Proceedings of the 8th International
IEEE Conference on Intelligent Transportation Systems
Vienna, Austria, September 13-16, 2005

Segment based Traffic Information Estimation Method Using Cellular Network Data

W. C. Mark Hsiao and S. K. Jason Chang

Abstract—**In traditional traffic information estimation method, the difference in positions is used to estimate traffic information. However, variation or flutter of positioning caused by imprecise mobile phone location accuracy may lead to the unstable measurement. Take the latest mobile phone location technology into consideration, this paper introduces a new segment based traffic estimation method. It is proved by simulation approach that segment based method performs well when location accuracy is within the length of a segment. Simulation results also demonstrate that segment based method is better than traditional method under all kinds of traffic condition except for the high location error and over saturation condition. Through simulation analysis, we conclude that enough sample size, i.e. larger vehicle generation rate, longer data collection interval, shorter location update interval, and larger mobile penetration rate are crucial factors to generate accurate traffic information.**

I. INTRODUCTION

INTELLIGENT Transportation System (ITS) relies on comprehensive and accurate traffic information for managing the road network and providing navigation services for the road users. Two major methods have been proposed to collect quantitative traffic information, including roadside vehicle detectors and vehicle probes. The former approach can easily calculate instantaneous traffic characteristics, but they are extremely costly to deploy and can only be installed in limited locations. The latter generally uses GPS/GSM technology based in-vehicle device to collect and report traffic information. In order to continuously collect traffic information, commercial vehicles are preferred to be employed for the latter method. However, commercial vehicles, e.g. trucks and coaches, cannot represent the general traffic, and this leads to a bias of real traffic information.

Since most drivers or passengers already hold mobile phones, using a mobile phone instead of a GPS/GSM technology based in-vehicle device to calculate traffic information provides more advantages, such as wider road network coverage compared to roadside vehicle detectors and higher in-vehicle device penetration compared with vehicle probes. Moreover, mobile phone location based traffic information system is expandable to provide value-added traffic information service and bring in additional revenue for cellular network operators. Both public and private sectors can obtain benefits from this approach, hence many research studies and operational tests [1, 2, 3] have been conducted to explore its feasibility.

Generally speaking, mobile phone location based traffic information system comprises of two major subsystems: mobile phone location positioning system and traffic information estimation system. The mobile phone location positioning system is in charge of positioning the mobile phone and generating the geographical coordinates of subscribers. The traffic information estimation system transforms the geographical coordinates into spatial locations on the road network and measures the traffic information. The primary positioning technologies adopted for earlier mobile phone location based traffic information systems are a triangulation based approach, including Angle of Arrival (AOA), Enhanced Observed Time Difference (E-OTD) and Observed Time Difference of Arrival (OTDOA). These systems depend on network overlays of additional radio transmissions, which require expensive special location measurement hardware integrated into the cell site base stations. Because of high implementation costs, these solutions are limited to test implementations. Researches also concluded that earlier wireless location technology based traffic monitoring systems are categorized as unsuccessful with the following reasons [4].

- Location accuracy of these systems is usually inaccurate to some degree.
- These systems rely on the cellular communication information of active subscribers. Consequently, sample size is too small to deduce traffic information.
- Map matching error is another issue. These position estimates may not lie directly on the right roadway.

Recently, several techniques for enhancing location accuracy have been introduced, including improvement to Enhanced Cell-ID (E-Cell-ID), Time Difference of Arrival (TDOA) and Advanced Forward Link Trilateration (AFLT). The prediction algorithms compute the location of the

Manuscript received January 14, 2005.

W. C. Mark Hsiao is with the Civil Engineering Department, National Taiwan University, Taipei, Taiwan, ROC (phone: +886-2-27365474, fax: +886-2-27368030, email: mark.hsiao@seed.net.tw)

S. K. Jason Chang is with the Civil Engineering Department, National Taiwan University, Taipei, Taiwan, ROC (phone: +886-2-23625920, fax: +886-2-23639990, skchang_2001@hotmail.com)

observation is able to provide a location positioning accurate to within 100 meters 67% of the time [5]. In the meanwhile network based passive and real-time bulk cellular network data interception technologies have been introduced to collect cellular communication information of the passive subscriber, hence enough sample data can be provided. Several commercial systems have been claimed to be successful based on this new-generation system architecture [5, 6].

For all these systems, the difference in the positions is used to measure traffic information (hereinafter distance based method). However, variation or flutter of the positioning caused by imprecise location accuracy leads to an unstable measurement. To conquer this problem, a segment based estimation method in contrast with the distance based estimation method is proposed in this paper. It improves the stability and accuracy of traffic information significantly.

