Published in: Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 International Conference of
Print ISBN: 978-1-4799-6984-5
DOI: 10.1109/ICAICTA.2014.7005914
http://dx.doi.org/10.1109/ICAICTA.2014.7005914

Statistically Optimum Virtual Trip Line for Real-Time Traffic Monitoring System

Fergyanto E. Gunawan
Binus Graduate Programs
Bina Nusantara University
Jakarta, Indonesia
f.e.gunawan@gmail.com, fgunawan@binus.edu

***Abstract*—The virtual trip line (VTL) has been proposed in conjunction with the floating car data method to minimize the potential breach of the probe vehicle privacy. However, a long VTL, as well as a short one, may have unintended consequences; thus, the VTL length should be optimized. This article intends to propose a method to establish the optimum length of the VTL. The optimal length of VTL is determined empirically based on the geo-location data collected by ten units of smart-phones. A custom-made application is designed and installed to these devices to track and record the movement of the probe vehicle. The recorded geo-location data are projected to the road GIS line to establish the projection distances. We demonstrate that the distances are distributed approximately following the normal distribution, from which the critical region of the optimum VTL length for a given significance level is established. Finally, we empirically evaluate the method, and demonstrate that for a given VTL length, the empirical probability of the probe vehicle crossing a VTL is close to the theoretical probability.**

Keywords—**Floating car data; virtual trip line; normal distribution, intelligent transportation system; probe vehicle**.

I. Introduction

A virtual trip line (VTL) is a line connecting two geo-location points, see Fig. 1, and intersects with a road segment at the location where the traffic is intended to be monitored. The VTL is used in conjunction with the floating car data (FCD) method. In the FCD method, the traffic data are collected by using a number of probe vehicles, which move within a traffic and are used to report the traffic status.

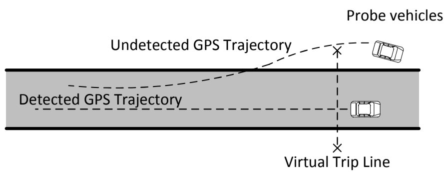

Fig. 1. Detected and undetected GPS tracked trajectories of vehicles.

To provide reliable traffic data using the FCD method, the number of the required probe vehicles, also known as the penetration rate, should be reasonably high depending on the number of vehicles in the traffic and the status of the roads. For highways, the penetration rate should be in order of 4% to 5% [1, 2]; but, for minor arterial and local roads, the required penetration rate is higher [2].

In the FCD method, the use of VTL is beneficial for a number of reasons but most importantly, is for managing data quality and privacy. The privacy concern can be minimized, because by the use of VTL, the data points are only identified through the ID of the VTL and not that of the mobile device [3]. The number of data can also be managed such that only the data at the necessary locations are reported.

The fact that revealing probe vehicle positions to an external traffic-monitoring server may compromise the probe vehicle identity and its mitigations has come to the attention of some researchers since 2000. For this purpose, the virtual acquisition points are proposed in the United States patent number 6,012,012 [4]. In addition, Hoh et al. [5] develop a novel time-to-confusion criterion to characterize privacy in a location dataset and propose an uncertainty-aware path cloaking algorithm that hides location samples in a dataset to provide a time-to-confusion guarantee for all vehicles. A year later, Hoh et al. [6] proposes the use of the virtual trip line and an associated cloaking technique. Kalnis et al. [7] propose the transformation of the spatial data based K-anonymity concept. Finally, Herrera et al. [3] report a field experiment of the FCD method where the data are collected at a set of VTLs.

As described above, the use of VTL is important in the context of the FCD method. However, none of the existing publications has discussed the optimal length of the lines. It is clear that a long VTL is undesirable because the line may intersect with other roads beside the road where the traffic is being monitored. A short VTL is also undesirable because the probe vehicle may cross the measurement area outside the VTL as illustrated in Fig. 1. Therefore, the length of the virtual trip line should be optimal in the sense that it has minimum length but can detect the traveling probe vehicle at a high probability. This work intends to propose a method to determine the optimal length of VTL and to verify empirically the proposed length.

