means that 760 VDs were acquired to fulfill such traffic-information collection requirements. This amounts to nearly \$US1 million for implementing this solution. In addition, VDs are easily affected by temperature fluctuation, moisture, and other factors; therefore, they require seasonal or annual maintenance.

Another approach is to use probe cars for reporting the traffic information in real time. By merging specific vehicles equipped with location and communication devices with the traffic stream, their travel time and average speed can be monitored during travel. However, the penetration rate of GPS-based probe cars must be adequately high to infer more accurate real-time traffic information. For example, 12% of the active probes are required in the total traffic volume to lower the speed error rate to less than 3% [3]. In addition, additional transmission costs are incurred when probe cars send back data through the air.

Based on real-time and accurate traffic information, the relying parties can make corresponding decisions to meet their specific goals. Traffic control centers can control the traffic signs, control gateways, or bulletin boards. Roadway users can make the appropriate route choice based on factors such as estimated travel time and congestion information. ITS applications that are even more advanced can predict traffic conditions and develop a corresponding strategy to prevent the occurrence of traffic congestion and other undesired traffic conditions.

Therefore, gathering accurate and real-time traffic information is the basis of a successful ITS. To obtain the information efficiently and effectively is crucial in the field of ITS.

*2.2. Traffic Information Estimation Methods Based on Cellular Networks.* Because the mobile phone penetration rate in Taiwan exceeds 100%, we can assume that every vehicle carries at least one mobile phone. Thus, the location of mobile phones can be traced. Several studies have investigated mobile positioning using CFVD, such as cell identification (ID), handover, and fingerprint positioning data [4].

The location of a mobile phone can be considered the location of a vehicle. In addition, MSs can be located using location methods. The positioning and time difference information of an MS can be retrieved to estimate speed [5, 6]. The method of using cell ID, handover information, and an FPA for locating and to estimate speed is explained in the following subsections.

*2.2.1. Cell ID-Based Method.* A cellular network is a radio network distributed over land areas called cells, and every cell is served by at least one base station (BS). When joined

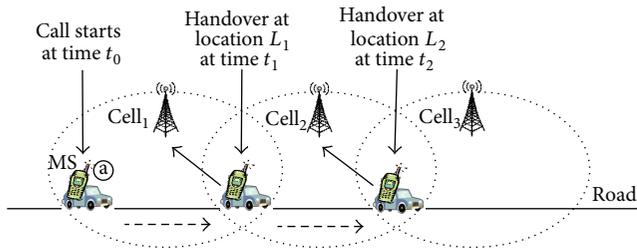


FIGURE 1: Diagram of the MS handover process.

together, these cells facilitate communication between MSs. Each BS has a unique identifier known as its cell global identity (CGI), which the BS broadcasts periodically.

The cell ID-based location method is defined in the technical LCS specification established by the 3rd Generation Partnership Project (53GPP) [7]. Network operators can record the MS cell ID to locate a BS in the database when an MS connects to a cellular network. However, low accuracy is a critical characteristic of cell ID for relatively large cells in global system for mobile communications (GSM) networks [8]. Generally, the cell ID positioning error ranges between approximately 1 and 5 km, especially in rural areas. This approach is unsuitable for speed estimation.

**2.2.2. Handover-Based Method.** Numerous studies have been conducted in this research area. Gundlegard and Karlsson reported handover location accuracies of below 20 m and 40 m in GSM and universal mobile telecommunications system (UMTS) networks, respectively. The study results showed good handover location accuracy in both GSM and UMTS networks. Additionally, the location accuracy in UMTS networks exceeded that in GSM networks [9]. Caceres et al. used LU events as a “virtual traffic counter” to measure the traffic flow when passing through two LAs for phones [10]. Regarding the estimates of travel time and speed, the average absolute relative travel time difference was 10.7% in [11]. According to survey results [9, 11–13], satisfactory travel time estimations in GSM and UMTS networks are feasible. Thiessenhusen et al. measured traffic speed according to double handover (DHO) events and compared average speeds based on cell phone probe data, floating car data, and loop detector data [14].

The coverage area of a BS is considered a cell. When an MS connects to a cellular network, the location of the cell ID can be recorded. If the communicating MS moves from one cell to another, the handover process is triggered. Figure 1 illustrates the process of collecting real-time traffic information by tracking the locations of MS. For example, the MS which is in the car (see (a) in Figure 1) go through the road, call starts at time  $t_0$  and connect to the Cell<sub>1</sub> in the meantime. When the MS continues moving along the road, entering the coverage area of Cell<sub>2</sub> at  $t_1$ , the MS connects to a new BS, and the handover process is executed. At this time, the radio channel is transferred from Cell<sub>1</sub> to Cell<sub>2</sub>.

Therefore, the moving situation of an MS can be determined according to the handover data, and the MS moving speed can be estimated. As shown in Figure 1, if more

than two handover processes are performed, then the first handover sequence tags (CGI<sub>1</sub>, CGI<sub>2</sub>) at  $t_1$  and the second handover sequence tags (CGI<sub>2</sub>, CGI<sub>3</sub>) at  $t_2$  can be retrieved. In addition, if the handover sequence tags at  $L_1$  (CGI<sub>1</sub>, CGI<sub>2</sub>) and  $L_2$  (CGI<sub>2</sub>, CGI<sub>3</sub>) are known, then the distance between the two positions  $D(L_1, L_2)$  can be determined, and the vehicle speed can be estimated according to (1). However, the location errors of some handover pairs are substantial, especially in rural areas. The lower the handover location accuracy and distance between two handover events are, the greater the speed estimation error is [15]. Consider

$$v = \frac{D(L_1, L_2)}{(t_2 - t_1)}. \quad (1)$$

**2.2.3. Fingerprint Positioning Algorithm.** Several studies have implemented network interface signal acquisition and analysis by tracking users’ MS location in GSM/general packet radio service (GPRS)/UMTS heterogeneous networks. An FPA was proposed for collecting and analyzing the RSS of MRs and determining the location of MSs to facilitate real-time traffic information estimations [16–20]. With current GSM/UMTS standards, when an MS is active, the base station controller (BSC) and radio network controller (RNC) receive MRs from the MS during calls. A k-nearest neighbor (kNN) based FPA is adopted to locate the MS according to the MR signals. If more than two pieces of location data exist for the same MS, then the vehicle speed can be computed [18]. To estimate speed, the location of the MS is estimated according to the RSS in the MR by using the FPA algorithm. The speed is then analyzed based on the distance and time difference of the MS, the steps of which are explained as follows.