The paper is organized as follows: Section II describes the definitions and approach of this paper. Section III outlines the proposed simulation method. Section IV presents results of simulation. Finally, Section V concludes the findings along with future research topics.

II. DEFINITIONS AND APPROACH

A. System Design

The operational process of the mobile phone location based traffic information system consists of the following five steps.

1) Step 1- data collection and positioning: Two kinds of signal data can be collected and positioned, including periodically polled/report data by inactive subscribers and randomly report data sent by active subscribers. The former data is useful for generating regular traffic measurements; the latter data is used as an auxiliary to detect traffic jams or incidents timely. These data is acquirable from the above-mentioned new-generation system.

2) Step 2- positioning subscriber: Traffic situations in freeways, highways, and streets are different in virtue of traffic restrictions of the underlying network and traffic pattern. For example, one-way streets and intersection stops are common in urban areas while stops on the freeway are rare and many stops are even abnormal. These characteristics are useful to locate subscribers on the precise road.

3) Step 3- positioning vehicle: To calculate the traffic information denoted by vehicular movement, non-vehicular data such as pedestrians need to be filtered, and multiple subscribers on the same vehicle also need to be extracted. By using pre-calibrated probability distribution of subscriber penetration, it is possible to generate vehicular data accurately.

4) Step 4- traffic information estimation: In general, traffic information is measured in terms of travel time, travel speed and/or incident. Further measurements like traffic flow, delay, OD trips are preferred by professional requirement.

5) Step 5- verification and feedback: Several criteria are used to verify the results of estimation, including valid numerical range, threshold of variation or fluctuation, and traffic rationality of measurement.

This paper focuses on Step 4 - the method of traffic information estimation, i.e. transforming the vehicular spatial location on the road network into the traffic information. To enhance the quality and accuracy of traffic information, this paper proposes a segment based method rather than the distance based method. This is like quantizing an analog signal to eliminate background noise, which is the concept behind a digital cellular which has cleaner voice than an analog cellular phone.

B. Segment based Model

An example as shown in Figure 1 is presented to describe the segment based method. Consider the following scenario:

- One single lane and unidirectional freeway,
- Normal traffic condition, cruising speed = 30 km/h,
- Signal data reported periodically, and;
- Maximum location error = 50 m.

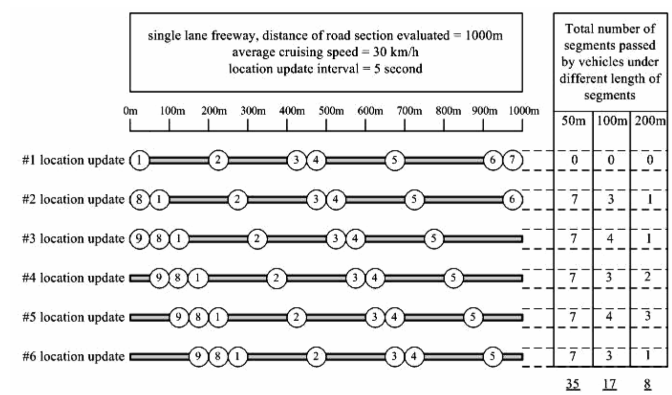


FIG. 1. SEGMENT BASED METHOD ILLUSTRATION

The circles in Figure 1 are representative of the maximum location error that the location system achieves on the road. Relative parameters are defined as follows.

Input parameters:

- LOR: length of road section under evaluation (meter)
- LUI: location updates or signal data report interval (sec)
- LOS: length of segment (meter)
- OTP: observed time period (second) i.e. traffic data collection interval

Data collection:

- SUM: summation of observed vehicles during observed time period (vehicles)
- TPS: total number of segments passed through by vehicles; as shown in Figure 1, TPS is 35, 17 and 8 accordingly for three different length of segment – 50/100/200 meters.

Output result:

- TTT: total travel time (second) = SUM*OTP
- TTD: total travel distance (meter) = TPS * LOS
- ATS: average travel speed (km/h) = TTD / TTT
- DST: density (vehicle/segment/time slot) = SUM / number of time period
- Traffic volume = DST * ATS

Table 1 shows the estimated average travel speed (ATS) is approximate to actual average travel speed (30km/h). The result is reasonable that the smaller the segment the less the quantization error and the more the estimation result is close to the actual status. However, different scenarios and input parameters result in different outcomes. To further investigate the factors affecting the results of the segment based method, a comprehensive simulation program is developed and elaborated in the next section.