II. Description of Floating Car Data System

The present infrastructure for the floating car data system is depicted in Fig. 2. The system has three main components, namely, a probe client, a web server, and a web client.

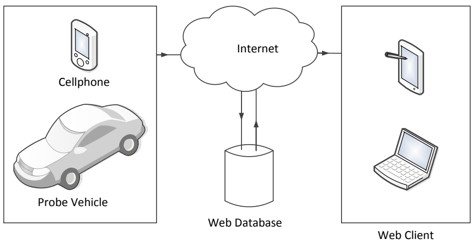

Fig. 2. Framework of traffic monitoring using the floating car data method.

Currently, the probe client utilizes a smartphone, which is a hand-held device that integrates the functionality of a mobile phone with other features but mainly with geo-location functionality. The probe client is attached to a probe vehicle and is used to measure the probe vehicle position, velocity, and heading. Finally, those data and related timestamps will be transmitted to a server via a wireless network.

In the current development, the probe client is an Android application installed in a mobile phone, and the phone will be attached to a probe vehicle. The Android application is designed according to the software architecture depicted in Fig. 3.

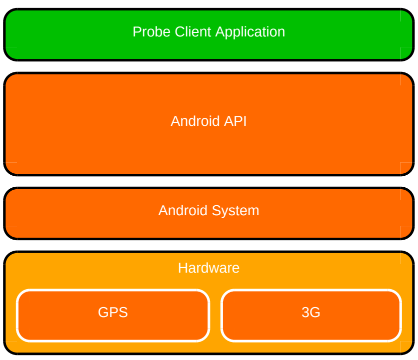

Fig. 3. The software architecture of the probe client application.

The application uses Android API and its underlying system to access the built-in GPS receiver unit. The GPS unit provides data related to the phone position, velocity, bearing, accuracy and timestamps. Those data are then preprocessed and finally are transmitted to the web server via the 3G network.

Meanwhile, the software architecture of the server side applications for FCD traffic monitoring is shown in Fig. 4.

The server side consists of web server, client, 3rd party web services, and probe subsystems. The web

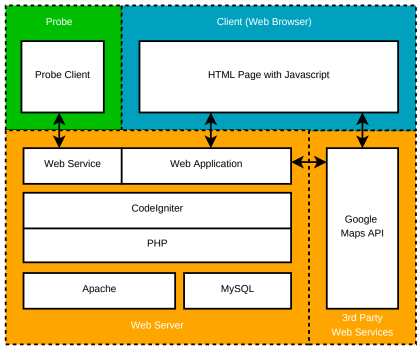

Fig. 4. The software architecture of the server side application.

server is the core system that collect, process, store, and provide traffic information. The system has two features: web service and web application. The web service communicates with probe clients to receive updates of the traffic data. The other feature, web application, is accessible from client web browser to retrieve traffic information. The client web browser is the only component accessible by end users. The web application will instruct a client web browser to load map provided by Google Maps API, and will overly traffic information on the map.

The web service and web application is a single system build on top of CodeIgniter framework, which uses PHP programming language. The use of CodeIgniter framework is preferred since it offers numerous technical advantages and organizational advantages, such as faster development and cleaner application structure, in comparison to developing native PHP application [9]. PHP is installed as a module on Apache web server and also integrated with MySQL database server.

III. Research Method

The optimal length of VTL is determined empirically based on the geo-location data recorded by means of ten units of smart-phones of similar type. In the current experiment, we use Samsung Galaxy Fame S6810. Those phones are installed with a custom-made application (see Fig. 5), which is designed to record the track or movement of a probe vehicle.

The application records the vehicle position (longitude and latitude) and the vehicle instantaneous velocity, and transmits the data to a designated server. A Global Navigation Satellite System (GNSS) provides those geo-location data. In the current investigation, the application also transmits the capture time and the accuracy level of the data.

