



12-2007

Evaluation of Impacts on Delay, Cycle-Length Optimization, Control Types, and Peak-Hour Factor with the Randomness of Traffic

Jan-Mou Li

University of Tennessee - Knoxville

Recommended Citation

Li, Jan-Mou, "Evaluation of Impacts on Delay, Cycle-Length Optimization, Control Types, and Peak-Hour Factor with the Randomness of Traffic. " PhD diss., University of Tennessee, 2007.
https://trace.tennessee.edu/utk_graddiss/225

This Dissertation is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Jan-Mou Li entitled "Evaluation of Impacts on Delay, Cycle-Length Optimization, Control Types, and Peak-Hour Factor with the Randomness of Traffic." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Lee Han, Major Professor

We have read this dissertation and recommend its acceptance:

Hamparsum Bozdogan, Arun Chatterjee, Tom Urbanik

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Jan-Mou Li entitled "Evaluation of Impacts on Delay, Cycle-Length Optimization, Control Types, and Peak-Hour Factor with the Randomness of Traffic." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Lee Han, Major Professor

We have read this dissertation
and recommend its acceptance:

Hamparsum Bozdogan

Arun Chatterjee

Tom Urbanik

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of
the Graduate School

(Original signatures are on file with official student records.)

**Evaluation of Impacts on Delay, Cycle-Length
Optimization, Control Types, and Peak-Hour
Factor with the Randomness of Traffic**

A Dissertation
Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

Jan-Mou Li
December 2007

Copyright © 2007 by Jan-Mou Li
All rights reserved.

DEDICATION

I would like to dedicate this dissertation to my wife Yu-Chen and our daughters Tsai-Lin (Irene), and Jr-Lin (Iris), since without their support, love, patience, encouragement and sacrifices this work would have not been possible.

This dissertation is also dedicated to my parents whose love and support have made all my endeavors possible.

ACKNOWLEDGEMENTS

I want to take this opportunity to express my most sincere respect and gratitude to all the people who helped to make this work possible. First, I want to thank my advisor, Dr. Lee D. Han, for his excellence guidance, inspiration and encouragement during this study, as well as his kindness and support throughout my entire stay at the University of Tennessee. It has been an honor to be a part of his research group. I will miss him as a mentor, role model and friend.

I would also like to thank my committee members: Dr. Tom Urbanik, Dr. Arun Chatterjee, and Dr. Hamparsum Bozdogan. Thank you all for your timely insights and willingness to share in this research.

I would also want to express my gratitude to my wife Yu-Chen and our daughters Tsai-Lin (Irene), and Jr-Lin (Iris), for their patience, love and encouragement through all these time. Without their consistent support and help, earning my Ph.D. degree would not have been possible. I would also like to thank my parents for their faith in me, their support and to help me be the professional that I am now.

I also gratefully acknowledge Dr. Way Kuo for his support in the GSSAP. I would like to extend my appreciation to my colleagues and friends, Dr. Fang Yuan, Dr. Zhongren Gu, Hal Millegan, Shaun Quayle, Stephanie Hargrove, Manoel M. Castro-Neto, Xiaoli Sun, Chunghao Chen, and TeYu Chien at the University of Tennessee. I will miss the wonderful time with all of you.

ABSTRACT

Some basic concepts about traffic which are correct in theory may be misinterpreted in practice. Such misinterpretations may lead to a different direction from the ideal operation. This four-part dissertation is designated to examine fundamental concepts in traffic operation, and to validate the impact of randomness on control delays, cycle-length optimization, control types, and the peak-hour factor.

Control delays experienced by drivers is a critical performance measure on interrupted-flow traffic which involves movements at slower speeds and stops on intersection approaches, as vehicles move up in the queue or slow down upstream of an intersection. Since the basic term of control delay in a signalized intersection was originally from queueing analyses within a cycle, results from such models may be inaccurate due to the neglect of inter-cycle traffic variation. Besides, traffic is rare varying on the clock. Therefore, the peak-hour factor will be inaccurate to a certain degree if peak periods are placed on the clock.

All parts of this dissertation, except the first and the last, are independent papers for different professional journals, and are summarized as follows. Part II of this dissertation, "Impacts of Inter-Cycle Demand Fluctuations on Delay", distinguishes between intra- and inter-cycle demand fluctuations and recognizes the potentially significant impact of delay underestimation when inter-cycle demand fluctuation is unaccounted for, as in all previous models. "Short or Long ... which is Better? A Probabilistic Approach towards Cycle Length Optimization" in the third part of this dissertation proposes a framework to determine the

optimal or near-optimal cycle length for signalized intersections based on the criterion with minimal control delays. The fourth part with title “A Trade-Off Framework for Determining the Best Control at an Intersection” in this dissertation uses the same criterion with minimal control delays to assist decision makers in the trade-off between signals and stop signs for an intersection. Part V of this dissertation, “Impacts of Misplaced Peak Intervals on PHFs”, argues about the significant difference among different ways to define the peak intervals, and distinguishes the differences between the “real” and “on the clock” peak-hour factors.

TABLE OF CONTENTS

PART I. INTRODUCTION	1
INTRODUCTION	2
Impacts of Inter-Cycle Demand Fluctuations on Delay	3
A Probabilistic Approach towards Cycle Length Optimization	6
A Trade-Off for Determining the Best Control at an Intersection	9
Impacts of Misplaced Peak Intervals on PHFs	12
REFERENCES FOR PART I	15
PART II. IMPACTS OF INTER-CYCLE DEMAND FLUCTUATIONS ON DELAY	
.....	19
ABSTRACT	21
INTRODUCTION	22
ANALYTICAL CONTEMPLATION	28
VERIFICATION WITH MONTE CARLO SIMULATIONS	33
Webster's Delay Model	33
HCM 2000 Delay Model	34
Some Assumptions	34
Incremental Queue Accumulation (IQA) Method	35
Simulation of Hypothetical Cases	37
Results and Discussions	39
CONCLUSIONS.....	42
REFERENCES FOR PART II	43

PART III. SHORT OR LONG ... WHICH IS BETTER? A PROBABILISTIC

APPROACH TOWARDS CYCLE LENGTH OPTIMIZATION.....	46
ABSTRACT	48
INTRODUCTION	50
THE ANALYTICAL FRAMEWORK	55
IMPLEMENTATION PROCEDURE	62
SOLUTION DEMONSTRATION	64
Hypothetical Scenario	64
Monte Carlo Simulation and Results	65
Sensitivity Analyses.....	65
Design for Extremities	67
Level of Service Expectations	69
Cycle Failure Probabilities	70
CONCLUSIONS.....	72
RECOMMENDATIONS.....	74
REFERENCES FOR PART III	75

PART IV. A TRADE-OFF FRAMEWORK FOR DETERMINING THE BEST

CONTROL AT AN INTERSECTION	78
ABSTRACT	80
INTRODUCTION	81
METHODOLOGY.....	89

SIMULATION RESULTS.....	95
Examination of the Exhibit 10-15 in HCM 2000	96
Comparisons among Signal, TWSC, and AWSC	97
An Elegant Case – Symmetric Traffic Flow Rate.....	101
Signalized Intersections Only	102
FINDINGS AND DISCUSSIONS.....	104
Examination of the Exhibit 10-15 in HCM 2000	104
Comparisons among Signal, TWSC, and AWSC	105
The Elegant Case – Symmetric Traffic Flow Rate.....	107
Signalized Intersections Only	108
CONCLUSIONS.....	110
REFERENCES FOR PART IV	112
PART V. IMPACTS OF MISPLACED PEAK INTERVALS ON PHFS.....	116
ABSTRACT	118
INTRODUCTION	119
METHODOLOGY.....	122
Definition of Peak-Hour Factor (PHF).....	122
Monte Carlo Simulation	124
Significance Tests – The Wilcoxon Signed-Rank Test	125
Measurement of the Difference - Mean Absolute Percentage Error (MAPE)	
.....	126
RESULTS	127

DISCUSSION.....	135
Misplaced Peak Intervals.....	135
Local Data are Recommended.....	136
CONCLUSIONS.....	138
REFERENCES FOR PART V	139
PART VI. CONCLUSIONS	141
CONCLUSIONS.....	142
Vita.....	145

LIST OF TABLES

Table 2.1. Mean Time-Headway under Different Demand Levels.....	39
Table 2.2. Average Delay Estimated by Different Models.....	40
Table 3.1. Sensitivity Analysis Results.....	67
Table 3.2. Percentiles of Demands Targeted and Associated Cycle Length.....	69
Table 3.3. LOS Probability as a Function of Cycle Length.....	70
Table 4.1. Average Control Delays for a Hypothetical Intersection with Through-Traffic.....	82
Table 4.2. Definitions of LOS for Signalized, TWSC, and AWSC Intersections..	84
Table 4.3. Average Control Delays for a Hypothetical Intersection with Left-Turn Traffic.....	97
Table 5.1. Results from 5000-run Monte Carlo Simulations with Truncated Normal Distributions.....	129
Table 5.2. Results from 5000-run Monte Carlo Simulations with Poisson Distributions.....	131
Table 5.3. Statistics from Real Data.....	134

LIST OF FIGURES

Figure 2.1. Queueing diagram and QAD with an initial queue at $t = 0$	24
Figure 2.2. Demand fluctuations seen at different time scales.....	27
Figure 2.3. Demand variations within a 1-minute cycle and for 15 minutes.....	29
Figure 2.4. Two cases of different demand patterns.....	31
Figure 2.5. Low-to-High and High-to-Low demand patterns.....	32
Figure 2.6. A case of oversaturation.....	37
Figure 2.7. Three different demand patterns with same average demand.....	38
Figure 3.1. Demand fluctuations observed at different time scales.....	54
Figure 3.2. Basic probabilistic relationship between cycle length and demand.....	58
Figure 3.3. Average delay as a function of cycle length.....	59
Figure 3.4. Expectation and distribution of average delay.....	66
Figure 3.5. Percentile of delay distribution and level of service.....	68
Figure 4.1. The Exhibit 10-15 in HCM 2000.....	82
Figure 4.2. Contours of Average Delay For Signal Control and TWSC.....	98
Figure 4.3. Contours of Average Delay For Signal Control and AWSC.....	99
Figure 4.4. Contours of Average Delay (in Seconds) For TWSC and AWSC	100
Figure 4.5. Average Delays (in Seconds) Under Symmetric Flow.....	102
Figure 4.6. Average Delays at Signalized Intersection under Different Saturation Flow Rate.....	103
Figure 5.1. Samples from (a) $N(2,0.6)$, (b) $N(2,1)$, (c) $(2,2)$	128

Figure 5.2. Samples from (a) Poisson(2), (b) Poisson(3) , (c) Poisson(4).....	130
Figure 5.3. Samples from Real Data.....	132
Figure 5.4. A sample with peak intervals from detector #5124 on January 2, 2007.....	132

PART I. INTRODUCTION

INTRODUCTION

Randomness is the nature of traffic. In addition, it varies over time and at different sites. Even though there are models to describe and analyze traffic, it is still a challenge to have a common recommendation for all kinds of conditions due to the randomness of traffic. Moreover, some basic assumptions, which are correct in theory, may be misinterpreted in practice. This four-part dissertation is designated to examine some fundamental concepts in traffic operation, and to evaluate the impact of randomness on delay, cycle-length optimization, control types, and peak-hour factor. In short, because of the randomness of traffic, inter-cycle demand fluctuation should be considered on the average delay estimation. Because of the randomness of traffic, the expectation of average delay should be considered on deciding the “just right” cycle length for a pre-timed operation. Because of the randomness of traffic, the average delay for the intersection should be considered on the trade-off among control types. Because of the randomness of traffic, the peak-hour factor should not be derived from an “on the clock” basis.

Control delays experienced by drivers is a critical performance measure on interrupted-flow traffic which involves movements at slower speeds and stops on intersection approaches, as vehicles move up in the queue or slow down upstream of an intersection. Since the basic term of control delay in a signalized intersection was originally from queueing analyses within a cycle, results from such models may be inaccurate because of the inter-cycle traffic variation. Part II of this dissertation, “Impacts of Inter-Cycle Demand Fluctuations on Delay”,

distinguishes between intra- and inter-cycle demand fluctuations and recognizes the potentially significant impact of delay underestimation when inter-cycle demand fluctuation is unaccounted for, as in all previous models. “Short or Long ... which is Better? A Probabilistic Approach towards Cycle Length Optimization” in the third part of this dissertation proposes a framework to determine the optimal or near-optimal cycle length for signalized intersections based on the criterion with minimal control delays. The fourth part with title “A Trade-Off Framework for Determining the Best Control at an Intersection” in this dissertation uses the same criterion with minimal control delays to assist decision makers in the trade-off between signals and stop signs for an intersection. Part V of this dissertation, “Impacts of Misplaced Peak Intervals on PHFs”, argues about the significant difference among different ways to define the peak intervals, and distinguishes the differences between the “real” and “on the clock” peak-hour factors. All of these parts are independent papers for different professional journals, and are summarized as follows.

Impacts of Inter-Cycle Demand Fluctuations on Delay

According to Newell (1965), the simplest models of traffic flow through intersections were considered by Clayton in 1941 and perhaps by other researchers even earlier. In these early queueing models, vehicles were assumed to arrive at regularly spaced time intervals with a mean time-headway of $1/q$, where q is the average flow rate over a certain time period. The vehicles form a queue during the red phase, R , at a traffic light and then during the subsequent green phase, G , depart at regularly spaced intervals with a time-

headway of $1/s$ until either the end of the green time or when the queue has fully dissipated. The assumption that traffic arrivals and departures are uniformly distributed is an important part of Webster's work (Webster 1958), which attempts to attribute the average vehicular delay at signalized intersection to three main components, or terms, i.e. uniform delay, random delay, and empirical errors. A very similar formulation for delay estimation is later employed by the 1985 edition (TRB 1985) and subsequent updates of the Highway Capacity Manual (HCM) (TRB 1994, 1997, 2000). The first term in each of these delay formulae represents uniform delay, which can be and is derived from simple queueing-based analysis. By assuming uniform arrivals within a signal cycle, or intra-cycle, and by ignoring the discrete nature of vehicles, traffic can be considered as a continuous flow arriving at a uniform rate of q . At some point in time the flow is dammed up for a period of R ; it is then released at a rate of s until the build-up has dissipated. Therefore, with all the simplicity in its algebraic form, the first term of Webster's delay model has stood the test of time.

Because neither the world nor traffic at a signalized intersection is deterministic, researchers have endeavored to introduce stochastic terms into delay models in order to estimate delay more realistically. To this end, the second term of Webster's model makes some allowance for the random nature of the arrivals. In a rather subtle and largely unnoticed manner, that the random nature of vehicular arrivals within a cycle (intra-cycle) and that among cycles (inter-cycle) can be considered identical and are, thus, represented with identical statistical distribution. In fact no delay model, Webster's or else, distinguished

inter-cycle and intra-cycle randomness until Han and Li (2007) while readdressing the cycle-length optimization problem with Monte Carlo simulations. One of the benefits of this implied assumption is one could simplify the analysis and treat the entire study period of, say, an hour as a single signal cycle with the same average demand of q throughout. The flip side, however, is the errors this assumption introduces when inter-cycle randomness exists. Since unused capacity at a signalized intersection cannot be carried over from one cycle to succeeding ones, if inter-cycle demand fluctuation exists, the delay model has to be formulated to address the factor of randomness beyond the boundary of a single signal cycle.

Many studies have analyzed the impact of fluctuating demand on average delay, but none has distinguished the randomness of demand within and among cycles. Akcelik and Roupail (1993) applied symmetrical triangular and parabolic functions to represent demand over the total flow period. Heidemann (1994) assumed the number of vehicles arriving during a time interval to follow the Poisson arrival process and the arrivals for different but equal-length time intervals to be identically and independently distributed (IID). However, he did not approach the subject from a signal-cycle perspective, and he did not consider non-identical distribution cases from one interval to the next. While many studies on delay at signalized intersections have considered demand fluctuation within a cycle (intra-cycle), they have often implicitly treated demand over multiple cycles (inter-cycle) to be the same and, consequently, have reduced the analysis for a longer period, e.g. 15 minutes, to a single cycle. Therefore, it has to be

distinguished between intra- and inter-cycle demand fluctuations and then recognized the potentially significant impact of delay underestimation when inter-cycle demand fluctuation is unaccounted for, as in all previous models.

Since unutilized capacity at a signalized intersection cannot be saved or carried over to be used by succeeding cycles when demand surges due to fluctuation, the pattern of inter-cycle demand variance is important. Simulation results demonstrate that different patterns of inter-cycle demand variance can result in different levels of average delay. The importance of inter-cycle demand variance on delay analysis is pointed out, especially under heavy traffic conditions. That is, not only the intra-cycle demand variance but also the inter-cycle demand fluctuation has a significant impact on the delay at a signalized intersection. Neglecting inter-cycle demand variance may lead to significant inaccuracy and, hence, suboptimal signal timing decisions.

A Probabilistic Approach towards Cycle Length Optimization

Since the introduction of automatic traffic signals in 1926, signal-timing optimization has become a classic problem, but not always a satisfactorily solvable one, mainly because of demand fluctuation over time. The fact is that if demand were constant or largely predictable, signal-timing optimization would be eminently solvable, indeed trivial. However, in practice, this problem is not at all easy to solve, and there has been little agreement among experts. There have been practitioners in favor of long cycle durations (e.g. more than 150 seconds) because less lost time is observed over a period, say an hour, of time; yet there are also proponents for snappier cycle durations (e.g. less than 60 seconds) to

avoid the buildup of queues. To help clarify this issue, this paper will examine how cycle length should be chosen, whether short or long or neither, for fixed time signals in isolated intersections when all other factors are unchanging. (Actuation and other adaptive means that vary cycle length in real time have the potential to address this issue but would also bring much complexity into this matter, so they are not here addressed.) Vehicular delay as a result of traffic signals has been commonly identified as a primary measure of motorists' perception of how well a signalized intersection operates. Therefore, timing optimization often involves the minimization of vehicular delay. It is common to balance volume to capacity, or v/c , ratios of critical movements for this purpose.

To keep matters simple, early studies often assumed uniform arrival at intersections (e.g. Clayton 1941; Wardrop, 1952; Webster, 1958; Newell, 1965). The first attempts at analytical models of fixed-cycle traffic signals were by Beckmann et al. (1956) and Newell (1956). Newell proposed a model in which arrival headways were independently and identically distributed (IID) random variables of arbitrary distribution while departures were regularly spaced during the green. It is noteworthy that these classic works were conducted based on queuing analysis within a single cycle and, hence, dealt only with intra-cycle demand fluctuations. This was the reason why some researchers, e.g. Han (1996), considered these early works to be based on time-stationary assumptions. Subsequent studies modeled delay using various approaches. Akcelik and Roupail (1993) considered symmetrical triangular and parabolic functions as variable demand functions during the analysis period and proposed

a delay model that was suitable for variable demand conditions. Heidemann (1994) assumed that the number of vehicles arriving during each time interval was considered to be stochastic, and the arrival distributions for different time intervals were assumed to be identical and stochastically independent. Han (1996) may have been the earliest to report that when traffic demands are different in successive time periods, the signal settings that are optimal for each individual period are only local solutions to the problem. He developed a sequential optimization technique to minimize the total intersection delay over successive periods by searching for the optimal signal timings. But one of the primary assumptions in his work was that the traffic demand, though varying from one period to the next, stayed constant in each individual period.

Although each was valuable, none of these studies formally addressed the effects of inter-cycle as well as intra-cycle fluctuations of traffic demand. The issue here is that inter-cycle fluctuation will affect the result of delay analysis because unused capacity that is due to a momentary drop in demand from one cycle is capacity that is lost forever and cannot be reclaimed to help future temporary surges in demand. To properly represent this in queueing analysis and, hence, obtain better delay estimation, there should be multiple piecewise segments instead of a single straight line when demand is aggregated over time. This representation of fluctuating demand would lead to different, most likely higher, delay values, and subsequently a different optimization scheme.

A probabilistic approach is employed to consider cycle-length optimization for isolated intersections and attempted to answer the question of whether

shorter or longer cycles are “better.” With mathematical formulations and Monte Carlo simulations, the authors established that certain “just right” cycle lengths could be derived following a five-step optimization framework. A hypothetical example was then presented to demonstrate how the framework functions with ensuing analyses and discussions on sensitivity of the solution, expected LOS, and potential cycle failures. The major contribution of this paper is the proposed framework for optimizing cycle length under stochastic inter- and intra-cycle demand levels based on the expectation function of delay. When deployed, this framework can aid traffic engineers in choosing the desirable cycle length for minimal delay or for any, reasonable, LOS requirements.

A Trade-Off for Determining the Best Control at an Intersection

To assign the right of way at an intersection is definitely a complex issue because numerous factors are involved. That is why engineering judgments or studies are necessary for such a situation, even though there are already massive amounts of research dedicated to traffic signal or stop signs respectively. Since both traffic signals and stop signs are supposed to serve users in a more efficient way, a framework for the trade-off between a traffic signal and a stop sign will be very useful for traffic engineers. However, very little literature mentioned the trade-off between signals and stop signs for an intersection, either qualitatively or quantitatively. There is an exhibition (Exhibit 10-15) to forecast the likely intersection control types in HCM 2000. Unfortunately, the reference for that exhibit is incorrect, and therefore it can not provide any further information in order to validate the trade-off decision. If the

traffic patterns will not change with different control types at an intersection, some of turning points on the exhibit will get confusing results, especially in higher traffic volume conditions. There are eight warrants (FHWA, 2004) for justifying traffic control signals in chapter 4C of the Manual on Uniform Traffic Control Devices 2003 edition (MUTCD 2003). The first three warrants are relative to vehicular volumes, and they may raise the same questions as the Exhibit 10-15 did, because the conditions are quite similar to those on the Exhibit 10-15 in HCM 2000, especially in Warrant 3.

Even though the thresholds of level of service (LOS) for different control types at an intersection are different, control delay is the same cornerstone of LOS for both signal control and stop-controlled intersections. Richardson (1987) proposed an iterative method and used the Pollaczek-Khintchine formula to estimate delays of AWSC intersections, based on an M/G/1 model of queuing process. Although the subject delay in his model is a function of subject, conflicting, and opposing flow rate, statistical analysis suggests that this model might provide a credible estimate of delay (Kyte and List, 1999). Eck and Biega (1988) concluded that four-way stop sign control at low-volume residential street intersections should be changed to two-way stop sign control, because the use of two-way stop sign control in place of four-way stop sign control minimizes delay and road user costs. Chan et al. (1989) proposed a response-surface model with four determinants, i.e. traffic volume, volume split, percentage of left-turns, and street width, to estimate average delay at an AWSC intersection. One of their findings is highly controversial in relation to that by Zion et al. (1989), that is, the

more imbalanced the volume split is, the smaller the delay. Zion et al. (1989) tested delay models, which proposed by Richardson (1987) and by Chan et al. (1989), with field data for AWSC intersections. What they found are that delay increases as the intersecting volume increases; intersections with balanced volumes have lower delays than those without; and the percentage of left turns has a noticeable effect on delay.

Besides AWSC, a two-way stop-controlled (TWSC) intersection is another, and maybe more efficient, type of assigning the right-of-way with the stop sign at the intersection. Byrd and Stafford (1984) examined the operational characteristics of traffic controls at low-volume, low-speed intersections with unwarranted four-way stop sign control. Then they suggested that unless an accident problem susceptible to correction by four-way stop sign control exists, the unwarranted use of four-way stop sign control results in unnecessary delay and road user costs to the driving public and that the intersection traffic control should be changed to two-way stop sign control. In order to clarify the trade-off among signal, AWSC, and TWSC, a trade-off framework to evaluate these three control types at an intersection is proposed in this paper. The average delay models for signalized and unsignalized intersections in HCM 2000 are used as the basis of the framework.

A hypothetical intersection with two-way, two-lane for each direction is examined by the framework. According to the simulations, the feasible areas for both TWSC and AWSC are not rectangle but polygon because they depend upon the contour of average control delay. Since the sensitivities of the average

control delay for different control types differ from each other with different traffic patterns, to facilitate the trade-off among different control types is the most important function for the framework. The sensitivities can be checked very easily through the framework. Based on results in the hypothetical intersection, where the two-way flow rate is less than 600 vph in the major street, 350 vph or less in the minor street, and 10% left-turn traffic for each direction, an AWSC may be a better choice than others since the average control delay for an AWSC is lowest and less sensitive than others in that scenario.

Impacts of Misplaced Peak Intervals on PHFs

Traffic in a road network is varying all the time and the variation is rarely on the clock. In most cases, analyses focus on the peak hour of traffic for a certain approach because it represents the most critical period for operations and has the highest capacity requirements. However, to define the peak hour as well as the worst 15 minutes in practice raises inaccuracy if the traffic variation was not treated properly. According to the HCM 2000 (TRB, 2000), the selection of an analysis period must consider the impact on design and operations of higher volume hours that are not accommodated. It also mentioned that the design for a smaller range, say a 5-minute interval, of the peak flow rate would result in substantial excess capacity during the rest of the peak hour; and the design for a larger range, say an 1-hour interval, of the peak flow rate would result in oversaturated conditions for a substantial portion of the hour. Since most of the procedures in the HCM 2000 are based on peak 15-minute flow rates, the peak hour factor is defined as the ratio of total hourly volume to the peak 15 minute

flow rate within the hour. Nevertheless, it did not mention what would happen if there is a higher peak 15-minute interval outside the peak hour. Such situations occurred in real data when they are closely examined.

Even though traffic varying over time is common sense, the variability of peak hour factor has been investigated recently. Tarko and Perez-Cartagena (2005) investigated the variability of PHF overtime and across locations, and found that the day-to-day variability is as strong as the site-to-site variability. They recommend that PHF be estimated on the basis of several days of vehicle counting to improve the precision of the average PHF estimate. Notwithstanding the spatial difference, even the variation of traffic within a day will not be the same within another day. That is, the peak hour for tomorrow may not start at the exactly same time as today.

For some reasons, practitioners employ the literal meaning of the peak-hour in several ways. Most of the time, they classify the peak hour on the clock, e.g. from 7 a.m. to 8 a.m. or 4:30 p.m. to 5:30 p.m. There is nothing wrong if the hourly, half-hourly, or even 15-minute traffic volume is the only data we had. But such an aggregation may shift the peak hour from the “real” one to a certain degree. When the resolution of data is increased, the difference between the peak hour on the clock and the “real” peak hour should be noticed. Most modern detectors can collect traffic data every thirty seconds. Therefore, the peak hour may start at 7:11:30 a.m. based on the data more precisely.

