CHAPTER IV

DATA ANALYSIS AND DISCUSSIONS

In this section, data analyses of obtaining travel times for link, route and OD were carried out. The analyses employed the research methodology were discussed in Chapter 3. Various traffic scenarios were simulated in Paramics Microsimulation model. The simulation yielded several data, including vehicle identification numbers, travel times, origins and destinations, and other traffic data for validation. The results from the simulation were installed in files, then transferred to worksheets for further analysis. Several statistical analyses on the corresponding data were carried out. The key examination was the quality of the travel time received from probe vehicles under various traffic and probe (market penetration) conditions. The analyses were divided into three parts: Link travel time, Route travel time, and OD travel time (Case study). The results of link travel times were discussed in section 4.1, the results of route travel times were discussed in Section 4.2 while section 4.3 discussed the Origin and Destination (OD) travel times.

4.1 LINK TRAVEL TIME

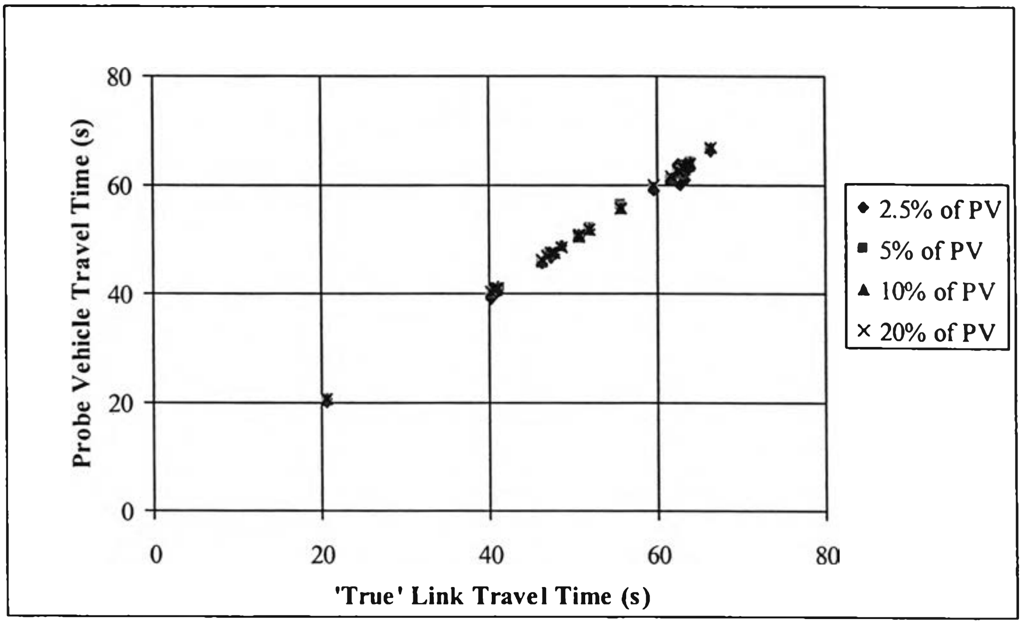
After processing the output from Paramics V5 according to the network building method, the 'true' and probe vehicle average travel times and their corresponding standard deviations were obtained for 45 minutes of simulation period. The 'true' average link travel time is the average of the travel time experienced by all vehicles that traversed the link during the simulation period, while the average probe vehicle travel time is the average travel time from only probes traveling on the link.

The accuracy of the link travel time obtained from the probe data is described by the magnitude of its standard deviation. Suppose that n_p probe vehicles traveling on a link and that each of these probe vehicles reports the time required to traverse the link. Then the standard deviation of the average reported link travel times from the probe vehicles is as follows:

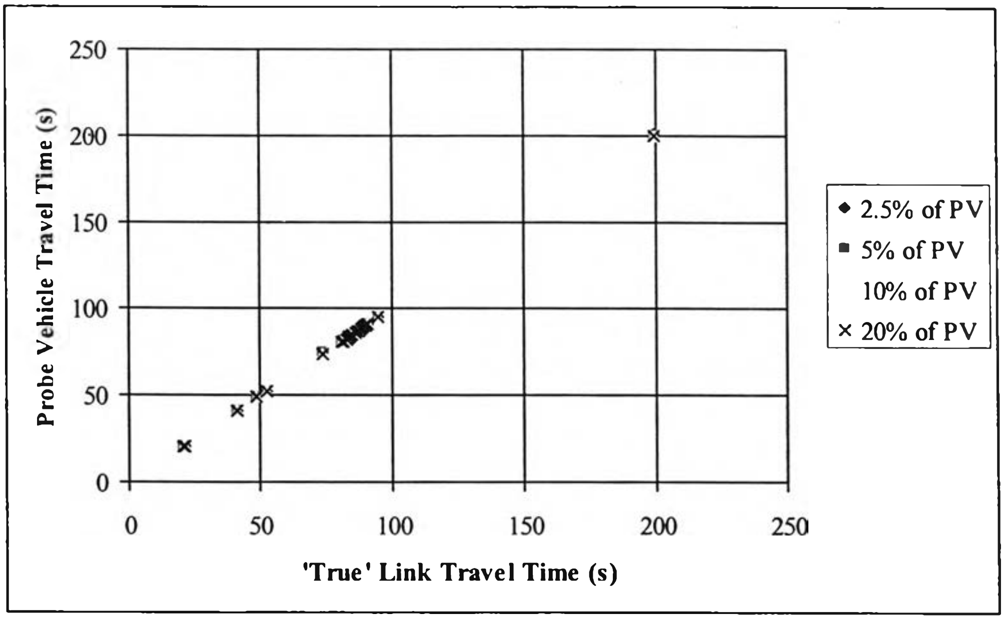
$$\sigma = \frac{\sigma_T}{\sqrt{n_p}} \tag{4.1}$$

where: σs = standard deviation of the mean link travel time reported by the probe vehicles
n_p = number of probe vehicles

The 'true' link travel time for every link in the network was analyzed from the travel time average of all vehicles passing the link. In case of the probe vehicle travel time, the ID of each vehicle for each link was matched with 200 bootstrap vehicles IDs. The average travel time obtained from the 'true' and the probe vehicles were able to be compared. The comparison of the 'true' travel time and the travel time by probe vehicles is shown in Figure 4.1. The standard deviation of the travel times and the average value of the travel times of probe vehicles are shown later in Figure 4.3.

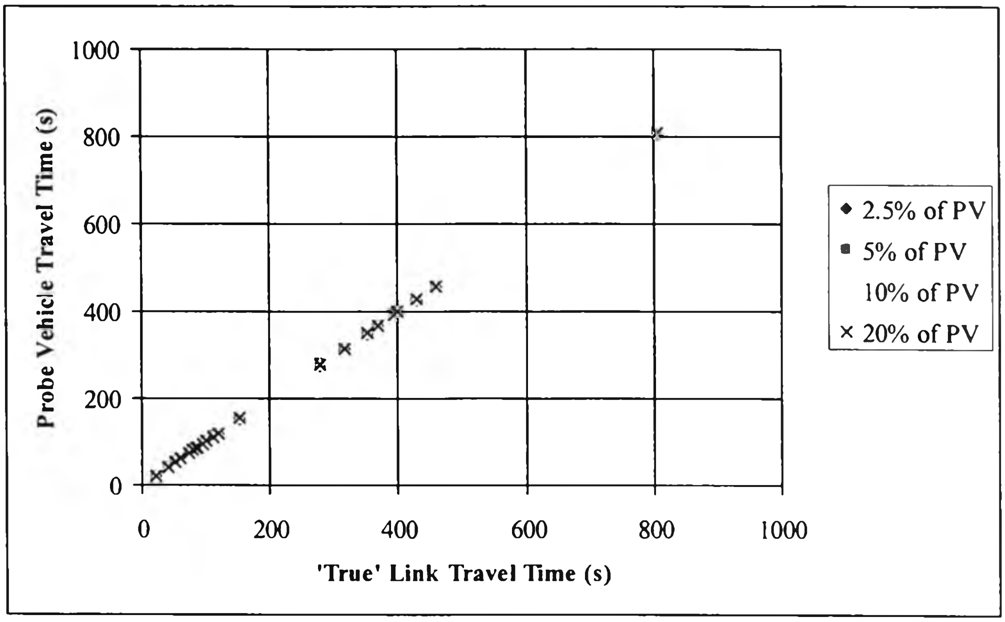


a) Traffic volume entering the network is 4000 veh/h

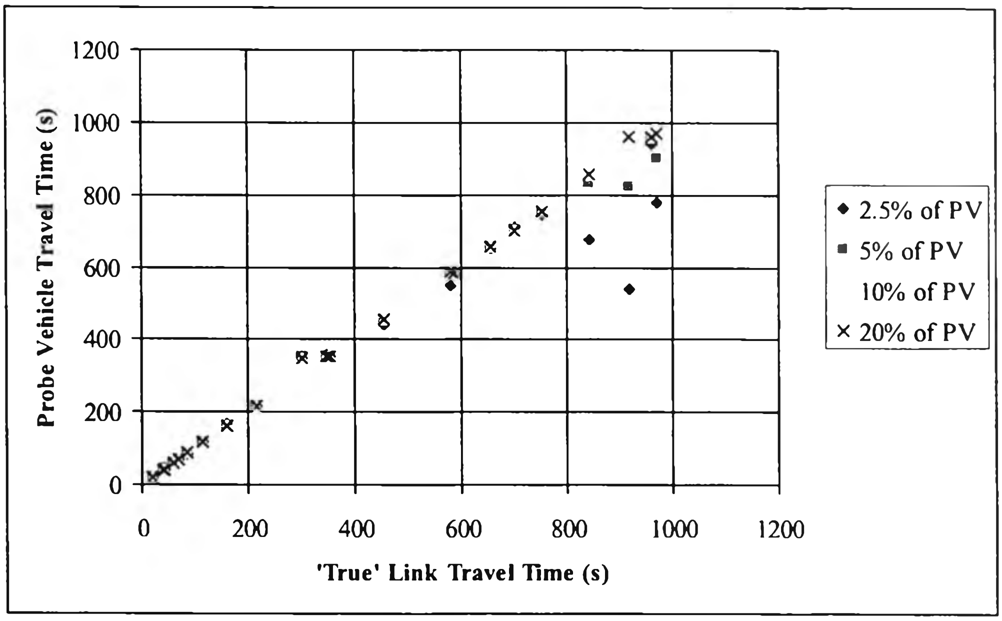


b) Traffic volume entering the network is 6000 veh/h

Figure 4.1 Average probe vehicle travel time (bootstrapped) versus 'true' link travel time.

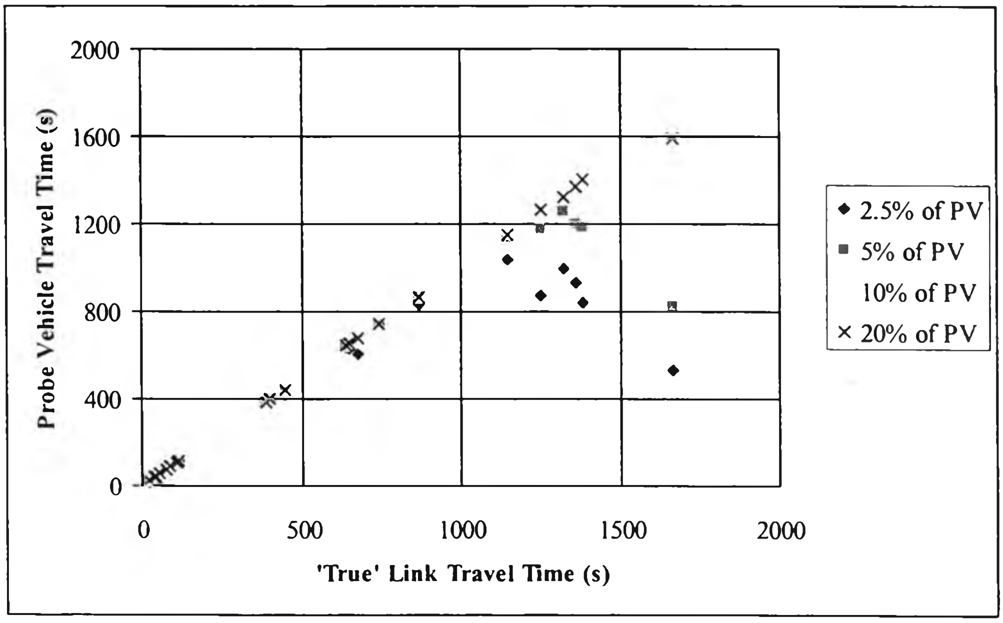


c) Traffic volume entering the network is 8000 veh/h



d) Traffic volume entering the network is 10000 veh/h

Figure 4.1 Average probe vehicle travel time (bootstrapped) versus 'true' link travel time (continued)