The data are recorded on January 3, 2014 where the probe vehicle is driven along Jakarta Inner-Ring Road (see Fig. 6). The road mainly has three vehicular lanes on each direction. During the test, the probe vehicle is maintained to move along the central vehicular lane

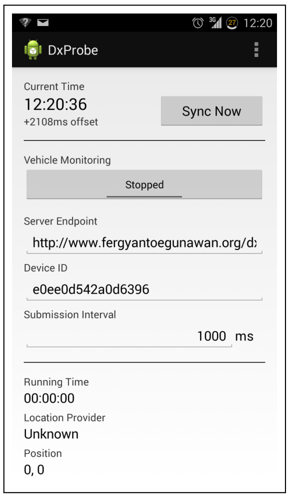

Fig. 5. The GUI of the client-side application for traffic monitoring using floating-car data system.

of the road. Several hundreds of thousand of data are obtained during the test.

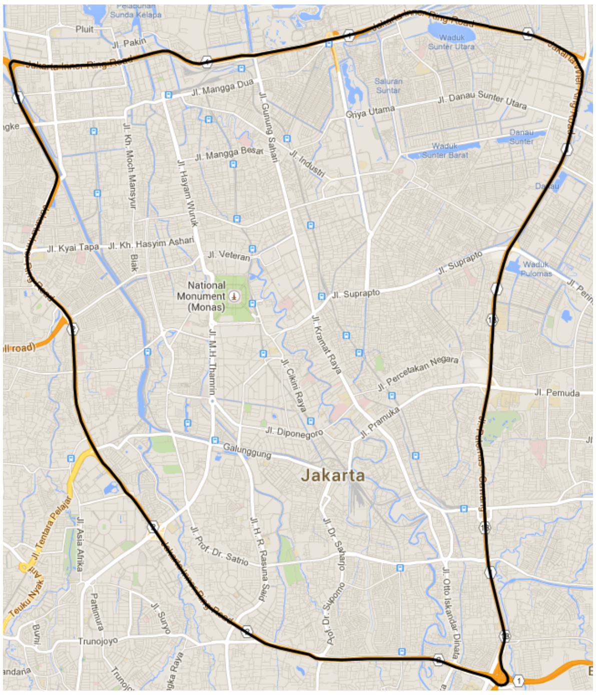

Fig. 6. The test site: Jakarta Inner-Ring Road.

Subsequently, the GPS location data are projected to the road GIS line using the algorithm explained in the following. Finally, the projection length is used to statistically establish the optimal length of VTL.

For a given GIS line-segment, and two GIS points, $\mathbf{x}_r^1$ and $\mathbf{x}_r^2$, we can establish the direction of the line segment by (see Fig. 7):

$$\hat{\mathbf{r}} = \frac{\mathbf{x}_r^2 - \mathbf{x}_r^1}{\|\mathbf{x}_r^2 - \mathbf{x}_r^1\|},$$

where $\|\square\|$ denotes the Euclidian norm. We project the GPS point $\mathbf{x}$ onto the line by the vector $\mathbf{d}$, which is:

$$\mathbf{d} = \left(\mathbf{x} - \mathbf{x}_r^1\right) - \left[\left(\mathbf{x} - \mathbf{x}_r^1\right) \cdot \hat{\mathbf{r}}\right] \hat{\mathbf{r}}.$$

The projection length is simply $\|\mathbf{d}\|$, and this algorithm is applied to all recorded data. The statistical distribution of the projection length $\|\mathbf{d}\|$ is established from which an optimal length of the VTL is determined.

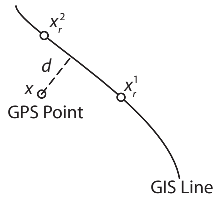

Fig. 7. Projection of a GPS point to a GIS line.

Furthermore, in order to determine the confidence interval, associated with the length of VTL, we should establish the statistical distribution of the projection lengths. We assume the data are normally distributed. The data are numerically and visually inspected for the distribution. The following criteria are used to evaluate the normality of the projection length data:

- Skewness and kurtosis z-values should be somewhere in the span of -1.96 to $+1.96$,
- The Shapiro-Wilk test p-value should be above 0.05, and
- Histogram and normal Q-Q plots should indicate normality.

