Assessing the Sensibility of Signal Timing Split Optimization In Addressing Congestion

James E. Clark, PE

(Presented at the Institute of Transportation Engineers (ITE) 2007 Annual Meeting and Exhibit, August 2007; Pittsburgh, PA; copy distributed with ITE permission.)

Abstract

The assessment of traffic signal timing split optimization methods and the inherent problems associated with our profession's traditional approaches to develop optimum timings are addressed. The goal is to review the basis of the split optimization objective functions and discuss the comparative results of a simple experiment that was performed. The focus is on the split optimization objective function, not about various software methods being used today (this was purposely done to concentrate on the theory of optimization objective functions and not perform a comparison of software). A brief history and underlying theory surrounding traditional approaches as it relates to Webster's method, V/C, total intersection delay, Highway Capacity Manual procedures, critical movement delay, etc. are discussed. The results of a simple controlled experiment are discussed that clearly illustrate the problems associated with traditional split optimization techniques, thus questioning the sensibility of these methods. Further, a more realistic performance measure and methodology to achieve optimum timings is discussed and recommended. A profound and compelling argument challenging the transportation engineering profession to rethink our methodology of determining optimum signal timing splits is presented. The author stresses the critical need for the transportation engineering profession to further investigate this important technical topic.

PURPOSE

The purpose of this paper is to focus on the objective function for traffic signal timing split optimization and present the comparative results of a related simple signal timing experiment. The paper is not about comparing specific traffic signal timing optimization software methods being applied today, but rather the underlying methods such software might employ. This was purposely done to concentrate on the theory of optimization objective functions and not perform a comparison of software.

INTRODUCTION

Although none of us were born yet, even during the horse and buggy days traffic in big cities was often heavy. Back then, police officers had to be stationed full time directing traffic at busy intersections. Thus, police officers became our de-facto first form of traffic control.

As for an actual traffic control device, some[1] consider the world's first traffic signal came into being before the automobile was in use and traffic consisted only of pedestrians, buggies, and wagons. Installed at an intersection in London in 1868, it was a revolving lantern with red and green signals. Red meant "stop" and green meant "caution". The lantern, illuminated by gas, was turned by means of a lever at its base so that the appropriate light faced traffic. On January 2, 1869, this crude traffic signal exploded, injuring the policeman who was operating it. One could say that traffic signals today can still fail (rather than explode), it is just the cause is traffic congestion instead of a gaseous vapor!

Even from these early times, folks have wrestled with providing the best ("optimum") allocation ("split") of green time among the major and minor traffic movements. As the automobile rapidly emerged onto the scene, traffic congestion and intersection complexity got steadily more complicated. As a result, determining optimum signal timings splits was not a straight forward task.

In the past, those who have determined optimum signal timing splits have used various techniques to do so. Some of the methods used to attempt to develop optimum signal timing splits are:

- **A traffic cop** – Perhaps the first form of traffic control, where a police officer is stationed in the intersection, acting as a referee, or optimizer, determining green time allocation based upon his observation of traffic conditions.

- **"Seat-of-the-pants"** – These are the rare individuals that just seem to have a gift at determining pretty good initial timings based upon their years of experience.

- **Field trial & error** – This is not to be mistaken for final fine-tuning of signal timings in the field; rather this is just guessing or experimenting with timings, in hopes of achieving optimum performance. This is a time consuming, risky and inefficient way to determine optimum splits.

- **Formula calculation by hand** – The manual application of mathematical deterministic expressions that are based upon traffic flow research data to calculate optimum splits.

- **Computer software** – Software, designed around mathematical and subjective expressions, has helped traffic engineers to quickly determine optimum signal timings even for an entire system of traffic signals.

All of the above techniques or methods have been known to succeed and fail at producing optimum signal timings. If posed the question, "what is the basis of the split optimization function being applied?" it is likely that many traffic engineers would not have a very informed response. More often than not, one might get that deer-in-the-headlight-look, because they simply do not know. Yet, without understanding the basis of a method's objective function and the traffic performance measure being used to determine timings, how do we know we are producing sensible optimum timing splits?

TRAFFIC SIGNAL SPLIT OPTIMIZATION METHODS

The Traffic Cop, "Seat-of-the-pants" and Field Trial & Error signal timing methods are not based upon mathematical expressions or traffic flow theory principles. Rather, these techniques are purely subjective methods in attempting to develop optimum timings, although they do reflect an attempt to achieve some sort of objective, albeit a subjective one. As such, the discussion of this paper will focus on the technical approach associated with the mathematical formula of the remaining two methodologies.

Current methods for signal timing split optimization can be broadly categorized in either one of two forms, based on the optimization function's focus.

1) **Volume to Capacity (V/C) Methods** – based on some form of volume to capacity ratio (V/C), usually to either simply balance the V/C of critical movements, or to theoretically provide an approximate minimization of total intersection delay.

2) **Movement Delay Methods** – based on specific evaluation of individual movement's delays, usually focused on minimizing critical movements' delays.

These are profoundly different approaches to determining optimum signal splits. Later, a simple experiment will clearly illustrate the differences.

It should be noted that some methods of split optimization explicitly optimize the total intersection delay (or something similar) through iterative procedures, and that for the purposes of this paper these methods fall into the V/C Methods category. Also, some other methods of intersection evaluation fall into the V/C category that might not be so obvious. These include the Critical Movement Analysis (CMA) method of Circular 212 and the Intersection Capacity Utilization (ICU) method.