The object of this paper is to investigate the impact of the misplaced peak hour and peak 15 minutes on the PHF. By comparing different methods locating

the peak intervals, the “on the clock” approach may provide an inaccurate estimation of PHF. Real traffic count data, which were collected by the Minnesota Department of Transportation, at a 30-second interval from over 4,000 loop detectors located around the Twin Cities Metro freeways, are also used for the analysis. It is shown that the PHFs by search are significantly different from those by ‘on the clock’, and that the peak intervals should be located to a more precise period with higher resolution data.

REFERENCES FOR PART I

Akcelik, R. and Roupail, N. M., 1993. Estimation of Delays at Traffic Signals for Variable Demand Conditions. *Transportation Research Part B* 27(2), 109-131.

Beckmann, M., McGuire, C. B., and Winston, C. B., 1956. Studies in the Economics of Transportation, Yale University Press, New Haven.

Byrd, M.N. and Stafford, D.B., 1984. Analysis of Delay and User Costs at Unwarranted Four-Way Stop Sign Controlled Intersections. *Transportation Research Record* 956, 30-32.

Chan, Y., Flynn, L.J., and Stocker, K.J., 1989. Volume-Delay Relationship at Four-Way-Stop Controlled Intersections: A Response-Surface Model. *ITE Journal*, 59 (3), 27-34.

Clayton A. J. H., 1941. Road Traffic Calculations. *Journal of Institute of Civil Engineers* 16, 247-284 and 558-594.

Eck, R.W, and Biega, J.A., 1988. Field Evaluation of Two-Way Versus Four-Way Stop Sign Control at Low-Volume Intersections in Residential Areas. *Transportation Research Record*, 1160, 7-13.

Federal Highway Administration (FHWA), 2004. Manual on Uniform Traffic Control Devices for Streets and Highways, 2003 Edition with Revision Number 1, Available online: <http://mutcd.fhwa.dot.gov/HTM/2003r1/html-index.htm> , Accessed: March 6, 2007.

Han, B. 1996. Optimising Traffic Signal Settings for Periods of Time-Varying Demand. *Transportation Research Part A* 30(3), 207-230.

Han, L. and Li, J.-M., 2007. Short or Long ... which is Better? A Probabilistic Approach towards Cycle Length Optimization. TRB 86th Annual Meeting

Compendium of Papers CD-ROM. TRB, National Research Council, Washington, DC.

Heidemann, D., 1994. Queue Length and Delay Distributions at Traffic Signals. *Transportation Research Part B* 28(5), 377-389.

Kyte, M. and List, G., 1999. A capacity model for all-way stop-controlled intersections based on stream interactions. *Transportation Research Part A*, 33(3-4), 313-335.

Newell, G. F., 1956. Statistical Analysis of the Flow of Highway Traffic through a Signalized Intersection. *Applied. Mathematics Quarterly* 13, 353-369.

Newell, G. F., 1965. Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light. *SIAM Review* 7(2), 223-240.

Richardson, A.J., 1987. A Delay Model for Multiway Stop-Sign Intersections. *Transportation Research Record*, 1112, 107-114.

Tarko, A. P., and Perez-Cartagena, R. I., 2005. Variability of Peak Hour Factor at Intersections. *Transportation Research Record* 1920, 125-130.

Transportation Research Board (TRB), 1985. Special Report 209: Highway Capacity Manual, 3rd ed., TRB, National Research Council, Washington, D.C.

Transportation Research Board (TRB), 1994. Special Report 209: Highway Capacity Manual, 3rd ed., 2nd rev. TRB, National Research Council, Washington, D.C.

Transportation Research Board (TRB), 1997. Special Report 209: Highway Capacity Manual, TRB, National Research Council, Washington, D.C.

Transportation Research Board (TRB), 2000. Special Report 209: Highway Capacity Manual, TRB, National Research Council, Washington, D.C.

Wardrop, J. G., 1952. Some Theoretical aspects of road traffic research. Proceedings of Institute of Civil Engineers II, Vol. 1, No. 2, pp.325-362.

Webster, F., 1958. Traffic Signal Settings, Road Research Technical Paper No. 39. Road Research Laboratory, Her Majesty's Stationery Office, London, U.K.

Zion, M., List, G.F., and Manning C., 1989. Testing Delay Models with Field Data for Four-Way, Stop Sign-Controlled Intersections. *Transportation Research Record* 1225, 83-90.

**PART II. IMPACTS OF INTER-CYCLE DEMAND FLUCTUATIONS
ON DELAY**

This part is a slightly revised version of a paper with the same title submitted to Journal of Transportation Engineering by Lee Han, Jan-Mou Li, and Tom Urbanik:

Han, L., Li, J.-M., and Urbanik, T., 2007. Impacts of Inter-Cycle Demand Fluctuations on Delay. Journal of Transportation Engineering.

My primary contributions to this paper include (1) development of the problem into a work relevant to my doctoral research study, (2) development of experimental setup, (3) most of the gathering and interpretation of literature, (4) performing the laboratory experiments, (5) interpretation and analysis of test results, (6) most of the writing.

ABSTRACT

This paper demonstrates that in addition to intra-cycle demand fluctuation, which is already a consideration in many delay models, inter-cycle demand variance also impacts average delay at signalized intersections. Webster-type delay models treat demand fluctuation over the whole analysis period, often 15 minutes or longer, as if it were just within a single cycle. Such an approach is fine if used judiciously, one might presume. However, results from Monte Carlo simulations with the Incremental Queue Accumulation (IQA) method indicate that Webster-type delay models will underestimate the average delay under heavy traffic conditions.

Since unutilized capacity at a signalized intersection cannot be saved or carried over to be used by succeeding cycles when demand surges due to normal fluctuation, better understanding of the patterns of inter-cycle demand variance is important. Simulation results demonstrate that different patterns of inter-cycle demand variance can result in different levels of average delay. A low-to-high demand pattern will cause a higher average delay than a high-to-low pattern would, even though the overall demand level is exactly the same. It is therefore clear that neglecting inter-cycle demand variance may lead to significant inaccuracy and, hence, suboptimal signal timing decisions.

INTRODUCTION

Queueing theory has been the primary basis of delay analysis at signalized intersections. According to Newell (Newell 1965), the simplest models of traffic flow through intersections were considered by Clayton in 1941 and perhaps by other researchers even earlier. In these early queueing models, vehicles were assumed to arrive at regularly spaced time intervals with a mean time-headway of $1/q$, where q is the average flow rate over a certain time period. The vehicles form a queue during the red phase, R , at a traffic light and then during the subsequent green phase, G , depart at regularly spaced intervals with a time-headway of $1/s$ until either the end of the green time or when the queue has fully dissipated.

The assumption that traffic arrivals and departures are uniformly distributed is an important part of Webster's work (Webster 1958), which attempts to attribute the average vehicular delay at signalized intersection to three main components, or terms, i.e. uniform delay, random delay, and empirical errors. A very similar formulation for delay estimation is later employed by the 1985 edition (TRB 1985) and subsequent updates of the Highway Capacity Manual (HCM) (TRB 1994, 1997, 2000).

The first term in each of these delay formulae represents uniform delay, which can be and is derived from simple queueing-based analysis. By assuming uniform arrivals within a signal cycle, or *intra-cycle*, and by ignoring the discrete nature of vehicles, traffic can be considered as a continuous flow arriving at a

uniform rate of q . At some point in time the flow is dammed up for a period of R ; it is then released at a rate of s until the build-up has dissipated. A tool in the form of queue accumulation diagram, QAD, as depicted in Figure 2.1, has been quite useful for such analyses. The first term of Webster's delay model, with all the simplicity in its algebraic form, has stood the test of time.

Because neither the world nor traffic at a signalized intersection is deterministic, researchers have endeavored to introduce stochastic terms into delay models in order to estimate delay more realistically. To this end, the second term of Webster's model makes some allowance for the random nature of the arrivals.

Webster further employed Monte Carlo simulations to devise a third term to fit a wide range of flow conditions. According to the description in Appendix 2 of Webster's report (Webster 1958), the randomness of the arrivals was assumed.

"Traffic is assumed to arrive at the intersection at random. In fact, the actual distribution obtained from observations on the road could be used but random traffic has the advantage that it can be generated artificially using tables of random numbers to derive the intervals between successive vehicles."

This implies, in a rather subtle and largely unnoticed manner, that the random nature of vehicular arrivals within a cycle (*intra-cycle*) and that among cycles (*inter-cycle*) can be considered identical and are, thus, represented with identical

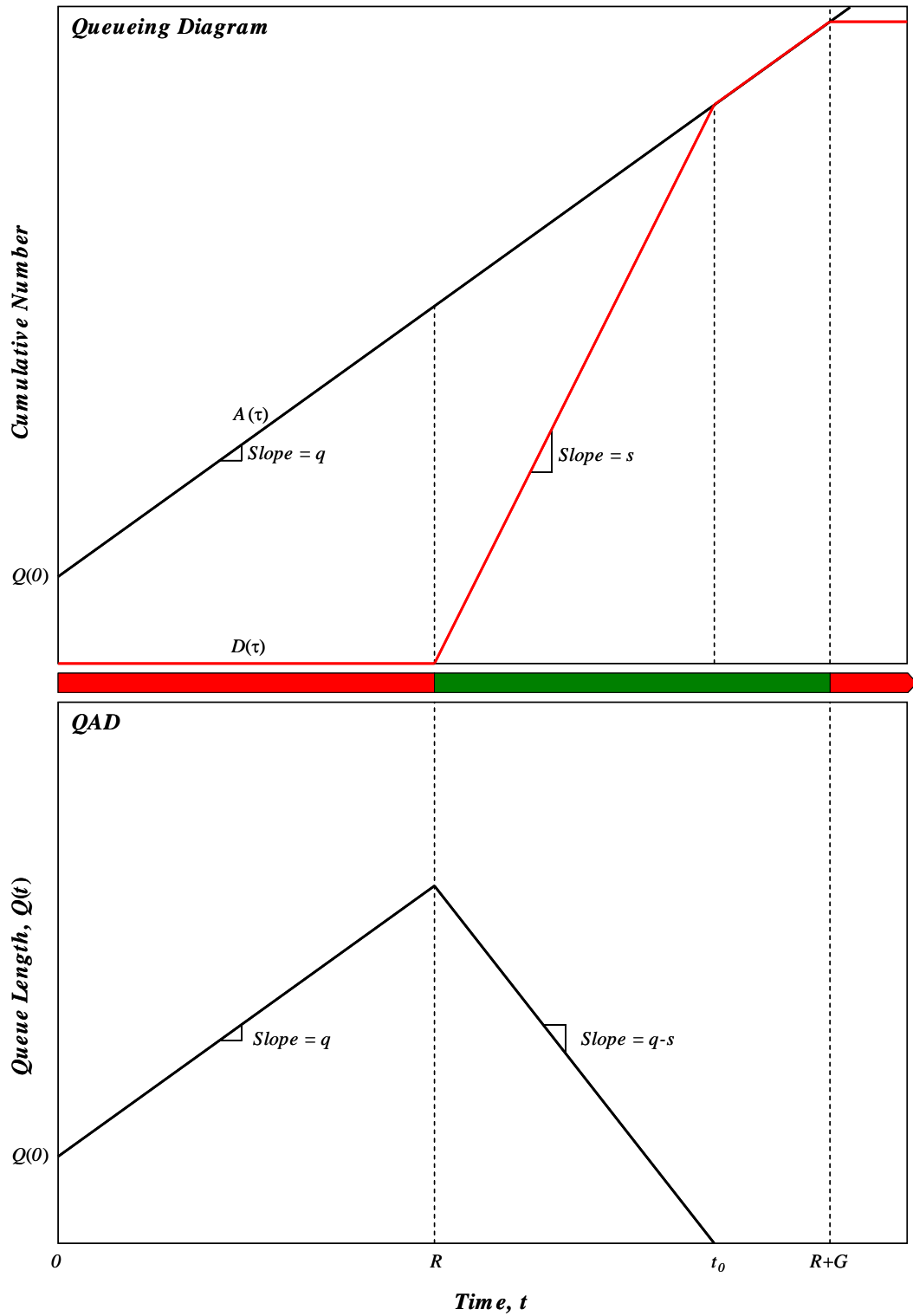


Figure 2.1. Queueing diagram and QAD with an initial queue at $t = 0$

statistical distribution. In fact no delay model, Webster's or else, distinguished inter-cycle and intra-cycle randomness until Han and Li (2007) while readdressing the cycle-length optimization problem with Monte Carlo simulations. One of the benefits of this implied assumption is one could simplify the analysis and treat the entire study period of, say, an hour as a single signal cycle with the same average demand of q throughout. The flip side, however, is the errors this assumption introduces when inter-cycle randomness exists. Since unused capacity at a signalized intersection cannot be carried over from one cycle to succeeding ones, if inter-cycle demand fluctuation exists, the delay model has to be formulated to address the factor of randomness beyond the boundary of a single signal cycle.

Many studies have analyzed the impact of fluctuating demand on average delay, but none has distinguished the randomness of demand within and among cycles. Akcelik and Roupail (1993) applied symmetrical triangular and parabolic functions to represent demand over the total flow period. Heidemann (1994) assumed the number of vehicles arriving during a time interval to follow the Poisson arrival process and the arrivals for different but equal-length time intervals to be identically and independently distributed (*iid*). However, he did not approach the subject from a signal-cycle perspective, and he did not consider non-identical distribution cases from one interval to the next. Han (1996) proposed a similar approach to handle time-varying demands where the overall analysis period (usually 1 hour) is divided into a sequence of sub-periods (usually 5 to 15 minutes) with traffic demands constant throughout all sub-periods. While

many studies on delay at signalized intersections have considered demand fluctuation within a cycle (intra-cycle), they have often implicitly treated demand over multiple cycles (inter-cycle) to be the same and, consequently, have reduced the analysis for a longer period, e.g. 15 minutes, to a single cycle.

This paper distinguishes between intra- and inter-cycle demand fluctuations (see Figure 2.2) and recognizes the potentially significant impact of delay underestimation when inter-cycle demand fluctuation is unaccounted for, as in all previous models.

The remainder of this paper presents the approach used to analyze the inter-cycle demand fluctuations; the Monte Carlo simulations performed, with detailed descriptions of various scenarios; the results, observations, and discussions of the analyses; and the conclusions.

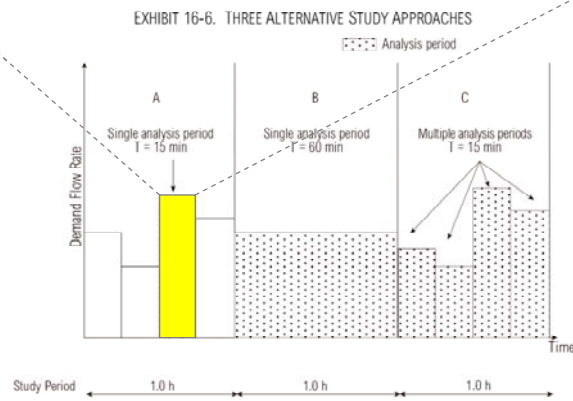
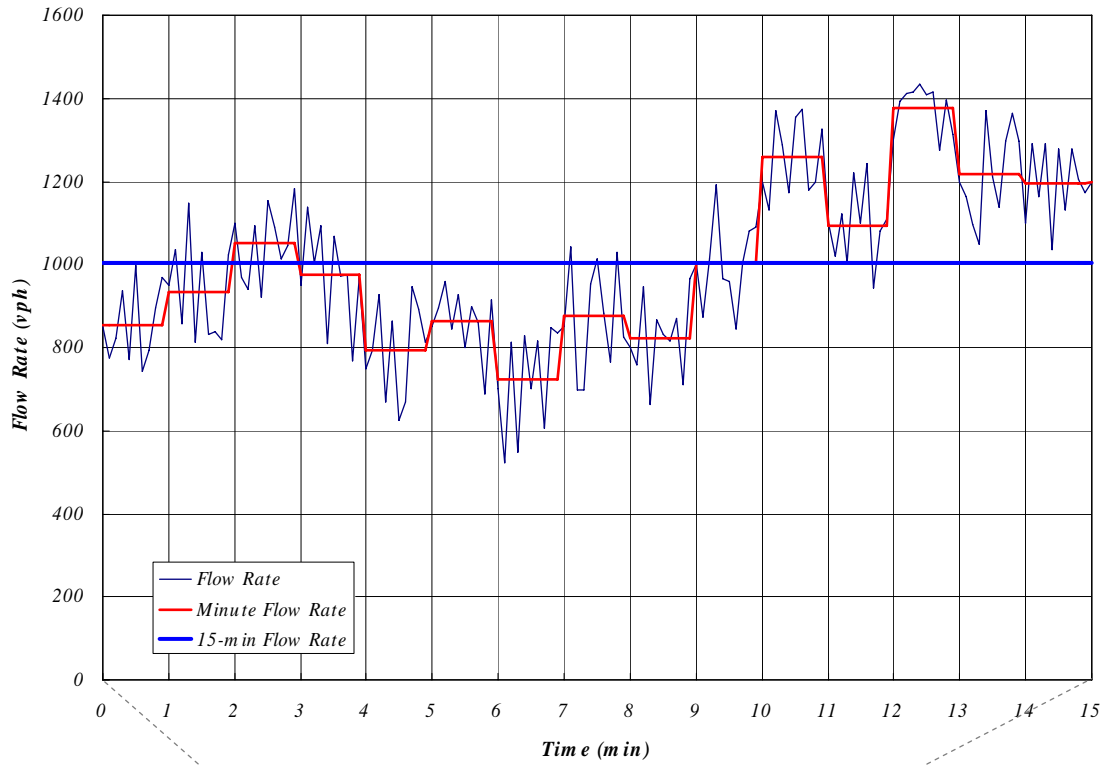


Figure 2.2. Demand fluctuations seen at different time scales
 (Adapted with modifications from TRB's HCM 2000 Exhibit 16-6)

ANALYTICAL CONTEMPLATION

Queueing analysis is employed to assess the impact of inter-cycle demand fluctuations on delay, in comparison with the case of intra-cycle demand fluctuations already studied by earlier researchers. Following Newell's fluid model, let $A(\tau)$ be the cumulative number of arrivals at time τ and let $D(\tau)$ be the cumulative number of departures at time τ . Then for the single cycle depicted in Figure 2.1, $A(\tau)$, $D(\tau)$, and $Q(\tau)$ can all be derived for any given τ within that single cycle.

Under previous assumptions, the total delay of all vehicles in the queue during the cycle, $R+G$, is the area under the QAD curve in Figure 2.1. This is what the first term in Webster's model was based on. When the randomness was added to the arrival, i.e. assuming q follows a certain kind of stochastic distribution, a Webster-type of delay model could be derived. Since Webster-type delay models were derived from a single cycle, the assumption of randomness was really for the entire analysis period. If the analysis period were 60 seconds, the arrival curve would look like the one in Figure 2.3(a); if the period were 15 minutes, the arrival curve (based on the assumption that the entire period had a single stochastic distribution and a fixed mean) would resemble the one in Figure 2.3(b).

Unfortunately, empirical data have shown that demand does not remain nearly stationary over a long period of 15 minutes. In fact, Figure 2.2 is closer to

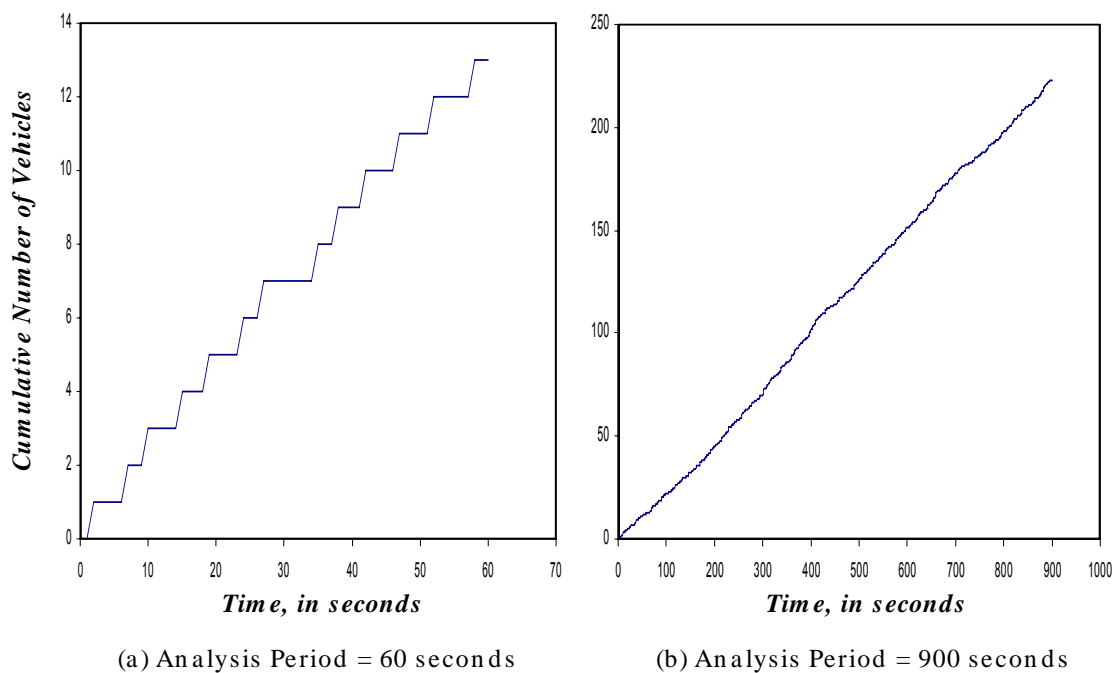


Figure 2.3. Demand variations within a one-minute cycle and for 15 minutes

what may actually occur. With an average flow rate of 1,000 vph over the 15-minute period, each individual minute (and perhaps cycle) will have a different average demand for that minute (or cycle) as a result of the stochastic fluctuation of demand over time.

The delay for 15 one-minute cycles, each with an identical demand rate of 1,000 vph, will differ (greatly in fact) from the delay for 15 one-minute cycles with non-identical demand rates of 861, 935, 1049...1203 vph, even though the average demand (over 15 minutes) for both cases is the same at 1,000 vph.

Another way to look at this problem is this: Let $A_1(\tau)$ represent the arrival curve in the first period; q_1 follows a certain distribution, say $N(\mu_1, s_1^2)$. Let $A_2(\tau)$

represent the arrival curve in the succeeding period, with q_2 following a slightly different distribution, e.g. $N(\mu_2, s_2^2)$. When the single cycle approach is employed to analyze the whole (two-cycle) period, there can be a third arrival function, $A_3(\tau)$. Even if q_3 also follows normal distribution, e.g. $N(\mu_3, s_3^2)$, it cannot be the summation of q_1 and q_2 . That is, μ_3 will not equal to $\mu_1 + \mu_2$, and s_3^2 will not equal to $s_1^2 + s_2^2$.

Figure 2.4 further illustrates this situation. Case 1 shows a common approach that basically extends the same average demand rate from a single cycle to the whole analysis period. In Case 2, the arrival rate is shown to have changed over time, even though the average arrival rates for both cases are identical over the whole analysis period. Considering average delay, the two cases will be different and may very well have different levels of service (LOS). The average delay in Case 2 will be larger than that in Case 1. In fact, Case 2 may experience some cycle failures towards the latter part of the analysis period.

The reason inter-cycle variance, as opposed to the intra-cycle variance that has been studied quite thoroughly, should be emphasized is this: The underutilized capacity cannot be “saved” for or carried over to succeeding cycles where the capacity would be needed when demand surges randomly. In order to obtain realistic estimates of average delay, one may average out the varying demand within a cycle by some statistical methods, but one cannot and should not do the same for varying demand in the case of inter-cycle fluctuations. That is, delay models which were derived from the queueing analysis within a single

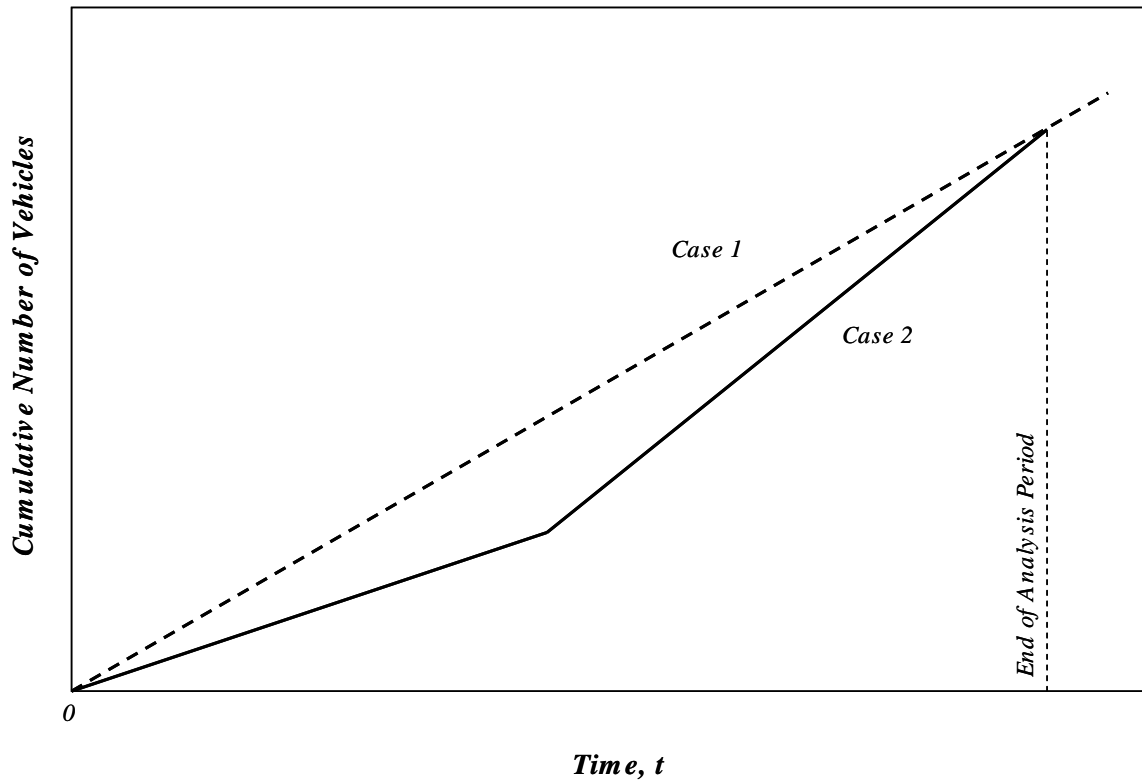


Figure 2.4. Two cases of different demand patterns

cycle, e.g. Webster and Webster-like models, should not be extended to an analysis period beyond a single cycle unless inter-cycle demand fluctuation is minimal to nonexistent. Failing this, the Webster type model, when misused, will underestimate average delay and potentially lead to incorrect LOS projection.