If the results of the previous step indicate that the project lengths are indeed following the normal distribution function, then we can relate the probability of the occurrence that the probe vehicle will cross the virtual trip lines and the VTL length. With the normal distribution assumption, we transform the projection lengths to the standard form by $z = (d - \mu)/\sigma$, where d is the length of the VTL, μ is the associated mean, and σ is the associated standard deviation. We assume that the optimal VTL length should be established such that for the 95% of the cases, the probe vehicle should cross the line, or mathematically speaking, $P(|z| < z^*) = 95\%$, or $z^* = 1.96$ (see Fig. 8).

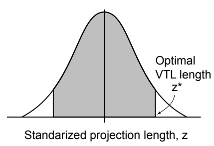

Fig. 8. The normal distribution of the projection length and determination of the optimal VTL length.

Finally, we verify the above results by the following scenario. We establish 88 VTLs across the Jakarta Inner Ring road. We drive a probe vehicle along the road and monitor the number of lines that are successfully detected by the probe vehicle.

IV. Research Results

As mentioned previously, the recorded geo-location data also contain the data level of accuracy. The number of the recorded data as a function of the accuracy level is depicted in Fig. 9. According to the accuracy level, the data seem to distribute following the Poisson distribution with a peak at 16 m of the accuracy level. Majority of the data, $\pm 70\%$, are recorded with the accuracy levels of 12 m and 16 m. The distribution in the figure should provide the characteristics of the accuracy the GPS unit utilized in the current experiment.

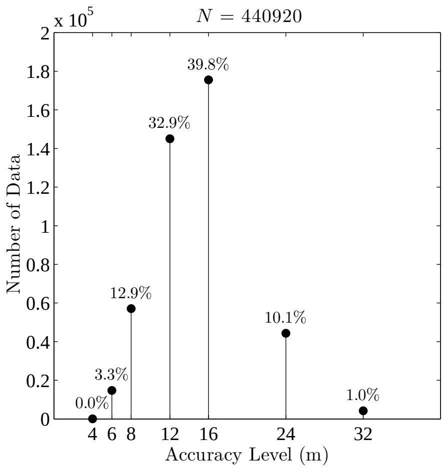

Fig. 9. The distribution of the number of data as a function of the accuracy level.

As described in Section III, the distribution of the projection length data at an accuracy level is assumed to be normally distributed. The assumption is evaluated with respect to the distribution skewness and kurtosis, the Shapiro-Wilk p-value, and some graphical instruments.

The z-values for the skewness and kurtosis are shown in the last two rows of Table I. For the accuracy levels of 4 m, 6 m, 8 m, and 32 m, the z-values of the skewness are 0.494, -1.232, 1.610, and 0.978, respectively. These z-values are within the range of -1.96 and 1.96. Thus, in this respect, the data are slightly skew but are approximately normally distributed. The data skew to the left of the mean value, or positively skew, except the case of 6 m level of accuracy where the data skew to the right, or negatively skew. Among the cases, the data for 4 m accuracy level have the smallest skewness and closely resemble the normal distribution. These conclusions are clearly in agreement with the histograms depicted in Fig. 10.

The z-values for the kurtosis are -0.148, 0.030, 0.070, and -0.943 for the accuracy levels of 4 m, 6 m,

		GPS Accuracy							
		4 m		6 m		8 m		32 m	
		Statistics	SE	Statistics	SE	Statistics	SE	Statistics	SE
Mean		-2.61E-01	0.285	-5.32E-02	0.704	-2.61E+00	0.412	9.56E-01	1.065
95% CI for Mean	Lower bnd	-8.40E-01		-1.49E+00		-3.45E+00		-1.21E+00	
	Upper bnd	3.19E-01		1.38E+00		-1.77E+00		3.12E+00	
5% Trimmed Mean		-2.69E-01		4.58E-02		-2.69E+00		8.42E-01	
Median		-6.74E-01		8.90E-01		-2.94E+00		-5.34E-02	
Variance		2.757		1.69E+01		5.77E+00		38.599	
Std. Deviation		1.66E+00		4.11E+00		2.40E+00		6.21E+00	
Minimum		-3.70E+00		-9.43E+00		-6.79E+00		-8.92E+00	
Maximum		3.71E+00		8.87E+00		2.67E+00		1.28E+01	
Range		7.42E+00		1.83E+01		9.45E+00		2.17E+01	
Interquartile Range		2.38E+00		5.56E+00		2.91E+00		8.95E+00	
Skewness		0.199	0.403	-0.496	0.403	0.649	0.403	0.394	0.403
Kurtosis		-0.117	0.788	0.024	0.788	0.055	0.788	-0.743	0.788
z-value of skewness		0.494		-1.231		1.610		0.978	
z-value of kurtosis		-0.148		0.030		0.070		-0.943	

TABLE I

The statistical description of the projection length data.