It must also be recognized that when V/C ratios approach 1.0 or higher, any optimization objective function will have difficulty determining the best signal splits since the results observed on the street will be bad regardless. In these cases it will be difficult to see the benefits of one method over another.

Volume to Capacity (V/C) Methods

Volume to Capacity (V/C) split optimization methods are primarily based on research conducted by F.V. Webster 50 years ago. The research findings of this effort can be found in the publication "*Traffic Signal Settings*, F.V. Webster, B. Sc., Ph.D., Road Research Technical Paper No. 39, London, Her Majesty's Stationery Office, 1958."[2] This research led Webster to focus primarily on 'degree of saturation (V/C) of critical movements' and 'average delay for the whole intersection' as the key areas to achieving optimum signal timing conditions at an intersection. Webster did some truly pioneering work that produced formula-based procedures for determining optimum signal timings and investigated the effect on delay associated with varying the signal cycle length. Some of the important results of Webster's efforts relative to this paper are illustrated in Figures 1, 2, 3 and 4 (these figures are copies of actual pages taken from Webster's Technical Paper No. 39).

SUMMARY OF PROCEDURE FOR SETTING TRAFFIC SIGNALS

In a period where the traffic flow is varying randomly about the mean, the procedure for obtaining optimum settings is as follows:

(i) Estimate the flow and saturation flow for each arm of the intersection.

(ii) Evaluate the ratio of flow to saturation flow for each arm, and select the y value for each phase (i.e. the maximum q/s value).

(iii) Add the y values together to give Y for the whole intersection.

(iv) Decide on all-red periods for pedestrians, turning traffic, etc. and estimate the lost time, R, due to this, e.g. if sequent ambers occur twice per cycle then $R=6$ seconds; if there are two all-red periods of 2 seconds each then $R=10$ seconds (see Fig. 6).

(v) Calculate the cycle time from equation (4):

$$c_0 = \frac{1{\cdot}5L + 5}{1 - Y}$$

where L is the total lost time per cycle, given by

$$L = nl + R$$

where n is the number of phases and l is the average lost time per phase due to starting delays.

(vi) Subtract the total lost time, L, from the cycle time giving the available green time and divide this in the ratio of the y values, i.e.

$$g_1 = \frac{y_1}{Y}(c_0 - L)$$

$$g_2 = \frac{y_2}{Y}(c_0 - L) \text{ etc.}$$

(vii) Add l seconds to each effective green time, $g_1, g_2 \ldots$ and subtract the amber period (3 seconds) to give the controller setting of green time.

Figure 1. Summary of Procedure for Setting Traffic Signals (from Webster's Research)
Figure 1 summarizes the formulae developed through Webster's research. Basically, Webster's procedural method suggests that the optimum green splits can be calculated based upon the use of equal V/C ratios for each critical movement at the intersection.

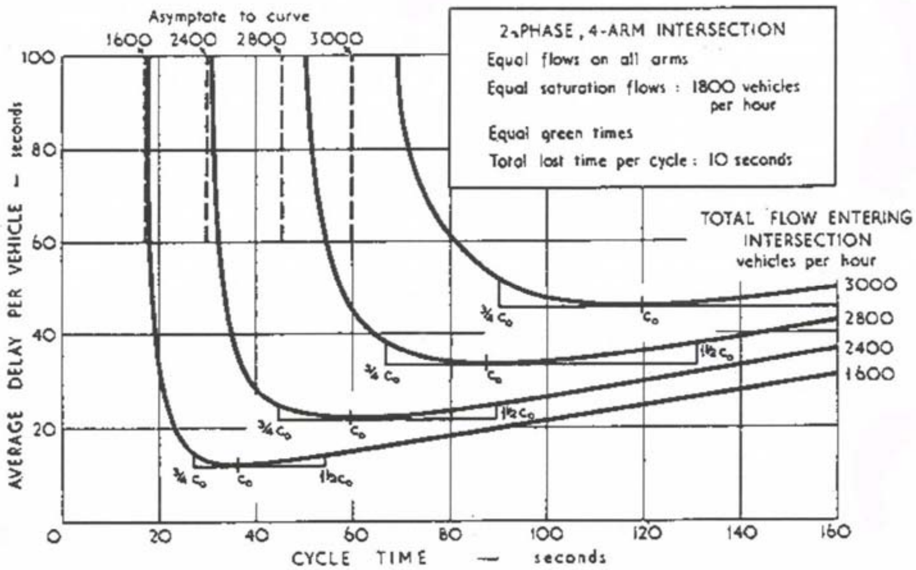

FIG. 7. EFFECT ON DELAY OF VARIATION OF THE CYCLE LENGTH

From graphs such as Fig. 7, it was found that in most practical cases the delay for cycle times within the range ¾ to 1½ times the optimum value is never more than 10 to 20 per cent greater than that given by the optimum cycle. This fact can be used in deducing a compromise cycle time when the level of flow varies considerably throughout the day. It would be better either to change the cycle time to take account of this, or, as is more common, to use vehicle-actuated signals. However, for a single setting of fixed-time signals the simple approximate method outlined below may be used.

Figure 2. Effect on Delay Associated with Varying the Signal Cycle Length (from Webster's Research)

Figure 2 (Webster's Figure 7 in Technical Paper No. 39) clearly illustrates that the average intersection delay (in seconds) per vehicle greatly increases with corresponding increases in the signal cycle length. The increase in delay shows the same pattern regardless of the traffic flow (in vehicles per hour) entering the intersection. As part of this particular research, Webster concludes, "…that in most practical cases the intersection delay for cycle times within ¾ to 1½ times the optimum value is never more than 10 to 20 per cent greater than that given by the optimum cycle."