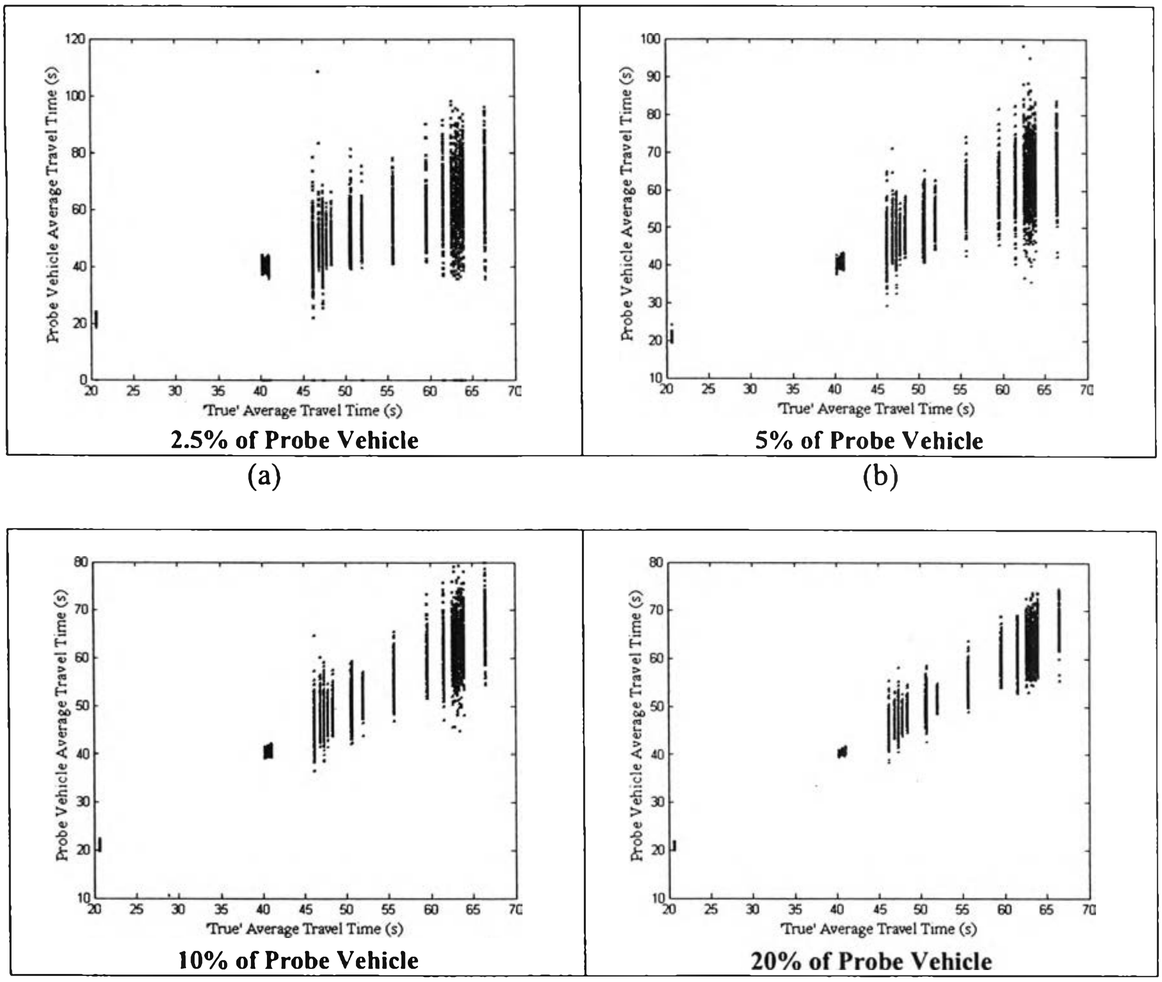


e) Traffic volume entering the network is 12000 veh/h

Figure 4.1 Average probe vehicle travel time (bootstrapped) versus 'true' link travel time (continued)

Figure 4.1 was constructed from the average link travel times reported by probe vehicles. The traffic conditions were broken into 5 traffic situations, depending on their congestion level (as indicated by the entering flow onto the test network). The average probe vehicle travel time for each link appearing in the figure was not the data from a single set of probe vehicles, rather it was the average of 200 bootstrapped travel times on that link. Thus, Figure 4.1 does not intend to show the difference of the average travel time from each (possible) set of probes, and the comparison with the true link travel time implies the ability of probe vehicles (at a particular percentage of probe vehicles) to give an accurate data, *on average*. Although unable to see the quality of the travel time (in terms of accuracy) of each reported data, Figure 4.1 reveals the effect of the amount of travel time data by probe vehicles (percentage of probes) to the resulting discrepancy of travel times, compared with the 'true' travel time.

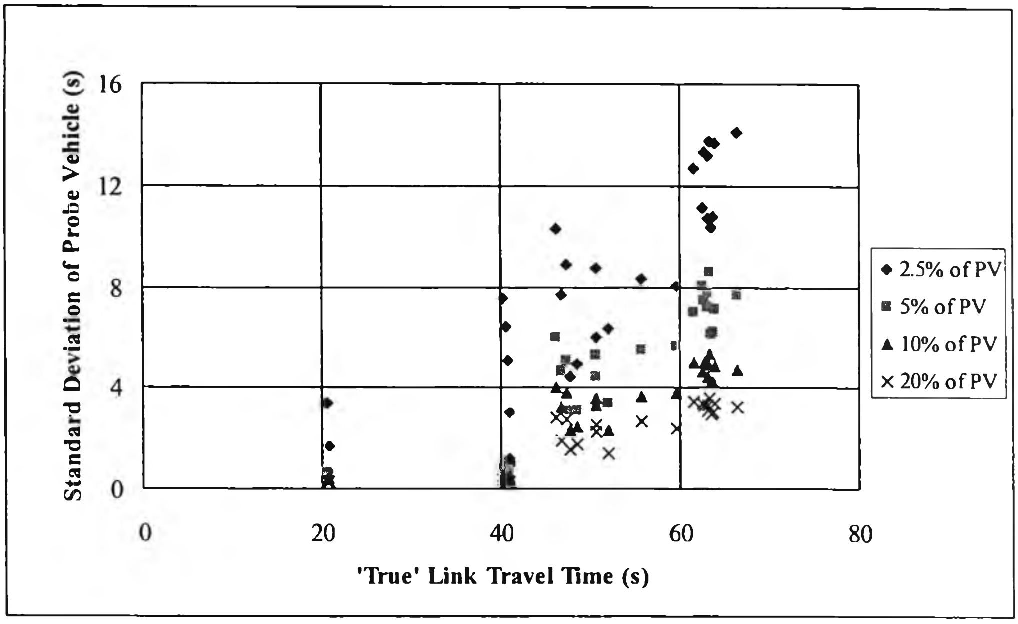
As shown in Figure 4.1, for every set of traffic volume, the travel times from probe vehicles are almost equivalent to the 'true' travel times. The plots of probe vehicle travel time versus 'true' link travel time agree well. This is very promising result, as it indicates the relations between accuracy of the travel times with different traffic volumes. This implies the obtainment of travel time accuracy by probe vehicles in a wide range of traffic volumes (situations). Therefore, the travel time from probe vehicles can be used for the representation of the 'true' travel time. However, when the level of service is poor, during the network starts to load with vehicles and exceeds the capacity (in other words, in saturated traffic conditions), the average travel time given by probe vehicle is not well equivalent to the 'true' travel time. As shown in Figure 4.1 d) and e), at 2.5% and 5% of probe vehicles, the average travel time from probe vehicles has underestimated the 'true' travel time. It is observed that the small amount of probe vehicles could result in the large discrepancy of travel time in congested conditions. When the number of probe vehicles is larger, then the reported average travel time would be closer to the 'true' travel time.



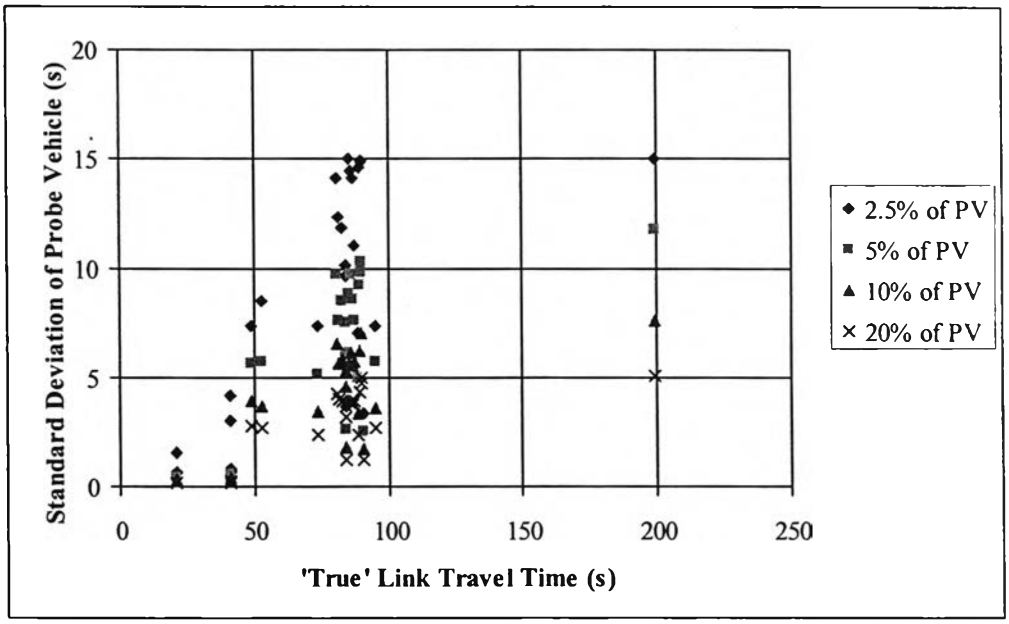
(a) (b) (c) (d)

Figure 4.2 Example of average probe vehicle travel times as a function of 'true' probe vehicle travel times from 200 bootstrapped for traffic volume 4000 veh/h

While the previous figures considered the average of 200 bootstrapped travel times on each link simultaneously, now the consideration of the travel time from 200 bootstrapped was made independently. As shown in Figure 4.2, the example plot of average travel times from 200 bootstrapped was illustrated to compare with the 'true' travel time for the probe percentage of 2.5%, 5%, 10% and 20% respectively at traffic volume of 4000veh/h. The purpose of the illustration of these figures was to give a picture on how the average travel time given by probe vehicle varies at 200 bootstrapped at various probe percentage. These figures demonstrate the close similarity between bootstrap predicted average travel times and 'true' average travel times. It can be seen that the variation of average travel time for every links is wider when having percentage of probe to be 2.5% and this variation becomes lower as the percentage of probe vehicle increases. This is consistent with intuition, as sample sizes are smaller, and thus variation of the average travel time must also be larger. The expected variation is zero if the probe vehicle percentage is 100% where in other words, the average travel time given by probe vehicle will definitely replicate the 'true' average travel time when assuming all vehicles to be probes. However, in reality, these variations may vary as the traffic pattern is affected by several factors such as the characteristics of the drivers, even if the flows are the same.

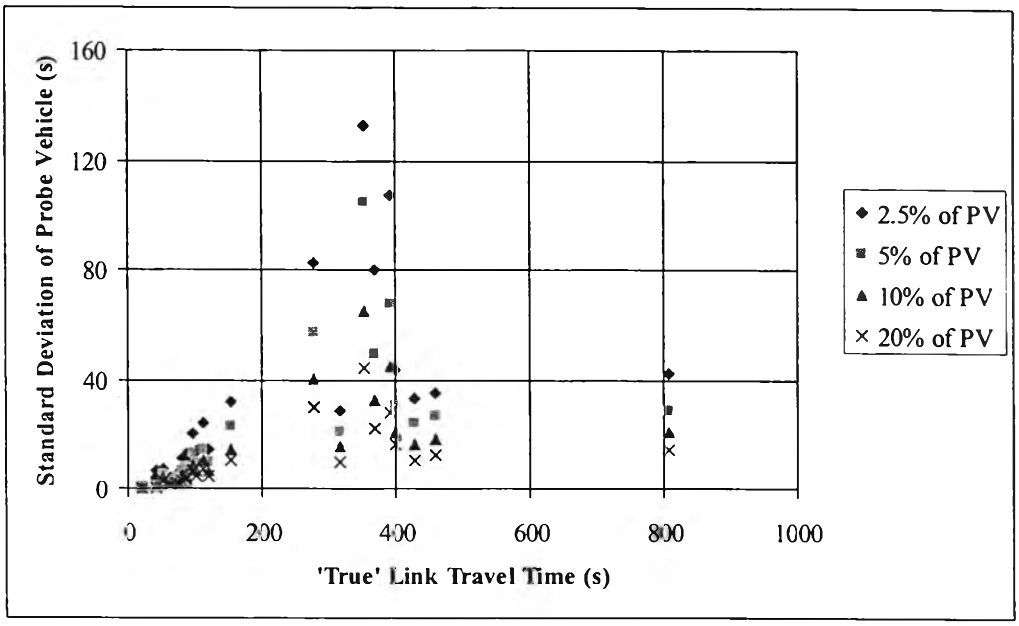


a) Traffic volume entering the network is 4000 veh/h

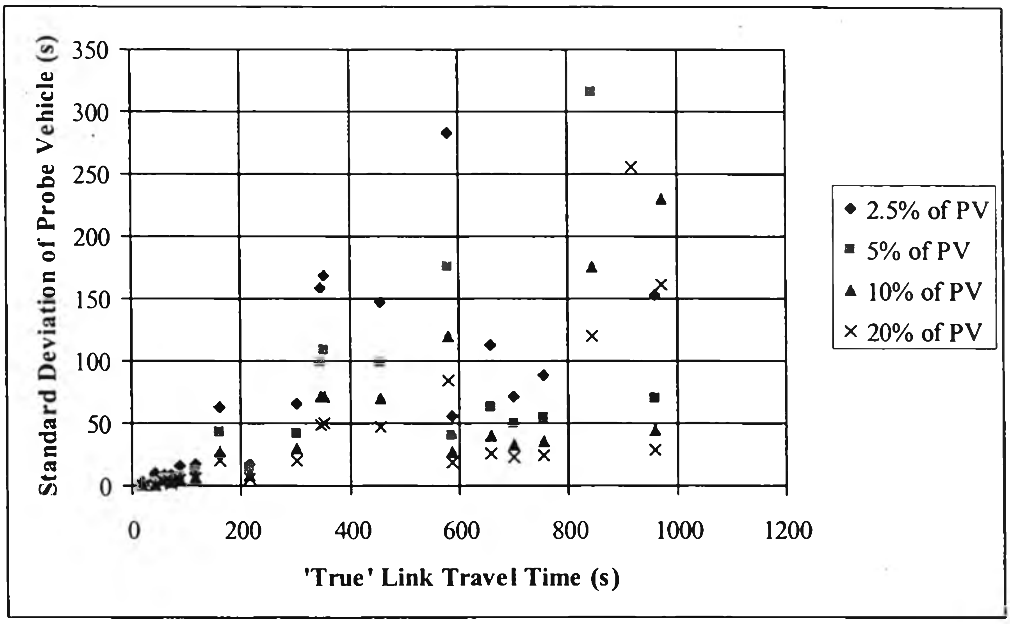


b) Traffic volume entering the network is 6000 veh/h

Figure 4.3 Standard deviation of travel times from probe vehicles compared to 'true' link travel times under various traffic conditions