A previous study presented a comparison of the location errors yielded when using the FPA, cell ID-based, and handover-based location methods [20]. The MS experiment was conducted on a 12.2 km long segment of Expressway No. 66 in Taiwan. Lin drove this segment 12 times with an MS device, and a global positioning system receiver was used to collect the signal data and geographic coordinates from the MS [20]. During these experiments, the location errors yielded by the FPA, cell ID-based, and handover-based location methods were compared. The results showed that the average error of the cell ID-based location method was 714.07 m and that of the handover-based location method was 293.54 m. Thus, the average location error was approximately 36.11 m, which is much better than that of other methods [20].

**2.2.4. Summary.** In the previous sections, the average location error was approximately 36 m, which is much better than that of the others. Therefore, using the FPA to determine the location is a suitable method for estimating speed. However, the FPA speed estimation method necessitates two location and time distance data sets. The sampling period was our primary research topic. In addition, we simulated the traffic conditions, including traffic flow and road length. VISSIM tool is appropriate for traffic simulations [21]. In Section 4 of this study, several reasonable parameters are adopted to simulate relevant traffic conditions.

TABLE 1: The definition of notation.

Parameter	Description
$\lambda$ (call/hr)	The call arrival rate
$\tau$ (hr)	The call interarrival time
$t$ (hr)	The call holding time
$1/\mu$ (hr)	The mean call holding time
$f$ (car/hr)	The traffic flow
$v$ (km/hr)	The average vehicle speed
$x$ (hr)	The time $x$ is between the preceding call arrival and entering the target road
$c$ (second)	The sampling period
$d$ (km)	The distance of target road
$r$ (count/hr)	The reporting data which we acquire per hour
$w$ (count/hr)	The waste computing data which we waste per hour

### 3. Mathematical Models for Estimating the Optimal Sampling Period

In this section, an analytical model is proposed for analyzing report rates based on communication behaviors, traffic conditions, and two consecutive fingerprint positioning locations for the same call to determine the optimal sampling period and estimate the vehicle speed. In Section 3.1, the design issues encountered in this study are explained. In Section 3.2, an overview of the proposed models is provided. In Section 3.3, the proposed models are discussed. The notations that appear in this section are presented in Table 1.

**3.1. Problem Definition.** When analyzing location samples to estimate speed, excessively short sampling periods may generate a substantial amount of data computation costs and increase server loads. Conversely, excessively long sampling periods may result in a small number of reporting data. If the reporting data are insufficient, the accuracy of the traffic information decreases. In the next section, a model for identifying the optimal sampling period is proposed.

**3.2. Proposed Mathematical Models.** In this section, we define the proposed models. We considered the communication behavior, traffic conditions, and two consecutive fingerprint positioning locations signals for the same call and then proposed the following two models. The reporting rate was defined as the quantity of reporting data acquired per hour. Another model was based on the waste computing rate, which was defined as the quantity of cost computing data wasted per hour. The results of these two models were compared to identify the optimal sampling period and determine the quantity of reporting data necessary to minimize the computation costs.

**3.2.1. Assumptions.** Figure 2 is a timing diagram showing vehicle movement and communication behavior. A preceding call (at  $t_0$  in Figure 2) was initiated before entering the target road (at  $t_1$  in Figure 2). The MS in the vehicle moving

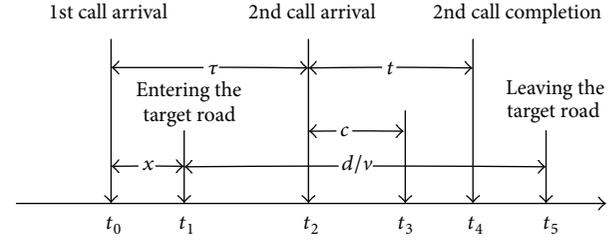


FIGURE 2: The timing diagram for vehicle movement and communication behavior.

along the target road performed the call setup (at  $t_2$  in Figure 2) and completion (at  $t_4$  in Figure 2) before leaving the target road (at  $t_5$  in Figure 2). Thus, the optimal sampling period  $c$  (at  $t_3$  in Figure 2) is between the call arrival and call completion times. The following assumptions are used in the model.

Each moving car has a cell phone.

- (i) The call arrival (including call origination and termination) process is assumed to be a Poisson process with average rate  $\lambda$  (call/hour) [22].
- (ii) The call interarrival time  $\tau$  is exponentially distributed with mean  $1/\lambda$  [22].
- (iii) The call holding time  $t$  is exponentially distributed with mean  $1/\mu$  [22].
- (iv) The traffic flow  $f$  and average vehicle speed  $v$  along the road can be obtained from the fixed vehicle detector (VD) data of the road.
- (v) The time  $x$  refers to the period between the arrival of the preceding call and entering the target road  $[t_0, t_1]$ .
- (vi) The time  $c$  refers to the period after the call arrival and is defined as the sampling period.
- (vii) The distance of the target road is  $d$ .

According to these assumptions, the output measures are the reporting rate  $r$  and the waste computing rate  $w$  for the target road. The call arrival data and the sampling period  $c$  of the MS are used to generate reporting data for speed estimation according to CFVD. The two proposed models are introduced separately in the next section.

**3.2.2. The Reporting Rate Model (RRM).** The reporting rate is defined as the quantity of reporting data acquired per hour during the sampling period  $c$ . The reporting rate is dependent on traffic flow, vehicle speed, call arrival rate, call holding time, sampling period, and road segment length (shown in Figure 2). Because many routing paths may exist between different road segments, only call arrival events and reporting events in the same road segment were collected to estimate vehicle speed. Moreover, the two fingerprint positions and the time difference before call completion can be obtained when the sampling period  $c$  is less than the call holding time  $t$ . Therefore, the vehicle speed can be estimated after computing the distance between the positioning locations and the time

difference with the sampling period  $c$ . The reporting rate can be expressed as

$$\begin{aligned}
r &= f \times \int_{x=0}^{\infty} \int_{\tau=x}^{x+(d/v)-c} \lambda e^{-\lambda\tau} \int_{t=c}^{\infty} \mu e^{-\mu t} dt d\tau dx \\
&= f \times \int_{x=0}^{\infty} \int_{\tau=x}^{x+(d/v)-c} \lambda e^{-\lambda\tau} e^{-\mu c} d\tau dx \\
&= f \times \int_{x=0}^{\infty} e^{-\mu c - \lambda x} \left(1 - e^{\lambda((cv-d)/v)}\right) dx \\
&= f \times \frac{1}{\lambda} e^{-\mu c} \left(1 - e^{\lambda((cv-d)/v)}\right) \\
&= \frac{f e^{-\mu c} \left(1 - e^{\lambda((cv-d)/v)}\right)}{\lambda}.
\end{aligned} \tag{2}$$