The impacts of inter-cycle demand variance can also be result from the patterns of the variance. Figure 2.5 shows two different demand patterns; each consists of two different arrival rates, i.e., μ_H and μ_L , within the analysis period, although the overall arrival rate for both cases is the same, μ_A . The average

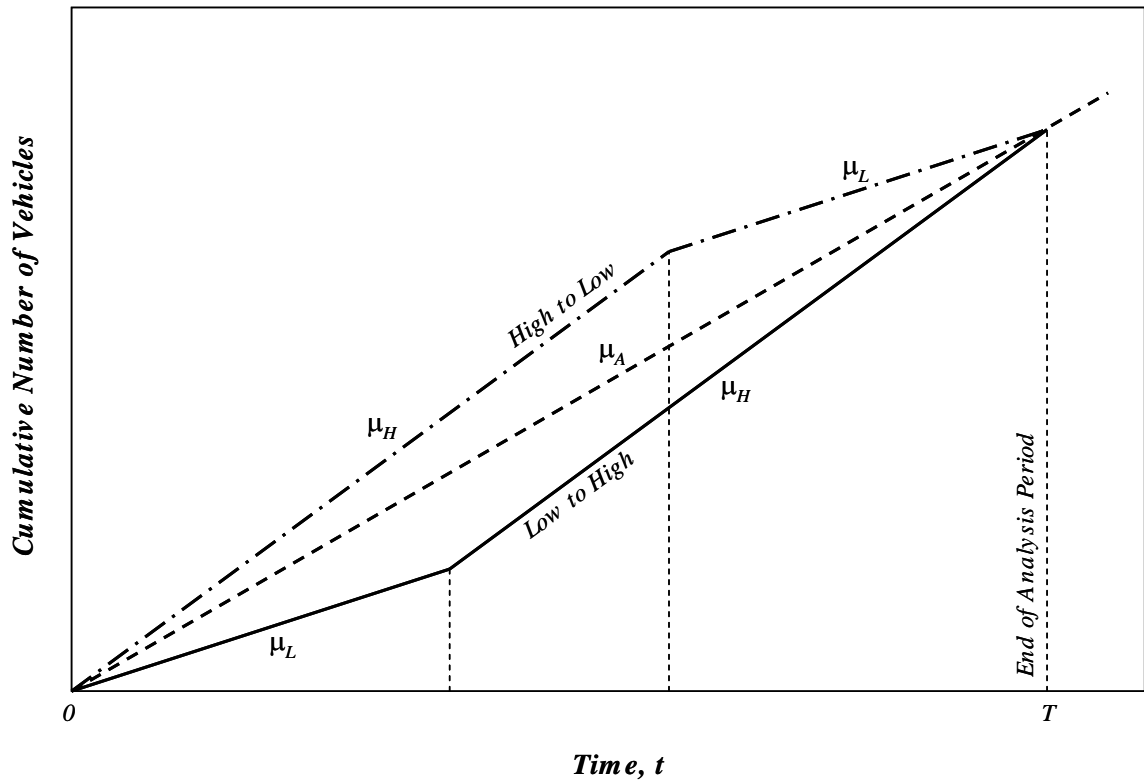


Figure 2.5. Low-to-high and high-to-low demand patterns

arrival rates are from low to high for one case and high to low for the other. If T equals the cycle length, then the average delay can be approximated by using μ_A as the demand for both cases. But if T is a longer duration of, say, 15 minutes or even an hour, which spans over many cycles, the average delay may be quite different.

VERIFICATION WITH MONTE CARLO SIMULATIONS

To verify the concerns posed in the previous section, several scenarios were designed to examine the impact of inter-cycle variation on average delay via Monte Carlo simulation. The results from the simulation are compared to those from Webster and from the HCM 2000 delay models, as detailed below. The Incremental Queue Accumulation (IQA) method was employed to calculate the delay within the system.

Webster's Delay Model

Webster's model, which is based on a single-cycle analysis, is expressed as follows:

$$d = \frac{f(1 - \lambda)^2}{2(1 - \lambda X)} + \frac{X^2}{2q(1 - X)} - 0.65 \frac{f \phi}{s q^2} X^{2+51} \quad (1)$$

where

d = average delay per vehicle;

ϕ = cycle length (s);

λ = proportion of the cycle, which is effectively green of the phase under consideration (i.e. g_e/ϕ);

g_e = effective green time (in seconds or s);

q = traffic demand;

s = saturation flow rate;

X = lane group demand/capacity, or v/c , ratio or degree of saturation;
 this is the ratio of the actual flow to the maximum flow that can pass
 through the intersection and is given by $X = q/\lambda s$.

HCM 2000 Delay Model

When an initial queue is nonexistent, the HCM 2000 model for average control delay per vehicle for a given lane group can be simplified as

$$d = \frac{0.5f_c \left(1 - \frac{g_e}{f} \right) \frac{c}{\phi}}{1 - \min(1, X) \frac{g_e}{f}} PF + 900T \left(X - 1 + \sqrt{(X - 1)^2 + \frac{8kIX}{cT}} \right) \quad (2)$$

where

- d = control delay per vehicle (in seconds);
- P = proportion of vehicles arriving on green;
- T = duration of analysis period (in hours);
- k = incremental delay factor;
- l = upstream filtering/metering adjustment factor;
- c = lane group capacity (vehicles per hour or vph);
- PF = uniform delay progression adjustment factor, which accounts for effects of signal progression.

Some Assumptions

Normal, Pearson Type III, and negative exponential time-headway distributions for high, intermediate, and low demand conditions, respectively (May 1990), were

used for the Monte Carlo simulation runs. Other assumptions include the following:

1. The site is an isolated signalized intersection of two one-way one-lane roads;
2. Arrivals in the two approaches are assumed to be similar so that delay in only one approach needs to be simulated;
3. A pre-timed, two-phases signal control is running with cycle length 60 seconds with an effective green time of 30 seconds, and an effective red of 30 seconds; and
4. At the onset of effective green time, queued vehicles discharge at a saturation flow rate, s , of 1,800 vph, or 0.5 vehicle/second.

Since traffic is assumed to arrive at the intersection randomly, further assumptions for the HCM 2000 delay model include these:

1. Arrival type is 3;
2. Each approach sustains a 4 second/cycle lost time;
3. Uniform delay progression adjustment factor $PF = 1$;
4. Incremental delay factor $k = 0.5$ for pre-timed controller settings;
5. No upstream filtering/metering exists, so the adjustment factor $I = 1$.

Incremental Queue Accumulation (IQA) Method

The Incremental Queue Accumulation (IQA) method originally proposed by Strong and Roupail (2006) was used to implement the HCM model with more flexibility. It extends the usability of the HCM to better reflect conditions commonly found in the field without the plethora of limiting assumptions that are

required by the current HCM 2000 method. This method suggests that equal-sized time slices be used, adding/subtracting the number of arrivals/departures during each time slice to the queue at the start of the time slice and resulting in the queue at the end of the time slice. Even though the concept of the IQA method is intuitive, some characteristics of this method are introduced here due to its novelty. The method

1. uses equal-sized time slices during the analysis period;
2. examines the queue accumulation every time slice;
3. calculates the uniform delay component;
4. is consistent with the model in HCM 2000 and Webster's model; and
5. is fully capable of handling variable arrival rates in different parts of the cycle;

The IQA method is a more generalized approach to calculating the queue accumulation area using multiple trapezoids, and it simplifies the calculation of trapezoids, which represent the periods of time during the cycle when the inflow and outflow rates are not changing. Because the boundaries of each time slice fall squarely on points where signal status and traffic flow rate change, IQA is considered suitable for this research and was used for the purpose.

To calculate average delay in oversaturated or successive cycle failure conditions, one has to estimate and project the delay for queued vehicles that could not depart by the end of the analysis period. At the end of the analysis period, or time T , as shown in Figure 2.6, it is evident that a non-trivial number of

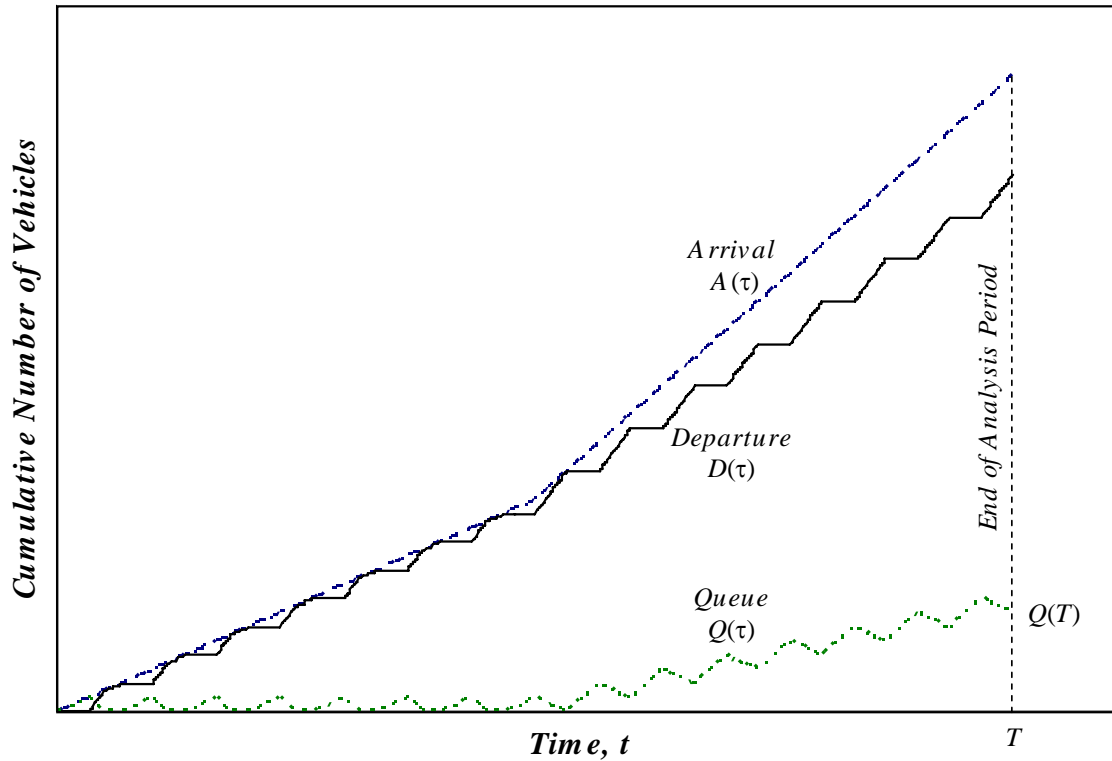


Figure 2.6. A case of oversaturation

queued vehicles, $Q(T)$, have to depart after T . The total delay for each of these queued vehicles was estimated based on their projected departure times.

Simulation of Hypothetical Cases

Traditional Webster-type delay models do not consider inter-cycle demand changes, even though many of them do consider intra-cycle demand fluctuation. This approach is fine if the analysis period is limited to a single cycle and is not extended to a longer period, or if demand holds relatively steady throughout the analysis period and is not like those in Figure 2.2. To test how Webster and HCM 2000 delay models may be “off” when inter-cycle demand fluctuation is a factor,

three very simple and, obviously, hypothetical demand patterns were designed for this purpose (see Figure 2.7). All of the three patterns share the exact average demand over the analysis period, with Case 1 representing the traditional straight-line approach showing no inter-cycle variance throughout the analysis period, while Cases 2 and 3 each have exactly one change in demand level during the analysis period. In Case 2, the mean of arrival rate for the first section is lower than that in the second section. In contrast, the mean of arrival rate for the first section of Case 3 is higher than that in the second section. To further simplify the analysis and simulation, the two sections in both Cases 2 and 3 were assumed to be of the same length of time. Simulated vehicle arrivals in

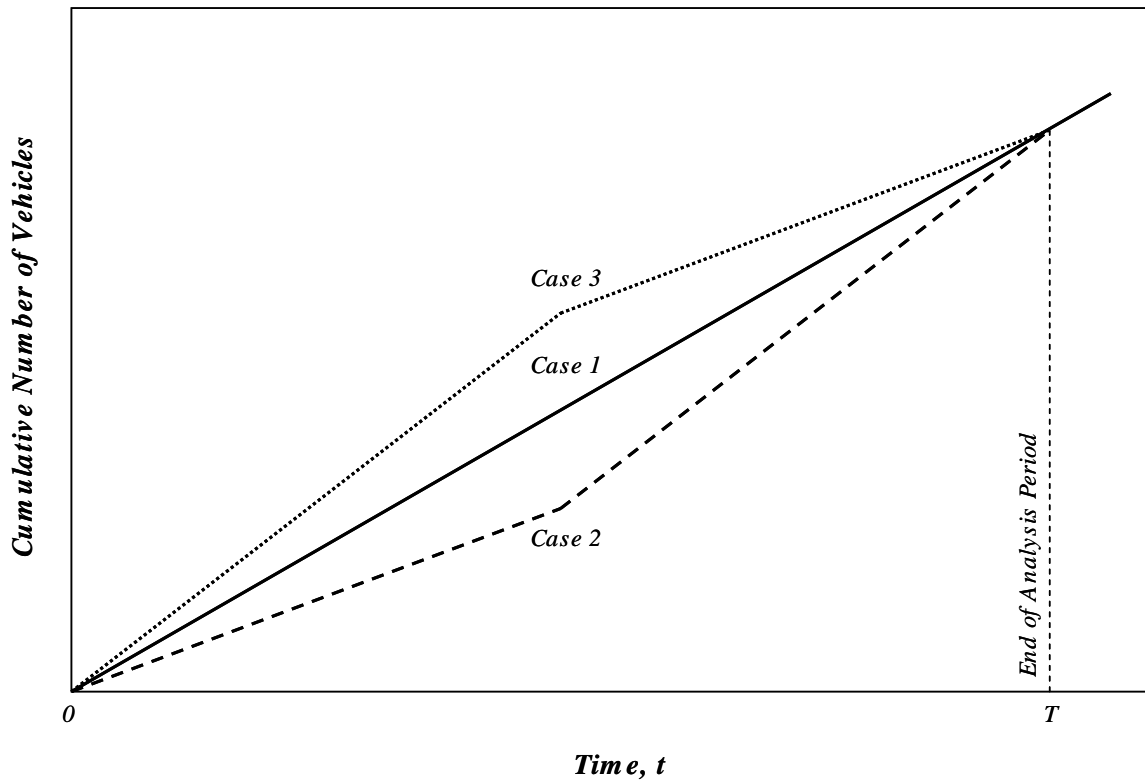


Figure 2.7. Three different demand patterns with the same average demand

each case were generated according to the mean-time headway as listed in Table 2.1. Three demand levels of 300, 600, and 900 vph were used to represent light, intermediate, and heavy traffic conditions, respectively.

For the simplest case of a longer-than-one-cycle analysis period, a two-cycle analysis was selected in which the first section mentioned previously would be the first cycle, and the second section is the second cycle. In addition, analysis periods of 15 minutes, which is typical to HCM 2000, 30 minutes, and 60 minutes were also used for comparison purposes.

Results and Discussions

Table 2.2 tabulates the results from Monte Carlo simulation, for Cases 1, 2, and 3, and from Webster and HCM 2000 models under the prescribed hypothetical conditions. The first impression is that neither the results from Case 1 nor those from the Webster model changed at all as the analysis period T increased from 2 minutes to 60 minutes. This verifies what was presented previously, that like Case 1, Webster model does not consider any inter-cycle demand fluctuations.

The results from Cases 2 and 3 do show higher levels of average delay than those from Case 1 as a result of a single inter-cycle demand change. The

Table 2.1. Mean Time-Headway (in seconds) under Different Demand Levels

DEMAND	CASE 1	CASE 2		CASE 3	
		1 ST	2 ND	1 ST	2 ND
		SECTION	SECTION	SECTION	SECTION
300 VPH	12	15	10	10	15
600 VPH	6	9	4.5	4.5	9
900 VPH	4	6	3	3	6

Table 2.2. Average Delay Estimated by Different Models

T (MIN)	DEMAND	CASE 1 (UNIFORM)	WEBSTE R*	HCM200 0	CASE 2 (LOW- HIGH)	CASE 3 (HIGH- LOW)
2	300 vph	6.00	9.00	9.98	6.30	6.30
	600 vph	10.20	11.25	14.67	12.11	12.60
	900 vph	14.93	>15.73	25.95	17.07	19.93
15	300 vph	6.00	9.00	10.00	6.16	6.44
	600 vph	10.20	11.25	15.15	12.73	13.06
	900 vph	14.93	>15.73	45.00	43.80	68.90
30	300 vph	6.00	9.00	10.00	6.30	6.30
	600 vph	10.20	11.25	15.20	12.97	12.97
	900 vph	14.93	>15.73	57.43	86.07	124.33
60	300 vph	6.00	9.00	10.00	6.30	6.30
	600 vph	10.20	11.25	15.22	12.97	12.97
	900 vph	14.93	>15.73	75.00	161.07	236.83

*The value of 15.73 seconds was calculated under a demand of 899 vph.

increases in delay, however, were not significant in light (an increase of merely 5%) and intermediate (an increase between 19 and 27%) flow conditions. The results did not worsen as T increased. The reason is that the fluctuation of demand from one cycle to the next, under light and intermediate traffic, never reached the same serious tandem cycle-failure situation as those in Figure 2.6. Therefore, the average delay never quite got out of control.

Results from Webster and HCM 2000 models, in general, are higher than those from the three cases under light and intermediate traffic. Under heavy traffic condition, HCM 2000 projects higher delay than Case 1, Webster, and even Cases 2 and 3 for $T = 0.0333$ hour. It is unclear why HCM 2000 yields significant higher delay than the other models, though.

Under heavy traffic, i.e. 900 vph, as T increases, results from Cases 2 and 3 reflect serious cycle-failures and, hence, increasingly undesirable delay levels, which eventually reached an increase of 979% for Case 2 and one of 1487% for Case 3 in comparison with Case 1, when $T = 1$ hour.

The fact that the resultant delay from HCM 2000 under heavy traffic increases as T goes from 2 to 60 minutes indicates some attempt to account for inter-cycle demand fluctuation. The values of the estimated delay, which are significantly lower than those from Cases 2 and 3 when T is large, may indicate that the simple inclusion of T in the model's second term in a linear fashion is insufficient; or, perhaps, the explanation is as simple as the result of oversimplification in the design of the two cases. More complicated and realistic cases will have to be designed to test this thoroughly.

Between Cases 2 and 3, it is clear that Case 2, which squandered away unused capacity during the first half of the analysis period, resulting in a 47% higher level of delay than that of Case 3, which had its cycle-failures in the first half of T , but had extra capacity in the second half available to accommodate the queued traffic.

CONCLUSIONS

Not only the intra-cycle demand variance but also the inter-cycle demand fluctuation has a significant impact on the delay at a signalized intersection.

Webster-type delay models treat the variance over the whole analysis period as if it were within a single cycle. Such an approach is fine if used judiciously.

Simulation results indicate, however, that this type of delay model will underestimate the average delay under heavy traffic conditions.

Since unutilized capacity at a signalized intersection cannot be saved or carried over to be used by succeeding cycles when demand surges due to fluctuation, the pattern of inter-cycle demand variance is important. Simulation results demonstrate that different patterns of inter-cycle demand variance can result in different levels of average delay. A low-to-high demand pattern will cause a higher average delay than a high-to-low pattern would, even though the overall demand level is exactly the same.

This paper points out the importance of inter-cycle demand variance on delay analysis, especially under heavy traffic conditions. Neglecting inter-cycle demand variance may lead to significant inaccuracy and, hence, suboptimal signal timing decisions. Further research is needed to investigate the patterns of inter-cycle demand variance in the real world and to revise existing delay models to handle inter-cycle demand fluctuations.

REFERENCES FOR PART II

Akcelik, R. and Roupail, N. M., 1993. Estimation of Delays at Traffic Signals for Variable Demand Conditions. *Transportation Research Part B* 27(2), 109-131.

Han, B. 1996. Optimising Traffic Signal Settings for Periods of Time-Varying Demand. *Transportation Research Part A* 30(3), 207-230.

Han, L. and Li, J.-M., 2007. Short or Long ... which is Better? A Probabilistic Approach towards Cycle Length Optimization. TRB 86th Annual Meeting Compendium of Papers CD-ROM. TRB, National Research Council, Washington, DC.

Heidemann, D., 1994. Queue Length and Delay Distributions at Traffic Signals. *Transportation Research Part B* 28(5), 377-389.

May, A. D., 1990. *Traffic Flow Fundamentals*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

Newell, G. F., 1965. Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light. *SIAM Review* 7(2), 223-240.

Strong, D. W. and Roupail, N. M., 2006. Incorporating Effects of Traffic Signal Progression into Proposed Incremental Queue Accumulation Method. TRB 85th Annual Meeting Compendium of Papers CD-ROM. TRB, National Research Council, Washington, D.C.

Transportation Research Board (TRB), 1985. Special Report 209: Highway Capacity Manual, 3rd ed., TRB, National Research Council, Washington, D.C.

Transportation Research Board (TRB), 1994. Special Report 209: Highway Capacity Manual, 3rd ed., 2nd rev. TRB, National Research Council, Washington, D.C.

Transportation Research Board (TRB), 1997. Special Report 209: Highway Capacity Manual, TRB, National Research Council, Washington, D.C.

Transportation Research Board (TRB), 2000. Special Report 209: Highway Capacity Manual, TRB, National Research Council, Washington, D.C.

Webster, F., 1958. Traffic Signal Settings, Road Research Technical Paper No. 39. Road Research Laboratory, Her Majesty's Stationery Office, London, U.K.

**PART III. SHORT OR LONG ... WHICH IS BETTER?
A PROBABILISTIC APPROACH TOWARDS CYCLE LENGTH
OPTIMIZATION**

This part is a slightly revised version of a paper with the same title presented in TRB 86th Annual Meeting, and also accepted to be published in Transportation Research Report by Lee Han and Jan-Mou Li:

Han, L. and Li, J.-M., 2007. Short or Long ... which is Better? A Probabilistic Approach towards Cycle Length Optimization. TRB 86th Annual Meeting Compendium of Papers CD-ROM. TRB, National Research Council, Washington, DC.

My primary contributions to this paper include (1) development of the problem into a work relevant to my doctoral research study, (2) development of experimental setup, (3) most of the gathering and interpretation of literature, (4) performing the laboratory experiments, (5) interpretation and analysis of test results, (6) most of the writing.

ABSTRACT

Traffic-signal timing would be a trivial undertaking if demand were constant and uniform. Once stochastic factors and demand fluctuation are taken into consideration, however, the optimization of signal timing becomes challenging if not impossible, even for an isolated, fixed-time signal. To answer the question of whether a longer cycle, e.g. more than 150 seconds, or a shorter one, e.g. less than 60 seconds, is better under fluctuating demand conditions, this paper employs a probabilistic approach to studying minimal average delay by the use of mathematical formulations and Monte Carlo simulations. The idea is to select a cycle length that is small enough to insure low delay and, hence, level of service, yet still provide adequate capacity to handle most of the fluctuating demand conditions.

A five-step framework is presented for carrying out the analyses, which are demonstrated using a hypothetical example. Subsequent sensitivity analyses, level-of-service assessment, and cycle failure rate estimation were conducted based on random demand and are presented herein. Conclusions of the paper include 1) the introduction of fluctuating demand level increases the average delay in general; 2) longer cycle lengths do not yield optimal delay results; and 3) with extremely short cycle lengths, delay is usually high due to a lack of capacity and, hence, guarantees frequent cycle failures. A major contribution of this paper is a proposed framework for optimizing cycle length under stochastic inter- and intra-cycle demand levels based on the expectation

function of delay. When deployed, this framework can aid traffic engineers in choosing the desirable cycle length for minimal delay or for any, reasonable, LOS requirements.

INTRODUCTION

Since the introduction of automatic traffic signals in 1926, signal-timing optimization has become a classic problem, but not always a satisfactorily solvable one, mainly because of demand fluctuation over time. The fact is that if demand were constant or largely predictable, signal-timing optimization would be eminently solvable, indeed trivial. However, in practice, this problem is not at all easy to solve, and there has been little agreement among experts. There have been practitioners in favor of long cycle durations (e.g. more than 150 seconds) because less lost time is observed over a period, say an hour, of time; yet there are also proponents for snappier cycle durations (e.g. less than 60 seconds) to avoid the buildup of queues. To help clarify this issue, this paper will examine how cycle length should be chosen, whether short or long or neither, for fixed time signals in isolated intersections when all other factors are unchanging. (Actuation and other adaptive means that vary cycle length in real time have the potential to address this issue but would also bring much complexity into this matter, so they are not here addressed.)

Vehicular delay as a result of traffic signals has been commonly identified as a primary measure of motorists' perception of how well a signalized intersection operates. Therefore, timing optimization often involves the minimization of vehicular delay. It is common to balance volume to capacity, or v/c , ratios of critical movements for this purpose. For example, Webster's optimal cycle length formulation (Webster, 1958) can be expressed as:

$$X_w = \frac{5 + 1.5L}{1 - \sum_i y_{ci}}, X_{min} \leq X_w \leq X_{max} \quad (1)$$

where

X_w is the optimal cycle length per Webster;

L is the total lost time;

$\sum_i y_{ci}$ is intersection critical flow ratio, i.e. the approach volume divided by

saturation flow rate, for critical movements or lane groups i ; and

X_{min} and X_{max} are the practical boundaries of cycle lengths.

To keep matters simple, early studies often assumed uniform arrival at intersections (Webster, 1958). According to Newell (1965), Clayton may have been the first to propose the earliest and simplest models of traffic flow through intersections (Clayton, 1941), and Wardrop seems to have been the first to report any calculations of random delays at signal-controlled intersections (Wardrop, 1952). The most extensive work on this subject was conducted by Webster (1958), who derived formulas for the average delay by fitting curves to data with Monte Carlo simulations.