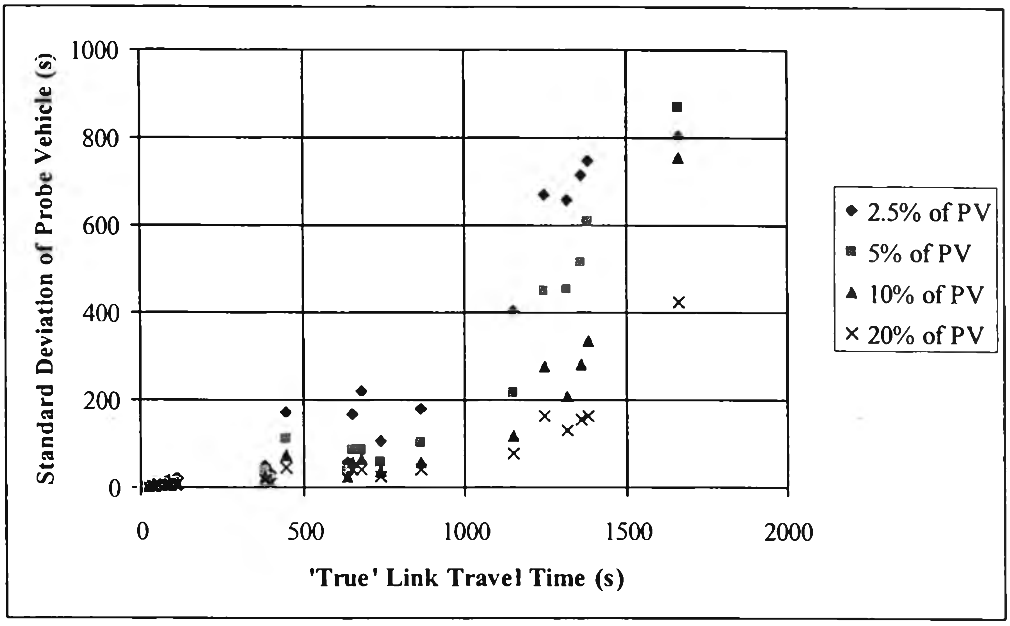


c) Traffic volume entering the network is 8000 veh/h



c) Traffic volume entering the network is 10000 veh/h

Figure 4.3 Standard deviation of travel times from probe vehicles compared to 'true' link travel times under various traffic conditions (continued)



e) Traffic volume entering the network is 12000 veh/h

Figure 4.3 Standard deviation of travel times from probe vehicles compared to 'true' link travel times under various traffic conditions (continued)

As mentioned earlier, the accuracy of the link travel time obtained from the probe vehicles is described by the magnitude of its standard deviation. It is a measure of the average value of the difference between probe vehicle travel time observations and the average value of the 'true' travel time observations. As shown in Figure 4.3, the variation of travel time standard deviation is larger as the link travel time increases. The standard deviation is a commonly used statistic to describe the variability of a random variable. The standard deviation increases with the increasing of probe vehicle average travel time. Heuristically, a larger number of probe reports imply a better accuracy in travel time prediction.

Comparing the number of probe vehicles, it is clear that the larger number of probes brings about the small deviations. In all traffic conditions, Figure 4.3 a) to e), the 2.5% percentage of probe vehicles give the largest variation in reported travel time, while the 20% probe vehicles gives the smallest.

Travel times of probe vehicles are less accurate with the increasing of traffic volume in the network. A telling statistic is the increase in standard deviation, indicating that there is a greater spread in the 'true' travel times when comparing it with the increasing of the traffic volume. The increases in standard deviation support the expectation when the traffic generally is cleared during green light and experiences no extraordinary delays. However, if the network has full loads of traffic, traffic can quickly back up into the preceding intersection and take multiple traffic signal cycles to dissipate the queue. In spite of that, the standard deviation primarily depends on the traffic flow and the traffic environment. In case of low traffic flow, most of the drivers can keep their desired speeds and the travel times vary as much as the speed preferences. On the other hand, higher flows make more vehicles hindered and the travel time to a higher extend determined by the surrounding traffic. However, comparing probe vehicles with all vehicles, the same tendency of average travel time and variation could be observed although there were few samples of probe vehicles.

To further examine the variance (standard deviation) of travel time from the probe vehicles at various traffic conditions, the coefficient of variation is considered. The coefficient of variation is basically the standardized form of standard deviation, and can be expressed as follow:

$$COV = SD/average \tag{2}$$

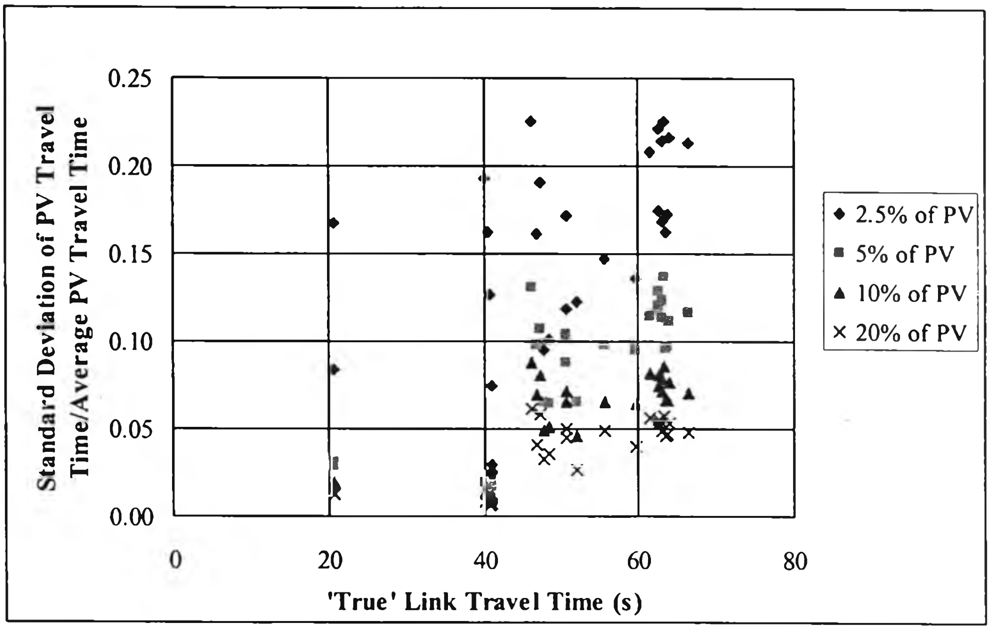
Where, COV = coefficient of variation

SD = standard deviation

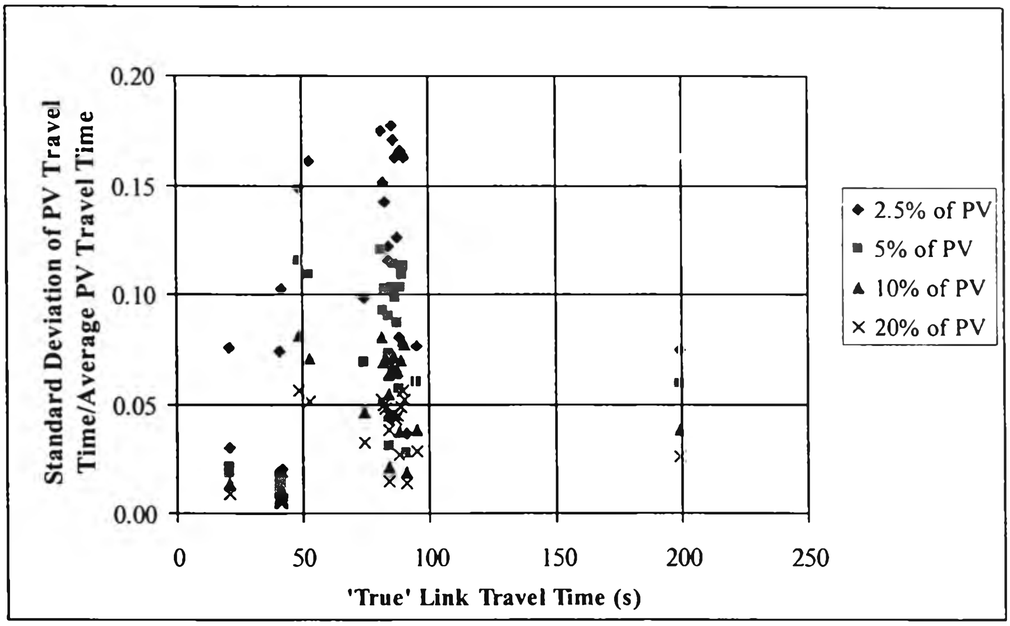
Average = average value

And the plots of COV are illustrated in Figure 4.4.

Considering Figure 4.4 a) through e), there is no pattern in the COV. It implies that the standard deviation (variation) of the travel times reported by probe vehicles is independent to the amount of link travel time.