Overall, Webster theorized that for optimum division of the optimum cycle time the degree of saturation or <u>V/C should be equal for all phases of the intersection</u>. Webster also concluded that equal V/Cs produce the <u>least delay for all vehicles using the intersection</u>. Figures 3 and 4 provide miscellaneous results and a summary of Webster's research.

MISCELLANEOUS RESULTS

Degree of saturation

It is shown in Appendix 4 that for optimum division of the cycle time the degree of saturation should be the same for all phases of the intersection. In this calculation, we have considered only one arm from each phase—the one with the highest q/s value. The degree of saturation for optimum settings of the controller appears to be independent of the amount of lost time per cycle, depending only on Y. It is given by equation (4·7) in Appendix 4 as

$$x_0 = \frac{2Y}{1+Y} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Average delay for the whole intersection

The average delay to all vehicles using an intersection has been deduced for optimum settings of the controller. The steps of the calculation are shown in Appendix 4 where the average delay per vehicle is given by equation (4·12) as

$$\bar{d} = \frac{c_0}{2}\left\{1 - \frac{\sum_1^{n'} y_r\ q_r}{YQ} + \frac{2n'Y^2}{LQ(1+Y)}\right\} \quad \text{approximately} \ldots\ldots (7)$$

where n' is the number of approaches to the intersection. This expression applies only to junctions where all arms of any one phase have approximately the same ratio of flow to saturation flow. The expression does not include the empirical correction term of equation (1), but this can be taken into account, approximately, by reducing $\bar{d}$ by about 10 per cent.

Figure 3. Miscellaneous Results (from Webster's Research)

Figure 3 (a copy of page 15 taken from Webster's Technical Paper No. 39), suggests that the optimum intersection signal timing split of the cycle length is achieved if all phases have equal V/C ratios.

SUMMARY

Delays εt intersections controlled by traffic signals have been investigated using an electronic computing machine to simulate traffic conditions.

A formula for the average delay per vehicle on a single approach to an intersection controlled by fixed-time traffic signals (or vehicle-actuated signals working on a fixed cycle because of heavy traffic demands) has been derived from the computed results.

Formulae have been deduced for the cycle time and green times which give the least delay to all vehicles using the intersection. Tables and formulae for queues and the number of stops and starts of vehicles have been obtained. These formulae and tables have been tested under actual road conditions with satisfactory results.

Figure 4. Summary (from Webster's Research)

Figure 4 (a copy of page 24 taken from Webster's Technical Paper No. 39), concludes that formulae developed as part of Webster's research produces the <u>least delay to all vehicles using the intersection</u>.

Later, through a simple controlled experiment use of this objective for optimization is reviewed more closely and questioned.

Summary of Volume to Capacity (V/C) Methods

Perhaps the important observations as a result of Webster's research could be summarized as follows:

- ♦ Explicit objective was to equalize the V/C ratio for the critical movements of all phases.
- ♦ Theoretically achieves the least delay for all vehicles using the intersection.
- ♦ Performance measure is critical movement V/C as a surrogate for total intersection delay.

Webster was a true pioneer, making significant contributions towards the advancement of the traffic engineering profession. The profession owes a lot of respect to Webster for his efforts in this area.

Movement Delay Methods

Movement Delay Methods have their roots in the Highway Capacity Manual[3] (HCM) procedures. In the United States, the HCM in one form or another is the basis for virtually all roadway capacity and operational analysis. The HCM has had a long time running, with its first publication in 1950, some 57 years ago!

The HCM has had six major editions: Edition 1 – 1950, Edition 2 – 1965, Edition 3 – 1985, Edition 4 – 1994, Edition 5 – 1997 and the current Edition 6 – 2000. The following is a brief summary of the important developments of the HCM as it relates to the critical movement delay method for determining optimum intersection signal timing splits.

1950 Manual:

- ♦ Basic concepts regarding intersection capacity.
- ♦ Number vehicles expressed in rate of flow relative to time and street width.
- ♦ Introduced saturation flow rate (referred to as basic capacity, i.e., 1,250 vehicles per 10 feet of width per hour of green).

1965 Manual:

- ♦ New term Load Factor, or "degree of utilization" per approach.
- ♦ New concept of Level of Service (LOS) as a function of signal load factor.
- ♦ Capacity analysis on an approach basis - there was no way to evaluate an entire intersection (all approaches). This reflected the importance of individual movements in contrast to the whole.

1985 Manual:

- ♦ "Stopped delay" per vehicle introduced as principal LOS determinant, replaced Load Factor as the MOE most relevant to the driver.
- ♦ Analysis based on "lane group", "approach" and intersection calculations.
- ♦ Incorporated signal control equipment type.

1994 Manual:

- ♦ Major changes to the 1985 version.
- ♦ Retained basic analysis methodologies/terminology.
- ♦ Significant changes:
    - Protected-Permitted left turn analysis.

- Application of progression factors.

1997 Manual:

- Major changes to the 1994 version.
- Retained basic analysis methodologies/terminology.
- Significant changes:
  - Oversaturated delay could be calculated.
  - “Total delay” (or control delay) used instead of stopped delay.

2000 Manual:

- A <u>major re-write</u> of the text was made.
- Methodology remained primarily unchanged.
- Most significant changes:
  - A comprehensive queue model was provided.
  - A new model for the effects of pedestrians and bicycles on saturation flow.
- Minor changes were made to the protected-permitted left turn model for shared lanes.