TABLE II

Shapiro-Wilk test results for the normality test of the projection length data.

Level of	Kolmogorov-Smirnov			Shapiro-Wilk		
accuracy (m)	Statistic	df	Sig.	Statistic	df	Sig.
4	0.117	34	0.200	0.982	34	0.824
6	0.162	34	0.024	0.951	34	0.133
8	0.180	34	0.007	0.943	34	0.073
32	0.111	34	0.200	0.954	34	0.146

8 m, and 32 m, respectively. These z-values are also within the range of -1.96 and 1.96; thus, the data are approximately normally distributed.

The Shapiro-Wilk test of normality hypothesizes that the data are normally distributed (H_0); meanwhile, the alternative hypothesis H_a is that the data are not normally distributed [8]. The null hypothesis H_0 is rejected if the associated p-value is below 0.05. In Shapiro-Wilk tests, the p-values are 0.824, 0.133, 0.073, and 0.146 for the cases of 4 m, 6 m, 8 m, and 32 m, respectively (see Table II). Thus, only the case of 8 m, the null hypothesis is rejected; for the remain cases, we accept the null hypothesis. We conclude that in term of Shapiro-Wilk test, the data are approximately normally distributed.

As the final assessment, we evaluate the Q-Q plot of the projection length data with respect to the standard normally distributed data. The results, depicted in Fig. 11, show that the data are closely distributed along the diagonal lines of the normal distribution.

Evaluation on these criteria—skewness and kurtosis in Table I, Shapiro-Wilk test in Table II, histograms in Fig. 10, and Q-Q plots in Fig. 11—leads us to conclusion that the projection length data, or the distance of the recorded location to the actual location, are approximately normally distributed with skewness of 0.199 (SE = 0.403), -0.496 (SE = 0.403), 0.406 (SE = 0.394), and 0.394 (SE = 0.403) for accuracy level of 4 m, 6 m, 8 m, and 32 m, respectively.

With the fact that the projection length is approximately normally distributed, we can now determine the VTL length that optimum statistically. In order to provide a reliable monitoring system, the VTL

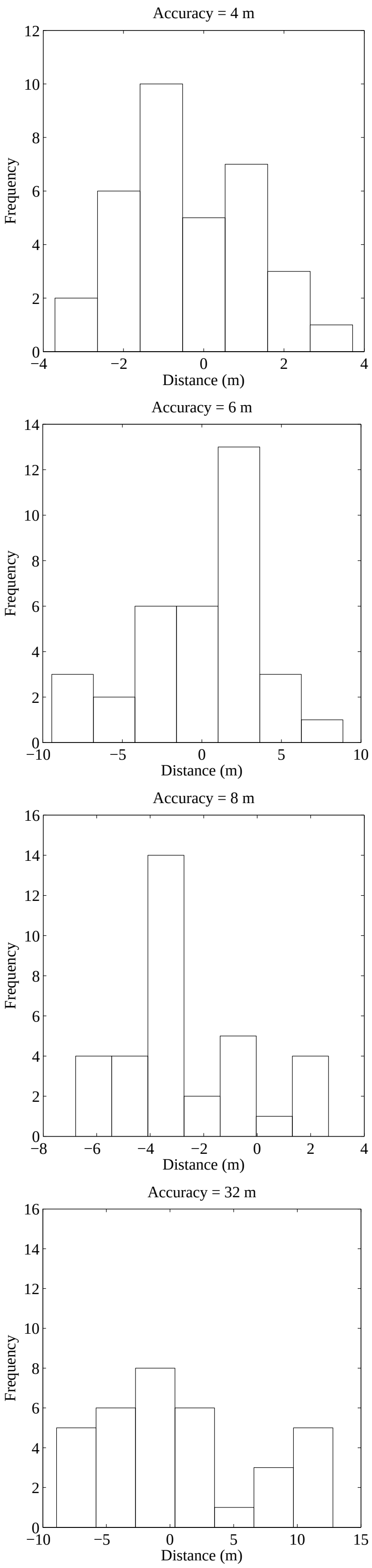

Fig. 10. Histograms of the projection length data.