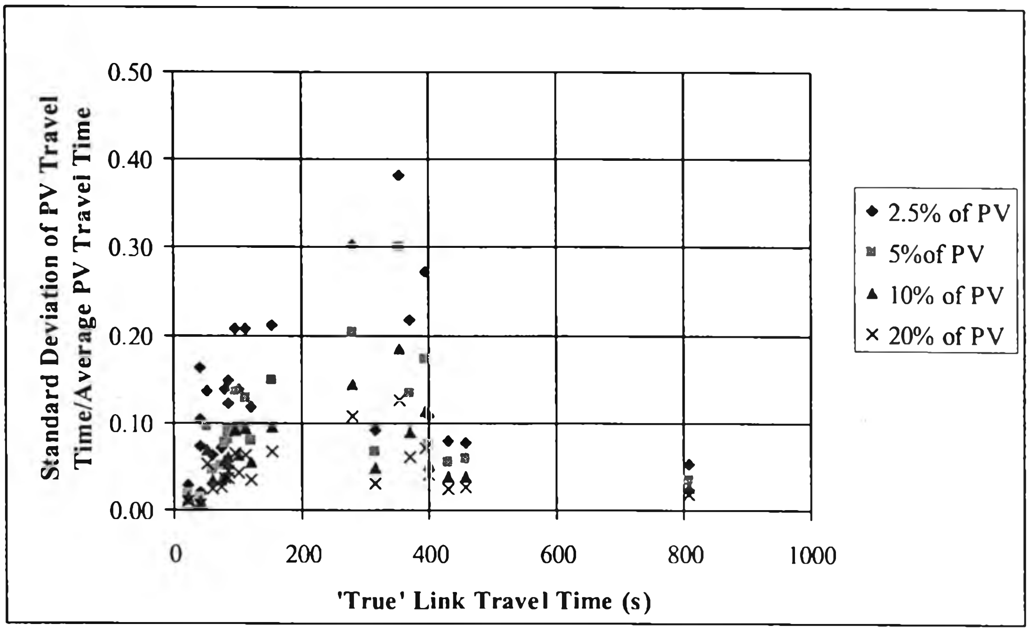


a) Traffic volume entering the network is 4000 veh/h

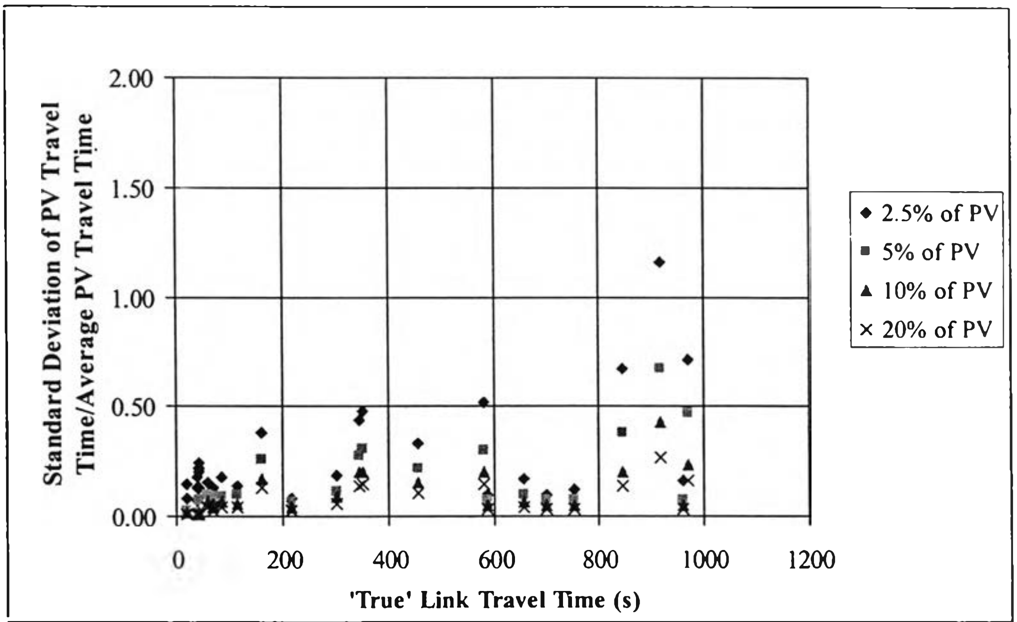


b) Traffic volume entering the network is 6000 veh/h

Figure 4.4 Coefficient of variation of the travel time estimates

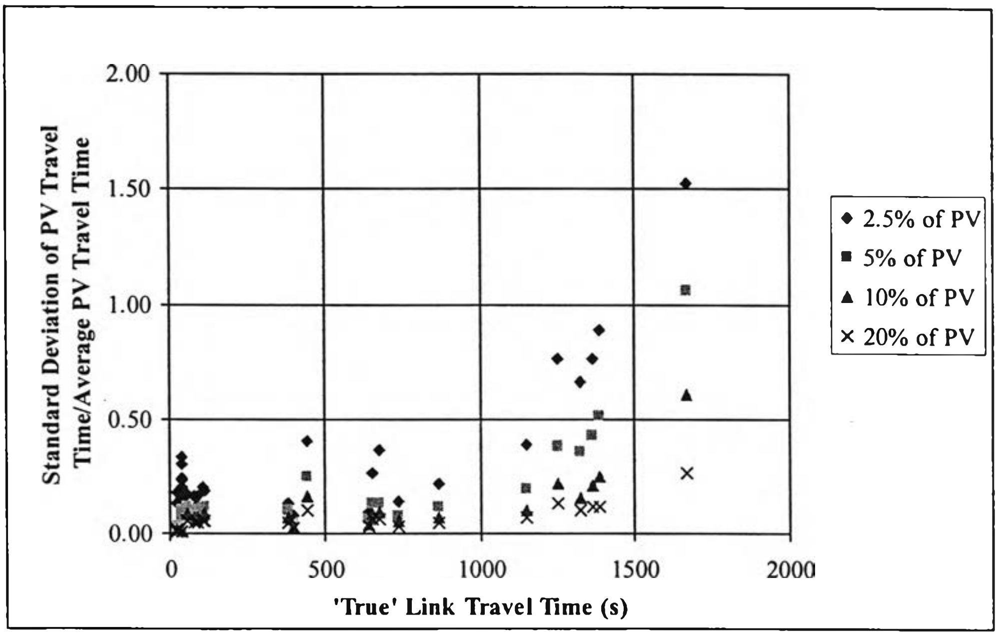


c) Traffic volume entering the network is 8000veh/h



d) Traffic volume entering the network is 10000veh/h

Figure 4.4 Coefficient of variation of the travel time estimates (continued)



e) Traffic volume entering the network is 12000veh/h

Figure 4.4 Coefficient of variation of the travel time estimates (continued)

Travel times given by probe vehicles were examined for each link. Three links were selected to exemplify the travel time results in different link types: entrance, exit, and internal links. Referring to Figure 3.4, link 4-5, 19-5 and 5-20 were selected for further travel time investigation. These 3 links represented the entrance, internal and exit link respectively. The summary of number of probe vehicles, the average travel time, standard deviation as well as the 'true' average travel time for each analyzed link is shown in Table 4.1.

Table 4.1 'True' and probe vehicle travel time comparison

a) Traffic volume entering the network is 4000 veh/h

Average for 'True'				Average for Probe Vehicle											
						No of PV			Travel Time (s)			Standard Deviation (s)			
Link	No of Vehicle	TT (s)	SD (s)	PV Percentage (%)	Link	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	SD TT
4-5	403	47.77	12.10	2.5	4-5	7	1	16	47.19	39.29	62.25	10.03	1.04	30.22	4.46
19-5	257	63.15	20.48		19-5	5	1	11	63.51	37.00	95.67	20.06	1.06	42.43	10.71
5-20	253	40.52	1.95		5-20	5	0	12	39.53	0.00	43.00	1.79	0.00	3.75	6.43
				5	4-5	14	6	23	47.47	40.19	56.42	11.19	1.33	22.15	3.10
					19-5	9	3	17	63.39	45.64	85.42	20.26	2.16	30.24	7.19
					5-20	9	3	16	40.56	39.00	42.25	1.89	0.29	3.18	0.69
				10	4-5	28	16	44	47.54	41.41	54.75	11.39	2.96	18.98	2.33
					19-5	18	7	29	62.91	48.65	76.16	20.22	7.93	25.97	4.93
					5-20	18	8	28	40.52	39.39	41.78	1.94	0.86	2.76	0.46
				20	4-5	55	37	74	47.70	43.95	52.61	11.75	5.54	16.69	1.56
					19-5	35	21	50	63.10	54.78	72.45	20.52	13.94	24.05	3.34
					5-20	35	21	47	40.53	39.70	41.24	1.94	1.26	2.42	0.31

b) Traffic volume entering the network is 6000 veh/h

Average for 'True'				Average for Probe Vehicle											
						No of PV			Travel Time (s)			Standard Deviation (s)			
Link	No of Vehicle	TT (s)	SD (s)	PV Percentage (%)	Link	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	SD TT
4-5	951	199.35	63.06	2.5	4-5	17	7	28	201.03	164.13	241.75	62.32	44.29	86.96	14.98
19-5	323	81.56	27.73		19-5	6	1	14	81.58	47.25	133.50	26.32	2.12	45.26	12.34
5-20	379	41.32	1.74		5-20	7	1	14	41.36	39.50	44.00	1.66	0.41	3.72	0.72
				5	4-5	34	19	54	199.73	161.85	224.27	61.89	48.68	78.17	11.78
					19-5	11	4	22	81.62	57.90	114.63	27.15	9.87	38.60	7.57
					5-20	13	6	24	41.37	39.85	42.93	1.68	0.42	2.88	0.50
				10	4-5	66	49	87	199.44	176.09	217.79	62.75	53.52	73.32	7.57
					19-5	22	8	35	81.26	65.06	100.06	27.37	16.80	33.80	5.64
					5-20	26	15	43	41.35	40.26	42.23	1.71	0.95	2.52	0.34
				20	4-5	128	104	156	199.10	184.03	210.76	62.95	54.48	70.05	5.10
					19-5	43	24	61	81.43	70.37	93.26	27.50	23.45	32.83	4.04
					5-20	51	36	67	41.32	40.78	41.94	1.74	1.27	2.19	0.20

Table 4.1 'True' and probe vehicle travel time comparison (continued)

c) Traffic volume entering the network is 8000 veh/h

Average for 'True'				Average for Probe Vehicle											
						No of PV			Travel Time (s)			Standard Deviation (s)			
Link	No of Vehicle	TT (s)	SD (s)	PV Percentage (%)	Link	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	SD TT
4-5	841	429.24	119.47	2.5	4-5	14	6	26	422.53	338.91	547.13	118.17	52.28	156.38	33.28
19-5	403	84.73	28.23		19-5	6	1	13	83.76	60.00	138.17	26.92	1.77	57.28	12.40
5-20	316	41.01	1.80		5-20	5	0	10	40.68	0.00	43.25	1.74	0.00	4.25	4.18
				5	4-5	27	16	43	425.46	366.00	492.31	119.62	91.98	146.44	23.75
					19-5	13	5	22	84.12	66.18	104.05	27.44	10.59	41.53	7.95
					5-20	10	2	18	41.00	39.00	42.33	1.79	0.35	3.32	0.62
				10	4-5	53	37	74	429.42	368.54	474.53	119.80	99.68	148.00	16.36
					19-5	26	14	37	84.39	72.50	96.71	27.99	19.88	37.37	5.11
					5-20	19	8	33	40.96	39.50	42.02	1.81	0.85	2.68	0.42
				20	4-5	102	82	122	429.68	400.25	457.69	119.70	104.72	132.33	10.63
					19-5	49	34	69	84.66	73.03	94.77	28.25	21.95	35.55	3.75
					5-20	38	25	53	40.99	40.23	41.71	1.81	1.31	2.33	0.28

d) Traffic volume entering the network is 10000 veh/h

Average for 'True'				Average for Probe Vehicle											
						No of PV			Travel Time (s)			Standard Deviation (s)			
Link	No of Vehicle	TT (s)	SD (s)	PV Percentage (%)	Link	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	SD TT
4-5	666	700.12	175.76	2.5	4-5	7	1	16	711.86	477.00	873.50	166.62	20.74	262.04	71.11
19-5	600	115.73	39.79		19-5	7	2	14	117.51	44.00	159.50	39.22	3.97	72.18	16.51
5-20	431	40.35	1.86		5-20	5	0	12	40.10	0.00	43.50	1.77	0.00	4.60	3.01
				5	4-5	14	6	26	705.89	560.25	847.50	172.24	50.87	237.18	50.62
					19-5	13	5	24	117.01	83.41	152.82	39.86	16.75	60.53	11.17
					5-20	9	3	18	40.38	38.75	42.14	1.82	0.69	3.12	0.64
				10	4-5	28	16	44	701.78	609.93	816.70	175.02	113.36	207.02	33.11
					19-5	26	11	37	116.62	92.85	140.72	39.78	28.16	51.68	6.95
					5-20	18	8	30	40.39	39.28	41.50	1.83	0.92	2.74	0.44
				20	4-5	55	37	75	700.13	618.62	762.35	175.18	150.46	202.88	23.14
					19-5	50	31	66	115.50	102.42	128.42	39.75	29.14	48.39	5.11
					5-20	36	19	52	40.39	39.45	41.12	1.85	1.27	2.33	0.30

Table 4.1 'True' and probe vehicle travel time comparison (continued)

e) Traffic volume entering the network is 12000 veh/h

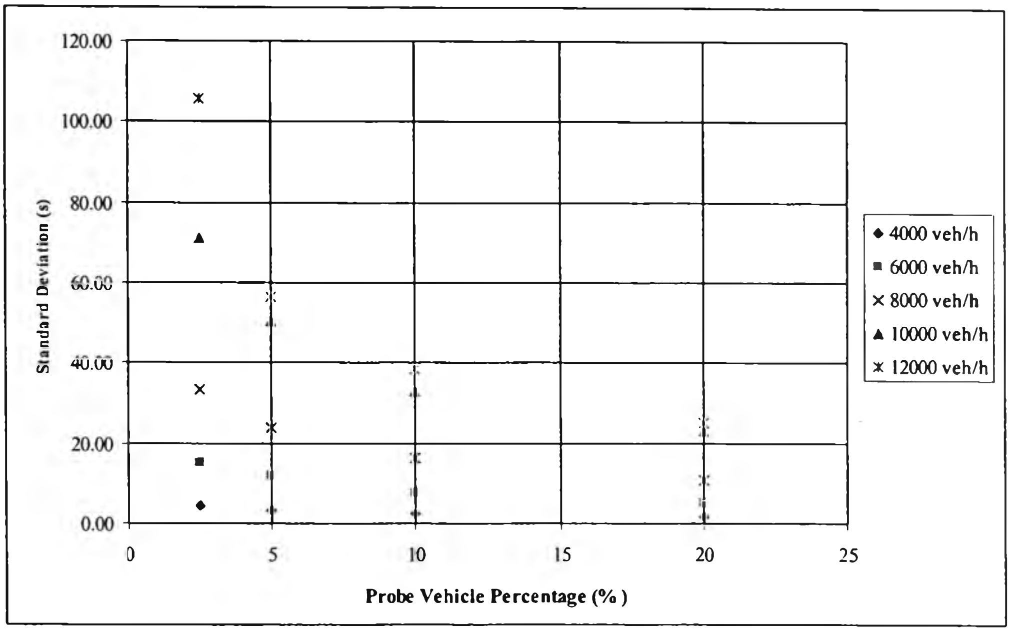
Average for 'True'				Average for Probe Vehicle											
						No of PV			Travel Time (s)			Standard Deviation (s)			
Link	No of Vehicle	TT (s)	SD (s)	PV Percentage (%)	Link	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	SD TT
4-5	639	740.59	173.73	2.5	4-5	5	0	12	738.91	0.00	1116.50	154.28	0.00	271.11	105.87
19-5	613	112.99	44.03		19-5	5	1	12	116.69	68.38	201.00	42.83	4.24	102.18	22.21
5-20	386	40.20	2.06		5-20	3	0	9	39.84	0.00	43.50	1.83	0.00	4.95	4.23
				5	4-5	11	3	19	738.29	585.06	872.43	167.96	49.61	243.99	56.46
					19-5	10	3	17	114.99	77.50	153.94	43.54	22.43	76.81	13.42
					5-20	6	1	15	40.20	37.50	42.17	2.02	0.29	3.89	0.95
				10	4-5	21	9	32	741.79	616.90	852.36	169.49	93.73	246.19	38.59
					19-5	20	11	34	114.17	83.30	143.13	43.59	25.89	62.28	9.34
					5-20	12	4	23	40.21	38.50	41.85	2.03	0.68	3.38	0.60
				20	4-5	41	25	59	740.70	671.98	804.94	170.41	132.25	213.07	25.08
					19-5	39	24	63	113.40	89.86	136.51	43.56	32.71	57.63	6.52
					5-20	24	14	37	40.20	39.13	41.44	2.05	1.37	2.66	0.41