The 1985 HCM delay equation was originally based on the delay concepts formulated by Webster in 1958. As the HCM evolved, the HCM three part delay equation for a given lane group has become perhaps the most significant tool available for determining optimum signal timing splits, as described later in this paper. The current HCM delay equation, taken from the 2000 HCM, is shown below.

Determining HCM Delay

The average control delay per vehicle for a given lane group is given by Equation16-9. Appendix A provides a procedure to measure control delay in the field.

$$d = d_1 (PF) + d_2 + d_3 \qquad (16\text{-}9)$$

where:

d = control delay per vehicle (s/veh);
d_1 = uniform control delay assuming uniform arrivals (s/veh);
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 (s/veh); 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 (s/veh) (detailed in Appendix F of this chapter)

Progression Adjustment Factor

$$PF = \frac{(1-P)f_{PA}}{1-\left(\frac{g}{C}\right)} \qquad (16\text{-}10)$$

where
PF = progression adjustment factor,

P = proportion of vehicles arriving on green,
g/C = proportion of green time available, and
f_{PA} = supplemental adjustment factor for platoon arriving during green.

Uniform Delay

$$d_1 = \frac{0.5C\left(1-\frac{g}{C}\right)^2}{1-\left[\min(1,X)\frac{g}{C}\right]} \qquad (16\text{-}11)$$

where

d_1 = uniform control delay assuming uniform arrivals (s/veh);
C = cycle length (s); cycle length used in pretimed signal control, or average cycle length for actuated control (see Appendix B for signal timing estimation of actuated control parameters);
G = effective green time for lane group (s); green time used in pretimed signal control, or average lane group effective green time for actuated control (see Appendix B for signal timing estimation of actuated control parameters); and
X = v/c ratio or degree of saturation for lane group

Incremental Delay

$$d_2 = 900T\left[(X-1)+\sqrt{(X-1)^2+\frac{8kIX}{cT}}\right] \qquad (16\text{-}12)$$

Where
d_2 = incremental delay to account for effect of random and oversaturation queues, adjusted for duration of analysis period and type of signal control (s/veh); this delay component assumes that there is no initial queue for lane group at start of analysis period;
T = duration of analysis period (h);
K = incremental delay factor that is dependent on controller settings;
I = upstream filtering/metering adjustment factor;
C = lane group capacity (veh/h); and
X = lane group v/c ratio or degree of saturation

Estimation of d_3

$$d_3 = \frac{1800Q_b(1+u)t}{cT} \qquad (F16\text{-}1)$$

Where

Q_b = initial queue at the start of period T (veh),
c = adjusted lane group capacity (veh/h),
T = duration of analysis period (h),
t = duration of unmet demand in T (h), and
u = delay parameter.

With the historical focus of the HCM on evaluating individual movement's levels of service (LOS), and the current focus of the HCM on evaluating individual movement delays to determine their LOS, methods for optimizing signal splits based on individual movement delay have a similar focus. These methods use iterative procedures rather than formula-based calculations to determine the signal splits which either balance the critical movement delays or prioritize certain movement delays as long as other movements do not exceed a certain allowed threshold of acceptable delay. V/C is not intrinsically a part of the optimization, and total intersection delay is primarily ignored in favor of the focus on individual movement performance.

<u>Summary of Movement Delay Methods</u>

The important observations of the movement delay split optimization methods can thus be summarized as follows:

- ♦ The HCM's performance measure has always been focused on individual movement performance, not intersection performance, so movement delay split optimization methods apply the same focus.
- ♦ The HCM method is now focused on evaluating the MOE that matters to the driver – individual movement delay, not V/C, so movement delay split optimization methods focus on individual movement delays.
- ♦ HCM has never focused solely on total intersection delay as the performance measure, so nor do movement delay split optimization methods.

Comparison of V/C Method versus Movement Delay Method

There are significant differences in using a V/C Method versus a Movement Delay Method to determine optimum signal timings. These fundamental differences between the two styles of optimization are profound.

> *Webster's and other **V/C Methods** focus specifically on a <u>volume to capacity</u> (V/C) derivative and on <u>total intersection delay</u> as the implied performance measure.*
> ***Movement Delay Methods** focus on (critical) <u>movement delay</u> and not on V/C or total intersection delay.*

To Webster's credit, he was a true pioneer in the advancement of signal timings development. He was the first to introduce a systematic approach to traffic control settings which was rational and feasible. Webster invented the concept of delay that could be calculated. His mathematical formula that forms the basis of signal timing optimization could be solved by using a slide rule (they did not have personal computers back then). All of this work may have been fine 50 years ago, but let's look at this methodology as it applies to developing optimum signal timings today.

The primary limitations to a V/C-based methodology relate to the explicit and implicit variables being optimized:

- ♦ V/C as the explicit optimization variable is the wrong MOE:
    - It is a measure that drivers cannot and do not perceive.
    - Drivers do not understand V/C ratio, their performance measure is delay.
    - Balanced V/C ratios do not mean balanced delay or queues.
    - Balanced V/C ratios do not imply equitable assignment of green.

- Total intersection delay as the implicit optimization variable masks the problem:
  - Delay of critical movements is inversely proportional to their volumes, so low volume critical movement delay might be ten times (or more) higher than the high volume movement delay!
  - High delay for movements with low volumes almost disappears from the intersection total due to the small weight that the low volume has on the total.
  - Drivers do not perceive total intersection delay, only their own delay.