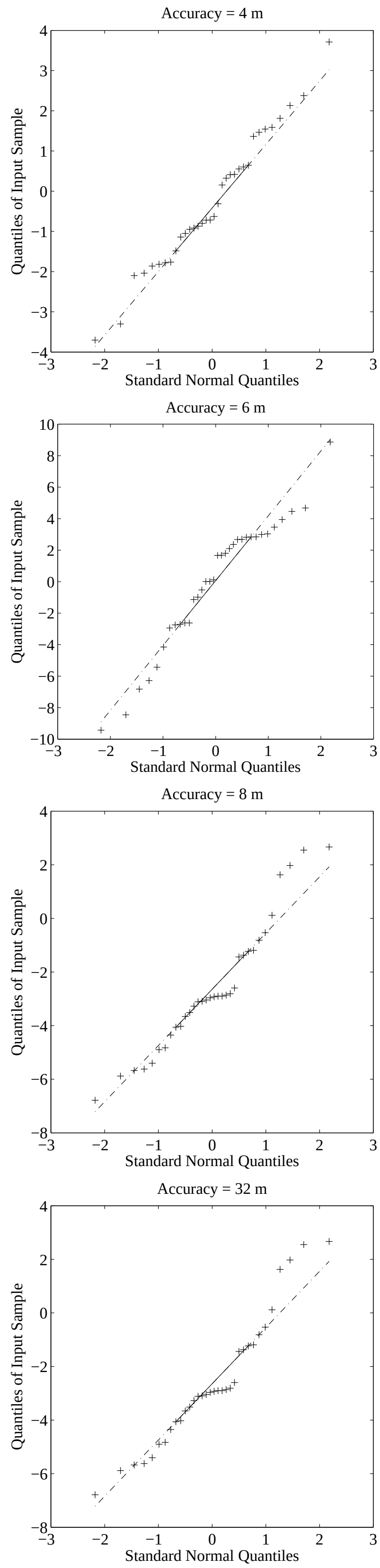

Fig. 11. Q-Q plot of the projection length data.

length should be determined based on the GPS data of the lowest possible measurement accuracy given by the equipment/smart-phone. For the present case, the smart-phone can provide the level of accuracy of 32 m on the worst case scenario. Therefore, the length is set at 95% confidence interval with $z = (d - \mu)/\sigma$, where $z = 1.96$, μ = 0.956 m, and σ = 6.21 m (see Table I). Finally, we obtain a VTL length of $d = \pm 13$ m. Therefore, with the total VTL length of 26 m, a probe vehicle will be detected by the VTL at 0.95 probability at the worst scenario or the lowest level of accuracy.

This above result is empirically verified by setting 88 VTLs along the Jakarta Inner-Ring road at nearly equaled distance and then, a probe vehicle is driven along the road to detect those VTLs. As a result, about 93% VTLs are detected during the verification test (see Fig. 12).