As indicated in Table 4.1, as the percentage of probe vehicle increases, the number of probe vehicles also increases. By statistical point of view, as expected, the more probe vehicles to generate the data, the closer the average travel time to the 'true' value will be. Besides, the range for the standard deviation decreases as the percentage of probe vehicle increases. From Table 4.1 a) for the entrance link (link 4-5), with the total of 403 vehicles passing the link, the travel time gathered is 47.77 seconds. The exit link (link 19-5) has 257 vehicles with the average travel time of 67.15 seconds and the internal link (link 5-20) has 253 vehicles with the average travel time of 40.52 seconds. Comparing these 'true' travel times to the average travel times given by probe vehicles, probe vehicles could yield similar travel times to the 'true' travel times of these three links. The same trend could be observed at all traffic conditions in Table 4.1 a) to e). For example, on link 4-5 the average link travel times (from 200 bootstrap samples) are 47.19, 47.47, 47.54, and 47.70 seconds for 2.5%, 5%, 10% and 20% of probe vehicles respectively, compared with 47.77 seconds of true travel times. Another example is the travel time on link 19-5 in the case of entering volume of 12000 veh/h. Compared to 112.99 seconds of 'true' travel time, the bootstrap travel times are 116.69, 114.99, 114.17, and 113.40 seconds for 2.5%, 5%, 10%,and 20% respectively. However, it is reminded that the average link travel times from probes are the "average" of 200 bootstraps. The closer averages from probes to the "true" link travel time as the percentage of probes increases indicate the (maximum) ability of a percentage of probes can give, in term of travel time accuracy.

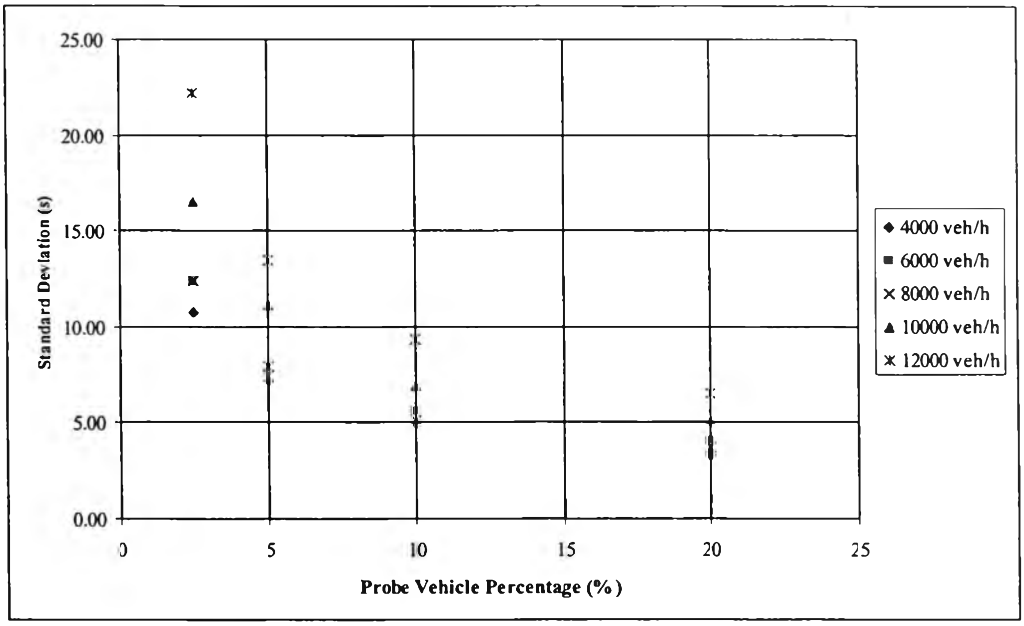
To look at the accuracy of individual reported travel time from a set of probe vehicles, each set of reported travel time from probe vehicles must be examined. The factors affecting the accuracy of the travel times are 1) number of probe vehicles on the links (number of reported travel times) and 2) average travel time values. From Table 4.1, the number of probe vehicles on each of three links varies, so does the average travel time values from them. Examining the number of probe vehicles on the links, it is eminent that the low percentage of probe vehicles creates the large variation of the number of travel time data for a link. For example, on link 5-20 at 2.5% probe vehicle and the entering volume of 4000 veh/h, the data indicate that one or more times the situation experiences no probe vehicle on the link. On link 4-5 and link 19-5, one or more times a single probe vehicle travels on the link(s). The fact that there is a variation in the number of probe vehicles (and travel time data) on a link would create the variation in average travel time. The discussion on the number of probes will be elaborated in the subsequent section.

The variation in average travel time value obtained from a set of probe vehicles can be examined from the range of travel time estimates (200 times) and the standard deviation of the travel time estimates. Table 4.1 summarizes the variation in terms of the minimum (Min) and maximum (Max) values of travel times obtained from each of 200 replications. The standard deviation of the travel time estimates is also displayed in the table as Standard deviation of the travel time estimates (SD TT). As expected, less number of travel time data (probe vehicles) could create a larger deviation in travel time estimates, compared to the 'true' travel time. To exemplify this, consider link 4-5 in Table 4.1 a) if probe vehicles account for 2.5% of the entire fleet, then the variation of travel time estimates obtained from these probes can be as large as +32 percent (-19 percent), or the range of the values constitutes approximately 50% of the travel time estimate. When the number of probe vehicles increases (20% probe vehicles), the range narrows down to approximately 6% of travel time estimates. A more simple consideration on the same point can be taken on the SD TT. The value of SD TT is smaller when the percentage of probes is higher. With the same example, travel time estimates on link 4-5 would vary considerable largely with 2.5% probe vehicles (4.46 seconds), and the variation becomes smaller to 1.56 with 20% probes.

It is good to know the reduction of standard deviation as the traffic volume increases. Figure 4.5 depicts the overview of the standard deviation as a function of probe vehicle percentage. As expected, the deviation of average travel time will be smaller given that the percentage of probe vehicle is higher. However, compared among the deviation of the average travel time for entrance, internal and exit link, as illustrated, it is seen that the deviation of the average travel time becomes the highest in the entrance link and the smallest in the exit link. For the entrance link, the travel time increases rapidly as the traffic volume increases. One of the reasons is that, at the entrance link, after the warming up period of 15 minutes, the network experiences waiting queues to pass the intersection. Some of the vehicles need to wait for a few cycle times in order to pass through that intersection. Besides, some vehicles start to queue back into the release zones. Therefore, these situations have caused the travel time to increase which may lead to the large dispersion of standard deviation. Conversely, in the internal link, we can consider it as a "stable" link. Since it is in the middle of the network, it may only get effects from the vehicles queuing at the intersection. Consequently, the deviation of the travel time is not as much as in the entrance link. For the exit link, the deviation of the travel time is varying not much even though the traffic volume increased. Free flow traffic can be expected in the exit link because all vehicles pass through all the intersections and the vehicle do not have any obstruction at the end of the network which can delay them to reach to their destination.

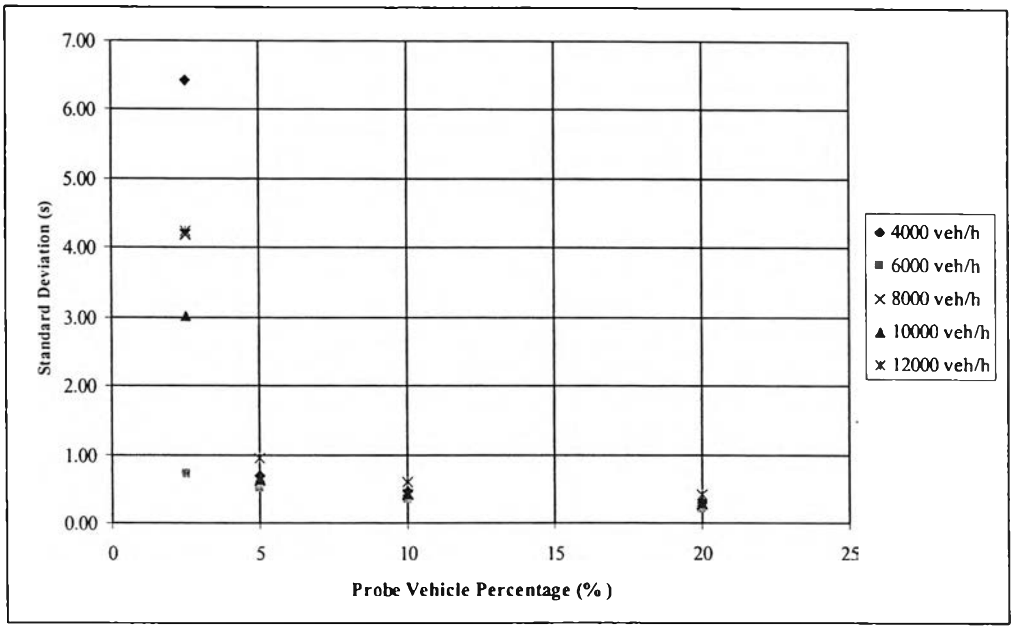


a) Link 4-5 (Entrance link)



b) Link 19-5 (Internal link)

Figure 4.5 Standard deviation as a function of probe vehicle percentage



c) Link 5-20 (Exit link)

Figure 4.5 Standard deviation as a function of probe vehicle percentage (continued)

Figure 4.5 a) shows the standard deviation of the travel time on an entrance link at various traffic congestion levels. The standard deviation of the travel time is lower when the congestion is lower. For example, the standard deviation of the travel time at low congestion condition (4000 veh/h entering flow) is 5 seconds while the standard deviation at high congestion is 105 seconds. Keep in mind that all congestion conditions have a variety of travel time values.

Figure 4.5 a) also implies the benefits that could be gained when the number of probe vehicles increases. Considering the low congestion condition, the increase in percentage of probe vehicles from 2.5 to 5 percent could bring about a great reduction in travel time variation (obtained from probes) by half (from 105 seconds to 58 seconds), at 12000 veh/h network volume. This trend in the reduction in standard deviation is observed in all traffic congestion level. When the percentage of probe is high, the percentage of reduction in travel time variation is less.

Considering Figure 4.5 b), link 19-5, an internal link, has different standard deviation from link 4-5, entrance link, in Figure 4.5 a). So does link 5-20, exit link, in Figure 4.5 c). The main difference among three links is that the exit link does not have as great an effect from higher congestion as the entrance and internal link. The exit link normally does not have a big variation in travel time. The variation basically comes from a variety of desired speed of individuals. Unlike the exit link, the entrance link travel time relies heavily on delay at the signal at downstream intersection. Thus, the absolute values of standard deviations of travel times should not be compared, rather the trend in reduction in standard deviation when the percentage of probes increases must be considered.

Figure 4.5 a)-c) yield the same conclusion that the reliability of travel time could be gained if the percentage of probes is increased. This is eminent in the case of the increase in number of probes at the low percentage of probe vehicles (e.g. 2.5 to 5%). The benefits are minimal on the exit link where the travel time variation is normally low.

4.2 ROUTE TRAVEL TIME

Route Travel Time is the travel time of a vehicle traversing from one end the other end of corridor. In this hypothetical network, the route travel times were determined from the travel from zone 1 to zone 2 (Eastbound, EB) and from the travel from zone 2 to zone 1 (Westbound, WB). Vehicles must proceed on six corridor links. Since some vehicles were assumed to be probe vehicles (PV), some of the O-D flow between zone 1 and zone 2 were selected to be probe vehicles and then their probe vehicle travel times were calculated. Note that the "true" route travel time is defined as the travel time calculated from *all* vehicles traversing on the O-D route. This resembles the travel time data obtained from vehicles when they complete the entire route. It is analogous to the 100% PV market penetration case. The data ignore travel times on links on the route.