When considering these limitations, traffic engineers may ponder as to what all this means. Essentially, a V/C methodology optimizes a variable drivers don't perceive and at the same time causes the low volume critical movements to experience unreasonably high delays with the potential for significant queuing as illustrated in Figures 5 and 6.

Figure 5. Low Volume Critical Movements (in this case, side street traffic as circled)

Figure 6. Low Volume Critical Movements (in this case, side street traffic as circled)

It is important to note that turning vehicles exiting from the high volume main street are usually considered low volume critical movements, yet their potential impact on traffic congestion can be significant. This issue can actually be the very cause of the traffic congestion we are trying to solve. For example, Figures 7 and 8 illustrate how the unreasonably high delay of the low volume critical movements (here the left turn vehicles from the main street), if not adequately served, can begin to spillover and effect the adjacent high volume critical through movements. In Figures 7 and 8, the low volume critical movement "delay" and corresponding queue spills over and is the very cause of the congestion, since the higher volume adjacent through lane(s) become blocked.

Figure 7. Low Volume Critical Movements (in this case, main line left turn vehicles, as circled, are spilling over into the adjacent through lanes)

When this type of traffic congestion occurs it is doubtful that frustrated drivers or anyone for that matter are even questioning what the V/C ratio or intersection delay is! Their performance measure is not V/C or intersection delay; rather it is simply the amount of delay they experience – movement delay. This is why movement delay is used as the primary MOE in all versions of the HCM since 1985. The key problem with a V/C-based approach is that an equitable balance is achieved for a variable that doesn't matter to a driver (V/C), and it can result in unreasonably high values of delay to individual movements, this being the variable that does matter.

SIMPLE CONTROLLED EXPERIMENT

To illustrate the pros and cons of various signal split optimization methods a simple experiment is described [4]. The results of this experiment will clearly demonstrate that signal timing split optimization is best achieved and most palatable to drivers when based upon critical movement delay rather than V/C.

A simple controlled experiment was created in order to easily demonstrate and illustrate several points through the use of two dimensional tables and graphs. By controlling just a few parameters a traffic engineer will be able to easily recognize the value of applying a critical movement delay approach over the traditional V/C methodology. If a more complicated problem was created for this example, the results would be quite similar, but much more difficult to see with the clarity that the two dimensional example provides.

Simple Controlled Experiment Description

Figure 8 provides a brief schematic description of the simple controlled experiment performed here. It is a simple intersection of two one way streets which eliminates the need to be concerned with conflicting left turns in the experiment. The main street approach (high volume critical movement) has one lane with 1,000 vehicles per hour and the minor street approach (low volume critical movement) has one lane with 100 vehicles per hour. The intersection is controlled by a 100 second cycle length under two phase operation, with each phase having a three (3) second clearance time.

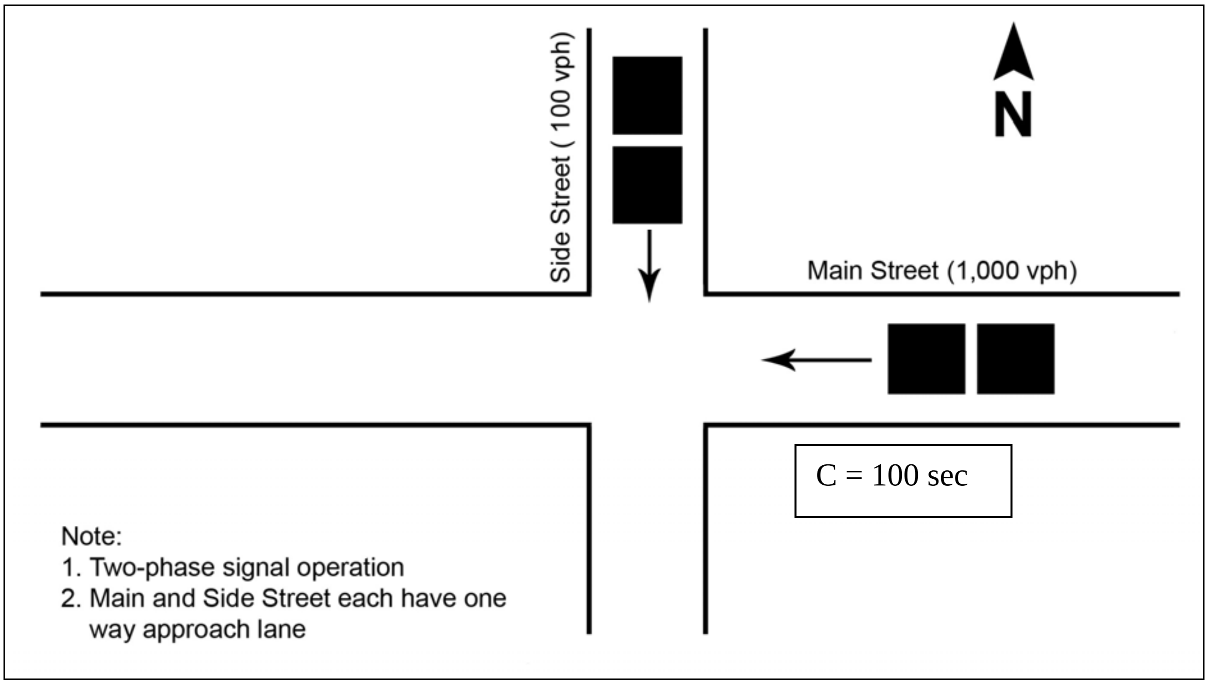

Figure 8. Schematic - Simple Controlled Experiment

Ten (10) separate split cases were developed and evaluated using the 2000 HCM. The results of this experiment are shown in Table 1 and in Figure 9. For each successive split case, five (5) seconds of green time were taken from the minor street green phase and given to the main street green phase. For each split case, the corresponding main street, minor street and intersection V/C ratios and delays were calculated with the HCM. For example, split case 1 had a main street green of 45 seconds, minor street green of 49 seconds (plus 6 seconds of clearance). This resulted in a main street V/C of 1.42, a minor street V/C of 0.13, an overall intersection V/C of 1.30, main street delay of 224 seconds per vehicle, minor street delay of 14 seconds per vehicle and overall intersection delay of 205 seconds per vehicle.