**3.2.3. The Waste Computing Rate Model (WCRM).** The waste computing rate is defined as the quantity of reporting data that cannot be acquired per hour but necessitate the computation of fingerprint position data; these computing data are waste costs. The proposed speed estimation method involves collecting the position locations based on call arrival events and reporting events. However, the reporting event depends on the sampling period  $c$ , which may be longer than the call holding time  $t$ . The second fingerprint position datum cannot be generated when the sampling period  $c$  is longer than the call holding time  $t$ . However, the first fingerprint position ( $t_2$  in Figure 2) must be collected and generated before call completion or the reporting event. This process of computing is called waste computing. The waste computing rate can be expressed as

$$\begin{aligned}
w &= f \times \int_{x=0}^{\infty} \int_{\tau=x}^{x+(d/v)-c} \lambda e^{-\lambda\tau} \int_{t=0}^c \mu e^{-\mu t} dt d\tau dx \\
&= f \times \int_{x=0}^{\infty} \int_{\tau=x}^{x+(d/c)-c} \lambda e^{-\lambda\tau} - e^{-\mu c} - \lambda e^{-\lambda\tau} d\tau dx \\
&= f \times \int_{x=0}^{\infty} e^{-\lambda x} (e^{-\mu c} - 1) \left[e^{-\lambda((d/v)-c)} - 1\right] dx \\
&= f \times \frac{1}{\lambda} (1 - e^{-\mu c}) \left[1 - e^{\lambda((cv-d)/v)}\right] \\
&= \frac{f (1 - e^{-\mu c}) \left(1 - e^{\lambda((cv-d)/v)}\right)}{\lambda}.
\end{aligned} \tag{3}$$

**3.2.4. The Optimal Sampling Period Derived from the RRM and WCRM.** We introduce two models proposed for the reporting rate and the waste computing rate, and the assumptions are shown in Figure 2. By analyzing the two proposed models, namely, the RRM and WCRM, we determined that, when the reporting rate equals the waste computing rate, several parameters can be adopted to identify the optimal sampling period, which can be expressed as (4). When the waste computing rate exceeds the reporting rate, the sampling period is not effective. Additionally, the costs incurred during

TABLE 2: The definition of abbreviations.

Abbreviation	Description
AER	Absolute average error rate
RER	Relative average error ratio
AS	Actual speed
ES	Estimated speed
AER'	Absolute error rate under optimal sampling period
C	Cost
RCR	Relative cost ratio

the optimal sampling period are lower than those incurred during shorter sampling periods. Moreover, the quantity of reporting data per hour necessary for cost minimization can be determined. Consider

$$\frac{f e^{-\mu c} \left(1 - e^{\lambda((cv-d)/v)}\right)}{\lambda} = \frac{f (1 - e^{-\mu c}) \left(1 - e^{\lambda((cv-d)/v)}\right)}{\lambda}. \tag{4}$$

**3.3. Summary.** The limitation of the two models proposed in this study, namely, RRM and WCRM, was that call arrivals must occur on the same target road. We assume that every vehicle has a mobile phone, and the road assumptions only restrict the freeway because standard roads or rural roads have numerous traffic lights and differing situations. Under these assumptions, the proposed models are suitable for identifying the optimal sampling period using the FPA. In the next section, we introduce the numerical analysis performed to optimize the sampling period. Simulations are subsequently conducted, the results of which are analyzed to verify the superiority of the optimized sampling period.

## 4. Numerical and Simulation Analyses

This section first presents numerical analyses and introduces the optimal sampling period. Second, a simulation environment was run and the simulation results for various cases are analyzed. A discussion of these results is provided at the end of this section. The abbreviations that appear in this section are defined in Table 2.

**4.1. Numerical Analysis.** This section presents an analysis of the optimal sampling period of the reporting rate, and an estimation of the reporting rates is addressed based on VD data. The speed reporting rate of the proposed method and that of the handover-based method are also compared.

**4.1.1. The Optimal Sampling Period for Speed Reporting Rate.** To demonstrate the proposed method, the following parameters were adopted to estimate the average speed rate:  $f = 300$  cars/h,  $\lambda = 1$  call/h,  $d = 5$  km, and  $1/\mu = 60$  s. Figure 3 shows the report rates  $r$  for different sampling periods  $c$ . The shorter the sampling period  $c$  is, the higher the number of report rates is. 03BC

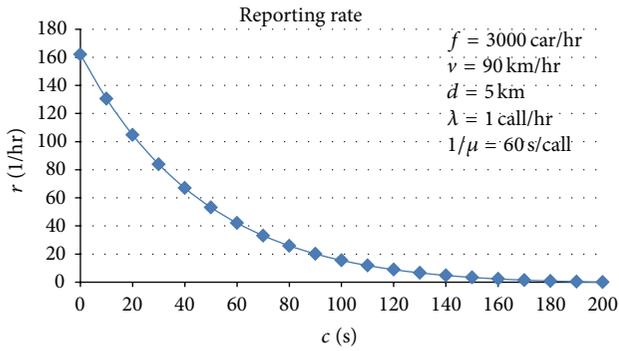


FIGURE 3: The reporting rates with different sampling periods.

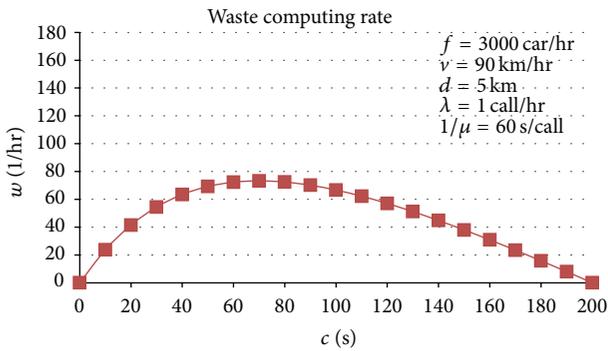


FIGURE 4: The waste computing rates for different sampling periods.