The first attempts at analytical models of fixed-cycle traffic signals were by Beckmann et al. (1956) and Newell (1956). Newell proposed a model in which arrival headways were independently and identically distributed (IID) random variables of arbitrary distribution while departures were regularly spaced during the green. It is noteworthy that these classic works were conducted based on

queuing analysis within a single cycle and, hence, dealt only with intra-cycle demand fluctuations. This was the reason why some researchers, e.g. Han (1996), considered these early works to be based on time-stationary assumptions.

Subsequent studies modeled delay using various approaches. Akcelik and Roupail (1993) considered symmetrical triangular and parabolic functions as variable demand functions during the analysis period and proposed a delay model that was suitable for variable demand conditions. Heidemann (1994) assumed that the number of vehicles arriving during each time interval was considered to be stochastic, and the arrival distributions for different time intervals were assumed to be identical and stochastically independent.

Han may have been the earliest to report that when traffic demands are different in successive time periods, the signal settings that are optimal for each individual period are only local solutions to the problem. He developed a sequential optimization technique to minimize the total intersection delay over successive periods by searching for the optimal signal timings. But one of the primary assumptions in his work was that the traffic demand, though varying from one period to the next, stayed constant in each individual period. Chang and Lin (2000) pointed out that conventional signal-control strategies were inadequate because the designed and “optimized” timing is only considered for the next single cycle after the current one, instead of for the entire congestion period. They proposed a timing-decision methodology that minimized total intersection delay during the entire oversaturated period, not just for a single cycle. However,

they did not specifically address demand variation over the entire congestion period.

Although each was valuable, none of these studies formally addressed the effects of inter-cycle as well as intra-cycle fluctuations of traffic demand. The issue here is that inter-cycle fluctuation will affect the result of delay analysis because unused capacity that is due to a momentary drop in demand from one cycle is capacity that is lost forever and cannot be reclaimed to help future temporary surges in demand. To properly represent this in queueing analysis and, hence, obtain better delay estimation, there should be multiple piecewise segments instead of a single straight line (e.g., 1,000 vph) when demand is aggregated over time (depicted in Figure 3.1). This representation of fluctuating demand would lead to different, most likely higher, delay values, and subsequently a different optimization scheme.

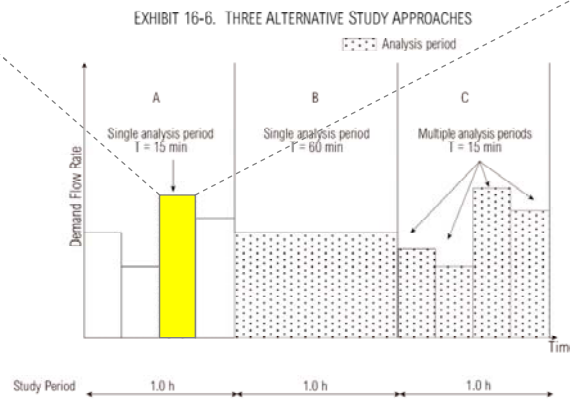
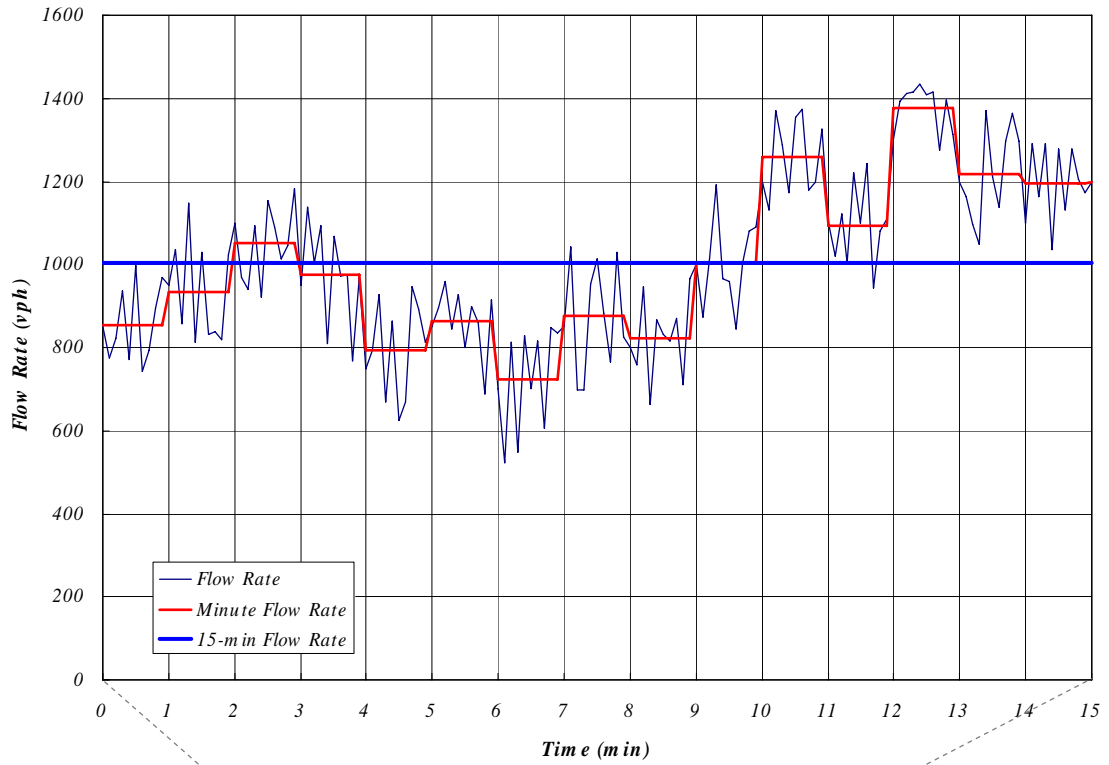


Figure 3.1. Demand fluctuations observed at different time scales
(HCM 2000 Exhibit 16-6, courtesy of TRB)

THE ANALYTICAL FRAMEWORK

The framework proposed herein uses an expectation function to quantify the variability of demand in the Highway Capacity Manual (TRB, 2000), or HCM, delay function. Each uncertain variable is considered as a random variable with an appropriate probability distribution, mean, and variance. Since delay is a function of the input variables, the resultant delay value also takes on the characteristics of a random variable with expectation and variance that can be calculated based on those of the input variables.

To ensure easy adoptability, this framework was designed to cope with demand fluctuation and to minimize average delay without resorting to any actuated hardware or algorithms. Since the basic relationships among delay, demand, and cycle length had been studied for some time, several models have been proposed and used to date. There are some differences in random terms and empirical correction terms in these models (e.g. Webster, 1958; Newell, 1956; Han, 1996; Akcelik and Roupail, 1993; and Heidemann, 1994), but the primary parts of the relationships are the same. Among these models, HCM 2000 delay function is the most commonly accepted and used:

$$d = d_1(PF) + d_2 + d_3 \quad (2)$$

where

d = control delay per vehicle (usually in seconds/vehicle or s/v);

d_1 = uniform control delay assuming uniform arrivals;

PF = uniform delay progression adjustment factor, which accounts for effects of signal progression;

- d_2 = incremental delay to account for effect of random arrivals and oversaturation queues, adjusted for duration of analysis period and type of signal control; this delay component assumes that there is no initial queue for lane group at start of analysis period; and
- d_3 = initial queue delay, which accounts for delay to all vehicles in analysis period due to initial queue at start of analysis period.

If there is no initial queue, then d_3 equals 0, and the function of average delay in HCM 2000 can be expanded as:

$$d = \frac{0.5\phi \left(1 - \frac{g_e}{\phi}\right)^2}{1 - \min(1, X) \frac{g_e}{\phi}} \frac{(1-P)f_{pa}}{1 - \frac{g_e}{\phi}} + 900T \left[X - 1 + \sqrt{(X-1)^2 + \frac{8kIX}{cT}} \right] \quad (3)$$

where

- d = control delay per vehicle (in seconds);
- g_e = effective green time (in seconds or s);
- ϕ = cycle length (s);
- X = lane group demand/capacity, or v/c , ratio or degree of saturation;
- P = proportion of vehicles arriving on green;
- f_p = supplemental platoon adjustment factor;
- T = duration of analysis period (in hour);
- k = incremental delay factor;
- l = upstream filtering/metering adjustment factor;
- c = lane group capacity (vehicles per hour or vph).

There are a few “hidden” variables, including demand, v , hidden within the degree of saturation or X , and cycle length, ϕ , within the term of lane group capacity, c . If the effective green time for a lane group is given by $g_e = \phi/2 - L$, and L represents the lost time, then the proportion (λ) of the cycle, which is effectively green of the phase under consideration, can be expressed as $\lambda = g_e/\phi = (\phi-2L)/2\phi$. If we represent the saturation flow rate for a subject lane group, then the lane group capacity will be $c = s g_e/\phi = s(0.5-L/\phi)$; and the degree of saturation can be expressed as $X = v/c = 2v/\phi s(0.5-L/\phi)$. Substituting these into Equation 3, one can see the relationship of delay as a function of cycle length and demand:

$$d = \frac{0.5\phi \times (\phi + 2L)}{2\phi - \min\left(1, \frac{2v\phi}{s(\phi - 2L)}\right) \times (\phi - 2L)} \times (1 - P) f_{pa} + \frac{900T}{s(\phi - 2L)} \left[2v\phi - s(\phi - 2L) + \sqrt{\left[2v\phi - s(\phi - 2L)\right]^2 + \frac{32kIv\phi^2}{T}} \right] \quad (4)$$

With a simple hypothetical example of traffic demand following normal distribution $N(720, 72^2)$, or a mean of 720 vphpl and a standard deviation of 72 vphpl, and a 50-50 split of the signal, Figure 3.2 illustrates the probabilistic distribution of average delay as a function of different cycle lengths. In general, as cycle length decreases, delay also decreases. However, since capacity also decreases with the cycle length, when the fluctuating demand exceeds the now nearly inadequate capacity, delay increases drastically.

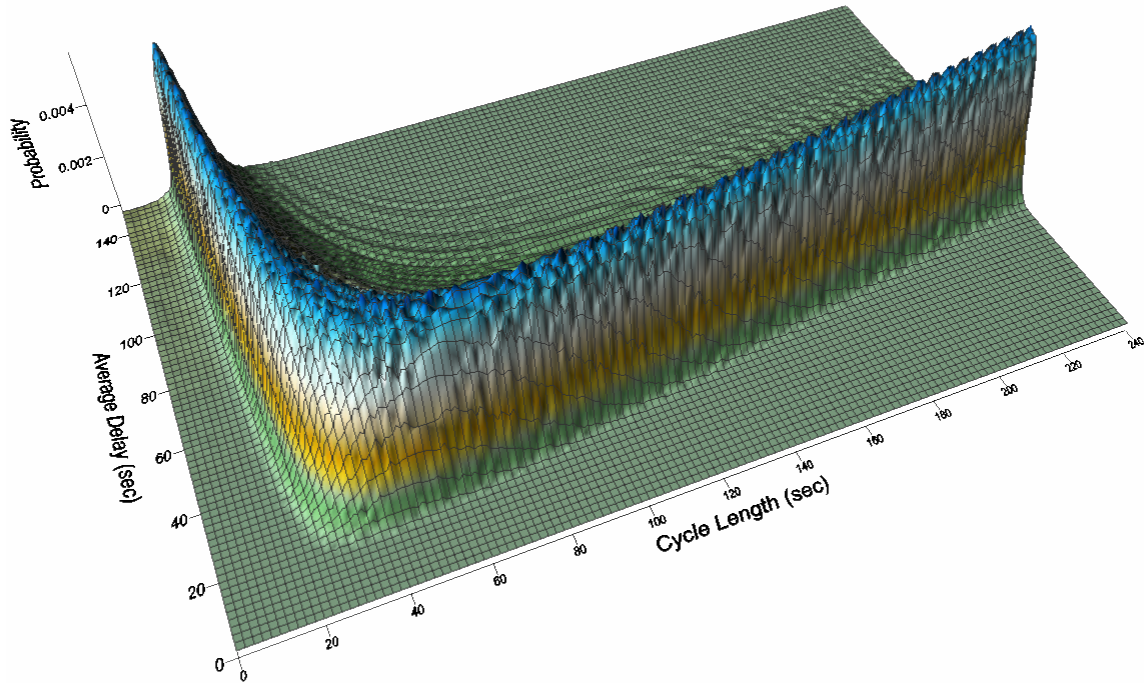


Figure 3.2. Basic probabilistic relationship between cycle length and demand.

The idea here, then, is to select a cycle length that is small enough to result in low delay yet will still provide adequate capacity to handle most of the fluctuating demand conditions.

According to the delay model, the particular cycle length ϕ^* that minimizes average delay under a certain demand v_c can be derived by setting the derivative of delay with respect to the cycle time equals zero:

$$\frac{\partial d}{\partial \phi} = 0 \tag{5}$$

If demand were constant, one could solve for the optimal cycle length for various demand values, as depicted in Figure 3.3, where each dashed line represents a

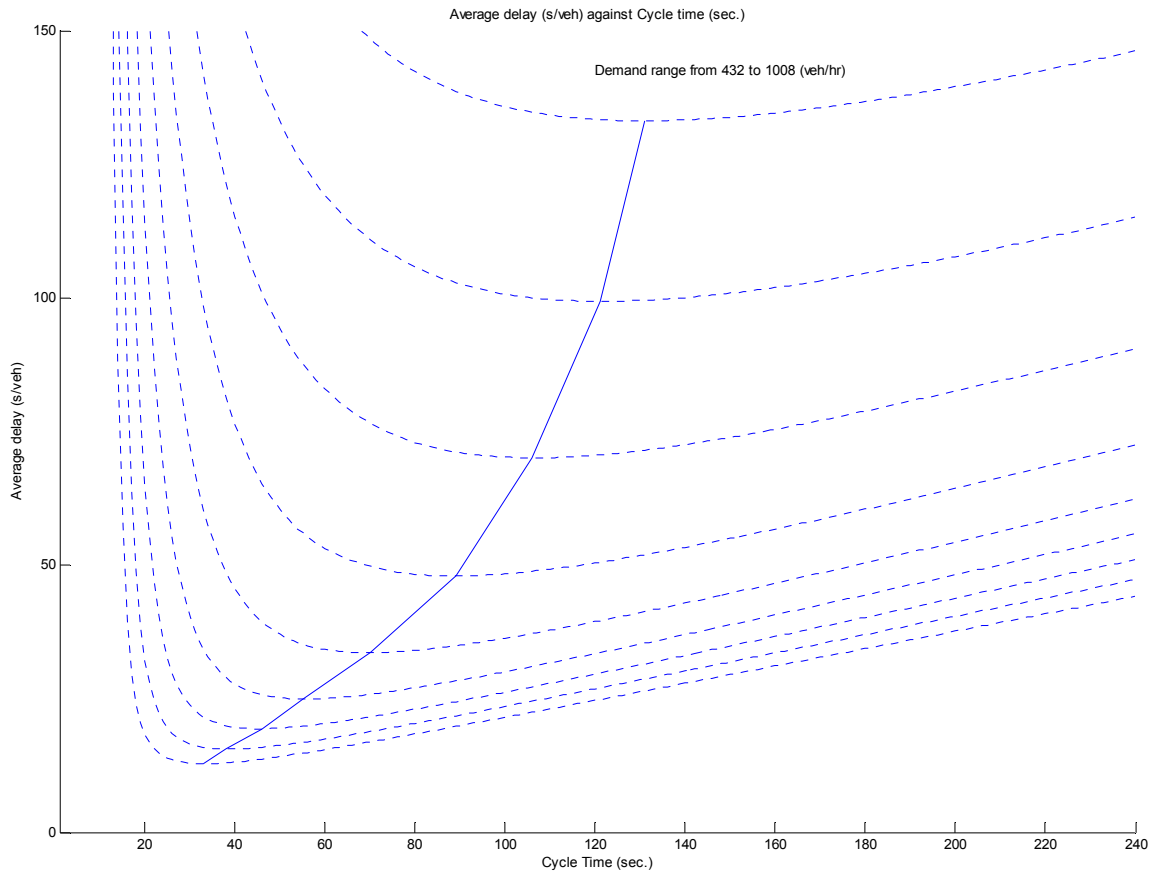


Figure 3.3. Average delay as a function of cycle length

demand level (between 432 and 1008 vphpl) and where the loci are optimal solutions attainable by solving an array of Equation 5. In reality, however, demand fluctuates over time, not only within a cycle but also among cycles. The optimal solution yielding the least amount of delay for a demand of 500 vph suddenly becomes woefully inadequate when the demand momentarily jumps to 1,000 vph for a minute or so. On the other hand, the optimal solution for a demand of 1,000 vph would unnecessarily exact a higher average delay if the demand were to drop momentarily to 500 vph. Under the assumption of fixed-

time signal plans, the traffic engineer might just set the cycle length based on the solution to Equation 5 and hope for the best. Yet a better alternative is to study the probabilistic nature of the fluctuating demand and then select a cycle length that will accommodate the demand at least, say, 95% of the time.

For any given cycle length selected by the traffic engineer, the expected average delay can be expressed in terms of the deterministic delay function and the probabilistic distribution of the fluctuating demand:

$$E[d/\phi = \phi_s] = \int_0^{\infty} f_d(\phi_s, v) f_p(v) dv \quad (6)$$

where

ϕ_s = a selected cycle length

v = demand, a random variable

$f_d(\phi, v)$ = the average delay as a function of cycle length and demand

$f_p(v)$ = the probability density function of demand, v

Now, as the traffic engineer tries out different cycle lengths, the expected average delay is now also a function of cycle length:

$$E[d] = \int_0^{\infty} f_d(\phi, v) f_p(v) dv \quad (7)$$

To optimize the average delay under varying demand conditions, one would solve Equation 8:

$$\frac{\partial E[d]}{\partial \phi} = 0 \quad (8)$$

The analytical solution is a complex one. But since the equation is a strictly convex function, as shown in Figure 3.3, the solution can be approached

numerically. A numerical analysis-based implementation procedure using this analytical framework is presented in the next section.

IMPLEMENTATION PROCEDURE

With a typical personal computer, numerical solutions to the cycle length optimization problem can be obtained. The following procedure, which consists of five easy steps, employs Monte Carlo simulation as a tool for the purpose. Similar approaches with other numerical tools would also work.

Step 1. Analysis Duration Selection. Traffic demand fluctuates throughout the day. Depending on the study period at hand, whether a.m. peak, p.m. peak, midday, Sunday noon, or other periods, different stochastic characters may manifest differently.

Step 2. Demand Fluctuation Assessment. In order to properly represent the stochastic nature of the traffic demand during the study period, one has to collect some field data and, subsequently, analyze the demand data with statistical techniques. Some statistical approaches, e.g. kernel density estimation (Goldberg, 1988), may be employed. Obviously, a better representation of the demand distribution is likely to lead to a better estimation of delay.

Step 3. Delay Model Selection. A number of delay models have been developed over the years. The most commonly used by practitioners is the average delay equation in HCM 2000. In addition to traffic demand, several parameters must be provided for this model. Since isolated intersection is the aim of this study, parameters associated with arrival

pattern, platooning, and upstream signals, i.e. P , f_{PA} , and I , should take on default values.

Step 4. Expected Delay Calculation. Equation 7 is the primary formula for this step. A simple series of Monte Carlo simulations can be employed to numerically approach the expectation and standard deviation of average delay. The cycle length associated with minimized average delay will also result through this process.

Step 5. Optimal Cycle Length Determination. After the cycle length for minimal average delay is identified, professional judgment still needs to be exercised. In particular, the traffic engineer should carefully choose a tolerance level of the cycle failure rate, say 1% of all cycles, before the implementation of the result. Otherwise, a higher ratio of cycle failure may occur.

The following section will demonstrate how a hypothetical scenario might be optimized.

SOLUTION DEMONSTRATION

Hypothetical Scenario

To simplify a rather complex process, the hypothetical case involves a single traffic signal standing at the intersection of two one-way, one-lane roads. The analysis is to be carried out for a.m. peak, roughly between 7:00 and 9:00 a.m., with multiple analysis periods of 15 minutes, i.e. $T = 0.25$. The demand for either direction is assumed to follow normal distribution with a mean of 720 vph and a standard deviation of 72, or $v \sim N(720, 72^2)$. The average delay function in HCM 2000 will be the delay model for this case with no initial queues, i.e. d_3 in Equation 2 equals 0.

The scenario assumes a saturation flow rate of 1,800 vphpl. The lost time per cycle per approach is assumed to be 4 seconds, per HCM 2000. Based on a queueing process with random arrivals and uniform service time equivalent to the lane group capacity, the incremental delay factor for the delay function was set to 0.5. Since the progression adjustment factor depends on proportion of vehicles arriving on green, proportion of green time available, and supplemental adjustment factor for platoon arriving during green, for random arrivals (arrival type 3), the metering adjustment factor (or upstream filtering), the platoon ratio, and the supplemental adjustment factor for platoon arriving during green were set to the default values of 1. Under these assumptions, the v/c ratio should fall between 0.72 and 1.28 for 99% of the time.

Monte Carlo Simulation and Results

Monte Carlo simulation, which has been widely employed for traffic analysis (e.g. Watling, 1996; Chen et al., 2002; and Tarko and Tracz, 2000), is used for this study. Even though the expectation function of average delay is complex in its mathematical form, its values can be calculated based on the parameters given herein. The basic logic for the calculation involves generating a random demand level, calculating average delay values for this particular demand level with an array of the cycle lengths under consideration, and repeating the process for numerous times. In fact, 100,000 demand samples were randomly generated, to mimic the actual distribution, for this purpose. The resultant relationship between delay and cycle is shown in Figure 3.4, where the convexity of the expectation function is evident. With extremely short cycle lengths (less than 40 seconds), delay is high due to a lack of capacity and, hence, it is guaranteed that there will be frequent cycle failures. The minimal average delay of about 37.5 seconds is reached at a cycle length of around 75 seconds.

Figure 3.4 also shows that an increase of cycle length beyond 75 seconds would push the delay value even higher. The numerous parallel vertical lines represent the results from 10,000 simulations under each cycle length. The thick line is the average delay of all outcomes for each cycle length case.

Sensitivity Analyses

It should be noted that the minimal average delay of 37.5 seconds at 75 seconds of cycle length is slightly higher than the HCM 2000 solution, via Equation 5, of

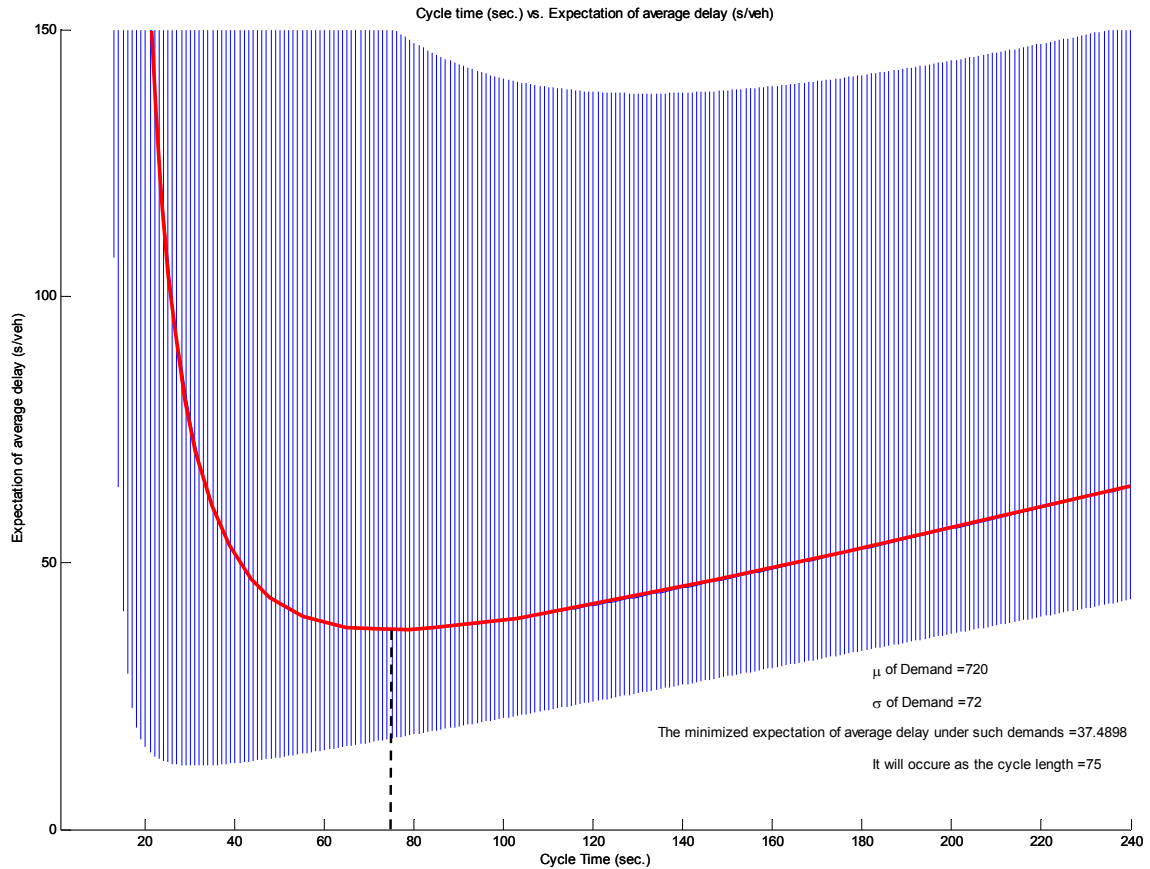


Figure 3.4. Expectation and distribution of average delay

33.5 seconds' average delay at a cycle length around 70 seconds if the demand were a constant 720 vph. The introduction of fluctuating demand level into Equation 5 increases the average delay in general.

A series of sensitivity analyses was conducted for different demand and variance levels. Table 3.1 summarizes cases with the same average demand of 720 vph but different standard deviations, i.e. 72, 81, 90, 99, and 108 vph. It is clear that increasing the magnitude of demand fluctuation, while the average

Table 3.1 Sensitivity Analysis Results

<i>Average Demand v, in vph</i>	<i>Std. Deviation of Demand σ, in vph</i>	<i>Expected Delay $E(d)$, in seconds</i>	<i>Cycle Length ϕ^*, in seconds</i>
720	0	33.5	70
720	72	37.5	75
720	81	38.5	76
720	90	39.5	77
720	99	40.5	78
720	108	41.5	79
720	90	39.5	77
810	90	59.6	95
900	90	89.3	113

demand stays the same, results in a higher expectation of delay and a slight increase in optimal cycle length.