TABLE I

SEGMENT BASED METHOD CALCULATION

Input Parameters			Data Collection		Output Result				
location update interval (LUI, unit: second)	length of segment (LOS, unit: meter)	observed time period (OTP: second)	summation of observed vehicles during observed time period (SUM, unit: vehicle)	total number of pass thru segments for all vehicles (TPS)	total travel time (TTT, unit: second) = SUM* observed time period	total travel distance (TTD, unit: meter) = TPS * LOS	average travel speed (ATS, unit: km/h) =TTD / TTT	density (SUM / number of time period)	volume (density * speed)
5	50	30	42	35	210	1,750	30	7	210
5	100	30	42	17	210	1,700	29	7	204
5	200	30	42	8	210	1,600	27	7	192

III. SIMULATION

The flow chart of the simulation program is illustrated as Figure 2. This program simulates each movement of real vehicles and emulates the positioning results of these real vehicles according to standard normal distribution with predefined location error as the value of standard error. After simulation, the average travel speed of real vehicles, distance based and segment based average travel speed of positioned vehicles are calculated for further performance evaluation.

We use one unidirectional single lane freeway with length of 2000 meters as the underlying network and assume that every vehicle has maximum one mobile phone.

Four kinds of scenarios are developed to evaluate the segment based method, including:

- General condition,
- Near saturation condition,
- Over saturation condition, and;
- Light traffic condition.

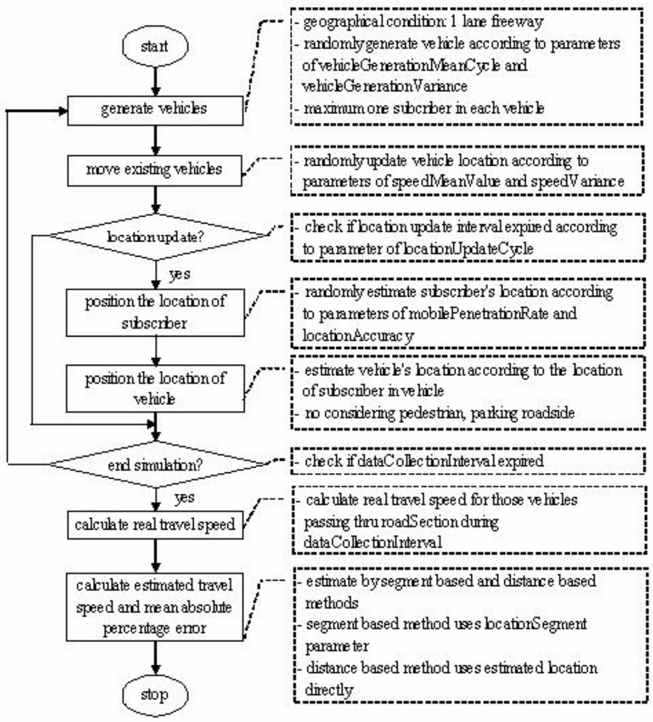


FIG. 2. SIMULATION FLOW CHART

Predefined program variables for each scenario are listed in Table 2.

TABLE 2

PREDEFINED PROGRAM VARIABLES

Variables	General Condition Scenario	Near Saturation Scenario	Over Saturation Scenario	Light Traffic Scenario
vehicleGenerationMeanCycle (second)	10	3	20	20
vehicleGenerationVariance (second)	5	2	5	5
speedMeanValue (km/h)	60	50	30	80
speedVariance (km/h)	20	20	10	20
dataCollectionInterval (second)	300	300	300	300
locationUpdateCycle (second)	5	5	5	5
locationSegment (meter)	500	500	500	500
locationAccuracy (meter)	50/100/250/500	50/100/250/500	50/100/250/500	50/100/250/500
mobilePenetrationRate (%)	100	100	100	100
roadSection (meter)	2000	2000	2000	2000

In the simulation program, vehicles are generated randomly by the interval x, where the probability density function of x is a uniform distribution, i.e. f(x)=1/(b-a), a=x=b, a=vehicleGenerationMeanCycle-vehicleGenerationVariance, b=vehicleGenerationMeanCycle+ vehicleGenerationVariance. Vehicles are generated randomly for each scenario to reflect the real situation.

Vehicles are also moved randomly by the speed x, where the probability density function of x is a uniform distribution, i.e. f(x)=1/(b-a), a=x=b, a=speedMeanValue-speedVariance, b=

speedMeanValue+speedVariance. Multiple speed readings are generated from the same vehicle as it traverses the network.