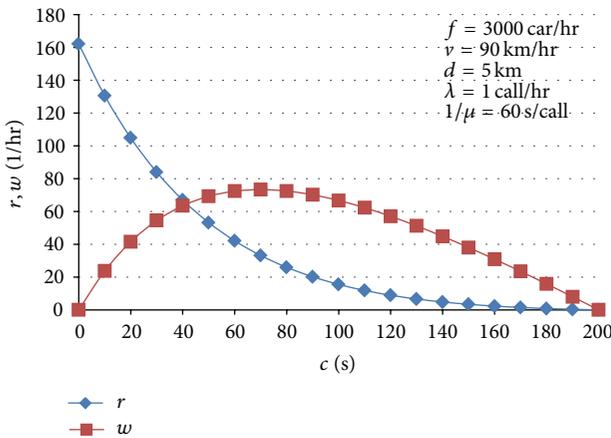


FIGURE 5: The optimal sampling time by report rates and waste computing rates.

Figure 4 shows the results of the WCRM for different sampling periods  $c$ . The longer the sampling period  $c$  was, the higher the waste computing rate. However, when the sampling period  $c$  was longer, the unconsidered target road was also longer; thus, the waste computing rate was lower.

The reporting rates were compared with the waste computing rates, and the point of the intersection for the two lines is the optimal period, as shown in Figure 5. The optimal sampling period  $c$  can be computed when the two equations are equal. The results show that  $c = 41.589$  s, and the report rate  $r$  was 64.32 counts/h. Figure 5 shows that the

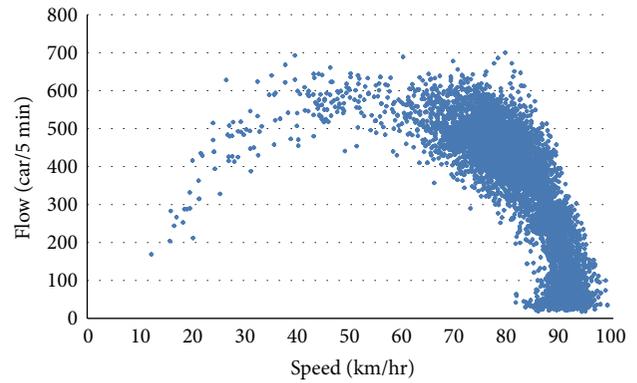


FIGURE 6: The real traffic information (average vehicle speeds and traffic flows) during February in 2008 from the fixed VD.

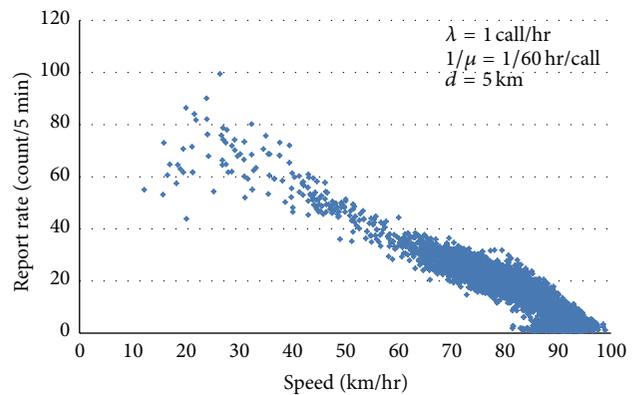


FIGURE 7: The speed reporting rates during February in 2008 by using VD logs.

optimal sampling period was poor when  $c > 41.589$  s and the waste computing rates exceeded the speed reporting rates. Because all data received were computed, when the waste data exceeded the useful data, the process was not efficient. Conversely, when  $c < 41.589$  s, the speed reporting rates clearly exceeded the computing rates. However, when the sampling period  $c$  was excessively short, the load was high and costly. Thus, the optimal sampling period  $c$  generated the lowest costs and server load with good efficiency. A speed reporting rate  $r = 64.32$  counts/h is sufficient to yield superior speed estimation results compared with those provided by the VD.

**4.1.2. Estimated Speed Reporting Rate Based on Vehicle Detector Data.** Regarding the reporting rate estimation, Figure 6 shows the real traffic information, including the traffic flow and average vehicle speeds measured every 5 min. These data were obtained in February 2008 from the fixed VD located at the 41.5 km milepost of Highway No. 1 in Taiwan. The following parameters were adopted based on historical data obtained from Chunghwa Telecom:  $\lambda = 1$  call/h,  $1/\mu = 60$  s, and  $d = 5$  km.

Figure 7 shows that the report rates are measured using the real traffic information obtained from VD logs. The average report rate was 16.35 for every 5 min.

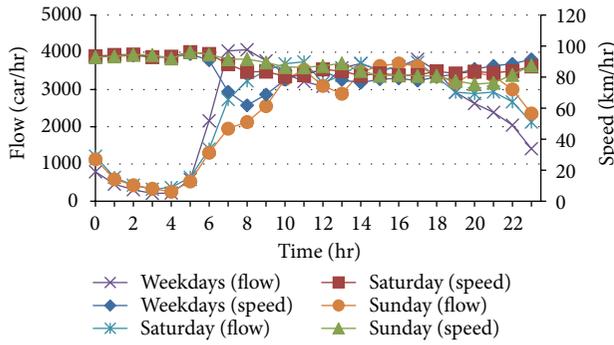


FIGURE 8: The real traffic information (average vehicle speeds and traffic flows) during July in 2008 from the fixed VD.

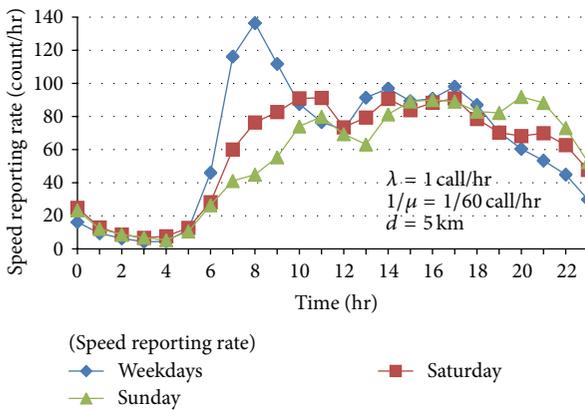


FIGURE 9: The estimated reporting rates based on VD data.

For the reporting rates during peak and nonpeak hours, the VD data from July 2008 was collected from the fixed VD at the 41.5 km milepost on Highway No. 1 in Taiwan. The average traffic flow and vehicle speed during weekdays, Saturdays, and Sundays are shown in Figure 8. The peak hour in this study is defined as a traffic flow higher than 3500 car/hr. The peak hours during the morning are 6:00–9:00 a.m., 9:00–11:00 a.m., and 10:00–11:00 a.m. on weekdays, Saturdays, and Sundays, respectively. During the afternoon, the peak hours are 4:00–6:00 p.m., 2:00–5:00 p.m., and 3:00–5:00 p.m. on weekdays, Saturdays, and Sundays, respectively. These values and the following parameters from (2) were adopted to estimate the average speed rate:  $\lambda = 1$  call/hr,  $1/\mu = 60$  sec, and  $d = 5$  km. Figure 9 shows that all the reporting rates during peak hours are higher than 60 count/hr. Therefore, the speed report can be announced every minute, and the proposed method can provide traffic congestion information immediately.