Holding the standard deviation the same, at 90 vph, and increasing the demand from 720 to 810 to 900 vph, the study saw significant increase in minimal average delay. It appears that minimal delay is more sensitive to average demand than the standard deviation, as one would expect. But delay does increase as a direct result of increase in either case.

Design for Extremities

Since each individual vertical line in Figure 3.4 represents the outcomes of 10,000 “runs,” a statistical understanding of the overall outcomes can be explored. Figure 3.2 illustrates the overall shape of the distributions, where each slice of the 3D form at a given cycle length shows the probabilistic distribution of delay based on fluctuating demand. As such, one could connect the 95-percentile points of individual slices and form a 95% line, as seen in Figure 3.5, representing delay as

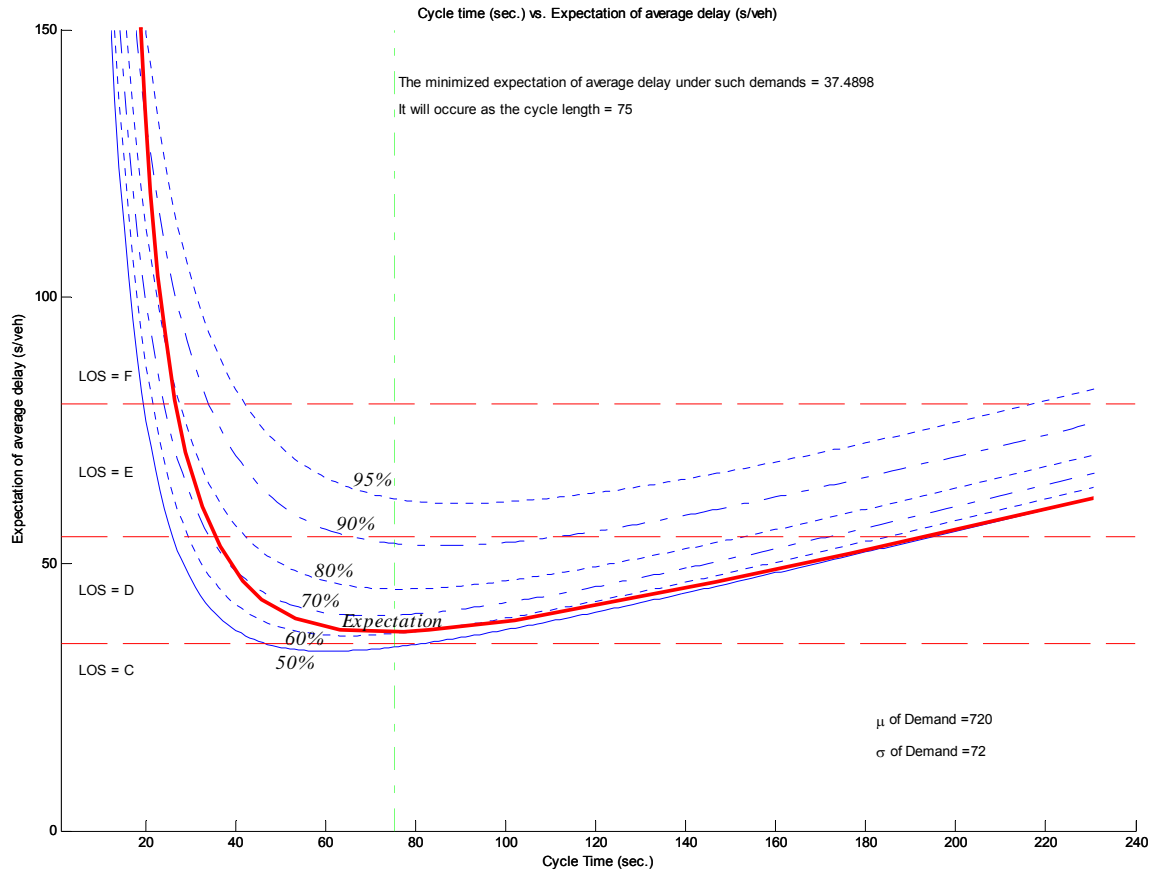


Figure 3.5. Percentile of delay distribution and level of service (for the hypothetical scenario)

a function of cycle length when demand has fluctuated to such a high level that only 5% of the time could it be any higher. Similarly, lines for 90%, 80%, and so on could also be constructed.

Some practitioners argue that a cycle length larger than the optimal one should be used to accommodate higher than usual demand conditions due to fluctuation. Figure 3.5 provides a tool for such purpose, as does Table 3.2. One could choose to accommodate demand at 95% or other levels if one is seriously worried about cycle failure under extreme demand surge conditions. It should be

Table 3.2. Percentiles of Demands Targeted and Associated Cycle Length
(for the hypothetical scenario)

<i>Percentiles of Demands</i>	<i>Cycle Length</i>	<i>Associated Expectation of Average Delay</i>
95 th	99	39.2
90 th	94	38.6
80 th	85	37.8
70 th	79	37.6
<i>Optimal Case</i>	75	37.5
60 th	74	37.5
50 th	70	37.6
40 th	65	38.0
30 th	62	38.4
20 th	57	39.5
10 th	52	41.3

expected, however, that a longer cycle length geared towards higher demand conditions will inevitably yield an average delay higher than the minimal one.

Level of Service Expectations

Since level of service (LOS) is also based on delay, or more precisely control delay, one can also gain a sense of how an intersection may operate in a probabilistic sense. For instance, if the traffic engineer has decided to go with a cycle length of 75 seconds, Figure 3.5 would suggest that the intersection LOS has a 52% probability to be equal or better than C, 39% probability to be D, and 9% to be E or worse.

This is quite different from solving Equation 5 and obtaining a singular delay of 33.5 seconds and a LOS of C, but perhaps more useful and insightful. If the engineer should decide to adopt different cycle lengths of, say, 45, 60, 90, or 105 seconds, the results from this analysis (see Table 3.3) would provide a more informative picture of the effects of cycle length choices.

**Table 3.3. LOS Probability as a Function of Cycle Length
(for the hypothetical scenario)**

LOS	$\phi = 45$	$\phi = 60$	$\phi = 75$	$\phi = 90$	$\phi = 105$
C or Better	49%	55%	52%	48%	39%
D	33%	33%	39%	45%	52%
E	14%	11%	9%	7%	9%
F	4%	1%	0%	0%	0%

For decision-making purposes, Table 3.3 also provides an overall perspective of various “*what-ifs*” as alternatives can be chosen to achieve different objectives. For example, if the traffic engineers were more concerned about reducing the probability for LOS E or worse conditions, he might choose a cycle length of 90 seconds instead of the optimal 75 seconds. Similarly, if a cycle length allowing the largest possible portion of the motorists to enjoy LOS C or better is desired, then the traffic engineer could adopt a cycle length of 60 seconds.

Cycle Failure Probabilities

Based on HCM 2000, cycle failure occurs when a given green phase does not serve queued vehicles and an overflow situation occurs. Although this may not always be a serious problem in theory, as suggested by HCM 2000 that “*individual cycle failures may begin to appear at LOS C,*” most traffic engineers would try to avoid the frequent occurrence of such events.

In the demonstration scenario, where the demand follows $N(720,72^2)$, the probability of cycle failure, when cycle length is set to 75 seconds, is about 0.5%. A similar scenario with a demand following $N(810,81^2)$ would have a much larger 8% probability of cycle failure, mainly due to the high average demand.

Perhaps the probability of cycle failure could be used, in addition to delay, as a secondary measure of effectiveness (MOE) in future traffic signal-timing analysis.

CONCLUSIONS

This paper employs a probabilistic approach to consider cycle-length optimization for isolated intersections and attempted to answer the question of whether shorter or longer cycles are “better.” With mathematical formulations and Monte Carlo simulations, the authors established that certain “just right” cycle lengths could be derived following a five-step optimization framework. A hypothetical example was then presented to demonstrate how the framework functions with ensuing analyses and discussions on sensitivity of the solution, expected LOS, and potential cycle failures. It is the intent of the authors to call attention to the following points:

- The introduction of fluctuating demand level increases the average delay in general;
- Longer cycle lengths of, say, more than 150 seconds in the presented case, do not yield optimal delay results; because the delay will increase with the cycle length due to the basic relation among demand, cycle length, and the delay (as shown in Figure 3.3).
- Very short cycles, much shorter than the optimal, will increase delay dramatically due to a lack of capacity. The consequence often includes cycle failure.
- A probabilistic approach to delay calculation, while more cumbersome than a straightforward fixed-demand and one-delay-value process, is more realistic and insightful; and

- The framework developed in the paper could be employed for better signal timing purposes.

A major contribution of this paper is a proposed framework for optimizing cycle length under stochastic inter- and intra-cycle demand levels based on the expectation function of delay. When deployed, this framework can aid traffic engineers in choosing the desirable cycle length for minimal delay or for any, reasonable, LOS requirements.

RECOMMENDATIONS

A rather simplistic hypothetical case was employed to demonstrate the utility of the proposed framework. To follow up this study, three future tasks have been identified. First, more realistic, and complex, signal control cases should be tested under this framework. Would the conclusions hold when pedestrians, left turn movements, and even multiple lanes compromise the promising results?

All delay calculations were based on HCM delay model, a commonly accepted one among many other delay models. Future studies using other models in comparison with real-world cycle length experiments would be desirable.

Finally, as identified in the paper, tabulation in the format of Table 3.3 can serve as a useful decision tool for traffic engineers. A closer look into how this tool may be used and improved is planned for the follow-up effort of this study.

REFERENCES FOR PART III

Akcelik, R. and Roupail, N. M., 1993. Estimation of Delays at Traffic Signals for Variable Demand Conditions. *Transportation Research Part B* 27(2), 109-131.

Beckmann, M., McGuire, C. B., and Winston, C. B., 1956. Studies in the Economics of Transportation, Yale University Press, New Haven.

Clayton A. J. H., 1941. Road Traffic Calculations. *Journal of Institute of Civil Engineers* 16, 247-284 and 558-594.

Chang, T.-H. and Lin, J.-T., 2000. Optimal Signal Timing for an Oversaturated Intersection, *Transportation Research Part B* 34(6), 471-491.

Chen, A., Yang, H., Lo, H.K., and Tang, W.H., 2002. Capacity Reliability of a Road Network: an Assessment Methodology and Numerical Results. *Transportation Research Part B*, 36(3), 225-252.

Goldberg, D. E., 1988. Genetic algorithms in search, optimization, and machine learning, Reading, Mass, Addison-Wesley Pub. Co.

Han, B. 1996. Optimising Traffic Signal Settings for Periods of Time-Varying Demand. *Transportation Research Part A* 30(3), 207-230.

Heidemann, D., 1994. Queue Length and Delay Distributions at Traffic Signals. *Transportation Research Part B* 28(5), 377-389.

Newell, G. F., 1956. Statistical Analysis of the Flow of Highway Traffic through a Signalized Intersection. *Applied. Mathematics Quarterly* 13, 353-369.

Newell, G. F., 1965. Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light. *SIAM Review* 7(2), 223-240.

Tarko, A. and Tracz, M., 2000. Uncertainty in Saturation Flow Predictions. Proceedings of 4th International Symposium on Highway Capacity.

Transportation Research Board (TRB), 2000. Special Report 209: Highway Capacity Manual, TRB, National Research Council, Washington, D.C.

Wardrop, J. G., 1952. Some Theoretical aspects of road traffic research. Proceedings of Institute of Civil Engineers II, Vol. 1, No. 2, pp.325-362.

Watling, D., 1996. Asymmetric Problems and Stochastic Process Models of Traffic Assignment, *Transportation Research Part B*, 30(5), 339-357.

Webster, F., 1958. Traffic Signal Settings, Road Research Technical Paper No. 39. Road Research Laboratory, Her Majesty's Stationery Office, London, U.K.

**PART IV. A TRADE-OFF FRAMEWORK FOR DETERMINING THE
BEST CONTROL AT AN INTERSECTION**

This part is a slightly revised version of a journal paper by Lee Han and Jan-Mou Li with the same title that will be submitted for review in 2007.

Han, L. and Li, J.-M., 2007. A Trade-Off Framework for Determining the Best Control at an Intersection. To be submitted for review.

My primary contributions to this paper include (1) development of the problem into a work relevant to my doctoral research study, (2) development of experimental setup, (3) most of the gathering and interpretation of literature, (4) performing the laboratory experiments, (5) interpretation and analysis of test results, (6) most of the writing.

ABSTRACT

Although there are warrants for signal, guidance and criteria for stop signs, how to determine an appropriate control type for an intersection is rarely discussed in literature. The trade-off can be accomplished by integrating information from several sources. Several models with different parameters, including traffic volume, can estimate the average control delay for an approach under a given condition. This paper proposes a framework based on the average control delay to determine the best control type for an intersection.

Since the sensitivities of the average control delay for different control types differ from each other in different traffic patterns, to facilitate the trade-off among control types is the most important function for the framework. Based on the simulations at the hypothetical intersection, an AWSC may be a better choice than a signal in which the two-way flow rate is less 600 vph in the major street, 350 vph or less in the minor street, and 10% left-turn traffic for each direction. It is recommended for the traffic condition not only because of the average control delay but also because of the sensitivity. It is worthwhile to notice that the sensitivities can be checked very easily through the framework.

INTRODUCTION

The most popular ways to assign the right of way at an intersection is to install either traffic signals or stop signs. It is definitely a complex issue because numerous factors are involved. That is why engineering judgments or studies are necessary for such an installation, even though there are already massive amounts of research dedicated to traffic signal or stop signs respectively. Since both traffic signals and stop signs are supposed to serve users in a more efficient way, a framework for the trade-off between a traffic signal and a stop sign will be very useful for traffic engineers. However, very little literature mentioned the trade-off between signals and stop signs for an intersection, either qualitatively or quantitatively.

Exhibit 10-15 (Figure 4.1) in Highway Capacity Manual 2000 edition (HCM 2000) (TRB, 2000) may be the most popular resource which can be referenced to make a decision about the control type at an intersection. According to the description in HCM 2000, it is used to forecast the likely intersection control types for future facilities. Unfortunately, the reference for that exhibit is incorrect, and therefore it can not provide any further information in order to validate the trade-off decision. If the traffic patterns will not change with different control types at an intersection, some of turning points on the exhibit will get confusing results, especially in higher traffic volume conditions. Table 4.1 shows the average control delays by Highway Capacity Software 2000 (HCS 2000) for a hypothetical

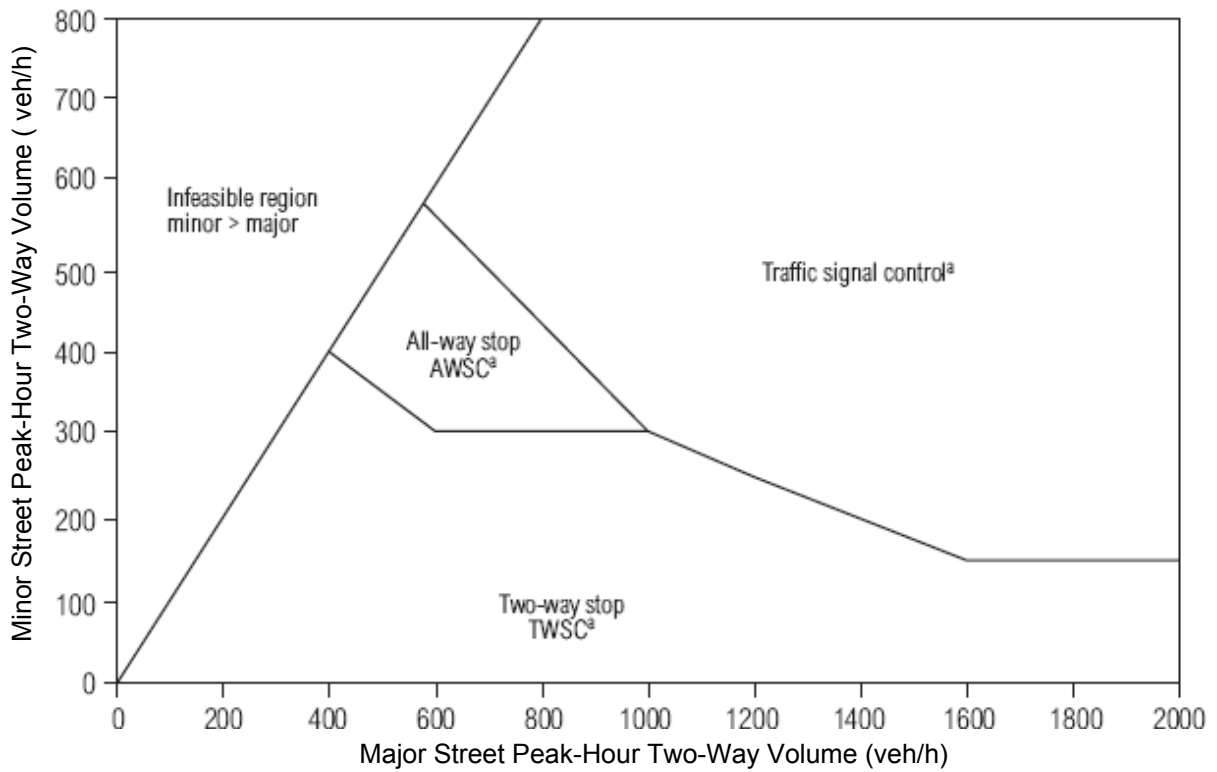


Figure 4.1. The Exhibit 10-15 in HCM 2000

Note: According to HCM 2000, it is adapted from Traffic Control Devices Handbook 1983 (FHWA, 1983), pp. 4–18.

Table 4.1. Average Control Delays for a Hypothetical Intersection with Through-Traffic Only

Type Volume**	Signal*	AWSC	TWSC
(2000,150)	170.50	241.77	151.80
(1800,150)	116.10	179.73	84.95
(1600,150)	67.00	121.12	47.20
(1000,300)	16.40	28.15	20.20
(600,600)	13.20	18.49	16.40
(600,300)	12.50	11.81	9.20
(400,400)	11.70	10.57	7.75

* The cycle length for the signal operation is 60 seconds.

** Peak-Hour two-way volumes for the major street and the minor street are shown in parentheses (major, minor).

intersection with through traffic only. Even though the cycle lengths are the same instead of optimal for different combinations of volumes, the average control delays for the signalized intersection are significantly lower than those with stop signs, when two-way traffic volumes in the major street are over 1000 and those in minor street are under 150.

There are eight warrants (FHWA, 2004) for justifying traffic control signals in chapter 4C of the Manual on Uniform Traffic Control Devices 2003 edition (MUTCD 2003). The first three warrants are relative to vehicular volumes. Since the manual has evolved over years, the warrants represents a threshold condition in the overall assessment of whether a traffic control signal may be justified based on a comprehensive engineering evaluation of the intersection's operations and safety benefits. They may raise the same questions as the Exhibit 10-15 did, because the conditions are quite similar to those on the Exhibit 10-15 in HCM 2000, especially in Warrant 3.

For both HCM 2000 and MUTCD 2003, delay is a very important measure on evaluating the performance of an intersection. Even though the thresholds of level of service (LOS) for different control types at an intersection are different (Table 4.2), control delay is the same cornerstone of LOS for both signal control and stop-controlled intersections. There are models (e.g. Webster, 1958; Akcelik and Roupail, 1993; TRB, 2000) to estimate the control delay at a signalized intersection. According to Han and Li (2007), all of these models agreed that the average delay will increase dramatically when the cycle length reduces below a

Table 4.2. Definitions of LOS for Signalized, TWSC, and AWSC Intersections

Level of Service	Signal	TWSC	AWSC
A	≤10	0-10	0-10
B	>10-20	>10-15	>10-15
C	>20-35	>15-25	>15-25
D	>35-55	>25-35	>25-35
E	>55-80	>35-50	>35-50
F	>80	>50	>50

* The unit for average control delay is second per vehicle

Source: Adapted from Highway Capacity Manual 2000 edition; data for Signal is from the Exhibit 16-2 in p. 16–2, data for TWSC is from the Exhibit 17-2 in p. 17–2, data for AWSC is from the Exhibit 17-22 in p. 17–32

certain threshold under a given demand. That is, optimal performance in a signalized intersection may be reached by a shorter cycle length, but not an extreme one, because some constraints, e.g. the lost time, may prevent extremely short cycle lengths from a better service (less delay) for an intersection.

According to McKinley (2001), one of the conditions for installing a traffic control signals is that an all-way stop-controlled (AWSC) intersection must be experienced increased delay and congestion. He also mentioned that the safest intersection control, assuming reasonable compliance with the law, is the AWSC, even though it is also the most inefficient type of intersection control in most cases. Even though extensive research is being done into the performance of stop-controlled intersections, arguing for “the most inefficient type” quantitatively is still a challenge. Sampson (1999) proposed the 4Q/6Q warrant to justify a signal. He argued the warrant based on queue is sensitive to a wide range of variables, e.g. site geometry and visibility, turning volume, and speeds.

However, such a warrant can not be used for planning purposes. That is why the average control delay is used as the criterion in this paper to compare the performance at an intersection between with signal control and with stop control.

Richardson (1987) argues that delays at an AWSC intersection are the result of a set of complex interactions between the flows on all approaches to the intersection. He proposed an iterative method and used the Pollaczek-Khintchine formula to estimate delays of AWSC intersections, based on an M/G/1 model of queuing process. Although the subject delay in his model is a function of subject, conflicting, and opposing flow rate, statistical analysis suggests that this model might provide a credible estimate of delay (Kyte and List, 1999). According to the analysis by Kyte and Marek (1989), Richardson's queuing model provides good estimates of vehicle delay for subject flow rates up to 400 to 450 vehicles per hour (vph), but the model will give poor results once the flow rates over this threshold.

Eck and Biega (1988) conducted a before-and-after analysis in order to evaluate the TWSC and AWSC at low-volume intersections in residential areas. In general, they concluded that four-way stop sign control at low-volume residential street intersections should be changed to two-way stop sign control, because the use of two-way stop sign control in place of four-way stop sign control minimizes delay and road user costs. Chan et al. (1989) proposed a response-surface model with four determinants, i.e. traffic volume, volume split, percentage of left-turns, and street width, to estimate average delay at an AWSC intersection. One of their findings is highly controversial in relation to that by Zion

et al. (1989), that is, the more imbalanced the volume split is, the smaller the delay. Zion et al. (1989) tested delay models, which proposed by Richardson (1987) and by Chan et al. (1989), with field data for AWSC intersections. What they found are that delay increases as the intersecting volume increases; intersections with balanced volumes have lower delays than those without; and the percentage of left turns has a noticeable effect on delay.

Besides AWSC, a two-way stop-controlled (TWSC) intersection is another, and maybe more efficient, type of assigning the right-of-way with the stop sign at the intersection. Byrd and Stafford (1984) examined the operational characteristics of traffic controls at low-volume, low-speed intersections with unwarranted four-way stop sign control. Then they suggested that unless an accident problem susceptible to correction by four-way stop sign control exists, the unwarranted use of four-way stop sign control results in unnecessary delay and road user costs to the driving public and that the intersection traffic control should be changed to two-way stop sign control.

Kyte et al. (1997a, 1997b) conducted research about the capacity and level of service at unsignalized intersections. In the second volume of the final report (1997b), they proposed the saturation headway at an AWSC intersection is dependent on the degree of conflict, the geometry, the directional movements of the interacting movements, and vehicle types. In the ninth chapter in the first volume of the final report (1997a), a comparison is made between the peak hour signal warrant of the MUTCD (FHWA, 1988) and the recommended capacity and LOS procedure. They did not reach a similar result to what Byrd and Stafford

(1984) did. Under their assumptions, they found that AWSC control is best applied when a balanced volume distribution on the major and minor street of intermediate magnitude is achieved; there is a fairly good correlation between the MUTCD (FHWA, 1988) signal peak hour warrants and the result obtained based on the operational assessment through the HCM (TRB, 1994) delay models. The most interesting achievement of their works is the figure related to optimum control type at an intersection based on the minimum average intersection delay and a five-second significant difference level. It looks just like the Exhibit 10-15 in HCM 2000, especially the area for AWSC.

Even though there are models to estimate the delay or capacity for stop-controlled intersections, some results were really confusing. Furthermore, there is still a shortage of the comparison of performance between stop-controlled and signal-controlled at an intersection. In order to clarify the trade-off among signal, AWSC, and TWSC, a trade-off framework to evaluate these three control types at an intersection is proposed in this paper. The average delay models for signalized and unsignalized intersections in HCM 2000 are used as the basis of the framework. Because combinations of traffic conditions at intersections are infinite, it is impossible to enumerate all possible combinations for the decision. Simulation results show that the intersection average delay will reach the lowest bound under symmetrical flows. Turning effect can be shown in sensitivity analyses with a certain model. Based on the framework, a stop sign is proper to cases of the two-way volume on both the major and minor street under about 500 vehicles per hour. This recommendation is quite different from the Exhibit 10-15.

The rest of this paper is organized as follows. The employment of methodologies, including delay models and sensitivity analyses, will be introduced and discussed in the coming section. Results from hypothetical scenarios, with no turns, left-turn involved, and approach-based, are shown in section three. Section four is designated to the findings according to the results, recommendations and discussions. Conclusions, such as a new trade-off framework, and no AWSC, are brought to the last section.

METHODOLOGY

While many methods are currently available to estimate the delays incurred at intersection approaches, either signal controlled or stop controlled, to achieve comparable results based on a common foundation is the most important criterion to choose proper models. That is, the definition and the calculation of the average control delay should be at the least same, even though assumptions and limitations may slightly differ from each model. Models in HCM 2000 and those in software packages (e.g. VISSIM) may meet this criterion in their own domain respectively, but they may not have the same foundation to calculate the average control delay in cases. Since models in HCM 2000 definitely meet the criterion, and may be the most popular methods to process the estimations in practice, they are used in this paper to illustrate the framework.