Table 1. Results of Split Optimization Experiment*

Split Case	Main Green	Minor Green	Main V/C	Minor V/C	Int V/C	Main Delay (sec/veh)	Minor Delay (sec/veh)	Int Delay (sec/veh)
1	45	49	1.42	0.13	1.30	224	14	205
2	50	44	1.28	0.15	1.18	159	17	146
3	55	39	1.16	0.16	1.07	107	20	99
4	60	34	1.07	0.19	0.99	67	24	63
5	65	29	0.98	0.22	0.91	40	28	39
6	70	24	0.91	0.27	0.85	24	32	25
7	75	19	0.81	0.34	0.81	16	38	18
8	80	14	0.80	0.46	0.77	10	46	14
9	85	9	0.75	0.71	0.75	7	68	12
10	90	4	0.71	1.59	0.79	4	369	37

* [Cycle Length = 100 seconds, two phase operation, 3 seconds clearance time per phase; main street 1,000 vehicles per hour, one lane, one-way; minor street 100 vehicles per hour, one lane, one-way.]

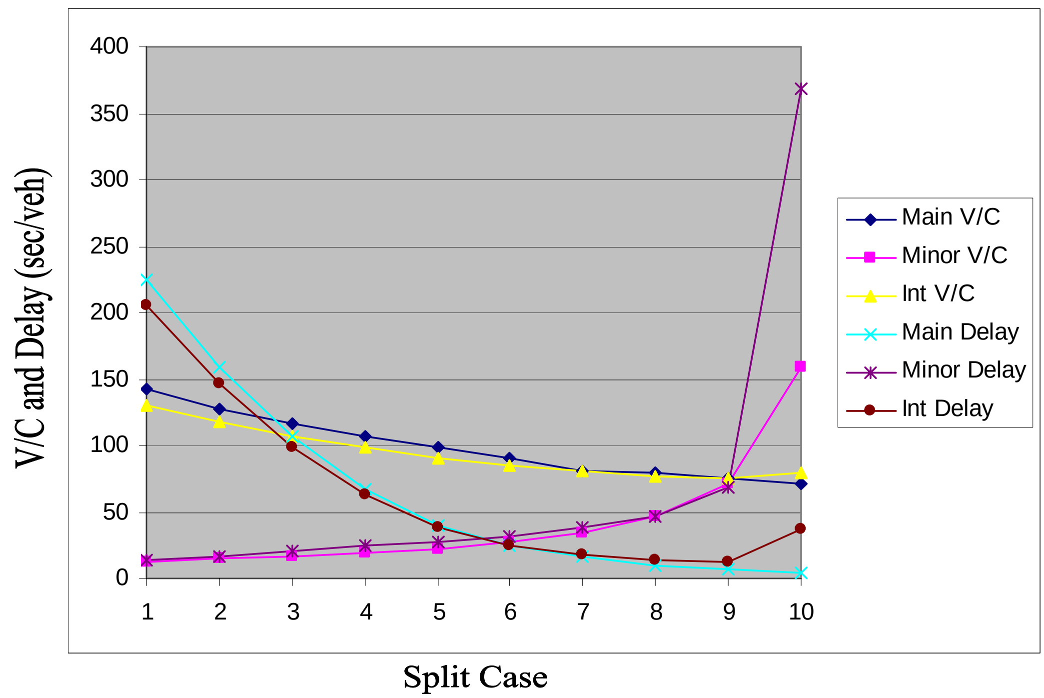

Figure 9. HCM Critical Movement Delay Experiment - Graph of V/C and Delay versus Split Case

Basic Results of Experiment

In reviewing the results produced in Table 1 and the corresponding graph of Figure 9, some interesting results are found. If we were to determine our optimum signal timing splits based upon a performance measure of lowest total intersection delay, it would suggest our optimum solution would be split case 9, since this case yields the lowest total intersection delay of 12 seconds per vehicle. Note also that as Webster's theory predicts, the V/C values are approximately equal for this case. However, if one was to drill down deeper and look at the corresponding main street versus minor street delays, it shows that the delays are significantly disproportional to one another. For split case 9, the corresponding minor street delay (68 seconds per vehicle) is approximately 10 times greater than the main street delay (7 seconds per vehicle)! Is this really a sensible and reasonable "optimal" solution? Is this a timing plan that the drivers in the minor traffic stream would find equitable? Is it a timing plan that can be defended as appropriate?

The corresponding graph reveals other interesting findings as well. If the traffic engineer was interested in more fairly balancing the delay among the main and minor streets, the optimum split arrangement, based upon the critical movement delay method would be somewhere between split cases 5 and 6. Regardless of what constitutes sensible "delay" among the critical movements, it is important to realize that the traditional way of focusing on average overall total intersection delay masks the problem of disproportionate delay among all movements. That is why it is not sensible to focus solely on total intersection delay as the performance measure, but also to be aware of critical movement delays.