4.1.3. *The Comparison of Speed Reporting Rate.* To demonstrate the proposed method, the following parameters were adopted to estimate the average report rate:  $f = 3000$  cars/h,  $d = 5$  km,  $\lambda = 1$  call/h,  $1/\mu = 60$  s, and  $c = 41.589$  s. The highway speed limit in Taiwan is 100 km/h. Figure 10 shows the report rates estimated with different average holding times. The results indicate that the report

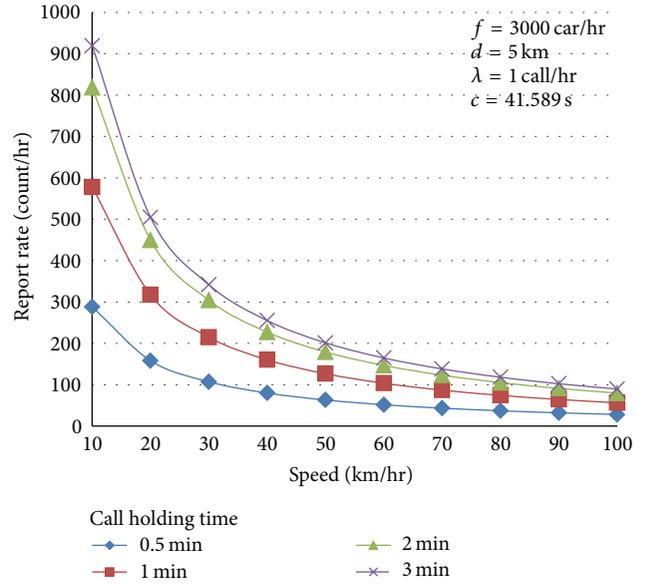


FIGURE 10: The report rates with different call holding times.

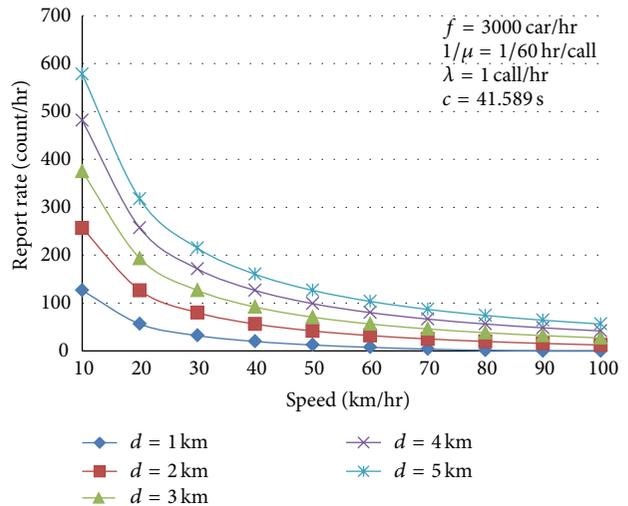


FIGURE 11: The report rates with different lengths of road segment.

rates can be determined using lower average vehicle speeds to provide traffic congestion information rapidly.

Figure 11 shows the report rates for different road segment lengths. The results indicate that the traffic information for longer road segments is reported at a higher rate.

The reporting rates of the proposed method ( $r$ ) and the reporting rates of the handover-based method ( $h$ ) were compared. The reporting rate model for the handover-based method was proposed and analyzed as (5) by [8]. The parameter  $D(L_1, L_2)$  is defined as the length of a road segment between two consecutive handover events (shown in Figure 1). The average  $D(L_1, L_2)$  on the freeway in Taiwan is about 1.5 km [8]. Therefore, the following parameters were adopted to estimate the average speed rate:  $\lambda = 1$  call/hr,  $1/\mu = 60$  sec,  $d = 5$  km, and  $D(L_1, L_2) = 1.5$  km.

TABLE 3: The comparison of the proposed method and handover-based method.

$v$	The proposed method ( $r$ )	Handover-based method ( $h$ )	$r/h$
10	580	0	
20	318	1	318.00
30	216	2	108.00
40	161	5	32.20
50	127	8	15.88
60	104	11	9.45
70	87	14	6.21
80	75	16	4.69
90	65	18	3.61
100	57	20	2.85

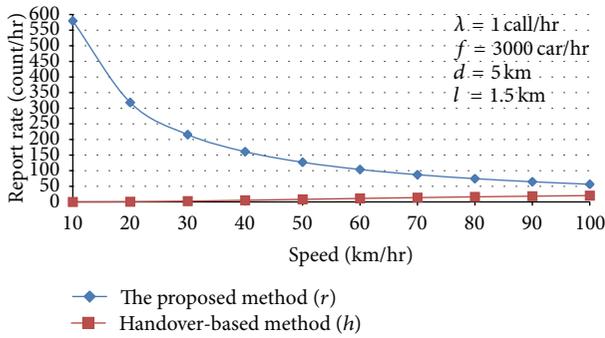


FIGURE 12: The comparison of the proposed method and handover-based method.

The reporting rates of these methods are shown in Table 3 and Figure 12, and the value  $r/h$  indicates the comparison of the reporting rate and computing cost. Although the proposed method required more computing cost, the reporting rates of the handover-based method were smaller when traffic was more congested. Therefore, the proposed method is more suitable for ITS than the handover-based method. Consider

$$h = \left( \frac{f\lambda}{\mu} \right) \times e^{(\mu \times D(L_1, L_2))/v}. \quad (5)$$

**4.2. Simulation.** We designed a simulation environment for simulating vehicle movement and MS communication. The results can be used to verify the optimal sampling period according to the speed estimation accuracy and the costs of speed estimation. Trace-driven experiments were designed in two parts, namely, an MS communication trace generation and vehicle movement trace generation. The two trace files were then combined to estimate traffic information based on cellular network data (as shown in Figure 13). The vehicle movement trace was obtained from the VisSim tool, which served as a traffic simulator as well as highway measure.