At each market penetration case (% PV), some vehicles were assumed to be PV. Then travel time determination was basically the average of travel times from all PVs. It is noted that the number of PV is random, based on %PV. It was natural that travel times from PVs varied due to their traffic and driving situations, thus variation in travel times could be seen (as indicated by range and standard deviation of travel times). Since the PVs were resampling, the bootstrap of PVs resulted in various number of PV in the same %PV case. In this study, the PVs were randomly selected for 200 times, producing also various travel times and their variations. While the link travel time consideration focused on the relationship between the travel time reported by probe vehicles and the 'true' travel time and the analysis showed the ability of each percentage of probes in giving accurate results, this section examined the route travel time, which was the total time traveling on several links. The closeness of the travel time by probe vehicles could be seen from bootstrap results, while the accuracy of individual reported travel time could be seen from standard deviations of the travel times.

In theory, the route travel time can be determined from at least two methods; the direct travel time measurement from (probe) vehicles that traverse on the entire route, or the estimation by the summation of link travel times (of the links on the route). The link travel times can be estimated from all vehicles on the links, or from selected vehicles that proceed in the considered directions. In the hypothetical network, two routes are under investigation. The East Bound route and the West Bound route are described previously in Section 3.2 . However, only the East Bound route was selected for the investigation.

In this following section, the number of probe vehicles that could be obtained from various percentages of probes was also examined. Route travel time was compared for East Bound route. Table 4.2 shows the results for average 'true' travel time as compared to the probe vehicle travel time while Table 4.3 shows the results for link by link basis.

Table 4.2 The average 'true' travel as compared to probe vehicle

	"True" Travel Time			PV Travel Time										
Traffic Volume (Veh/h)	**Avg TT (min)**	**Total Vehicle**	**SD**	**PV Percentage (%)**	**Avg TT (min)**			**No of PV**			**SD (min)**			**SD of Travel Time**
					Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	
4000	4.17	100	0.40	2.5	3.56	0.00	5.38	2	0	6	0.27	0.00	0.91	1.47
				5	4.11	0.00	4.97	4	0	9	0.35	0.00	1.05	0.56
				10	4.18	3.78	4.55	7	1	14	0.38	0.13	0.78	0.15
				20	4.17	3.94	4.46	14	5	23	0.39	0.20	0.62	0.10
6000	7.54	124	0.73	2.5	6.65	0.00	9.06	2	0	6	0.53	0.00	2.14	2.51
				5	7.43	0.00	8.94	4	0	12	0.65	0.00	1.70	1.14
				10	7.58	6.79	8.77	9	2	19	0.71	0.27	1.30	0.26
				20	7.58	6.99	8.05	17	8	30	0.72	0.46	1.06	0.17
8000	21.93	91	2.24	2.5	17.23	0.00	26.60	1	0	4	1.22	0.00	4.81	9.17
				5	20.87	0.00	26.23	3	0	7	1.83	0.00	4.73	4.76
				10	21.95	19.27	25.34	6	1	13	2.11	0.12	4.21	1.08
				20	21.92	20.14	24.34	11	4	20	2.18	0.93	3.74	0.77
10000	31.07	15	1.27	2.5	4.68	0.00	32.97	1	0	2	0.03	0.00	2.07	11.19
				5	8.90	0.00	32.97	1	0	2	0.05	0.00	2.07	14.15
				10	14.93	0.00	32.97	1	0	3	0.27	0.00	2.36	15.59
				20	22.70	0.00	32.97	2	0	5	0.64	0.00	2.36	13.87
12000	32.39	15	1.53	2.5	10.96	0.00	36.58	1	0	4	0.11	0.00	3.91	15.35
				5	16.25	0.00	36.58	1	0	4	0.34	0.00	5.13	16.18
				10	23.86	0.00	36.58	2	0	6	1.07	0.00	6.02	14.27
				20	29.30	0.00	36.58	3	0	9	1.75	0.00	4.52	9.65

Table 4.3 The link by link average travel time and standard deviation

Traffic Volume (Veh/h)	Link	No of Vehicle	TT (min)	SD (min)
4000	2-3	301	0.79	0.31
	3-4	397	0.87	0.21
	4-5	403	0.8	0.20
	5-6	424	0.81	0.21
	6-7	349	0.78	0.24
	7-8	243	0.34	0.03
	Total	**2117**	**4.38**	
6000	2-3	615	0.87	0.41
	3-4	882	1.51	0.22
	4-5	951	3.32	1.05
	5-6	881	1.4	0.61
	6-7	694	1.4	0.22
	7-8	391	0.35	0.02
	Total	4414	8.86	
8000	2-3	499	4.64	3.65
	3-4	591	13.4	2.19
	4-5	841	7.15	1.99
	5-6	821	5.27	1.76
	6-7	782	1.21	0.30
	7-8	432	0.35	0.03
	Total	**3966**	**32.1**	
10000	2-3	439	10.9	3.01
	3-4	465	16	3.61
	4-5	666	11.7	2.93
	5-6	874	3.59	0.91
	6-7	785	1.17	0.24
	7-8	374	0.35	0.02
	Total	**3603**	**43.7**	
12000	2-3	390	11.3	3.70
	3-4	515	14.4	4.29
	4-5	639	12.3	2.90
	5-6	1020	1.49	0.53
	6-7	908	1.47	0.56
	7-8	457	0.34	0.02
	Total	**3929**	**41.4**	

Table 4.2 has very similar consideration as Table 4.1, but examining the route, rather than link, travel time. The travel time data reported by probe vehicles for the entire route yield similar conclusion as the link travel time reports shown in Table 4.1. The bootstraps of route travel time show that on average the ability of probe vehicles in giving accurate results. Again, as the sample of probe vehicle increases, the travel time is also close to the 'true' one. As refer to Table 4.2, for the traffic volume 4000veh/h, there were 100 vehicles traveling along the route which gave the 'true' travel time of 4.17 minutes. For traffic volume 6000ve/h, the 'true' travel time was 7.54 minutes with 124 vehicles traveled along the route. The average travel times given by probe vehicle for the traffic volume 4000veh/h are 3.56, 4.11, 4.18 and 4.17 minutes respectively for 2.5%, 5%, 10% and 20% of probes. As compared to the probe average travel time, still, when having probe vehicle of 20% percent with the average number of probe vehicle 14, the travel time gathered is almost the same as the 'true' travel time. Similar the case of link travel time, it is reminded that the average route travel times from probes are the "average" of 200 bootstraps. The closer averages from probes to the "true" link travel time as the percentage of probes increases indicate the (maximum) ability of a percentage of probes can give, in term of travel time accuracy.

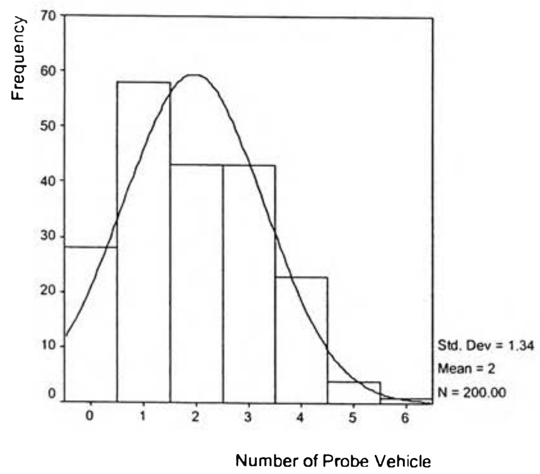
When comparing this to the average travel time from the summation of link, as indicate in Table 4.3, it is reasonable because the link summation gives travel time of 4.38 minutes with the total number of vehicle traveled is about 2117. However, taking into consideration of these two sources of average travel time, the probe vehicle based gives a better result which is closer to the 'true' average travel time. Note that the traffic volume along the link toward the end of the route is consistent but at the end of the link, fewer vehicles completed the journey to zone 2. Most of it turned at the intersection which consequently gave a low number of vehicles at the end of the link. Since the left turn, right turn and through movement might experience a larger difference in their time to traverse the intersection, it made more sense to measure the travel time based on route (a corridor consisting of multiple links) rather than link. The difference in the prediction performance can be attributed to the variance of the probe vehicle reports. Adding average link travel time together certainly propagates the variance of the total travel time of the route. As a result, with larger variance on travel time estimates, the link based average travel times are more likely to produce less satisfactory results from the 'true' value.

The same pattern of average travel time could be observed with traffic volume of 6000veh/h as well as at 8000veh/h. However, as the traffic volume increased which also lead to the diminishing of level of service, the performance of both, the probe vehicle travel time and summation of the link are completely poor. The results are not readily apparent. Theoretically, it is expected that the prediction of the average travel times from probe vehicle gets more and more close to the "true" value with the increasing number probes. It is also expected that in the case of having a high traffic volume, more probe vehicle can be sample. However, the results based on the analysis do not display such a trend, as observed from Table 4.2. Instead, it can be seen that the number of vehicles that can be passed throughout the route are relatively decreasing as the traffic volume increased which also depict the increasing of the congestion level. The reason is two folds. First, for the purpose of analysis, average travel time was taken during 45 minutes of simulation which means after consideration of the 'warming up' time. By taking into account the data that had been generated, when all vehicles were just entering the network, the congestion was not built up yet. However, after the network was fully loaded, the queue was assembled and continuously occurred until the end of the simulation time. Therefore, vehicles were still in the network during the simulation stop. The discrepancy in the travel times on this route was also as a result of the fact that the intersections were simulated using pre-timed control with split phases, with the first priority given to the vehicles traveling from zone 1 to zone 2 (the analysis route). Since this route was a major route with high traffic volume, the resulting of the signal setting leaded to the queue and the queue did not dissipate until the end of the simulation. Thus, vehicles spilled back into the release zones. Consequently, only few vehicles could complete the travel to their destination.

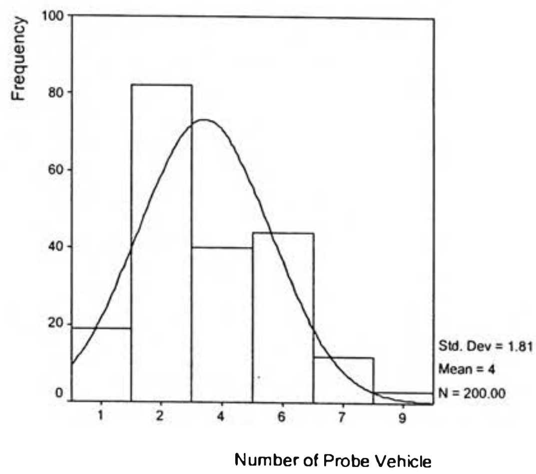
The accuracy of these probe travel times with respect to the "true" average travel time plays a very important role here. As mentioned earlier, probe vehicles were sampled from the vehicle population in the simulation output. Particularly, for the heavy flow condition more probe vehicles are needed to ensure accurate estimates. However, due to circumstances mentioned above, the resulting average travel time from probe vehicles would result in large variation from the "true" value. In the analysis, the standard deviation of the 'bootstrap' was consistent for traffic volume 4000, 6000 and 8000veh/h where it was decreased as the sample of probes increased. Nevertheless, an unpredictable change was observed in traffic volume 10000 and 12000 veh/h. The number of probes to gather the data decreased as the volume increased and there was a huge range of average on travel time on each link compared to 'true' value as in indicates Table 4.3. It was generally assumed that route was the addition of the travel times on its consisting links. However, for a probe based data collection system in which the number of reports was rather limited, this link based estimation might not be reliable. Again, the same reason is raised where at the intersection where the probe needed to make a turn, it might experience serious backup and consequently it might report a much longer travel time than the "true" average travel time on the links.

As seen in the results on average, tests on all traffic networks with different congestion levels show similar results where travel time averages by probe vehicle give rather close to the 'true' values. Therefore, it can be concluded that route travel time prediction would produce a better result than the link by link method. While the increasing of probe vehicle percentage could somewhat improve the prediction accuracy for route travel time, this improvement is not significant where we could not expect the increasing of number probe vehicle in the increasing of traffic volume (high congested network). Therefore, by having link by link average travel times, the observation could provide similar estimates of travel time and might shed some light on the implementation of travel time information system without a high market penetration of such devices.

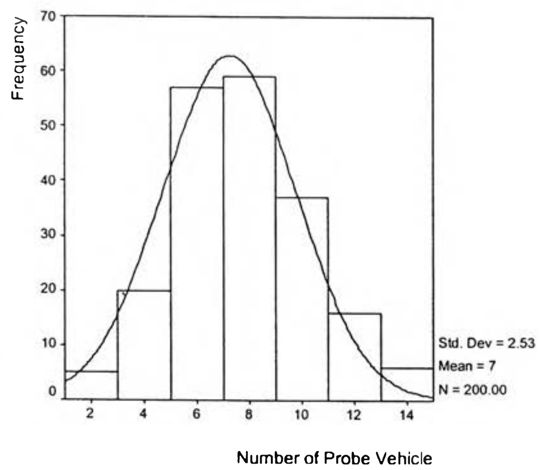
To elaborate in the chance of having a number of probe vehicles at a percentage of probe vehicles (market penetration), the detailed analysis was carried out from bootstrap data. At each percentage of probes available, it was a probability that a number of probes would travel on the route. The implications from the analysis of 200 times of bootstraps reveal that the number of probes that could give the travel time information varies. The amount of the probe vehicles not only depends on the percentage of probes, but also the amount of traffic (volumes) on the links. Figure 4.6 illustrates the probability of having probes on the route at various traffic and percentage of probes conditions.