If the controlled experiment were made to be more realistic, with two-way streets and conflicting turns, the same results would be observed – they would simply be difficult to display on a two-dimensional table and graph. However, one would immediately see that when a solution is based on balanced V/C for the critical movements, the phases with low volume critical movements would receive only a small amount of time and result in extreme delays (inversely proportional to the volumes) and the high-volume phases would receive a large amount of time resulting in disproportionately low delays. The only positive things that could be said about such a final result is that 1) the V/C for the critical movements was balanced and 2) the intersection delay was approximately minimized, but this assessment would fall on deaf ears for those drivers sitting in the long queues of the low volume phases.

Extended Results of Experiment

The simple controlled experiment was extended to determine optimum signal timing splits using two V/C Methods and two Movement Delay methods and the results were compared. Table 2 illustrates the results of this comparison. The top portion of Table 2 shows the results of the initial 10 split case experiment performed earlier. The bottom portion of Table 2 compares the V/C Methods (first two rows) to the Movement Delay Methods (last two rows).

Table 2. Comparative Results: V/C Methods versus Movement Delay Methods

Split Case	Main Green	Minor Green	Main V/C	Minor V/C	Int V/C	Main Delay (sec/veh)	Minor Delay (sec/veh)	Int Delay (sec/veh)
1	45	49	1.42	0.13	1.30	224	14	205
2	50	44	1.28	0.15	1.18	159	17	146
3	55	39	1.16	0.16	1.07	107	20	99
4	60	34	1.07	0.19	0.99	67	24	63
5	65	29	0.98	0.22	0.91	40	28	39
6	70	24	0.91	0.27	0.85	24	32	25
7	75	19	0.81	0.34	0.81	16	38	18
8	80	14	0.80	0.46	0.77	10	46	14
9	85	9	0.75	0.71	0.75	7	68	12
10	90	4	0.71	1.59	0.79	4	369	37
Split Case	Main Green	Minor Green	Main V/C	Minor V/C	Int V/C	Main Delay (sec/veh)	Minor Delay (sec/veh)	Int Delay (sec/veh)
Equal V/C	85.5	8.5	0.75	0.75	0.75	6	74	13
Min Int delay	84.4	9.6	0.76	0.67	0.75	7	63	12
Equal Delay	67.7	26.3	0.94	0.34	0.88	30	30	30
Priority Delay	76.7	17.3	0.83	0.37	0.79	14	40	16

Reviewing the second portion of Table 2 reveals the following results. The first split case (Equal V/C) was based upon Webster's explicit objective to equalize the V/C ratios for all phases. The Equal V/C row suggests that the main street green be set equal to 85.5 seconds and the minor street green set equal to 8.5 seconds. It can be seen that the V/C ratio is the same for the main street, the minor street and the intersection as a whole, all with a V/C value of 0.75. Based upon the analysis, overall average intersection delay was found to be 13 seconds per vehicle, while the main and minor street average delays are 6 and 74 seconds respectively.

In the second portion of Table 2, the second row (Min Int Delay) is based upon a Webster-type objective that explicitly attempts to minimize total intersection delay. This method suggests that the main street green be set equal to 84.4 seconds and the minor street green set equal to 9.6 seconds. As expected by Webster's theory, the resulting V/C ratios were close to being equal, with the main street, the minor street and the intersection as a whole at 0.76, 0.67 and 0.75, respectively. Based upon this approach, overall average intersection delay was found to be 12 seconds per vehicle, while the main street and minor street average delays are 7 and 63 seconds respectively. Not surprising, both Webster-type V/C methods produced similar optimum signal timing splits.

The third and fourth rows of the second portion of Table 2 are based upon Movement Delay Methods. If the traffic engineer wanted to balance the delay for the major and minor street movements (Equal Delay, third row), such a Movement Delay Method suggests that the main street green be 67.7 seconds and the minor street green be 26.3 seconds. This yields a main, minor and overall intersection average delay all exactly equal to 30 seconds per vehicle. For such a result the V/C ratios are nowhere close to being balanced, since V/C is not directly relevant to the objective for producing optimum splits - the main street V/C is 0.94, the minor street V/C is 0.34 and the intersection as a whole was 0.88.

The fourth row (Priority Delay) is a Movement Delay Method that allows the traffic engineer to establish sensible delay limits based upon what the engineer regards as reasonable. In this example the engineer assumes that no critical movement should have a delay greater than 40 seconds per vehicle and that the main street should otherwise be prioritized. Using such a Movement Delay Method, this suggests that the main street green be equal to 76.7 seconds and the minor street green be equal to 17.3 seconds. This produces an overall average intersection delay of 16 seconds per vehicle, while the main and minor street average delay are 14 and 40 seconds per vehicle, respectively. As before, the V/C ratios are still nowhere close to being balanced, since V/C balancing has nothing to do with producing optimum splits according to this optimization scheme - the main street V/C is 0.83, the minor street V/C is 0.37 and the intersection as a whole was 0.79.

Findings

In reviewing the initial results of Table 2 the traffic engineer could easily be coerced by historical methods to select the suggested optimum splits achieved from either of the V/C Methods, since they both produced the lowest overall average intersection delay of around 12 seconds per vehicle. Yet, is this a really an "optimum" solution when the minor street low volume critical movement delay is 9-10 times higher than the main street high volume critical movement delay and clearly at an unacceptable level by any standard? A V/C method of any form for determining optimum splits masks this disproportion amount of delay among the critical movements by focusing attention on the V/C and intersection delay. Is this really an optimum and sensible solution?