**4.2.1. Simulation Assumptions.** We considered a highway scenario characterized by the Wiedemann psychophysical car-following and lane-changing model, where the average

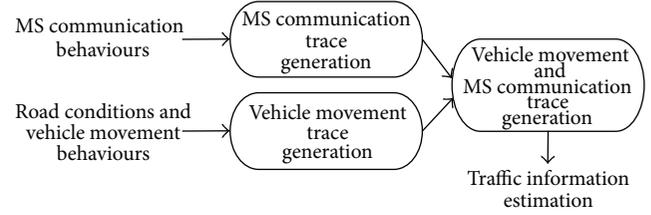


FIGURE 13: Trace-driven simulation for estimating traffic information based on cellular network data.

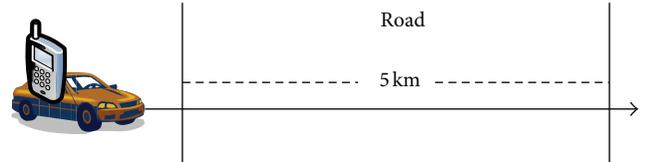


FIGURE 14: The lengths of road in simulation environment.

vehicle speeds were 90 km/h in Case 1 and 30 km/h in Case 2. The target road during the simulation experiments measured 5 km (as shown in Figure 14). The simulation time was 3600 s. We assumed that each car possessed one MS and adopted the following parameters to generate random numbers according to the assumption in Section 3 for estimating the reporting rate:  $f = 3000$  cars/h,  $\lambda = 1$  call/h, and  $1/\mu = 60$  s.

**4.2.2. Simulation Environment.** In real-life environments, the traffic situation is not predictable. Thus, we used the VisSim tool to simulate a random traffic situation. Different parameters can be adopted using the VisSim tool, including variations in traffic flow, vehicle speed, freeway mode, number of roads, road segment length, vehicle type, and simulation time. After the simulation was performed using the VisSim tool, data on each vehicle were collected; for example, data on the vehicle speed, road segment entry time, and road segment exit time were collected. Moreover, we used the EXCEL software to generate random communication behavior, such as the call arrival and call holding times. We subsequently compared the two sets of data to identify which cars contained an MS with call arrivals and to determine the call holding time duration. All assumptions are described in Section 4.2.1. The next section introduces the performance metrics used to verify the optimal sampling period. Section 4.2.3 presents the design of two case studies using the VisSim tool and different parameters. Finally, Section 4.2.4 presents a discussion of the different performance metrics of various cases. Table 4 lists the simulation environment data.

**4.2.3. Performance Metrics.** To evaluate the optimal sampling period, we adopted three indices. These indices were required to verify the optimal sampling period in the simulation environment. Consequently, the absolute error rate (AER), relative error ratio (RER), and cost were used as the main indicators in this study. The AER refers to the error rate with the actual speed under different sampling periods. The RER refers to the AER of different sampling periods compared

TABLE 4: Simulation environment.

Hardware	
CPU	Intel core i3
Memory	23 MB
Software	
O.S.	Windows 7
Development tool	Vissim
Statistical tool	Excel

with the AER of the optimal sampling period. Cost was defined as the total data required for speed estimations. The performance metrics are shown in Table 5.

**4.2.4. Simulation Cases.** Two simulation cases were developed for the environment. For Case 1, we designed a free-flow scenario involving an average vehicle speed of 90 km/h, traffic flow of 3000 cars/h, road segment length of 5 km, and a simulation duration of 3600 s. For Case 2, we designed a traffic congestion situation involving an average vehicle speed of 30 km/h, traffic flow of 4500 cars/h, road segment length of 5 km, and a simulation duration of 3600 s. These two cases included the conditions of the entire freeway. The cases selected for simulation are shown in Table 6.

**4.2.5. Results and Analyses.** During the experiment, three performance metrics of each case were adopted for comparison.

AER is the metric for the speed error rate of different sampling periods. RER is the metric for the speed error rate of different sampling periods in relation to the optimal sampling period. Cost is the metric for obtaining superior results in relation to the optimal sampling period and comparing the cost performance. The performance of each case was subsequently analyzed.

**Case 1 (free flow).** In Case 1, the absolute error rate and relative error ratio for each sampling period can be determined (as shown in Table 7). Regarding the absolute error rate, the results of a shorter sampling period are more favorable. The optimal sampling period was 41.589 s, and the absolute error rate was 2.87%. These results are superior to those yielded using the handover method and cell ID method for estimating speed [20]. The absolute error rate was 2.73% for 1 s and 2.82% for 20 s. This indicates that when the sampling period is shorter than the optimal sampling period of 41.589 s, superior results can be obtained. The relative error ratio of the other sampling periods was compared to that of the optimal sampling period. The results indicated that, when the sampling period exceeds the optimal sampling period of 41.589 s, a larger error rate is obtained, which is not efficient. However, the results for the sampling period of 20 s were considerably superior to those of the optimal sampling period of 41.589 s.

According to the metrics AER and RER, the results of sampling periods shorter than the optimal sampling period of 41.589 s are superior. The cost metric reflects the efficiency for

sampling periods shorter than the optimal sampling period of 41.589 s. The cost is estimated according to the total data necessary for estimating speed during different sampling periods. Table 8 shows the total data required for estimating the speed, cost, and relative error ratio of different sampling periods.

Efficiency can be estimated according to the cost and error ratio relative to the optimal sampling period. We first considered the sampling periods that were shorter than the optimal sampling period of 41.589 s. For the sampling period of 20 s, costs were higher than those of the optimal sampling period of 41.589 s (31 counts/h), but the error rate improved compared with the optimal sampling period by only 1.74%. This indicates that the efficiency was low. For the sampling period of 1 s, costs were higher than those of the optimal sampling period of 41.589 s (78 counts/h), but the error rate improved compared with the optimal sampling period by only 4.88%. This indicates that the costs are extremely high. Considering the sampling period of 20 s, the costs increased 31 counts/h, and the error rate improved by only 1.74%. For a sampling period of 60 s, the costs were reduced 19 counts/h, and the error rate increased by 6.62%. This indicates high inefficiency.

**Case 2 (traffic congestion).** In Case 2, the absolute error rate and relative error ratio for each sampling period were determined (as shown in Table 9). Regarding the absolute error rate, the results indicate that, for shorter sampling periods, superior outcomes can be obtained. The optimal sampling period was 41.589 s, and the absolute error rate was 1.34%. These results were superior to those obtained using the handover method and the cell ID method for estimating speed [20]. The absolute error rate was 1.30% for 1 s and 1.33% for 20 s. This shows that, when the sampling period is shorter than the optimal sampling period of 41.589 s, superior results can be obtained. The relative error ratio of the other sampling periods was subsequently compared with that of the optimal sampling period. The results show that, when the sampling period exceeds the optimal sampling period of 41.589, the error rate is greater, which does not indicate efficiency. The results for the sampling periods of 1 s and 20 s were significantly superior to those for the optimal sampling period of 41.589 s.