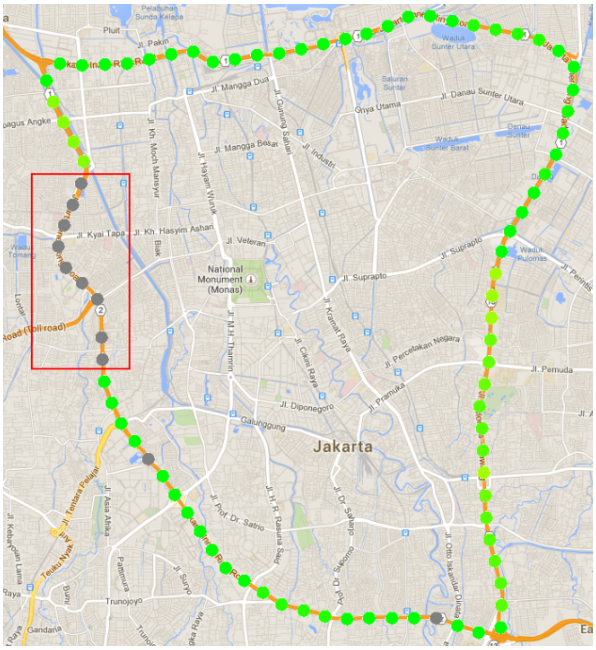

Fig. 12. The detected VTLs in the verification test. VTLs within the rectangle are excluded.

V. Conclusions

The floating car data (FCD) is rather new traffic monitoring method and it has a number of advantages in comparison to the traditional monitoring methods. The FCD method utilizes probe vehicles within traffic to monitor the traffic and it requires the probe vehicles to reveal their position and velocity to a designated monitoring server. Thus, the probe vehicle privacy may be compromised. One simple solution for this problem is to use the method in conjunction with the virtual trip line (VTL). In this scheme, each probe vehicle only reports the traffic status when the vehicle crosses the lines, avoiding a continuous monitoring of the probe vehicle position. However, a long VTL, as well as a short one, may have unintended consequences. In this article, we propose a method to determine the optimum length of the VTL statistically. Firstly, we demonstrate that the distances between the recorded geo-location data and the actual probe vehicle positions are distributed approximately according to the normal distribution. Then, we establish the critical region, and the associated VTL length, for a given level of significance. Finally, we empirically evaluate the method, and demonstrate that for a given VTL length, the empirical probability of the probe vehicle crossing a VTL is close to the theoretical probability.

Acknowledgment

We offer our highest appreciation to Dir. General Higher Education of Rep. Indonesia for the support via the Competitive Res. Grant No.: 017.A/DRIC/V/2013.

References

[1] Kumud K Sanwal and Jean Walrand. Vehicles as probes. Technical report, University of California, Berkeley, UCB-ITS-PWP-95-11, 1995.

[2] Xiaowen Dai et al., A simulation evaluation of a real-time traffic information system using probe vehicles. In *Intelligent Trans. Sys., Proc. IEEE*, vol 1, 2003.

[3] Juan C. Herrera et al., Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Trans. Res. Part C: Emerging Technologies*, 18(4):568–583, 2010.

[4] G. Fleck and R. Mertens. Method and system for determining dynamic traffic information, January 4 2000. US Patent 6,012,012.

[5] Baik Hoh et al. Preserving privacy in gps traces via uncertainty-aware path cloaking. In *Proceedings of the 14th ACM Conference on Computer and Communications Security*, CCS '07, pages 161–171, New York, NY, USA, 2007. ACM.

[6] Baik Hoh et al. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, MobiSys '08, pages 15–28, New York, NY, USA, 2008. ACM.

[7] Panos Kalnis et al. Preventing location-based identity inference in anonymous spatial queries. *IEEE Trans. on Knowl. and Data Eng.*, 19(12): 1719–1733, December 2007.

[8] Samuel S Shapiro et al. An analysis of variance test for normality (complete samples). *Biometrika*, 52:591–611, 1964.

[9] I P Vuksanovic et al., "Use of web app. frameworks in the development of small app.," in *MIPRO, IEEE proc. of the 34th int. conv.*, 458–462, 2011.

ICAICTA International Conference on Advanced Informatics: Concepts, Theory and Applications

IEEE Advancing Technology for Humanity

Certificate of Honour and Appreciation

presented to:

FERGYANTO GUNAWAN

for outstanding contribution

as

PRESENTER

at the First International Conference on Advanced Informatics:
Concepts, Theory, and Applications

Bandung, Indonesia, August 20 – 21, 2014

ICAICTA International Conference on Advanced Informatics: Concepts, Theory and Applications

Dr. Achmad Imam Kistijantoro
General Chair of the ICAICTA 2014