Alternatively, should the engineer be focused on the critical movement delays and select the Equal Delay solution, or be even more insightful and select the Priority Delay solution? In either case, the basic premise of this paper is that if we measure the quality of service of signal control using the delay of individual movements, then we should use this measure as the basis of our split optimization, and thus select a Movement Delay Method for optimization, not a V/C Method.

CONCLUSIONS

A comparison of signal timing split optimization objective functions was performed. Through a simple controlled experiment, it was found that when compared to Movement Delay Methods, the traditional Webster-type V/C Methods produce a significantly disproportionate share of delay among the critical movements. It is not uncommon to have such an imbalance of delay among the critical movements exceeding tenfold. This suggest that the traditional V/C Methods used to determine optimum splits results in the major high volume critical movements enjoying low delays at the expense of the low volume critical movements. Again, it is important to note that turning vehicles exiting from the high volume main street may be considered low volume critical movements, yet their potential impact on traffic congestion can be significant. These phenomena can actually be the very cause of the traffic congestion we are trying to solve. As such,

……major high volume movements enjoy disproportionately low delays……

……at the expense of the low volume movements which experience high delays ……..

……………………………. (drivers' thoughts tongue-in cheek) ……..

……and even the lower volume critical turning movements exiting from the high volume main street can experience significant delay and be the cause of further traffic congestion……

[Circled area - represents low volume critical movement left turn vehicles in the main line facility beginning to spill back into adjacent through lanes.]

……Is a Webster-type V/C Method of split optimization really reasonable?

………………………… (drivers' thoughts tongue-in cheek) ……..

As a result of the analysis and assessment in this paper, the author concludes that:

- ♦ V/C is not an appropriate objective function for determining optimum splits.
- ♦ Signal timing split optimization based upon a Webster-type V/C method can produce unreasonably high delays for the low volume critical movements.
- ♦ To base optimum splits on an optimization of movement V/C or overall total intersection delay is the wrong performance measure.
- ♦ The current definitive measure of intersection performance is HCM delay for individual movements, so this should be reflected in the split optimization process.
- ♦ Signal timing split optimization should be based upon individual (critical) movement delays if we are to achieve sensible and defensible signal timings for all movements.

Challenge to the Profession and Recommendations

Given the limitations of a V/C-based split optimization methodology, the challenge to the traffic engineering profession is to determine:

- Is the basis of a V/C-based split optimization methodology really a valid objective for proclaiming that optimum signal timing splits are achieved?
- What should the split optimization objective function be based upon - HCM movement delay?
- How is the profession going to achieve this objective?

In response to this challenge, it is recommended that the Institute of Transportation Engineers:

- Investigate this technical topic further,
- Involve the appropriate traffic flow theorists, experts and practitioners to resolve this issue, and
- Provide guidance to the profession regarding the development of optimum signal timing splits.

ACKNOWLEDGEMENTS

The author would like to acknowledge the assistance and contribution of several entities and people without whom this effort and the resulting paper would not have been possible. Thanks to several communities throughout Fairfax County Virginia (Centreville, Chantilly, Falls Church, the Town of Herndon and Tyson's Corners) for providing real examples of the kinds of traffic problems this paper explores; to the Virginia Section of the Institute of Transportation Engineers (VASITE) for the opportunity to initially present this technical subject (especially, Mr. Christopher Daily of VHB, Inc. for requesting that I present this topic at the 2006 VASITE Annual Meeting); to Dr. Robert Edelstein of DMJM Harris for providing many resources in support of the initial effort; to Mr. Amit Sidhaye of DMJM Harris for his hard work that was essential for the timely completion of the research and preparation of materials for the initial presentation at the VASITE 2006 Annual Meeting.

Thanks to Tindale-Oliver & Associates, Inc. – Mr. Steven Tindale for supporting me in the development of this paper and to Ms. Christina O'Neal for helping me complete this effort on time; to Mr. Thomas Brahms, Executive Director of Institute of Transportation Engineers and Mr. Warren Hughes of VHB, Inc. for their encouragement and strong support to continue making technical contributions to the transportation engineering profession.

A special thanks to Mr. Dennis Strong, President of Strong Concepts, who provided essential theoretical support for the discussion of the various analytical models and optimization strategies covered by this paper. Respectfully, Mr. Strong's professional dedication, knowledge and ability to explain traffic flow theory and practice are unmatched. As the immediate past chair of the Signals subcommittee for the TRB Highway Capacity Committee involved with the publication of the three most recent HCM editions, and as an internationally-recognized traffic engineering lecturer for Universities throughout the U.S.A. and the world, he provided continuous expert review of the technical aspects of the paper without which this effort could not have been completed.

REFERENCES

1 - www.ideafinder.com/history/inventions/trafficlight.htm.
2 - "*Traffic Signal Settings*, F.V. Webster, B. Sc., Ph.D., Road Research Technical Paper No. 39, London, Her Majesty's Stationery Office, 1958."
3 – Highway Capacity Manual 2000, Transportation Research Board, January 2000.
4 – A short course, "Designing Optimized Traffic Signals and Systems", Strong Concepts, 2007.

AUTHOR INFORMATION:

James E. Clark, PE
Director of Traffic Operations and
Intelligent Transportation Systems
Tindale-Oliver & Associates, Inc.
1595 S. Semoran Blvd.
Building 7, Suite 1540
Winter Park, FL 32792 USA
Telephone: 407-657-9210
Fax: 407-657-9106
Email: jclark@tindaleoliver.com