According to the metrics AER and RER, the results of sampling periods shorter than the optimal sampling period of 41.589 s are superior. The cost metric reflects the efficiency for sampling periods shorter than the optimal sampling period of 41.589 s. The cost is estimated according to the total data necessary for estimating speed during different sampling periods. Table 10 shows the total data required for estimating the speed, cost, and relative error ratio for different sampling periods.

Efficiency can be estimated according to the cost and error ratio relative to the optimal sampling period. We first considered the sampling periods that were shorter than the optimal sampling period of 41.589 s. For the sampling period of 20 s, costs were higher than those for the optimal sampling period of 41.589 s (115 counts/h), but the error rate improved compared with that of the optimal sampling period by only

TABLE 5: Performance metrics.

Absolute error rate (AER)	The error rate with the actual speed under the different sampling period. The formula (*) is shown below: $AER = \frac{ES - AS}{AS} \times 100\% \quad (*)$
Relative error ratio (RER)	The AER with different sampling periods which is compared to the AER'. The formula (**) is shown: $RER = \frac{AER - AER'}{AER'} \times 100\% \quad (**)$
Cost (C)	The total data needed for speed estimation.
Relative cost ratio (RCR)	The total data needed for speed estimation which is compared to the optimal sampling period.

TABLE 6: Cases designed for simulation.

	Experiment with
Case 1 Free flow	(i) Traffic flow 3000 car/hr
	(ii) Average vehicle speed 90 km/hr
	(iii) Road segment 5 km
	(iv) Simulation time 3600 seconds
	(v) Freeway mode.
	Experiment with
Case 2 Traffic congestion	(i) Traffic flow 4500 car/hr
	(ii) Average vehicle speed 30 km/hr
	(iii) Road segment 5 km
	(iv) Simulation time 3600 seconds
	(v) Freeway mode.

TABLE 7: AER and RER for Case 1.

Sampling period (sec)	Absolute error rate (AER)	Relative error ratio (RER)
1	2.73%	4.88%
20	2.82%	1.74%
<b>41.589</b>	<b>2.87%</b>	<b>0</b>
60	3.06%	-6.62%
80	3.10%	-8.01%
100	3.24%	-12.89%
120	3.31%	-15.33%

Bold values refer that the optimal sampling period was 41.589 s. And we explain in Section 4.2.5.

0.75%. This indicates that the efficiency was low. For the sampling period of 1 s, costs were higher than those for the optimal sampling period of 41.589 s (244 counts/h), but the error rate improved compared with that of the optimal sampling period by only 2.99%. This indicates that the cost was extremely high. For example, in the sampling period of 20 s, the cost increased by only 115 counts/h, and the error rate was improved 0.75%. For a sampling period of 60 s, the cost decreases by only 67 counts/h and the error rate increased by -3.73%. This indicates high inefficiency.

4.3. Discussion. In this section, the results for Case 1 (free flow) and Case 2 (traffic congestion) are compared. The differences in cost and error ratio between the two cases

TABLE 8: Cost and RER for Case 1.

Sampling period (sec)	Cost (count/hr)	Relative cost ratio (%)	Relative error ratio (RER)
1	132	144%	4.88%
20	85	57%	1.74%
<b>41.589</b>	<b>54</b>	<b>0</b>	<b>0</b>
60	35	-35%	-6.62%
80	22	-59%	-8.01%
100	20	-63%	-12.89%
120	9	-83%	-15.33%

Bold values refer that the optimal sampling period was 41.589 s. And we explain in Section 4.2.5.

TABLE 9: AER and RER for Case 2.

Sampling period (sec)	Absolute error rate (AER)	Relative error ratio (RER)
1	1.30%	2.99%
20	1.33%	0.75%
<b>41.589</b>	<b>1.34%</b>	<b>0</b>
60	1.39%	-3.37%
80	1.40%	-4.48%
100	1.41%	-5.22%
120	1.39%	-4.48%

Bold values refer that the optimal sampling period was 41.589 s. And we explain in Section 4.2.5.

are shown below. The costs for the free-flow scenario were considerably lower than those for the traffic congestion scenario. This was because, for the traffic congestion scenario of Case 2, the vehicle speed was 30 km/h; thus, the number of call arrivals was higher, and vehicles remained on the road segment for longer durations. Consequently, substantially more data were obtained for Case 2 compared with Case 1. Table 11 shows the cost differences between the two cases. When the sampling period is shorter, the server load is higher. However, when the sampling period is excessively long, the proposed method is ineffective because the amount of data obtained is insufficient. The error ratio differences between the two cases are shown in Table 12.

The experimental results showed that excessively short sampling periods generate substantial data costs and increase server loads regardless of the case examined. Conversely,

TABLE 10: Cost and RER for Case 2.

Sampling period (sec)	Cost (count/hr)	Relative cost ratio (%)	Relative error ratio (RER)
1	458	144%	2.99%
20	329	54%	0.75%
<b>41.589</b>	<b>214</b>	<b>0</b>	<b>0</b>
60	147	-31%	-3.37%
80	83	-61%	-4.48%
100	53	-75%	-5.22%
120	39	-82%	-4.48%

Bold values refer that the optimal sampling period was 41.589 s. And we explain in Section 4.2.5.

TABLE 11: Cost for Cases 1 and 2.

Sampling period (sec)	Free flow (count/hr)	Traffic congestion (count/hr)
1	132	458
20	85	329
<b>41.589</b>	<b>54</b>	<b>214</b>
60	35	147
80	22	83
100	20	53
120	9	39

Bold values refer that the optimal sampling period was 41.589 s. And we explain in Section 4.2.5.

TABLE 12: Relative error ratio for Cases 1 and 2.

Sampling period (sec)	Free flow (%)	Traffic congestion (%)
1	4.88%	2.99%
20	1.74%	0.75%
<b>41.589</b>	<b>0</b>	<b>0</b>
60	-6.62%	-3.37%
80	-8.01%	-4.48%
100	-12.89%	-5.22%
120	-15.33%	-4.48%

when the sampling period is substantially longer than the optimal sampling period, the efficacy and efficiency are lower. Furthermore, when the sampling period exceeds 120 s, the data reporting rate is lower than 12 counts/h, less than that of the VD on the freeway. The results show that, at lower speeds, the reporting rate is substantially higher, and the higher the reporting rate is, the more accurately the traffic information reflects real-time conditions. Thus, the optimal sampling period can provide accurate, real-time traffic information with lighter server loads.