According to HCM 2000 (TRB, 2000), the control delay involves movements at slower speeds and stops on intersection approaches, as vehicles move up in the queue or slow down upstream of an intersection. The definition and use of control delay, including initial deceleration delay, queue move-up time, stopped delay, and final acceleration delay, is consistent between traffic signals and stop signs in HCM 2000. Drivers frequently reduce speed when a downstream signal is red or there is a queue at the downstream intersection approach. Control delay requires the determination of a realistic average speed for each roadway segment. Any estimate of the average travel speed on urban streets implies the effects of control delay. At two-way stop-controlled and all-

way stop-controlled intersections, control delay is the total elapsed time from a vehicle joining the queue until its departure from the stopped position at the head of the queue. The control delay also includes the time required to decelerate to a stop and to accelerate to the free-flow speed.

Twenty-five parameters in three categories (geometric, traffic, and signalization conditions) are required to conduct an operational analysis for signalized intersections, if the approach in HCM 2000 is employed. Eighteen of these parameters have suggested default values. The saturation flow rate, which is used in the model, is the flow in vehicles per hour that can be accommodated by the lane group assuming that the green phase were displayed 100 percent of the time. It is the most complicated parameter, because twelve other parameters, excluding the base saturation flow rate per lane, are related to this parameter. Therefore, the most geometric conditions and part of the traffic conditions in a site may be properly reflected by the saturation flow rate.

The values derived from the delay model in HCM 2000 represent the average delay experienced by all vehicles that arrive in the analysis period, and it is worthwhile to notice that, including delays incurred beyond the analysis period when the lane group is oversaturated. It is special because a general queueing analysis will not consider the delay incurred beyond the analysis period, whether the system is under-saturated or oversaturated. The compact form of the delay model for signalized intersections is expressed by Equation 1. It is used to estimate the average control delay per vehicle for a given lane group.

$$d = d_1(PF) + d_2 + d_3 \quad (1)$$

where

d = control delay per vehicle (usually in seconds/vehicle or s/v);

d_1 = uniform control delay assuming uniform arrivals;

PF = uniform delay progression adjustment factor, which accounts for effects of signal progression;

d_2 = incremental delay to account for effect of random arrivals and oversaturation queues, adjusted for duration of analysis period and type of signal control; this delay component assumes that there is no initial queue for lane group at start of analysis period; and

d_3 = initial queue delay, which accounts for delay to all vehicles in analysis period due to initial queue at start of analysis period.

If an initial queue is nonexistent, then d_3 equals 0; and the delay model for signalized intersections in HCM 2000 can be expanded as:

$$d = \frac{0.5\phi \left(1 - \frac{g_e}{\phi}\right)^2}{1 - \min(1, X) \frac{g_e}{\phi}} \frac{(1-P)f_{pa}}{1 - \frac{g_e}{\phi}} + 900T \left[X - 1 + \sqrt{(X-1)^2 + \frac{8kIX}{cT}} \right] \quad (2)$$

where

d = control delay per vehicle (usually in seconds/vehicle or s/v);

g_e = effective green time (in seconds or s);

ϕ = cycle length (s);

X = lane group demand/capacity, or v/c , ratio or degree of saturation;

P = proportion of vehicles arriving on green;

f_p = supplemental platoon adjustment factor;

- T = duration of analysis period (in hour);
- k = incremental delay factor;
- l = upstream filtering/metering adjustment factor;
- c = lane group capacity (vehicles per hour or vph).

The stop-controlled approaches are referred to as the minor street approaches at TWSC intersections; and those are not controlled by stop signs are referred to the major street approaches. Average control delay for a minor movement is a function of the capacity of the approach and the degree of saturation. According to HCM 2000 (TRB, 2000), the analytical model used to estimate control delay at TWSC intersections assumes that the demand is less than capacity for the period of analysis. If the degree of saturation is greater than about 0.9, average control delay is significantly affected by the length of the analysis period. Based on the recommendation of HCM 2000, the analysis period may be fifteen minutes in many cases. If demand exceeds capacity during a fifteen-minute period, the delay results calculated by the procedure may not be accurate. In such a case, the period of analysis should be lengthened to include the period of oversaturation. The control delay model for TWSC intersections in HCM 2000 can be expressed as Equation 3.

$$d = \frac{3600}{c_{m,x}} + 900T \left[\frac{v_x}{c_{m,x}} - 1 + \sqrt{\left(\frac{v_x}{c_{m,x}} - 1 \right)^2 + \frac{\left(\frac{3600}{c_{m,x}} \right) \left(\frac{v_x}{c_{m,x}} \right)}{450T}} \right] + 5 \quad (3)$$

where

d = control delay per vehicle (usually in seconds/vehicle or s/v);

v_x = flow rate for movement x (veh/h);

$c_{m,x}$ = capacity of movement x (veh/h);

T = analysis time period (in hour);

In Equation 3, the constant, 5 s/veh, accounts for the deceleration of vehicles from free-flow speed to the speed of vehicles in queue and the acceleration of vehicles from the stop line to free-flow speed.

AWSC intersections require every vehicle to stop at the intersection before proceeding. Since every driver has to stop, the judgment as to whether to proceed into the intersection is a function of traffic conditions on the other approaches. Flows at AWSC intersections are determined by a consensus of right-of-way that alternates between the north-south and east-west streams (for a single-lane approach) or proceeds in turn to each intersection approach (for a multilane approach). The headways between consecutively departing subject approach vehicles depend on the degree of conflict between these vehicles and the vehicles on the other intersection approaches, and also depend on the vehicle type and the turning maneuver. The degree of conflict is a particular concept in taking headways into account at AWSC intersections. It is a function of the number of vehicles faced by the subject approach vehicle and of the number of lanes on the intersection approaches. The control delay model for AWSC intersections in HCM 2000 can be expressed as Equation 4.

$$d = t_s + 900T \left[(x-1) + \sqrt{(x-1)^2 + \frac{h_d x}{450T}} \right] + 5 \quad (4)$$

where

d = control delay per vehicle (usually in seconds/vehicle or s/v);

t_s = service time (s);

x = degree of utilization ($vh_d/3600$);

v = flow rate the approach (veh/h);

h_d = departure headway (s);

T = analysis time period (in hour);

Since an iterative process to calculate the departure headway is used, with the initial value of 3.2 seconds, the calculations are repeated until departure headway for each lane change by less than 0.1 second from the previous iteration.

SIMULATION RESULTS

Even though a few minor differences in the calculation of control delay exist between models for signalized and unsignalized intersections in HCM 2000, the models are good enough to be a foundation of the trade-off decision because the definition of control delay is consistent among these models. The control delay with models in HCM 2000 will be estimated by two ways here. One is to use the mathematical form of models with proper parameter settings in MATLAB, which is a software package for computation, visualization, and programming. Another is to employ the Highway Capacity Software (HCS 2000), which implements the procedures defined in the HCM 2000. Both of these two ways should reach the same estimations if the input parameters are exactly the same. In comparison, HCS 2000 is more like a calculator, easy to use but hard to do simulation with wide range variation of parameters.

Since there are many factors which may affect the control delay, the traffic flow rate is concerning most in the framework. In order to distinguish the impact of the traffic flow rate on the control delay from different control types, a hypothetical site with two-way, two-lane for both directions is used to illustrate the framework. During the simulation, parameters except the traffic flow rate remain the same. For more comparable results, two-way volume is used to specify the traffic flow rate. The range of two-way volume will vary from 100 veh/h to 600 veh/h for the minor street, and from 100 veh/h to 800 veh/h for the major street. To consider when volumes exceed this range is unnecessary in this framework

because there is no way to reduce the average control delay in a stop sign controlled intersection. That is, signal control is the only option for such intersections.

Five scenarios are simulated to show the sensitivities: through traffic only, 5 %, 10%, 15%, and 20% left-turn traffic. In the case of signal control, the cycle length is always 60 seconds with equal split and four seconds in Yellow. The criterion for decision making is the average control delay for the whole intersection. For the cases in TWSC intersections, a weighted average control delay based on the flow rate is used for the whole intersection since left-turn traffic causes delay in the major street. Two primary assumptions for all simulations in this paper are that there is no initial queue, and the traffic flow rate will not change with different control types.

Examination of the Exhibit 10-15 in HCM 2000

At the beginning of this paper, the through-traffic-only case in Table 4.1 shows there are problems in the Exhibit 10-15 in HCM 2000. It is worthwhile to have more complicated tests to validate the problems in the exhibit. Table 4.3 consists of four sub tables which represents 5%, 10%, 15%, and 20% left-turn traffic respectively for some critical volumes in the exhibit. Peak-Hour two-way volumes for the major and the minor street are shown in parentheses as (major, minor), and equal traffic volume for both ways are assumed. For example, (2000, 150) represents there are 2000 vehicles per hour for both ways, i.e. 1000 veh/h for each way, in the major street, and 150 vehicles per hour both ways, i.e. 75 veh/h for each way, in the minor street.

Table 4.3. Average Control Delays for the Intersection with Left-Turn Traffic

5%LT	Signal	AWSC	TWSC	10%LT	Signal	AWSC	TWSC
(2000,150)	253.60	243.08	N/A	(2000,150)	382.00	243.32	N/A
(1800,150)	170.70	180.87	N/A	(1800,150)	270.80	181.07	N/A
(1600,150)	97.30	122.08	130.85	(1600,150)	168.50	122.22	N/A
(1000,300)	17.20	28.64	36.75	(1000,300)	18.30	28.69	60.15
(600,600)	13.30	18.59	30.10	(600,600)	13.50	18.69	48.10
(600,300)	12.60	11.87	14.20	(600,300)	12.70	11.89	15.45
(400,400)	11.80	10.60	12.15	(400,400)	11.80	10.63	12.85

15%LT	Signal	AWSC	TWSC	20%LT	Signal	AWSC	TWSC
(2000,150)	493.30	245.67	N/A	(2000,150)	556.70	246.98	N/A
(1800,150)	361.70	183.14	N/A	(1800,150)	431.60	184.29	N/A
(1600,150)	239.30	123.96	N/A	(1600,150)	301.10	124.92	N/A
(1000,300)	20.40	29.19	109.25	(1000,300)	24.40	29.24	184.20
(600,600)	13.80	18.80	76.40	(600,600)	14.20	18.90	112.30
(600,300)	12.90	11.95	17.15	(600,300)	13.20	11.96	19.00
(400,400)	11.90	10.66	13.60	(400,400)	12.00	10.69	14.45

Notes:

1. The cycle length for the signal operation is 60 seconds.
2. Peak-Hour two-way volumes for the major street and the minor street are shown in parentheses (major, minor).

The average control delays for the whole intersection, which are estimated under a range of traffic volumes with the three control types at the hypothetical intersection, are shown in cells. Over capacity in the minor street causes some N/A in the tables since the HCS 2000 can not perform a calculation under such situations. It is extremely hard to find a gap for vehicles in the minor street to go through or join the traffic in the major street, even though the delays in the major street are still low in those cases.

Comparisons among Signal, TWSC, and AWSC

The contours of the average control delays, in seconds, for the whole intersection in the hypothetical site are shown in Figure 4.2. Two types of control at the intersection are considered in this figure. Solid contour lines here represent the

average control delay which resulted from the signal control; the two-way stop sign brought about the dot contour lines. Four sub figures which illustrated different percentage of left-turn traffic are shown in Figure 4.2 for sensitivity analyses. Each sub figure here consists of different traffic flow rate in the major and minor street with a certain percentage of left-turn traffic.

The contour intervals for different control types are different here because the sensitivities are different in the plotted area. For the two-way stop sign control, the intervals, from 10 to 50, are based on the definition of level of service

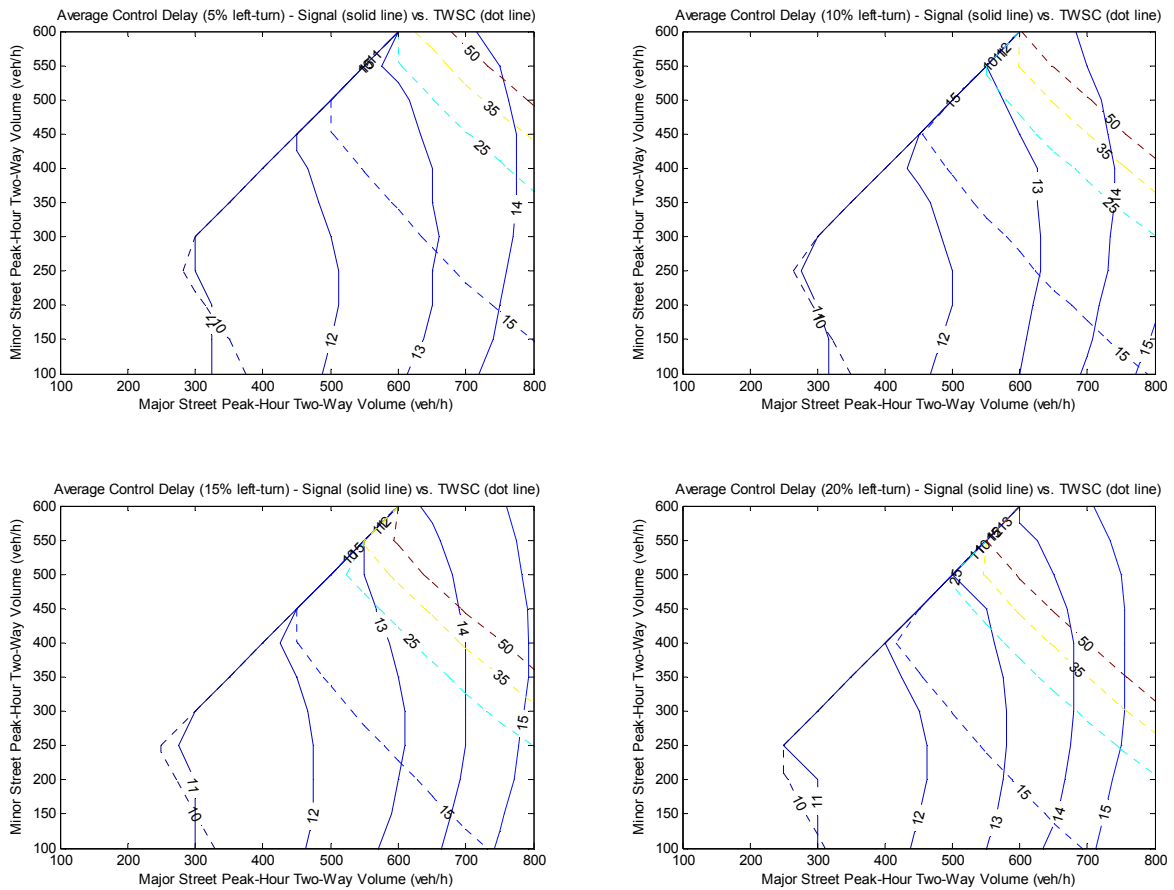


Figure 4.2. Contours of Average Delay For Signal Control and TWSC

in Table 4.2. Constant interval is used in the cases of signal control because those contours, from 10 to 14, can cover almost the whole area in the figure.

The comparison between signal control and AWSC are shown in Figure 4.3 which consists of four sub figures with different percentage of left-turn traffic. They are similar to those in Figure 4.2 but with AWSC instead TWSC. Dashed contour lines in this figure set represent the average control delay which results from the AWSC, while those affected by the signal control are presented in solid contour lines. It is worthwhile to notice that different contour intervals are used for different control types due to the differences of sensitivity. For the all-way stop sign control, the intervals, from 10 to 50, are based on the definition of level

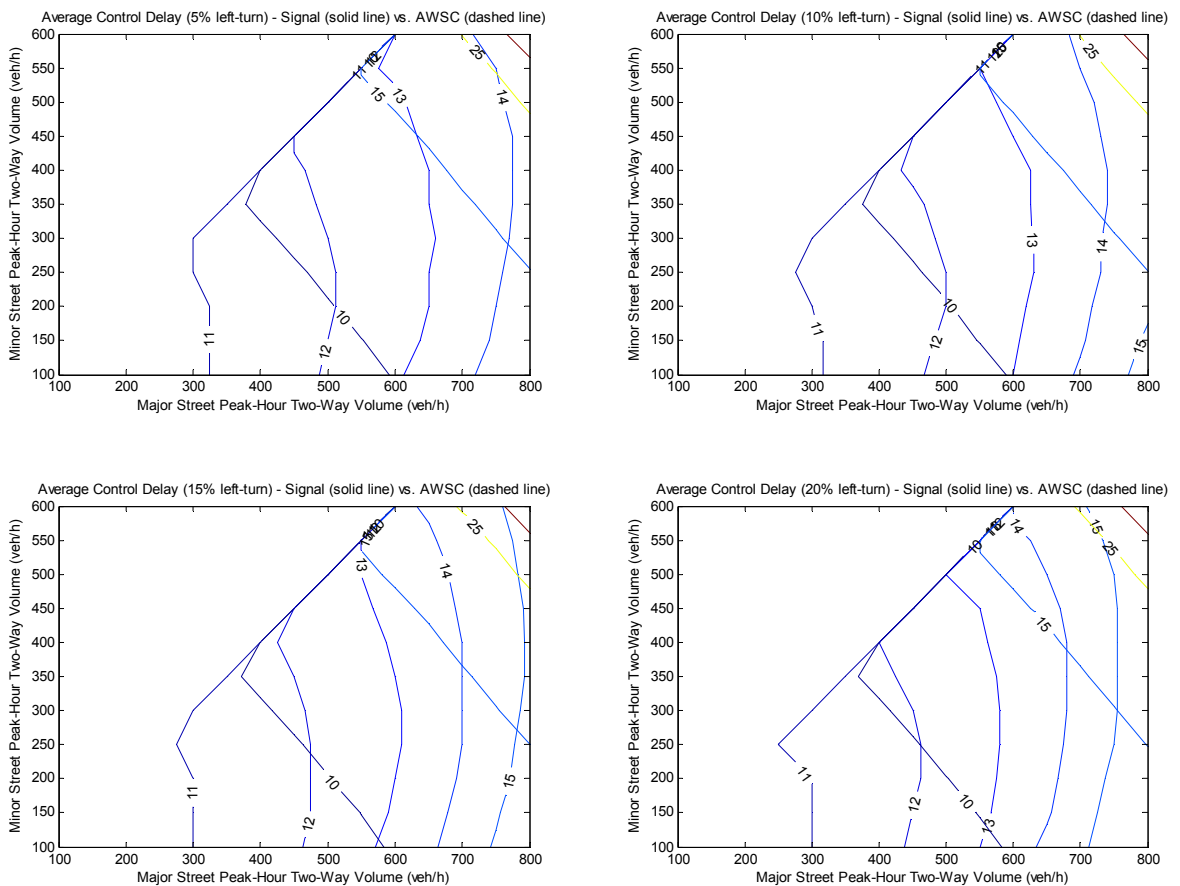


Figure 4.3. Contours of Average Delay For Signal Control and AWSC

of service in Table 4.2. Constant interval is used in the cases of signal control for the same reason as it was mentioned for Figure 4.2.

Figure 4.4 is designated for the comparison between TWSC and AWSC. Four sub figures represent the contour lines under different percentage of left-turn traffic, just as Figures 4.2 and 4.3 did. Even though the range of contours for TWSC (from 10 to 50) and AWSC (from 10 to 25) are different, the contour intervals are exactly the same in Figure 4.4 because the definition of level of service, either for TWSC (dot lines) or for AWSC (dashed lines), are exactly the same.

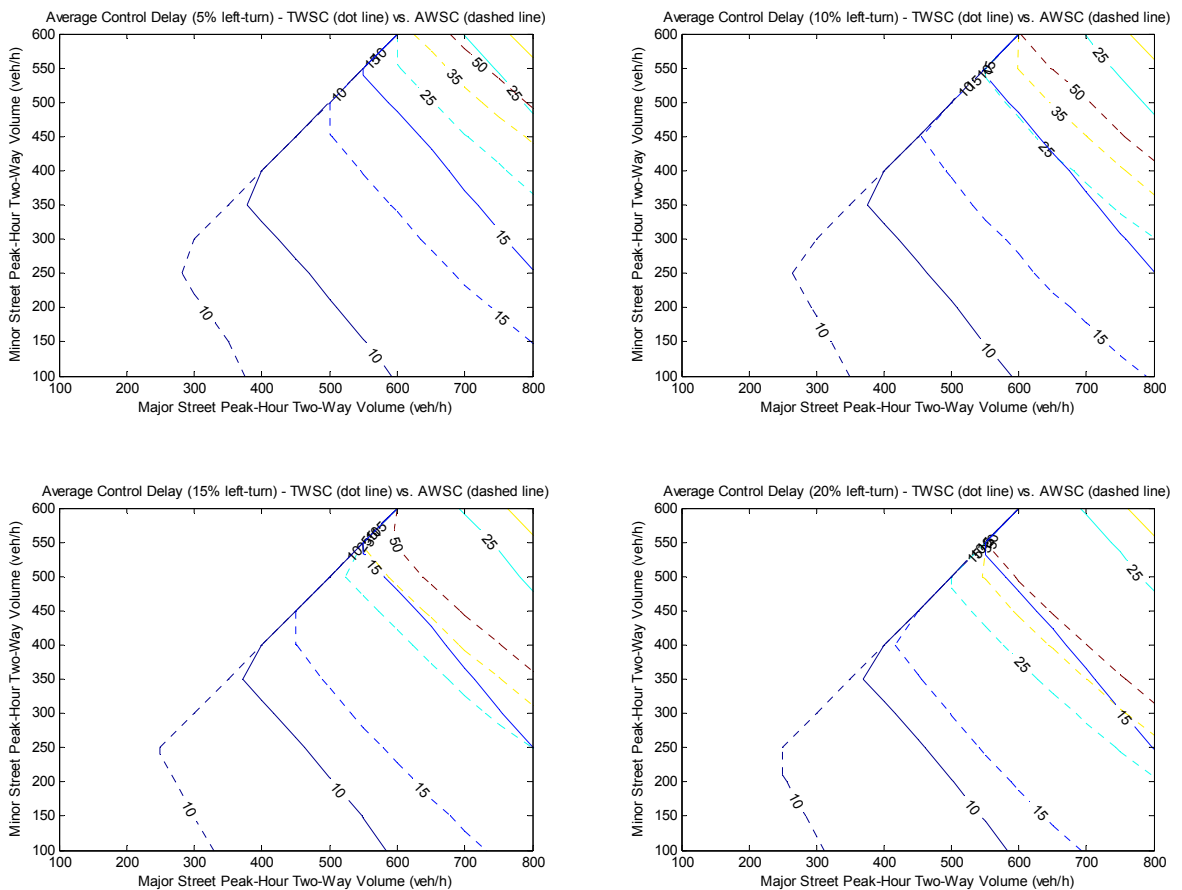


Figure 4.4. Contours of Average Delay (in Seconds) For TWSC and AWSC

An Elegant Case – Symmetric Traffic Flow Rate

In Figures 4.2, 4.3, and 4.4, there is always a line segment in each sub figure which consists of the average control delays under a symmetric traffic flow rate condition. That is, the traffic flow rate in both the major and the minor street are exactly the same, including the percentage of left-turn traffic. Although such a situation is exceptionally rare in reality, it is quite interesting in theory because the average control delays are always higher than it when the traffic flow rates in the major and minor street are not symmetric. In short, it is the lowest bound of the average control delay for given traffic flow rates in the major and the minor street.

Figure 4.5 shows all the lowest bound of the average control delay for each control type, with different percentages of left-turn traffic in four sub figures. For the consistency in labeling different control types, the dashed lines represent the delay by AWSC, the solid lines are for the signal control, and the dot lines are for the TWSC. For a closer observation on the intersection of those average control delay functions in Figure 4.5, the range of traffic flow rate is from 100 veh/h to 550 veh/h in all of the sub figures. It is appropriate enough for the decision making because all functions for the average control delay here are monotonic increasing.

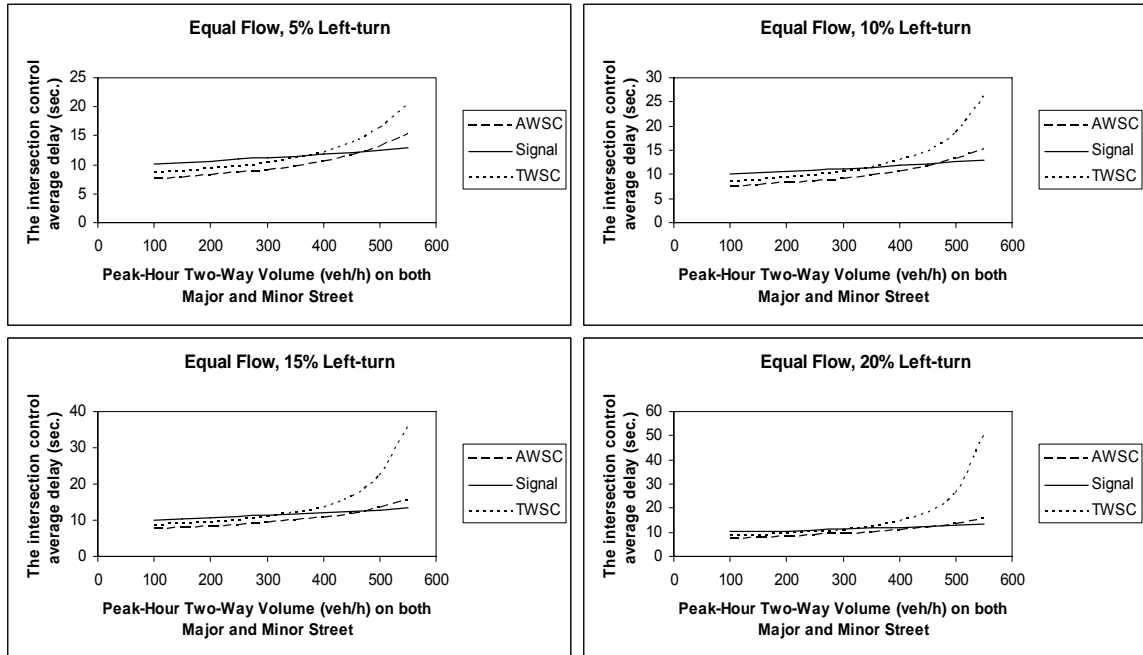


Figure 4.5. Average Delays (in Seconds) Under Symmetric Flow

Signalized Intersections Only

The signal control is the only option which may improve the performance by adjusting parameters to accommodate the traffic flow. Different sites may have different geometric and traffic conditions. However, most of those conditions may be reflected properly in the parameter of saturation flow rate. The suggested value of saturation flow rate for an ideal lane in HCM 2000 is 1900 vehicles per hour. The more complicated conditions exist in a site, the lower value of saturation flow rate should be used for the estimation. The percentage of left-turn traffic may not be an issue any more because the smaller value of saturation flow rate already reflected it properly.