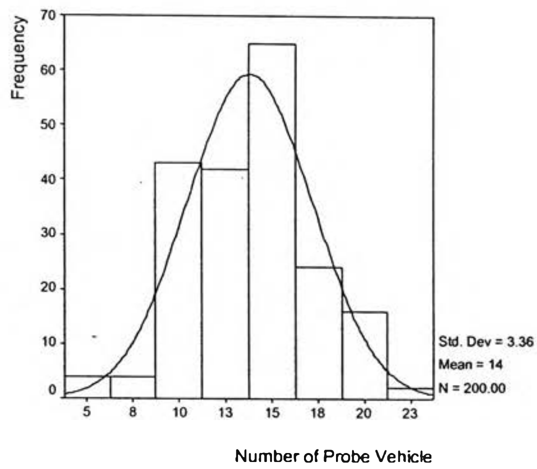
2.5% of Probe Vehicle



5% of Probe Vehicle

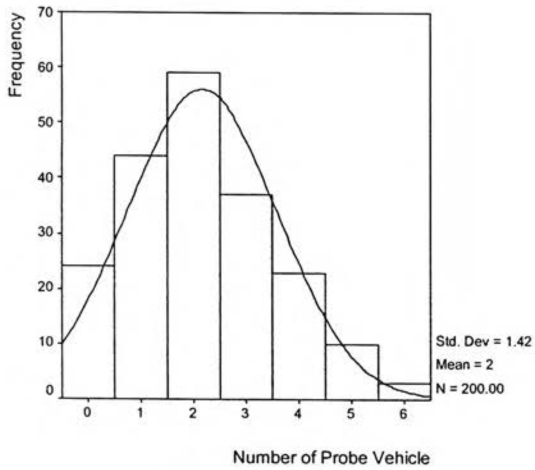


10% of Probe Vehicle

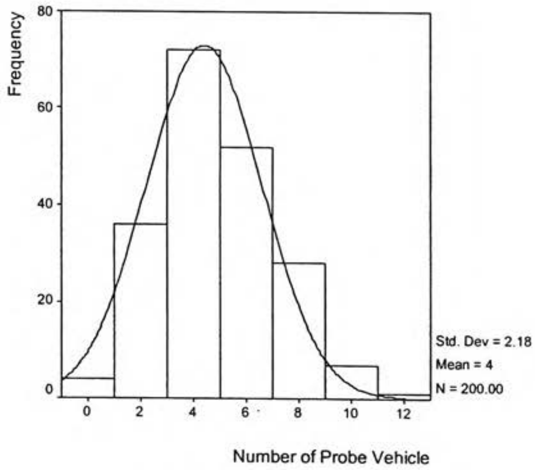


20% of Probe Vehicle

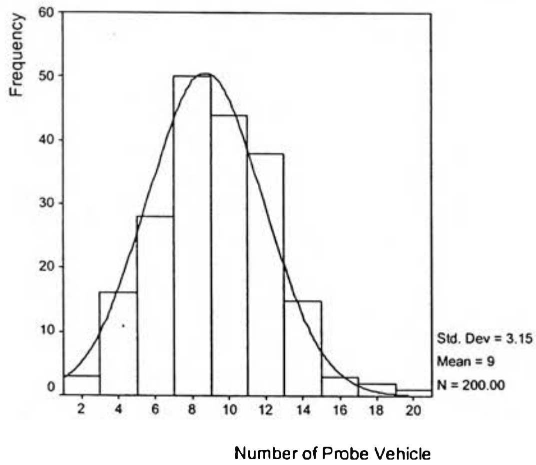
a) Traffic volume entering the network is 4000 veh/h



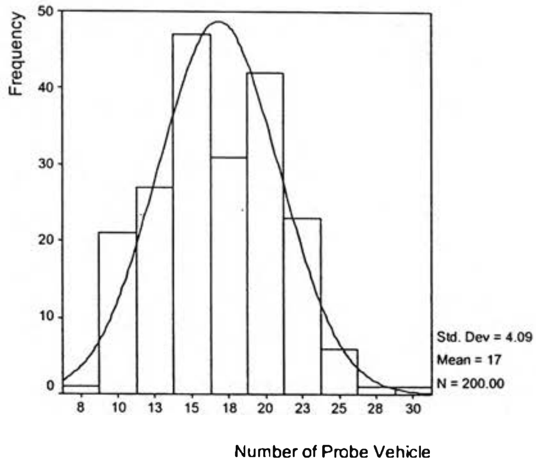
2.5% of Probe Vehicle



5% of Probe Vehicle

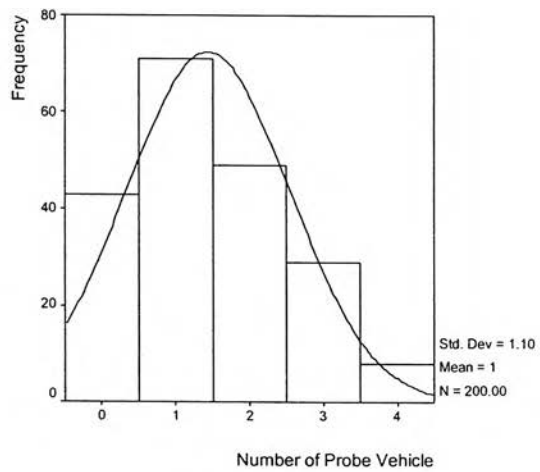


10% of Probe Vehicle

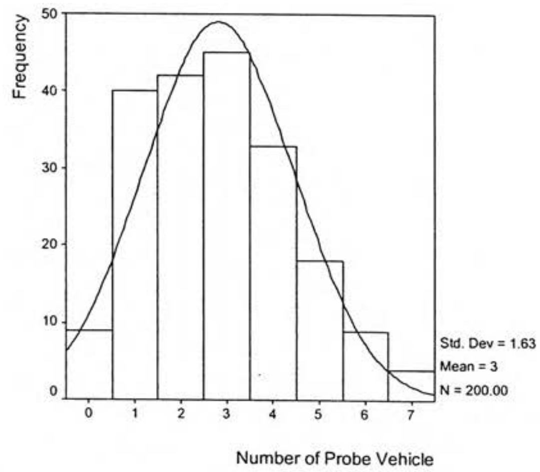


20% of Probe Vehicle

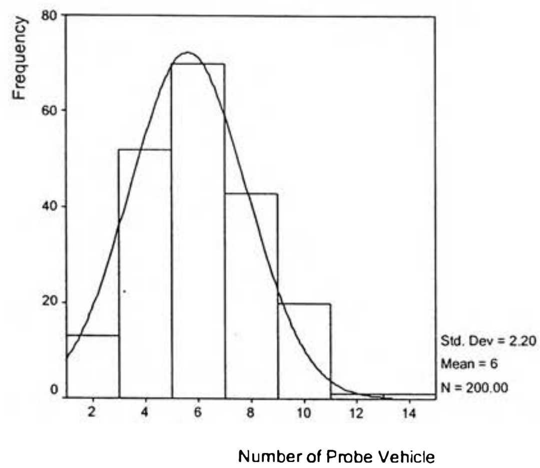
b) Traffic volume entering the network is 6000 veh/h



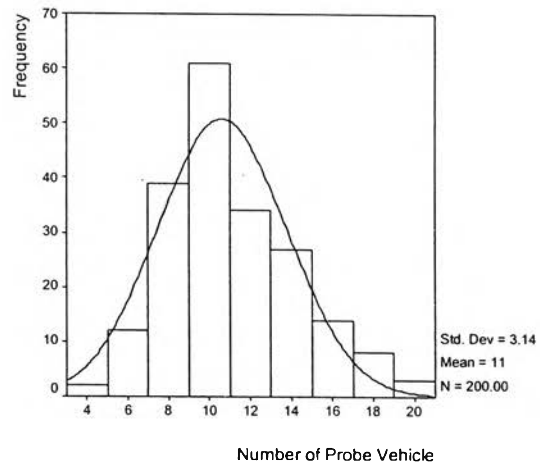
2.5% of Probe Vehicle



5% of Probe Vehicle

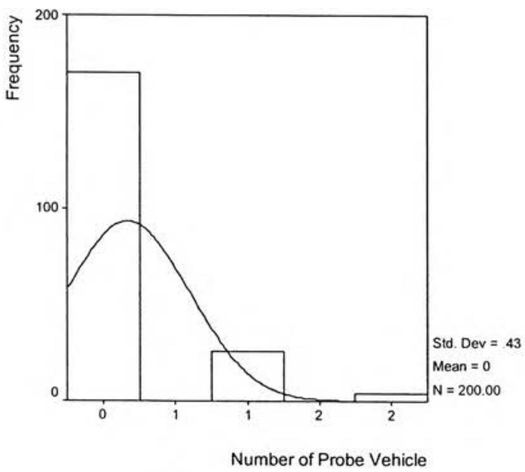


10% of Probe Vehicle

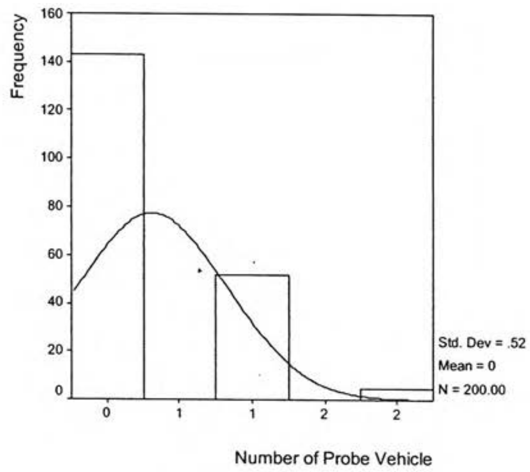


20% of Probe Vehicle

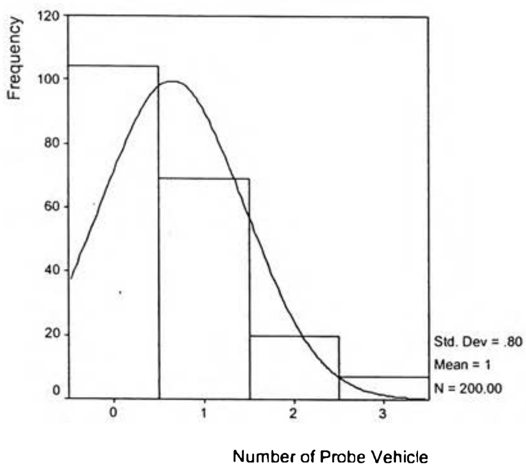
c) Traffic volume entering the network is 8000 veh/h