For every locationUpdateCycle, vehicles on the road are selected by the sampling rate of mobilePenetrationRate and are positioned by locationAccuracy. Where locationAccuracy means the location error compared to real location, it is defined as normal distribution with 67 percentile of error at 50/100/250/500 meters. In the simulation program, we use Box-Muller [7] transformation method to generate standard normal random quantities with 50/100/250/500 meters being the one sigma value for each scenario. The locations of positioned vehicles at each time step are recorded for further traffic information estimation using distance based and segment based method.

Variable locationSegment defines the length of each segment used for segment based calculation. When dataCollectionInterval expired, travel speed of distance based method and segment based method are calculated for those vehicles passing through road section. Therefore, the sample size for each scenario is equal to dataCollectionInterval/ vehicleGenerationMeanCycle. And ten runs of simulation for each scenario are conducted to generate average measurement.

IV. RESULTS OF SIMULATION

A. Location Accuracy Impact Analysis

Simulation results of each scenario are shown as Figure 3 to Figure 6 respectively. Where real.speed means the actual space mean speed, seg.speed means estimated speed by segment based method, seg.mape means mean absolute percentage error of segment based method, dst.speed means estimated speed by distance based method, and dst.mape means mean absolute percentage error of distance based method.

Take seg.mape as an example, mean absolute percentage error (MAPE) =

$$\sum_{i=1}^{N}[\frac{|real.speed(i)-seg.speed(i)|}{real.speed(i)}]\times 100/N$$

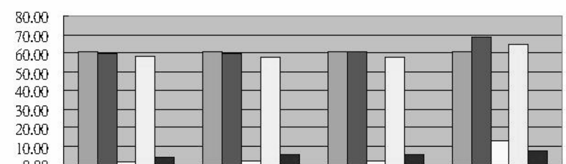
, where i means the i_{st} run of simulation, N=10.

It is obvious to conclude that the segment based method is better than the distance based method under all kinds of scenarios except for the high location error and over saturation condition. This result is reasonable. Since the segment speed is triggered only when the measurement value (with error) is over location accuracy (e.g. 50/100/250/500meter) mark, the probability of triggering the segment meter is the probability of X>x where X is normal distribution. And given the same mean value, a normal distributed random number has a larger probability of X>x if X has a larger sigma. So the MAPE of segment based method should increase with the location error. Especially when two times of location accuracy (diameter of location error circle) is larger than segment, the probability of triggering the segment meter is more distinct. This leads to higher MAPE than distance based method as shown in those scenarios where both location error and segment length are 500m.

The same as over saturation scenario, where vehicles move slowly and sample size is limited, thus the probability of triggering the segment meter is relatively higher than distance base method.

Although the MAPE of segment based method is higher than distance based method in over saturation condition, the MAPE of distance based method is still too high. That means mobile location based traffic information is not reliable for limited sample size condition like over saturation scenario, no matter segment based or distance based method are employed.

Through the simulation analysis, we can say that the MAPE of segment based method is only 1/3 of that from distance based method when the location error is within the length of segment.

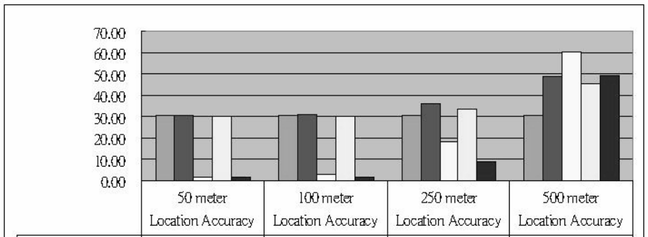


	50 meter Location Accuracy	100 meter Location Accuracy	250 meter Location Accuracy	500 meter Location Accuracy
real.speed (km/h)	60.42	60.42	60.41	60.42
seg.speed (km/h)	59.60	59.45	60.74	68.27
seg.mape (%)	1.44	1.69	1.75	12.98
dst.speed (%)	57.95	57.13	57.30	64.33
dst.mape (%)	4.09	5.46	5.17	7.20