In order to learn about the sensitivity and the range of the average control delay in signalized intersections, four sub figures with different situation flow

rates are shown in Figure 4.6. Parameters for the pre-timed signal operation, except saturation flow rate, will not change with different traffic flow rates.

Constant contour intervals make the sensitivity more observable. It is worthwhile to remind again that the hypothetical site is an intersection with two-way, two-lane for both directions. Otherwise, the values of average control delay in the figures may be higher than those in real situation.

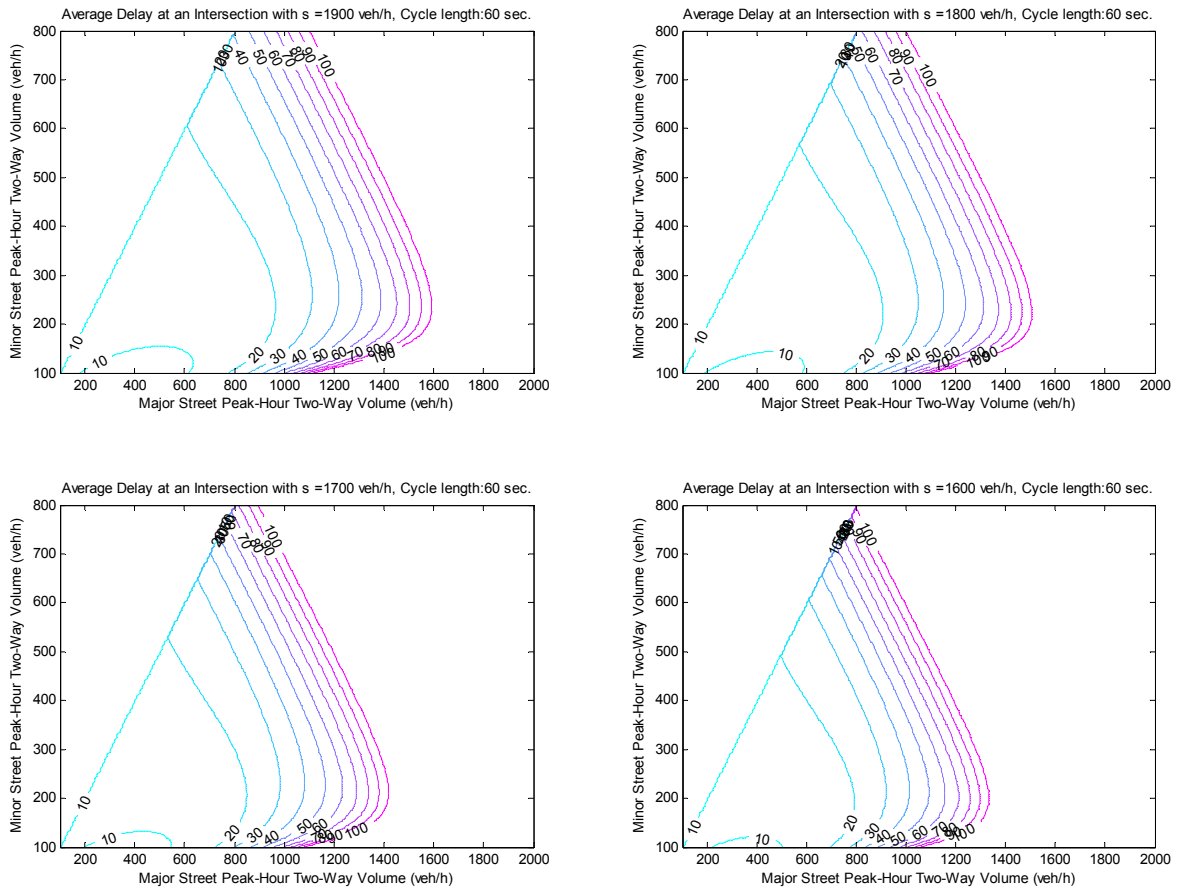


Figure 4.6. Average Delays at Signalized Intersection under Different Saturation Flow Rate

FINDINGS AND DISCUSSIONS

Examination of the Exhibit 10-15 in HCM 2000

According to Tables 4.1 and 4.3, the Exhibit 10-15 in HCM 2000 is inappropriate to determine an appropriate control for an intersection. For example, the recommended area for TWSC should be adjusted and shrunk dramatically for the hypothetical site. Among the three control types, only in a through-traffic case does a TWSC bring the lowest average control delay for the whole intersection. But this case is exceptionally rare in reality. If there is left-turn traffic, a TWSC may lead to some over-capacity situations in the minor street. In practice, those situations mean that it is extremely hard to find a gap for vehicles in the minor street to go through or join the traffic in the major street, even though the delays in the major street are still acceptable.

Furthermore, the Exhibit seems to promise that the user can make the decision just on the basis of the traffic volume on the major and minor street. Without any consideration of turning traffic, it will not realistic even if all other factors are neglected. There is also no information about the signal operations, which may optimize the performance under a given traffic flow rate. Since the Exhibit is too simple to be used, a framework based on the average control delay for the whole intersection is suggested to determine the best control type at an intersection. Such a framework may not be able to provide the exactly optimal solution for a site, but the searching space of the optimum can be shrunk dramatically by the framework.

Comparisons among Signal, TWSC, and AWSC

The most important observation on all of Figures 4.2, 4.3, and 4.4 should be the nonlinear sensitivities of the average control delay by signal, TWSC, and AWSC respectively. Since the sensitivities are not linear, making a trade-off in the higher sensitivity area should be more carefully thought out. Different sensitivities for different control types can be distinguished easily by these figures. Even with different percentages of left-turn traffic, the signal control for the hypothetical site still has the most gentle and lowest sensitivity within the plotted area. The average control delay by TWSC is more sensitive than the other two control types, especially in higher traffic flow rate areas. The sensitivity of average control delay by AWSC is more sensitive than that by signal control but less sensitive than that by TWSC.

The trade-off among control types for the hypothetical site can be easily decided upon by such a framework. After one determines the traffic flow rate in both the major and minor street, and the percentage of left-turn traffic, then the most appropriate option will be suggested based on the average control delay. The framework is also useful when the level of service (LOS) is used as the criterion, because boundaries of the related LOS are shown in each figure. For example, the area LOS F, where the average control delay exceeds 50 seconds, for the TWSC can be found on the upper-right corner in each sub figure of Figure 4.2.

According to Figure 4.4, the LOS will be C for the whole intersection, if a TWSC is used in where the two-way flow rate is less 650 vph in the major street,

450 vph or less in the minor street, and 10% left-turn traffic for each direction. It is worthwhile to notice that the feasible area is not a rectangle but a polygon because it depends upon the contour of average control delay. For the same or a better LOS, an AWSC could be used where the two-way flow rate is less 750 vph in the major street, 550 vph or less in the minor street, and 10% left-turn traffic for each direction. If the installation and maintenance costs do not matter, based on Figure 4.2, LOS B or better can be anticipated by using a signal control within the testing traffic flow rate range.

Since the sensitivities of the average control delay for different control types differ from each other in different traffic patterns, to facilitate the trade-off among different control types is the most important function for the framework. Based on Figure 4.3, an AWSC may be a better choice than a signal where the two-way flow rate is less 600 vph in the major street, 350 vph or less in the minor street, and 10% left-turn traffic for each direction. Again, the feasible area is not a rectangle but a polygon. The reason an AWSC is recommended for the traffic condition is that the average control delay for an AWSC is less sensitive than a signal in the intersection. Also, the sensitivities can be checked very easily through the framework.

The possibility of optimization by signal control may be an interesting topic, because signal control is the only control type which may achieve an optimal operation by adjusting the attributes, e.g. cycle length or green time split. That is, the framework uses a pretimed signal operation, but does not consider other alternatives with more complicated signal operations. However, since the

installation and maintenance costs for the signal operation are much higher than for stop signs, the optimization of signal control should be considered only after the stop sign is clearly not providing a good performance at the intersection.

Since the average control delay by TWSC is more sensitive than the others, TWSC should be used in very limited situations. Although vehicles in the major street may not stop in most cases under the TWSC, those in the minor street may wait for a long time. Actually, the average control delays for vehicles in the minor street are almost twice as long as the delays for the whole intersection in higher traffic volume situations. That is why the average control delay for the whole intersection, instead of a certain approach, is used as the criterion. In the hypothetical site, it is better to assign the right of way to every direction when the traffic flow rate in the minor street exceeds 600 vehicles per hour. If the TWSC is used in such a case with 5% left turn traffic, the average control delay for vehicles in the minor street will exceed 50 seconds.

The Elegant Case – Symmetric Traffic Flow Rate

Because of the monotonic increasing patterns in the average control delays for all three control types, the case of symmetric traffic flow rate is a very useful and elegant tool. Under such a case, the traffic flow rate in both the major and the minor street are exactly the same, including the percentage of left-turn traffic. According to Figures 4.2, 4.3, and 4.4, the cases with symmetric flow rate always reach the minimum of average control delay along a given traffic flow rate in the minor street. It is exceptionally rare in reality, but it is very useful for decision making. If the traffic flow pattern will not change with different control types, such

a case can be used to decide the lowest bound and the basic relation among control types.

The character of “lower volume, lower sensitivity” is shown in Figure 4.5. It also verifies that AWSC is better to use in symmetric traffic cases, about 500 vehicles per hour with 5% left turn traffic for each direction at the hypothetical site. The results are similar to those of Kyte and Marek (1989). Delay increases at a very slow rate for low traffic flow rates, up to subject flow rates of 400 to 500 vph. At this point, delay begins to increase exponentially, especially in the case of TWSC. Under the symmetric traffic cases, AWSC may be a better control type than TWSC; otherwise, according to Figure 4.4, TWSC will be recommended, just as Byrd and Stafford (1984) did.

Signalized Intersections Only

Figure 4.6 is another basic tool to facilitate the trade-off. Although the sensitivities of the average control delay are different in each sub figure, those contours which do not cross over each other indicate the pattern of monotonic increase when the traffic flow rate in the minor street is over 200 vehicles per hour. This pattern is important because it verifies that the case of symmetric or near symmetric traffic flow may have the minimal average control delay once the traffic flow rate in the minor street is known. The shape of the contours may change slightly when the parameters of signal operation are adjusted for optimization.

In the hypothetical site, the average control delay is more sensitive and the area between 10 and 20 is also shrunk when the saturation flow rate is decreasing.

This phenomenon supports that the traffic flow pattern can be reflected by the saturation flow rate because it is consistent with Figures 4.2 and 4.3. That is, the turning effect reduces the saturation flow rate in the hypothetical site.

CONCLUSIONS

Since there is a lack of necessary information, the exhibit 10-15 in HCM2000 is inappropriate to determine the best control type for an intersection. The trade-off among signal control, TWSC, and AWSC which is based on traffic volume only will lead to an improper decision. Several models with different parameters, including traffic volume, can estimate the average control delay for an approach under a given condition. A framework based on the average control delay is proposed to determine the best control type for an intersection.

A hypothetical intersection with two-way, two-lane for each direction is examined by the framework. Based on the simulation results, the LOS will be C for the whole intersection, if a TWSC is used in where the two-way flow rate is less 650 vph in the major street, 450 vph or less in the minor street, and 10% left-turn traffic for each direction. It is worthwhile to notice that the feasible area is not a rectangle but a polygon because it depends upon the contour of average control delay. For the same or a better LOS, an AWSC could be used where the two-way flow rate is less 750 vph in the major street, 550 vph or less in the minor street, and 10% left-turn traffic for each direction. If the installation and maintenance costs do not matter, LOS B or better can be anticipated by using a signal control within the testing traffic flow rate range.

Since the sensitivities of the average control delay for different control types differ from each other in different traffic patterns, to facilitate the trade-off among different control types is the most important function for the framework.

Based on the simulation results for the hypothetical intersection, an AWSC may be a better choice than a signal where the two-way flow rate is less 600 vph in the major street, 350 vph or less in the minor street, and 10% left-turn traffic for each direction. Again, the feasible area is not a rectangle but a polygon. The reason an AWSC is recommended for the traffic condition is that the average control delay for an AWSC is less sensitive than a signal in the intersection. Also, the sensitivities can be checked very easily through the framework.

REFERENCES FOR PART IV

Akcelik, R. and Roupail, N. M., 1993. Estimation of Delays at Traffic Signals for Variable Demand Conditions. *Transportation Research Part B* 27(2), 109-131.

Byrd, M.N. and Stafford, D.B., 1984. Analysis of Delay and User Costs at Unwarranted Four-Way Stop Sign Controlled Intersections. *Transportation Research Record* 956, 30-32.

Chan, Y., Flynn, L.J., and Stocker, K.J., 1989. Volume-Delay Relationship at Four-Way-Stop Controlled Intersections: A Response-Surface Model. *ITE Journal*, 59 (3), 27-34.

Eck, R.W, and Biega, J.A., 1988. Field Evaluation of Two-Way Versus Four-Way Stop Sign Control at Low-Volume Intersections in Residential Areas. *Transportation Research Record*, 1160, 7-13.

Federal Highway Administration (FHWA), 1983. Traffic Control Devices Handbook. United States Department of Transportation, Washington, D.C.

Federal Highway Administration (FHWA), 1988. Manual on Uniform Traffic Control Devices for Streets and Highways, United States Department of Transportation, Washington, D.C.

Federal Highway Administration (FHWA), 2004. Manual on Uniform Traffic Control Devices for Streets and Highways, 2003 Edition with Revision Number 1, Available online: <http://mutcd.fhwa.dot.gov/HTM/2003r1/html-index.htm> , Accessed: March 6, 2007.

Han, L. and Li, J-M., 2007. Short or Long ... which is Better? A Probabilistic Approach towards Cycle Length Optimization. TRB 86th Annual Meeting

Compendium of Papers CD-ROM. TRB, National Research Council, Washington, DC.

Kyte, M. and List, G., 1999. A capacity model for all-way stop-controlled intersections based on stream interactions. *Transportation Research Part A*, 33(3-4), 313-335.

Kyte, M. and Marek, J., 1989. Estimating Capacity and Delay at a Single-Lane Approach, All-Way Stop-Controlled Intersection. *Transportation Research Record* 1225, 73-82.

Kyte, M., Tian, Z., Mir, Z., Hameedmansoor, Z., Kittelson, W., Vandehey, M., Robinson, B., Brilon, W., Bondzio, L., Wu, N., and Troutbeck, R., 1997a. NCHRP Web

Document 5: Capacity and Level of Service at Unsignalized Intersections: Final Report, Vol. 1—Two-Way Stop-Controlled Intersections. Available online: <http://www.nap.edu/books/nch005/html> , Accessed: March 6, 2007.

Kyte, M., Tian, Z., Mir, Z., Hameedmansoor, Z., Kittelson, W., Vandehey, M., Robinson, B., Brilon, W., Bondzio, L., Wu, N., and Troutbeck, R., 1997b. NCHRP Web

Document 6: Capacity and Level of Service at Unsignalized Intersections: Final Report, Vol. 2—All-Way Stop-Controlled Intersections. Available online: <http://www.nap.edu/books/nch006/html> , Accessed: March 6, 2007.

McKinley, D.W., 2001. Traffic Signals. Traffic Control Devices Handbook, Pline, J.L. ed., pp.261-361, Institute of Transportation Engineers, Washington, D.C.

- Richardson, A.J., 1987. A Delay Model for Multiway Stop-Sign Intersections. *Transportation Research Record*, 1112, 107-114.
- Sampson, J.D., 1999. Queue-Based Traffic-Signal Warrants: The 4Q/6Q Warrant. *ITE Journal* 69(4), 30-36.
- Transportation Research Board (TRB), 1994. Special Report 209: Highway Capacity Manual, 3rd ed., 2nd rev. TRB, National Research Council, Washington, D.C.
- Transportation Research Board (TRB), 2000. Special Report 209: Highway Capacity Manual, TRB, National Research Council, Washington, D.C.
- Webster, F., 1958. Traffic Signal Settings, Road Research Technical Paper No. 39. Road Research Laboratory, Her Majesty's Stationery Office, London, U.K.
- Zion, M., List, G.F., and Manning C., 1989. Testing Delay Models with Field Data for Four-Way, Stop Sign-Controlled Intersections. *Transportation Research Record* 1225, 83-90.

PART V. IMPACTS OF MISPLACED PEAK INTERVALS ON PHFS

This part is a slightly revised version of a journal paper by Lee Han and Jan-Mou Li with the same title that will be submitted for review in 2007.

Han, L. and Li, J.-M., 2007. Impacts of Misplaced Peak Intervals on

PHFs. To be submitted for review.

My primary contributions to this paper include (1) development of the problem into a work relevant to my doctoral research study, (2) development of experimental setup, (3) most of the gathering and interpretation of literature, (4) performing the laboratory experiments, (5) interpretation and analysis of test results, (6) most of the writing.

ABSTRACT

The peak-hour factor (PHF), which represents the relationship between the busiest 15-min flow rate and the fully hourly volume, is applied to determine design-hour flow rates. An inadequate PHF may result in substantial excess capacity the rest of the time or result in oversaturated conditions for a substantial portion during the peak hour. Although several default values are suggested in the Highway Capacity Manual 2000 for different traffic conditions, local data are still recommended to use for a precise estimation, because the traffic varies, depending on time and site. Furthermore, the peak intervals can hardly be located on the clock. The impacts of misplaced peak intervals on the PHF are investigated with simulations and real data in this paper. By comparing different methods locating the peak intervals, the “on the clock” approach may provide an inaccurate estimation of PHF. According to the results, based on the Wilcoxon Signed-Rank Test and the mean absolute percentage error, the misplacements did occur and impact the PHFs, if either lower resolution data or “on the clock” approaches were used. It is recommended that the PHF should be calculated with searching the peak intervals through local, higher resolution data for obtain the most accurate estimation.

INTRODUCTION

Traffic in a road network is varying all the time and the variation is rarely on the clock. In most cases, analyses focus on the peak hour of traffic for a certain approach because it represents the most critical period for operations and has the highest capacity requirements. Since the annual average daily traffic (AADT) is used for planning applications, the peak hour factor (PHF) is one of the three important factors to convert the hourly volumes into the volume rate during the busiest 15 minutes of the hour. However, to define the peak hour as well as the worst 15 minutes in practice raises inaccuracy if the traffic variation was not treated properly.

According to the Highway Capacity Manual 2000 edition (HCM 2000) (TRB, 2000), the selection of an analysis period must consider the impact on design and operations of higher volume hours that are not accommodated. It also mentioned that the design for a smaller range, say a 5-minute interval, of the peak flow rate would result in substantial excess capacity during the rest of the peak hour; and the design for a larger range, say an 1-hour interval, of the peak flow rate would result in oversaturated conditions for a substantial portion of the hour. Since most of the procedures in the HCM 2000 are based on peak 15-minute flow rates, the peak hour factor is defined as the ratio of total hourly volume to the peak 15 minute flow rate within the hour. However, it did not mention what would happen if there is a higher peak 15-minute interval outside

the peak hour. Such situations occurred in real data when they are closely examined.

Traffic patterns vary in response to local travel habits and environments. It indicates the need for local data on which to base informed judgments. Even though traffic varying over time is common sense, the variability of peak hour factor has been investigated recently. Tarko and Perez-Cartagena (2005) investigated the variability of PHF overtime and across locations, and found that the day-to-day variability is as strong as the site-to-site variability. They recommend that PHF be estimated on the basis of several days of vehicle counting to improve the precision of the average PHF estimate. Notwithstanding the spatial difference, even the variation of traffic within a day will not be the same within another day. That is, the peak hour for tomorrow may not start at the exactly same time as today.

For some reasons, practitioners employ the literal meaning of the peak-hour in several ways. Most of the time, they classify the peak hour on the clock, e.g. from 7 a.m. to 8 a.m. or 4:30 p.m. to 5:30 p.m. There is nothing wrong if the hourly, half-hourly, or even 15-minute traffic volume is the only data we had. But such an aggregation may shift the peak hour from the “real” one to a certain degree. When the resolution of data is increased, the difference between the peak hour on the clock and the “real” peak hour should be noticed. Most modern detectors can collect traffic data every thirty seconds. Therefore, the peak hour may start at 7:11:30 a.m. based on the data more precisely.

The object of this paper is to investigate the impact of the misplaced peak hour and peak 15 minutes on the PHF. By comparing different methods locating the peak intervals, the “on the clock” approach may provide an inaccurate estimation of PHF. With 5,000 simulations in each of the truncated Normal distribution and the Poisson distribution, the varying locations of peak 15-min intervals during a peak hour are examined. Real traffic count data, which were collected by the Minnesota Department of Transportation, at a 30-second interval from over 4,000 loop detectors located around the Twin Cities Metro freeways, are also used for the analysis. It is shown that there are impacts on the PHF by the misplacement, and the phenomenon of which the higher peak 15-minute interval is outside the peak hour occurred. By the Wilcoxon Signed-Rank Test, the PHFs by search are significantly different from those by ‘on the clock’. The results show that the peak hour should be located to a more precise period with higher resolution data. Otherwise, extra errors should be considered, and then the PHF can have a better estimation on the traffic situation.

METHODOLOGY

Definition of Peak-Hour Factor (PHF)

According to HCM 2000, the peak-hour factor (PHF) represents the variation in traffic flow within an hour. PHF is the ratio of total hourly volume to the peak flow rate within the hour, computed by Equation 1:

$$PHF = \frac{\text{Hourly volume}}{\text{Peak flow rate (within the hour)}} \quad (1)$$

If 15-min periods are used, PHF is the ratio of total hourly volume to four times the highest 15-min volume within the peak hour. Under such a circumstance, the PHF may be computed by Equation 2:

$$PHF = \frac{V}{4 \times V_{15}} \quad (2)$$

where

PHF = peak-hour factor,

V = hourly volume (veh/h), and

V_{15} = volume during the peak 15-min interval of the peak hour (veh/15 min).

How to locate the peak hour or the peak 15 minutes is the questionable part, even though the definition of PHF is quite straight. “On the clock” is a method to locate them. That is, the peak hour always covers an entire hour, e.g. from 7 a.m. to 8 a.m., because the aggregation of traffic volumes during the hour

is higher than the other 23 hours. Also, the peak 15-min interval will be from 12 to 3, 3 to 6, 6 to 9, or 9 to 12 by this method. With such a method, the peak hour and related peak 15-min traffic volume can be computed by Equations 3 and 4. The advantage of this method is that it can be operated with the aggregation of traffic volumes every 15 minutes. It is good for historical periods when there was no way to collect data more precisely.

$$V_p = \max(V_i), \quad i = 0, 1, \dots, 23 \quad (3)$$

$$V_{p15} = \max(V_j), \quad k = 1, 2, 3, 4 \quad (4)$$

where

V_p = the peak-hour traffic volume in a day (veh/h),

V_i = the i^{th} hourly traffic volumes on the clock in a day (veh/h),

V_{p15} = the peak 15-min traffic volume within the peak hour (veh/15 min),

V_j = the j^{th} quarterly traffic volumes within the hour on the clock (veh/15 min).

Another method to locate the peak hour or peak 15 minutes is to shift the aggregation every time interval with higher resolution data. If data were collected every 30 seconds, then each aggregation of 120 such data points can represent an hour, and each aggregation of 30 such data can represent a 15-minute interval. That is, there are 2760 possible starting points of the peak hour, and 90 possible starting points of the peak 15-min interval within the peak hour. Assuming the 30-second data are given, the peak hour and related peak 15-min traffic volume can be computed by Equations 5 and 6 with this strategy:

$$V_p = \max \left(\sum_i^{i+119} v_i \right), \quad i = 1, 2, \dots, 2761 \quad (5)$$

$$V_{p15} = \max \left(\sum_j^{j+29} v_j \right), \quad j = 1, 2, \dots, 91 \quad (6)$$

where

V_p = the peak-hour traffic volume in a day (veh/h),

v_i = the i^{th} 30-second traffic volume in a day (veh/30 sec),

V_{p15} = the peak-15 min traffic volume within the peak hour (veh/15 min),

v_j = the j^{th} 30-second traffic volume within the hour (veh/30 sec),

Monte Carlo Simulation

Monte Carlo simulation (Fishman, 1996; Robert and Casella, 1999) is a stochastic technique based on the use of random numbers and probability statistics to investigate complex problems. The primary components of a Monte Carlo simulation include a probability distribution function, a random number generator, and a sampling rule. In this study, the technique is used to generate time headways to simulate the arrival within a peak hour, in order to observe the varying locations of the peak 15-min interval during the peak hour. With 5,000 runs in each of the truncated Normal distribution and the Poisson distribution, a large number of cases show the significant differences between different PHF computing methods.

Significance Tests – The Wilcoxon Signed-Rank Test

Since two methods to compute the PHF with the same data are considered, paired-samples T test might be a popular way to compare means. However, because the source population from which the differences have been drawn can not be assumed to have a normal distribution, the Wilcoxon signed-rank test (Sheskin, 2007) is the more appropriate approach to use here instead of the paired-samples T test. The Wilcoxon signed-rank test considers information about both the sign of the differences and the magnitude of the differences between pairs, if the two variables are similarly distributed, the number of positive and negative differences will not differ significantly. Assumptions for the Wilcoxon signed-rank test include:

- Each pair of values is drawn independently of all other pairs.
- Each difference between a pair comes from a continuous population and is symmetric about a common median.

The Wilcoxon signed-rank test, sometimes called the Wilcoxon matched-pairs test, is a nonparametric test, and begins by transforming each instance of difference into its absolute value. The instances without difference are deleted and then those differences are ranked in an ascendant sequence with the related signs. Assuming S is the smaller of the sum of ranks either for positive differences or for the negative differences, if S is equal or less than the critical value in a table of the distribution of Wilcoxon signed-rank test, then the two samples differ from each other significantly.

Measurement of the Difference - Mean Absolute Percentage Error (MAPE)

Mean absolute percentage error (also known as MAPE) is a measure of accuracy in a fitted value in statistics, specifically the trend in a time series analysis (e.g. Nikolopoulos et al., 2007). It usually expresses accuracy as a percentage and can be computed by Equation 7.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad (7)$$

where

A_i = the i^{th} based (real) value,

F_i = the i^{th} referred (forecasting) value, and

n = the sample size.

The mean absolute percentage error (MAPE) is also often useful for reporting the difference between samples, because it is expressed in generic percentage terms which are strictly positive. Since the randomness of traffic and the PHF is a ratio dependent upon the computational method, MAPE can distinguish the difference between those PHF values by different methods.