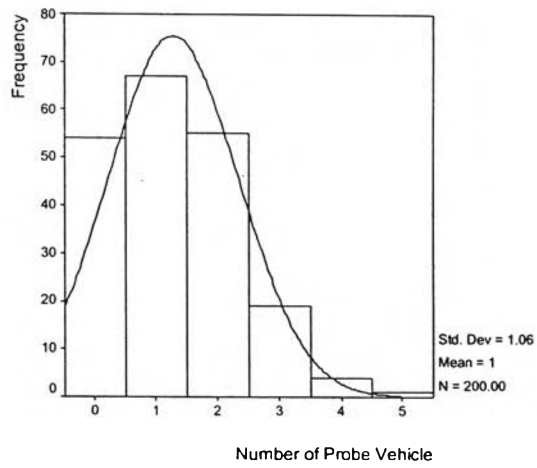
2.5% of Probe Vehicle



5% of Probe Vehicle

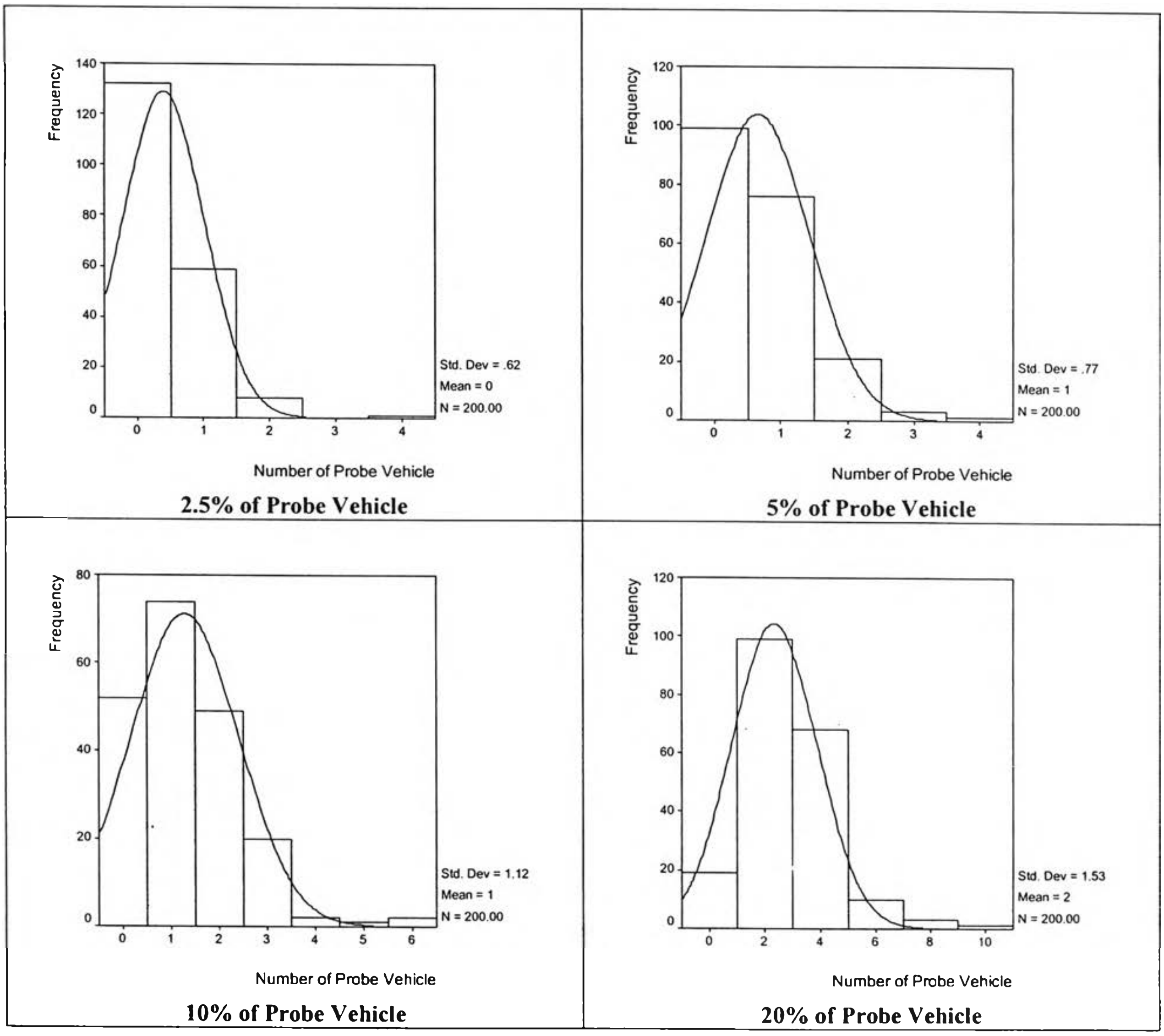


10% of Probe Vehicle



20% of Probe Vehicle

d) Traffic volume entering the network is 10000 veh/h



e) Traffic volume entering the network is 12000 veh/h

Figure 4.6 The distribution of number of probe vehicle on the route

Figure 4.6 a)-e) depict the distribution of number of probes on the route. These figures were constructed from the 200 times of bootstraps, representing various probabilities of having an amount of probe vehicles in a given percentage of probes (market penetration) available. The bar plot in each graph represents the frequency distribution from the data, while the curve plot represents the best-fitted known curve. The curve also represents the statistical trend of the distribution. By interpreting the curves, Figure 4.6 a) through e) yield similar conclusion: as the percentage of probes increases, the chance of having more number of vehicles increases. This can also be seen from the average (mean) number of probes in each case. However, this interpretation is just valid in the case of low traffic levels.

Perhaps the crucial consideration is on the percentage of times that there is no travel time data (probe vehicles) in the selected route. This can be observed from the first bar (of the raw data) reading of the curve data in each graph. For example, the probability of having no travel time data for the case of entering volume of 4000 veh/h (Figure 4.6 a) and 2.5% probe is approximately 20 percent. The probability decreases to 10, 8, and 2 percent at 5%, 10% and 20% probe condition, respectively. When the traffic volume increased such as in 10000veh/n or 12000veh/h, where level of service F could be found, a different trend is observed. The data cannot be retrieved in the sections of the route network which perform particularly poorly given that the level of congestion is high. The problem is that the probe vehicle gets trapped in the queue and may not be able to complete the trip which may lead to the situation where no data will be produced.

From these figures, an indication can be made where a higher variation of probe vehicle travel time occurs when there is no probe vehicle to provide the travel time data for each probe vehicle sample. Besides, as expected, the higher percentage of probe vehicle gives quantitative point of view of the accuracy of average travel time where the travel time provided is almost similar to the 'true' value. However, the same trend of average travel time and variation could be observed, although there are few samples of probe vehicles. These interpretations can be made in low traffic levels where no congestion is built until the end of the simulation. As shown in Figure 4.6 a) to e), a high difference could be observed when comparing the probability of having no probes in traffic volume 4000veh/h and 12000 veh/h. These differences are dramatic and speak clearly to the variation in the probe vehicle average travel time where indirectly lead to the level of quality in the travel time estimation.

Further analysis was done to explore the impacts of having bootstrap to get the variation of travel time from probe vehicle. In this part, the impacts of travel time generated by probe vehicle from one large sample to the route were analyzed. As in the case of link travel time, a simple, straightforward bootstrap method was applied to develop approximate confidence intervals for average travel time from a probe vehicle for different percentage of probe vehicles in different traffic situations. Confidence intervals generated by this method were based on the same source of variation as travel time estimates currently generated from Paramics V5 Microsimulation Software.

Table 4.4 The 95% confidence interval for the average route probe vehicle travel time for traffic volume 4000 veh/h, 6000 veh/h, 8000 veh/h, 10000 veh/h, 12000 veh/h (interval bounds are in minutes)

		4000 veh/h	6000 veh/h	8000 veh/h	10000 veh/h	12000 veh/h
2.5% of PV	Lower Bound	3.3539	6.2958	15.9481	3.1236	8.8203
	Upper Bound	3.7633	6.9968	18.5067	6.2436	13.1006
5% of PV	Lower Bound	4.0359	7.2714	20.2101	6.9268	13.9959
	Upper Bound	4.1921	7.5885	21.5372	10.8723	18.5094
10% of PV	Lower Bound	4.1594	7.5474	21.7981	12.7512	21.8721
	Upper Bound	4.2006	7.6210	22.0992	17.1002	25.8510
20% of PV	Lower Bound	4.1593	7.5538	21.8122	20.7646	27.9543
	Upper Bound	4.1863	7.6012	22.0271	24.6321	30.6453

Table 4.5 The average bootstrap travel time and standard deviation

Traffic Volume (veh/h)		2.5% of PV	5% of PV	10% of PV	20% of PV
4000	Avg TT (min)	3.56	4.11	4.18	4.17
	SD	1.47	0.56	0.15	0.10
6000	Avg TT (min)	6.65	7.43	7.58	7.58
	SD	2.51	1.14	0.26	0.17
8000	Avg TT (min)	17.23	20.87	21.95	21.92
	SD	9.17	4.76	1.08	0.77
10000	Avg TT (min)	4.68	8.90	14.93	22.70
	SD	11.19	14.15	15.59	13.87
12000	Avg TT (min)	10.96	16.25	23.86	29.30
	SD	15.35	16.18	14.27	9.65

As shown in Table 4.4, by including confidence intervals in reports of probe vehicle results, analysts can have comparable information for both, the probe vehicle average travel time to the 'true' one, which later on can provide understanding what confidence on the travel time value can be gained from probe vehicle data for travel time information. Table 4.5 shows the average travel time of 200 bootstrapped results.The 'true' average travel time for the route are 4.17 minute, 7.54, 21.93 , 31.07 and 32.39 minutes respectively for traffic volume 4000, 6000, 8000, 10000 and 12000veh/h. In general, the bootstrap distribution is centered somewhat below the 'true' value estimation. For example, for traffic volume 4000veh/h, the 'true' average travel time is 4.17 minutes. The 200 bootstrapped yield the average travel times of 4.17 minutes with the standard deviation of 0.10 minute. The 95% confidence intervals cover the range of 4.16 and 4.19 minutes or about plus or minus 0.24 percent from the 'true' average travel time at the 20% percentage of probe vehicle. The confidence interval for 10% of probe vehicles using the same traffic volume range between 4.04 and 4.19 minutes or about plus or minus 3.24 percent from the 'true' average travel time. The range is wider while decreasing the probe vehicle percentage.

For traffic volume of 12000veh/h, the confidence interval yields a span range between 27.95 and 30.65 minutes. In this example, the lower bound for the 95% confidence intervals is roughly 15.90 percent below the 'true' average travel time estimation, respectively. The upper bounds for the same intervals are about 5.68 percent higher. These results were obtained from a sample of 20% of probe vehicles.

Adding approximate confidence intervals to the average travel time from the bootstrap is worthwhile in addition to the standard analysis from previous sections. Showing a likely range of impact estimates provides a great deal more information than the typical available. In general, the 95% confidence interval ranges about plus or minus 20% from the 'true' value for percentage of probe vehicle 20%. However, the confidence intervals range is very huge for the network with high congestion.

4.3 CASE STUDY (ORIGIN DESTINATION (O-D) TRAVEL TIME)

In order to illustrate a simulation analysis of the potential of vehicle probes to provide O-D information, an example network was required. However, in order to provide meaningful results, the network characteristics were developed to be as realistic and objective as possible. Therefore, to illustrate the potential improvement in travel time estimates, a real network configuration consisting of some part of the Chulalongkorn University surrounding area was selected. The main focus was not replicating the real world in terms of the exact traffic volume and signal configuration, rather the replication the traffic conditions and also the road network configuration to see whether the data generated were valuable or not to be apply in real world. This application could be extensions of the link and route travel time approach application described in the previous sections but in this time, the proportion of vehicles traveling between each Origin-Destination (OD) pair was defined using real network configuration. The distribution of vehicles was then dependent on the proportion of vehicles on each OD path where the turning movements of these vehicles needed to be taken into consideration each time when the vehicles make to access the link.

The first definition of average OD travel time was the summation of average travel time from all vehicles that entered the network and travel on the link without knowing where the vehicles were heading to. Therefore, in this case, the effects of turning movement at the intersection were disregarded. Within the simulation analysis, the truth was defined when data for 100% vehicles existed in the link. Thus, in this case, the entire finite population of vehicles was measured. However, for the other definition, the OD travel time was defines as the travel time from vehicles that travel from and to OD. This was consistent as the case of route travel time where the probe vehicles traveled from one zone to another zone.

Table 4.6 The average travel time, standard deviation and number of vehicles traveling from and to each OD pair

OD	7-5	8-5	9-6	8-11
Average Travel Time (min)	18.07	14.82	3.78	12.62
Standard Deviation	4.98	9.35	1.24	7.68
No of Vehicle	181	143	96	1602

Table 4.7 The average travel time, standard deviation and number of vehicles travel on link by link basis

OD	7-5	8-5	9-6	8-11
Average Travel Time (min)	15.61	22.83	4.33	11.84
Standard Deviation	1.37	1.56	0.15	0.74
No of Vehicles	588	584	508	1817

Table 4.6 and Table 4.7 contain the results of average travel time, standard deviation and number of vehicles for OD and link by link basis. In Table 4.6, OD 7-5 gives the highest travel time because most of the vehicles was stuck in the queue which lead them of having longest travel time. However, referring to Table 4.7, OD 8-5 gives the highest travel times because this is the longest route in network. For this OD, there is a probability for the vehicle to get into the destination using a different route. However, in this analysis, in case of link by link analysis (population), the travel times was gathered when assuming the vehicles took a non-congested route. Therefore, the consideration of the level of congestion should take into account and this would affect the amount and accuracy of travel time information. In general, from these results, it is evident that the OD pair analysis consistently underestimated the population travel time (link by link basis). In particular, when considering travel time from all those vehicles traveling from one zone to another zone (OD travel time), the resulting mean travel times are much smaller than the population mean travel times. This is expected as the population of vehicles required to make left or right turn and consequently experience much greater delay than the sample of vehicles traversing the link to get to the destination.

However, generally, two observations can be made from these tables. First, the average travel times given by these two sources differ within a small range. Second, it is clear that when having a lot of vehicles, the standard deviation become smaller. The standard deviations (or variability), in all OD sets, clearly do not vary that much given as the number of vehicles to get the travel time data is higher. The standard deviation is thus a good measure of the performance of the estimates. Again, as in the previous cases, the mainline is matched best.

It is important here to remind that these calculations are dealing not with sampled dataset, as it would be the case for a probe vehicle based-estimation. The connotation that can be made about these results is regarding whether the accuracy of the combination of route travel time and the link by link travel time to be applied in the real world. It is good to know that if only link travel times could be obtained in the entire network, one could consider this as a good indicator of OD travel time or not. This information is better than no other information (such as route travel time) is available. From the results, if consideration is made from the entire population to get travel time link by link, it represents the real travel time that are experienced by all vehicles that exactly travel from and to the OD. In other words, it is clear when having all vehicles in the link, the O-D estimates are rather reliable. While the above results seem to paint a rather bleak picture, it is noteworthy that all population average travel times are inside the range of 20% off from the OD travel time basis. This is a good start as in the real world, it is very difficult to reliably estimate the travel time due to the high variability in the traffic stream that results from the flow conditions coupled with the interrupted flow nature of signalized links. The presence of non-interrupted flow makes the estimation of link travel times intrinsically much more reliable. Thus, the analysis has shown that even though in some circumstances the population travel time can appear to provide unreliable results, one can always obtain some indication on the quality of the data by providing even some vehicles in the network to track the travel time data. Some variation of the deviation may be caused by variations in traffic volume, although not so large.

Therefore, in general context, the use of probe vehicle travel time for measuring link by link data depends upon the context. For example, in congested routes where the location or congestion areas are known, the retrieval of travel times on link by link basis would be of higher value (a good estimation of average travel time on a particular location). In the case of random impacts and wide variations in congestion, the probe vehicle would be of higher value taking the appropriate sample size in account.