FIG. 3. SIMULATION RESULT OF GENERAL CONDITION

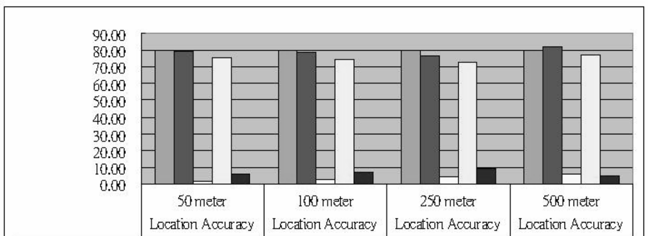
	50 meter Location Accuracy	100 meter Location Accuracy	250 meter Location Accuracy	500 meter Location Accuracy
real.speed (km/h)	50.49	50.49	50.49	50.49
seg.speed (km/h)	51.06	51.26	50.60	56.94
seg.mape (%)	1.12	1.52	1.43	12.78
dst.speed (km/h)	48.57	47.53	46.82	52.39
dst.mape (%)	3.80	5.87	7.28	3.81

FIG. 4. SIMULATION RESULT OF NEAR SATURATION



	50 meter Location Accuracy	100 meter Location Accuracy	250 meter Location Accuracy	500 meter Location Accuracy
real.speed (km/h)	30.43	30.43	30.43	30.43
seg.speed (km/h)	30.47	31.31	36.01	48.91
seg.mape (%)	1.48	3.02	18.38	60.76
dst.speed (km/h)	30.03	30.20	33.14	45.41
dst.mape (%)	1.46	1.82	9.02	49.27

FIG. 5. SIMULATION RESULT OF OVER SATURATION



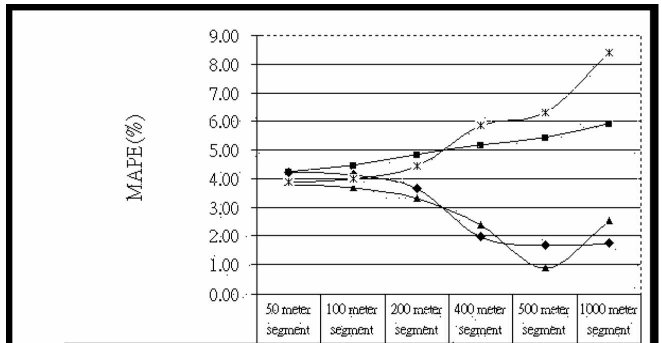
	50 meter Location Accuracy	100 meter Location Accuracy	250 meter Location Accuracy	500 meter Location Accuracy
real.speed (km/h)	80.36	80.36	80.36	80.36
seg.speed (km/h)	79.72	78.52	76.73	82.45
seg.mape (%)	1.08	2.49	4.52	5.65
dst.speed (km/h)	76.06	74.57	72.74	77.56
dst.mape (%)	5.36	7.20	9.48	5.27

FIG. 6. SIMULATION RESULT OF LIGHT TRAFFIC

B. *Segment Length Impact Analysis*

Figure 3 to 6 include those scenarios of higher location error. However, if location error is limited to within predefined segment length, than segment based method can attain more accurate result. Hence, it is not proper to conclude that distance method seems to be more robust.

The above simulation is conducted with segment length of 500m. It will be interesting to know the relationship between segment length and location accuracy. Figure 7 shows the MAPE between different segment length and location accuracy. It is concluded that the MAPE of segment base method is significant better than distance base when segment length is larger than two times of location accuracy (diameter of location error circle). It is reasonable since the probability of triggering the segment meter is lower than distance base method under this condition. This finding becomes the precondition to implementing a segment based mobile traffic information system.

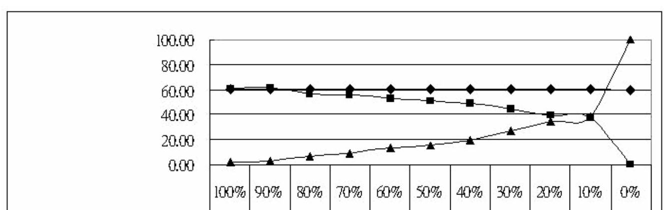


	50 meter segment	100 meter segment	200 meter segment	400 meter segment	500 meter segment	1000 meter segment
seg.mape (%) (100 meter location accuracy)	4.22	4.13	3.67	1.98	1.69	1.74
dst.mape (%) (100 meter location accuracy)	4.25	4.49	4.85	5.20	5.46	5.94
seg.mape (%) (200 meter location accuracy)	3.80	3.69	3.33	2.40	0.91	2.55
dst.mape (%) (200 meter location accuracy)	3.90	4.01	4.43	5.85	6.30	8.42

FIG. 7. SIMULATION RESULT OF SEGMENT LENGTH

C. *Mobile Penetration Rate Impact Analysis*

In addition, it is found that the sample size and location accuracy are the two major factors affecting the results of the mobile phone location based traffic information system. We realize that the sample size is determined by the vehicle generation rate, data collection interval, location update interval, and mobile penetration rate. The above simulations are conducted with the assumption of 100% mobile phone penetration rate. However, we might be interested to know the impact of mobile phone penetration rate. Different mobile phone penetration rates, then, are simulated under the general traffic condition scenario with 250m location accuracy to evaluate the influence of sample size to traffic information measurement. As Figure 8 shows, a 40% penetration rate will be tolerable to measure traffic information. 70% penetration rate can provide an even more accurate result.