RESULTS

In order to distinguish the impact of misplacing the peak 15 minutes on the value of PHF, results from hypothetical cases are examined first. The truncated Normal and the Poisson distributions are employed to generate time headways within a hypothetical peak hour. According to May (1990), a Normal distribution with mean 2 seconds and standard deviation 0.6 is used for a high traffic volume case. Since headway less than 0.5 second is unreasonable, the Normal distribution will be truncated on grounds of this criterion. The possibility below 0.5 second in such a distribution will be 0.0062. Furthermore, two more Normal distributions with the same mean 2 seconds, but different standard deviations, 1 and 2, are used for comparisons. They are also truncated once headway is lower than 0.5 seconds; the possibility of the truncation is 0.0668 for the case with standard deviation 1, and 0.2266 for 2.

Samples of the arrival, with different Normal distributions used to generate headways, are demonstrated in Figure 5.1. In the figure, subfigure (a) is an example following the Normal distribution with mean 2 seconds and standard deviation 0.6; subfigure (b) is an example following the Normal distribution with mean 2 seconds and standard deviation 1; and subfigure (c) is an example following the Normal distribution with mean 2 seconds and standard deviation 2. The period for the simulation is an hour, say from 7:00 a.m. to 8:00 a.m., since the peak-hour traffic is simulated. The arrival (dot lines in the figure) during this

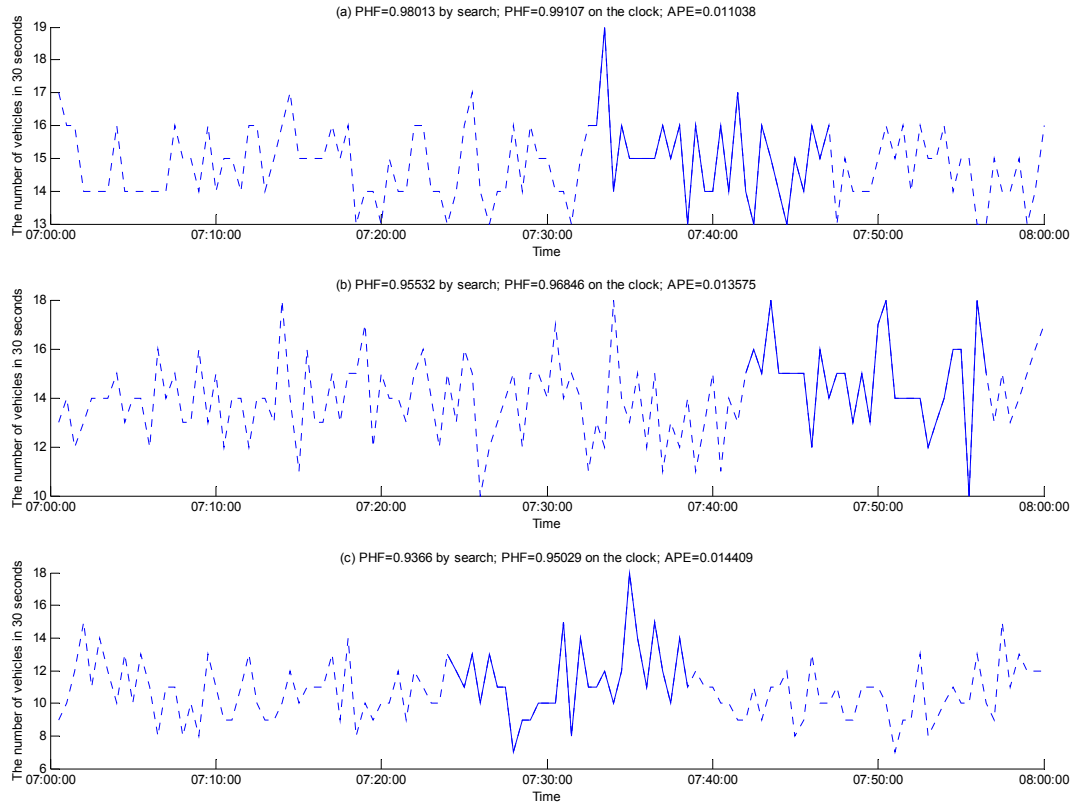


Figure 5.1. Samples from (a) $N(2,0.6)$, (b) $N(2,1)$, (c) $N(2,2)$

period is aggregated every 30 seconds. Solid lines in the figure represent the ‘real’ peak 15-min flow rates within the peak hour by search. The PHF by search is computed with Equation 6, to find the peak 15-min traffic volume, and Equation 2, to do the computation of PHF. On the other hand, the PHF on the clock is computed with Equation 4, to find the peak 15-minute traffic volume, and Equation 2, to do the computation of PHF.

Statistics for the simulations with Normal distributions are listed in Table 5.1. The values in the 2nd and 3rd column are the average PHF of the 5000 runs with search and ‘on the clock’ methods respectively. The fourth column is the

Table 5.1. Results from Monte Carlo Simulations with Truncated Normal Distributions

Samples	PHF by search	PHF on the clock	MAPE	Asymp. Sig. (2-tailed)
N(2, 0.6)	0.9708	0.9860	0.0081	0.000
N(2, 1)	0.9688	0.9798	0.0112	0.000
N(2, 2)	0.9542	0.9704	0.0166	0.000

mean absolute percentage error of PHF by the two methods with 5000 simulations. Values in the last column are all zero, which is the p-value of the Wilcoxon Signed-Rank Test.

Another kind of distribution, the Poisson, is commonly used (e.g. Little, 1961; Tarko and Perez-Cartagena, 2005) to generate the traffic counts within a certain period. The Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate, and are independent of the time since the last event. Samples of the arrival, with different Poisson distributions used to generate headways, are demonstrated in Figure 5.2. In the figure, subfigure (a) is an example following a Poisson distribution with mean 2 seconds; subfigure (b) is an example following a Poisson distribution with mean 3 seconds; and subfigure (c) is an example following a Poisson distribution with mean 4 seconds. Since the peak-hour traffic is simulated, the period for the simulation is also an hour. The arrival (dot lines in the figure) during this period is aggregated every 30 seconds. Again, solid lines in the figure are the 'real' peak 15-min flow rates within the peak hour by search. The computations of PHFs with Poisson distributions are the same as those with Normal distributions.

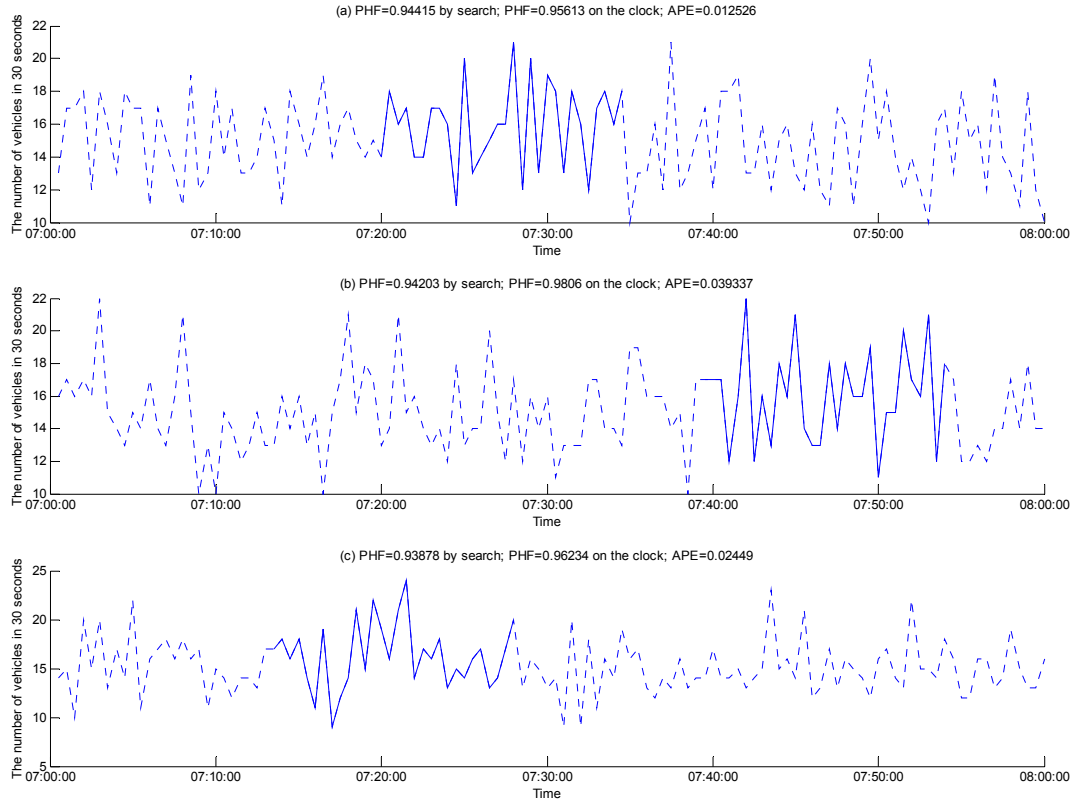


Figure 5.2. Samples from (a) Poisson(2), (b) Poisson(3) , (c) Poisson(4)

Statistics for the simulation with Poisson distributions are listed in Table 5.2. The structure is the same as Table 5.1. Values in the 2nd and 3rd column are the average PHF of the 5000 runs with search and 'on the clock' methods respectively. The fourth column is the mean absolute percentage error of PHF by the two methods with 5000 simulations. Values in the last column are all zero, which is the p-value of the Wilcoxon Signed-Rank Test.

Traffic data on every Tuesday, Wednesday, and Thursday (from January 2, 2007 until April 5, 2007), collected by detectors #2437, #2473, and #5124, are used for the analysis. These detectors were chosen randomly from over 4,000

Table 5.2. Results from 5000-run Monte Carlo Simulations with Poisson Distributions

Samples	PHF by search	PHF on the clock	MAPE	Asymp. Sig. (2-tailed)
Poisson($\lambda=2$)	0.9525	0.9706	0.0185	0.000
Poisson($\lambda=3$)	0.9516	0.9695	0.0184	0.000
Poisson($\lambda=4$)	0.9508	0.9687	0.0184	0.000

loop detectors located around the Twin Cities Metro freeways; and the data were collected every 30 seconds by the Minnesota Department of Transportation.

Figure 5.3 demonstrates the variation of traffic over April 5, 2007, according to detector #2437. It is interesting to notice that there are dramatic drops in traffic within both morning and evening peak hours. Such a drop may affect the location of peak 15-min intervals, especially when the search algorithm is employed.

The misplaced peak hour and peak 15-min intervals can be observed in Figure 5.4. The data are from detector #5124 on January 2, 2007. Four cases with different combinations of the location of peak intervals are considered; subfigure (a) demonstrates the situation with both the peak hour and the peak

15-min interval searched in the day; subfigure (b) shows the situation with a searched peak hour and the peak 15-min interval searched within the peak hour; subfigure (c) illustrated the situation by locating the peak hour on the clock and searching the peak 15-min interval within the peak hour; subfigure (d) depicted the situation including both the peak hour and the peak 15-min interval are on the clock. Even though the period for all of the subfigures is an hour, the start and end time in subfigures (a) and (b) obviously differ from the others. For

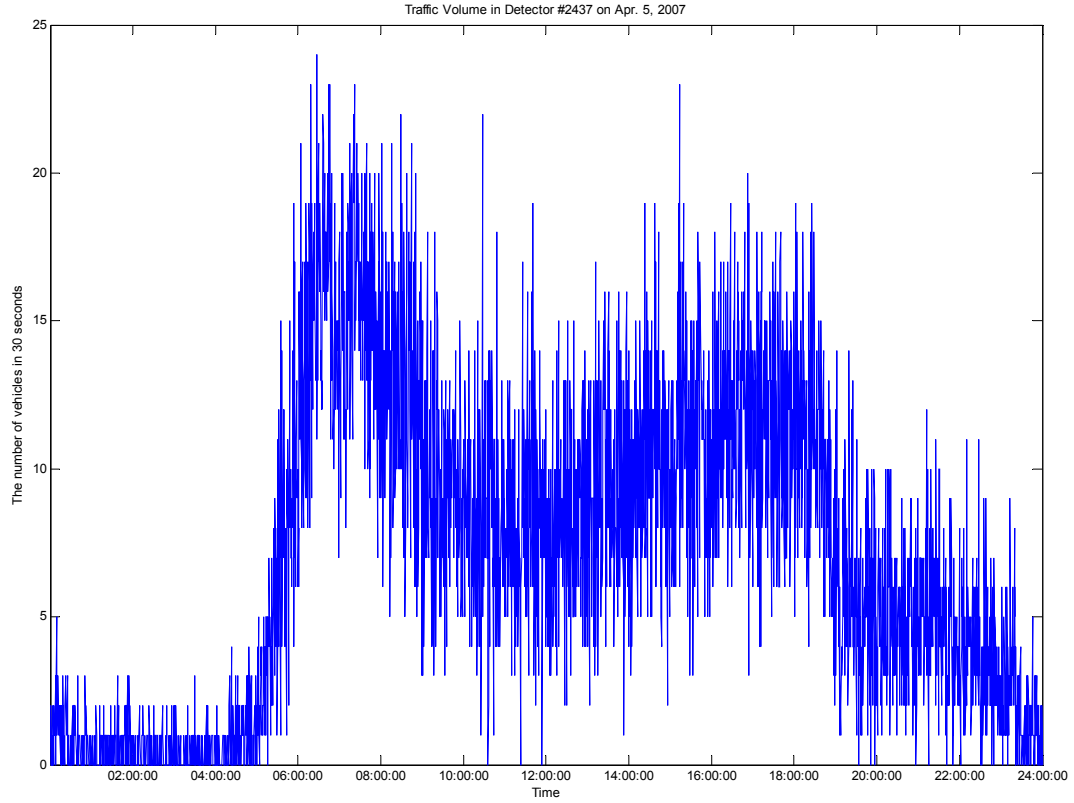


Figure 5.3. Samples from Real Data

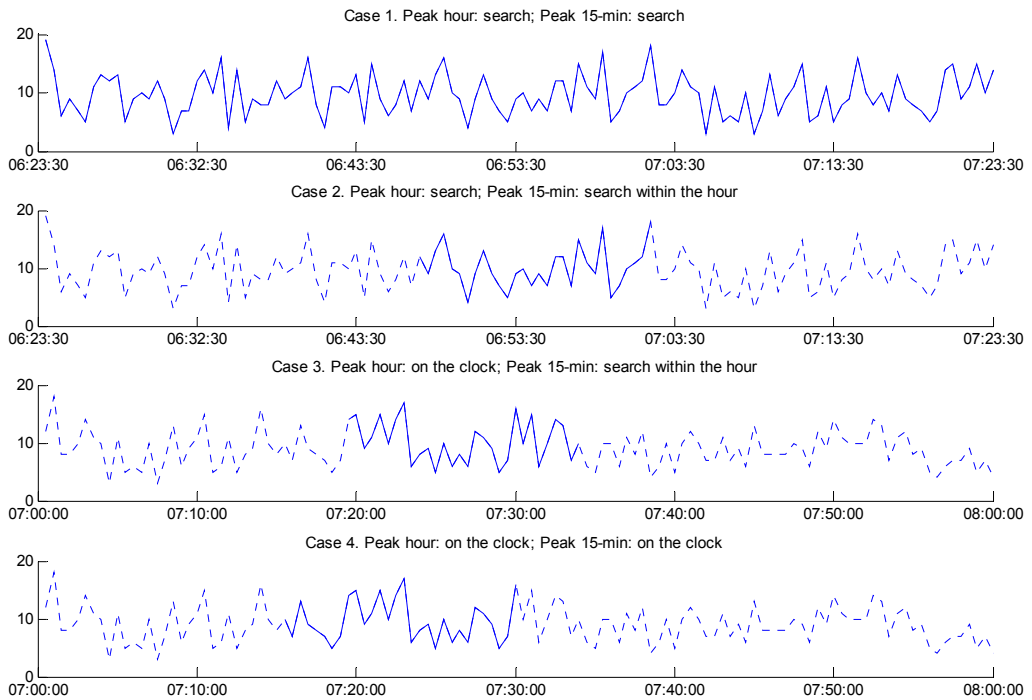


Figure 5.4. A sample with peak intervals from detector #5124 on January 2, 2007

consistency with Figures 5.1 and 5.2, the dot lines in Figure 5.4 represent the arrival during this period, aggregated every 30 seconds, and the solid lines are the peak 15-min flow rates within the peak hour. There is a solid line for the whole period in subfigure (a), which indicates the peak 15-min interval is not entirely located within the peak hour.

Results from all forty-two days' data are shown in Table 5.3. The first four rows of data represent the different cases mentioned in the previous paragraph. Values in the 1st row (as titled PHFR) are the PHFs computed in case (a); case (b) is in the 2nd row (as titled PHFRin); case (c) is in the 3rd row (as titled PHFAin); and case (d) is in the 4th row (as titled PHFA). Based on the comparison with the PHFs in case (d), the rest of the six rows are listed. That is, values in the 5th row (as titled MAPE 1) are the MAPEs between case (a) and case (d); the MAPEs between case(b) and case (d) are in the 6th row (as titled MAPE 2); the MAPEs between case(c) and case (d) are in the 7th row (as titled MAPE 3). With the same sequence, the p-values of the Wilcoxon Signed-Rank Test, representing 2-tailed asymptotic significance, are listed from the 8th row to the 10th row.

Table 5.3. Statistics from Real Data

	#2437	#2473	#5124
PHFR	0.9159	0.8801	0.9083
PHFRin	0.9195	0.8810	0.9106
PHFAin	0.8714	0.8312	0.8900
PHFA	0.9079	0.8638	0.8975
MAPE 1	0.0224	0.0511	0.0290
MAPE 2	0.0196	0.0502	0.0312
MAPE 3	0.0288	0.0376	0.0196
p-value 1	0.000	0.036	0.041
p-value 2	0.000	0.025	0.018
p-value 3	0.000	0.000	0.000

Note:

1. APE1: PHFA vs. PHFR; APE2: PHFA vs. PHFRin; APE3: PHFA vs. PHFAin
2. p-value 1: PHFA vs. PHFR; p-value 2: PHFA vs. PHFRin; p-value 3: PHFA vs. PHFAin

DISCUSSION

Misplaced Peak Intervals

According to the results shown in the previous section, the misplaced peak intervals did occur and impact the accuracy of PHFs. The most powerful proof is the p-values of the Wilcoxon Signed-Rank Test between the PHFs computed by different methods, although the MAPEs and average PHFs computed by different methods show small differences in value. The p-values indicate that there are statistically significant differences between the results by different methods. Such misplacement occurs for the peak 15-min interval and also the peak hour. Furthermore, the phenomenon of another higher peak 15-min interval not entirely located within the peak hour is observed in the real data; and these are not rare instances, 4 days on detector #2437, 1 day on #2473, and 5 days on #5124 (during 42 days).

The misplacement can also be observed directly by the different PHFs computed by different methods. If there is no misplaced peak interval, neither the peak hour nor the peak 15 minutes, the PHFs should be the same. Since the peak hour and peak 15 minutes can hardly start and end on the clock, the peak intervals, especial the peak 15-min interval, should be located by search instead of located on the clock. Otherwise, misplacement occurs and impacts the precision of PHF.

Local Data are Recommended

Values of the average PHF on the clock are higher than those by search in hypothetical cases, but they are lower in the cases with real data. There is nothing wrong with this situation; it just indicates the traffic patterns varying over time and on different sites, as shown by Tarko and Perez-Cartagena (2005).

Mathematically, the PHF by search should be smaller than that on the clock because the “real” peak 15-min traffic volume should be larger than that on the clock. Since the peak hour located by different methods may not be at the same place with the real data, to compare the values bases on different grounds may not provide a clearer insight. When the values are compared by pair, i.e. PHFR vs. PHFRin or PHFA vs. PHFAin, higher values are observed in the cases of ‘on the clock’ and within the hour by search.

The values of MAPE indicate the average differences between a pair of PHFs. Although it does not provide information about whether it is exceeded or shortened, it is still good to notice that there are errors between the pair. It is really hard to provide an overall estimation of the difference for cases because the traffic patterns are quite different. For example, values on the 7th row (as titled MAPE 3) in Table 5.3 show the difference between the searched and “on the clock” peak 15 minutes within an “on the clock” peak hour, they can range from 0.0196 to 0.0376 in different sites. It might be that the reasons for several default or recommended PHF values can be found in HCM 2000. For example, the statement of “PHFs in urban areas generally range between 0.80 to 0.98” can be found on page 8-9; the default value for PHF in Exhibit 10-12 is 0.92, but 0.90

in Exhibit A10-1; and it is 0.95 for congested conditions, 0.92 for urban areas, and 0.88 for rural areas on page 13-12 when there is an absence of field measurements.

The PHF is applied to determine design-hour flow rates, whether the design-hour is measured, established from the analysis of peaking patterns, or based on modeled demand. An inadequate design may result in substantial excess capacity the rest of the time or result in oversaturated conditions for a substantial portion during the peak hour. Local traffic data with higher resolution should be used to compute the PHF according to the results. When the 30-second interval data is used, traffic variation can be observed more clearly. Then it is obvious that there is no better way to get an estimation of the flow rate than using local, higher resolution data.

CONCLUSIONS

Since the peak intervals, either the peak hour or the peak 15 minutes, can hardly be located on the clock, the misplacement may bring an inaccurate PHF. By comparison with searching the “real” peak intervals, the misplaced peak intervals did occur and impact PHFs, if either lower resolution data or “on the clock” approaches were used. All p-values of the Wilcoxon Signed-Rank Test show that there are statistically significant differences between different methods to locate the peak intervals. Results from simulations in which traffic is simulated within a peak hour clearly address the impact of misplaced peak 15-min intervals on the PHF. Real data from 3 detectors randomly chosen from over 4000 detectors in the Twin Cities Metro area were also examined with four combinations of different methods to locate the peak intervals. All statistics support the recommendation of using local, higher resolution data to compute the PHF.

REFERENCES FOR PART V

- Fishman, G. S., 1996. Monte Carlo: Concepts, Algorithms, and Applications, Springer, New York.
- Little, J. D. C., 1961. Approximate Expected Delays for Several Maneuvers by Driver in a Poisson Traffic. *Operations Research* 9, 39-52.
- Nikolopoulos, K., Goodwin, P., Patelis, A., and Assimakopoulos, V., 2007. Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches. *European Journal of Operational Research* 180(1), 354-368.
- Robert, C.P., and Casella, G., 1999. Monte Carlo Statistical Methods. Springer-Verlag.
- Sheskin, D. J., 2007. Handbook of parametric and nonparametric statistical procedures, Fourth Edition. New York, Chapman & Hall/CRC Press.
- Tarko, A. P., and Perez-Cartagena, R. I., 2005. Variability of Peak Hour Factor at Intersections. *Transportation Research Record* 1920, 125-130.
- Transportation Research Board (TRB), 2000. Special Report 209: Highway Capacity Manual, TRB, National Research Council, Washington, D.C.

PART VI. CONCLUSIONS

CONCLUSIONS

In order to validate the impact of randomness on the average delay, cycle-length optimization, control types, and the peak-hour factor, this dissertation developed four individual investigations, examining some fundamental concepts in traffic operation. Each investigation results in a paper, which is included in this report. Since traffic varies over time and at different sites, it is a challenge to have a common recommendation for all kinds of conditions. State-of-the-art contributions to the profession as presented in each of the papers are summarized as follows.

In the first paper, “Impacts of Inter-Cycle Demand Fluctuations on Delay”, the importance of inter-cycle demand fluctuations on delay estimation are figured out, especially under heavy traffic conditions; since the unutilized capacity at a signalized intersection cannot be saved or carried over to be used by succeeding cycles. This paper demonstrates that different patterns of inter-cycle demand variance can result in different levels of delay estimation. It also points out that delay will be underestimated if Webster-type delay models are used, because those models treat the variance over the whole analysis period as constant for every signal cycle during the period.

In the second paper, “Short or Long ... which is Better? A Probabilistic Approach towards Cycle Length Optimization”, a five-step optimization framework to derive certain “just right” cycle lengths for a pre-timed signal operation is established. The probability of cycle failure as a secondary measure of effectiveness in traffic signal-timing analysis is proposed according to

simulation results in the paper. It also illustrates that longer cycle lengths may not yield optimal delay results; and very short cycles will increase delay dramatically due to a lack of capacity. The probabilistic approach to delay calculation, while more cumbersome than a straightforward fixed-demand and one-delay-value process, was found to be more realistic and insightful.

In the third paper, “A Trade-Off Framework for Determining the Best Control at an Intersection”, the trade-off framework based upon the traffic pattern to determine the “best” control type for an intersection is established. The average delay for the intersection is proposed to be a primary measure of effectiveness in the trade-off among different control types for an intersection. In order to facilitate such a decision, the sensitivities of the average delay for different control types at a hypothetical intersection under different traffic patterns are illustrated in this paper. In comparison to warrants in MUTCD, the framework provides a more realistic and insightful way to decide the control types for an intersection.

In the fourth paper, “Impacts of Misplaced Peak Intervals on PHFs”, it is identified that the “on the clock” PHF may be improper to determine design-hour flow rates; because a design based on a higher flow rate may result in substantial excess capacity during the rest of the peak hour; instead, a design based on a lower volume may result in oversaturated conditions for a substantial portion of the hour. Based on the Wilcoxon Signed-Rank tests, there are significant differences among different ways to define peak intervals. In order to reach a more proper PHF, it is recommended to use local, higher resolution

traffic count data. That is, the aggregation of higher resolution data may eliminate the variation and then lead to an improper PHF.

VITA

Jan-Mou Li was born in Taichung, Taiwan on July 14, 1968. He enrolled the National Taiwan Ocean University and earned his BS degree in Oceanography. Although he did not earn another degree in computer engineering, he did dedicate himself to computer programming and engineering during his undergraduate. In 1994, He earned his MBA in management of technology with a concentration on project management at National Chiao Tung University, which is ranked among top three universities in Taiwan. After the two-year official obligation and work in computer profession over five years, he realized the importance of transportation and joined the National Chiao Tung University again. He entered the doctoral program in Civil Engineering at the University of Tennessee, Knoxville in summer 2005 and pursuing his research in transportation. His research during the program involved validations on fundamental concepts in traffic operations. He officially received his doctoral degree in civil engineering in August, 2007.