## 5. Conclusions

Using CFVD to measure and forecast real-time traffic information by anonymously sampling mobile phone positions for ITS has become a crucial technique. However, a high

sampling frequency generates a substantial load for ITS servers, and traffic information cannot be provided immediately with long sampling periods. The two analytical models are proposed to analyze the effective reporting rates based on communication behavior, traffic conditions, and two consecutive fingerprint positioning locations for the same call to determine the optimal sampling period and estimate vehicle speed. The numerical analysis results indicated that when the sampling periods exceed the optimal sampling period, the waste computing rate exceeds the reporting rate and the costs of the optimal sampling period are lower than those of shorter sampling periods.

The experimental results show that the optimal sampling period was 41.589 s when the average call holding time was 60 s and the average speed error rate was only 2.87%. When the sampling period was longer than 41.589 s, the reporting rates were low and efficacy was poor. However, if the sampling period was shorter than 41.589 s, the costs were high and the efficacy only slightly improved. Thus, the overall efficiency was considerably lower. Using the FPA method proposed in this study, the optimal sampling period can be used to estimate traffic information immediately for ITSs. Furthermore, different optimal sampling periods can be adopted for different roads or communication behaviors to provide drivers with immediate speed information.

## Acknowledgments

This study was supported by the National Science Council of Taiwan under Grant nos. NSC 102-2410-H146-002-MY2, NSC 102-2410-H-009-052-MY3, NSC 102-2410-H009-028-MY2, NSC 101-2410-H146-004, and NSC 101-2420-H-009-004-DR. This work was also supported by Aiming for the Top University Program of the National Chiao Tung University and Ministry of Education of Taiwan.

## References

- [1] M. Bshara, U. Orguner, F. Gustafsson, and L. van Biesen, "Robust tracking in cellular networks using HMM filters and Cell-ID measurements," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1016–1024, 2011.
- [2] D. Middleton and R. Parker, *Vehicle Detector Evaluation*, Publication FHWA/TX-03 /2119-1, Texas Transportation Institute, Texas Department of Transportation, FHWA, 2002.
- [3] R. L. Cheu, C. Xie, and D. H. Lee, "Probe vehicle population and sample size for arterial speed estimation," *Computer-Aided Civil and Infrastructure Engineering*, vol. 17, no. 1, pp. 53–60, 2002.
- [4] C. H. Chen, C. C. Lo, and H. F. Lin, "The analysis of speed-reporting rates from a cellular network based on a fingerprint positioning algorithm," *South African Journal of Industrial Engineering*, vol. 24, no. 1, pp. 98–106, 2013.
- [5] A. Lala, E. D. Luca, and M. Iovieno, "Velocity handling in location services," WO 03/045075, 2003.
- [6] M. Zanelli, "Traffic monitoring system and method," US Patent 6505114, 2001.
- [7] 3GPP. 3GPP Technical Specification Group (TSG), Services and System Aspects, TS 22.07 1, Location Services (LCS) , Service description, Stage 1 (Release 11), version 11.0.0, 2012.

- [8] M. F. Chang, C. H. Chen, Y. B. Lin, and C. Y. Chia, "The frequency of CFVD speed report for highway traffic," *Wireless Communications and Mobile Computing*, 2013.
- [9] D. Gundlegård and J. M. Karlsson, "Handover location accuracy for travel time estimation in GSM and UMTS," *IET Intelligent Transport Systems*, vol. 3, no. 1, pp. 87–94, 2009.
- [10] N. Caceres, J. P. Wideberg, and F. G. Benitez, "Deriving origin-destination data from a mobile phone network," *IET Intelligent Transport Systems*, vol. 1, no. 1, pp. 15–26, 2007.
- [11] H. Bar-Gera, "Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel," *Transportation Research C*, vol. 15, no. 6, pp. 380–391, 2007.
- [12] N. Caceres, J. P. Wideberg, and F. G. Benitez, "Review of traffic data estimations extracted from cellular networks," *IET Intelligent Transport Systems*, vol. 2, no. 3, pp. 179–192, 2008.
- [13] J. L. Ygnace, C. Drane, Y. B. Yim, and R. de Lacvivier, "Travel time estimation on the San-Francisco bay area network using cellular phones as probes," PATH Working Paper UCB-ITS-PWP-2000-18, University of California, Berkeley, 2000.
- [14] K. U. Thiessenhusen, R. P. Schafer, and T. Lang, *Traffic Data from Cell Phones: A Comparison with Loops and Probe Vehicle Data*, Institute of Transport Research German Aerospace Center, Berlin, Germany, 2003.
- [15] C. H. Chen, C. C. Lo, and H. F. Lin, "The analysis of speed-reporting rates from a cellular network based on a fingerprint-positioning algorithm," *South African Journal of Industrial Engineering*, vol. 24, no. 1, pp. 98–106, 2013.
- [16] S. P. Kuo and Y. C. Tseng, "A scrambling method for fingerprint positioning based on temporal diversity and spatial dependency," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 678–684, 2008.
- [17] M. Bshara, U. Orguner, F. Gustafsson, and L. van Biesen, "Fingerprinting localization in wireless networks based on received-signal-strength measurements: a case study on wimax networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 283–294, 2010.
- [18] B. Y. Lin, C. H. Chen, and C. C. Lo, "A novel speed estimation method using location service events based on fingerprint positioning," *Advanced Science Letters*, vol. 4, no. 11-12, pp. 3735–3739, 2011.
- [19] C. H. Chen, B. Y. Lin, C. H. Lin, Y. S. Liu, and C. C. Lo, "A green positioning algorithm for Campus Guidance System," *International Journal of Mobile Communications*, vol. 10, no. 2, pp. 119–131, 2012.
- [20] C. H. Chen, B. Y. Lin, H. C. L. Chang, and C. C. Lo, "The novel positioning algorithm based on cloud computing—a case study of intelligent transportation systems," *Information*, vol. 15, no. 11, pp. 4519–4524, 2012.
- [21] Planung Transport Verkehr AG, *VISSIM User Manual—V. 5.20*, Planung Transport Verkehr AG, Karlsruhe, Germany, 2009.
- [22] C. H. Chen, H. C. Chang, C. Y. Su, C. C. Lo, and H. F. Lin, "Traffic speed estimation based on normal location updates and call arrivals from cellular networks," *Simulation Modelling Practice and Theory*, vol. 35, no. 1, pp. 26–33, 2013.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