	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%
real.speed (km/h)	60.42	60.42	60.42	60.42	60.42	60.42	60.42	60.42	60.42	60.42	59.95
seg.speed (km/h)	60.74	61.21	56.65	55.39	52.41	50.92	48.63	44.30	39.67	37.32	0
seg.mape (%)	1.75	3.35	6.26	8.34	13.26	15.73	19.53	26.68	34.35	38.26	100

FIG. 8. SIMULATION RESULT OF MOBILE PENETRATION RATE

V. CONCLUSION

The new-generation mobile phone location based traffic information system adopts a network based passive cellular network data interception technology and improved location technology, hence enough sample data can be provided and accurate positioning can be achieved. Although this new-generation system architecture provides a much more efficient approach than ever, the tasks of transforming mobile phone location into vehicular location as well as measuring traffic information are still full of research opportunities. Considering traffic restriction and historical traffic pattern of underlying road network as well as statistical analysis of local mobile phone users, it is possible to transform mobile phone location into vehicular location accurately. These topics are worthy of future research.

As for traffic information measurement, estimating travel time and speed using periodical signal data and estimating incident and delay using random sub-second signal data are two different research topics. This paper focuses on the former one and proposes a segment based approach instead of the conventional distance based approach. This approach eliminates the fluctuation of imprecise and unstable location accuracy and it proves that the segment based method is better than the distance based method under all kinds of scenarios except for the high location error and over saturation condition. If sample size is enough under general traffic scenario, the MAPE of segment based method is only 1/3 of that from distance based method. Through the simulation analysis, it is also found that even a limited mobile phone penetration rate is enough to measure traffic information.

Simulation experience shows that sample size and location accuracy are two critical factors to mobile phone location based traffic information system. Sample size is relevant to the vehicle generation rate, the data collection interval, the location update interval, and the mobile penetration rate. According to previous simulation analysis, sample size is not enough for over saturation scenario and low mobile penetration rate condition, so as to result in imperfect MAPE. Thus a larger vehicle generation rate, a longer data collection interval, a shorter location update interval, and a larger mobile penetration rate are all crucial to generate accurate traffic information. These factors become the keystones to designing mobile phone based traffic information system.

To make the results closer to the system implementation world, additional simulations to explore the effect on accuracy of road environments different from one single lane and unidirectional freeway conditions are necessary to be performed in the future.

REFERENCES

[1] Cayford, R. and Johnson, T., “Operational Parameters Affecting the Use of Anonymous Cell Phone Tracking for Generating Traffic Information.” Proceedings of the Transportation Research Board 82nd Annual Meeting. Transportation Research Board, Washington, D.C., 2003.

[2] Ygnace, J-L., Remy, J.G., Bosseboeuf, J.L., and Da Fonseca, V., Travel Time Estimates on Rhone Corridor Network Using Cellular Phones as Probes: Phase 1 Technology Assessment and Preliminary Results. INRETS Report. Arcueil, France. November, 2000.

[3] Yim, Y.B.Y. and Cayford, R., Investigation of Vehicles as Probes Using Global Positioning System and Cellular Phone Tracking: Field Operational Test. Report UCB-ITS-PWP-2001-9. California PATH Program, Institute of Transportation Studies, University of California, Berkely, CA, 2001.

[4] Smith, B.L., Zhang, H., Fontaine, M, and Green, M., Cellphone Probes As an ATMS Tool, Smart Travel Lab Report No. STL-2003-01, June, 2003

[5] McGeough J., Wireless Location Positioning Based on Signal Propagation Data, Marcus Evans Wireless Location Positioning Conference, San Francisco, 22 January 2003

[6] Applied Generics Ltd., www.appliedgenerics.com, Accessed May. 2004.

[7] Box, G. E. P. and Muller, M. E. "A Note on the Generation of Random Normal Deviates." Ann. Math. Stat. 29, 610-611, 1958.