ROBUST QUEUE-BASED TRAFFIC SIGNAL CONTROL

Lin XIAO

Hong K LO[*]

Department of Civil and Environmental Engineering

The Hong Kong University of Science and Technology

Clear Water Bay, Hong Kong

ABSTRACT

Traffic signal remains an active research topic for many years. A critical aspect of traffic signal analysis is modelling a signal plan's delay performance, in particular when traffic arrivals are stochastic. In this study, based on the reliability framework in Lo (2006), we develop a robust control scheme for general arrival distributions while considering the effects of residual queue from a previous cycle. The proposed strategy incorporates the notion of a "queue-based signal switching rule" into the probabilistic or reliability framework. Instead of aiming to minimize vehicle delay directly, this control scheme aims to maintain vehicle queues within permissible ranges. While doing this, the expected delay is also reduced. Some numerical results are included to demonstrate the benefit of this approach as compared with Webster's approach and our earlier reliability-based approach.

Keywords: robust control, reliability, traffic signal control

[*] Corresponding author. Tel.: +852 2358 8742; fax: +852 2358 1534.
E-mail address: cehklo@ust.hk (Hong K. Lo).

BACKGROUND

Traffic signal control remains an active research topic for many years. Traffic signals ensure safety by separating conflicting vehicular and pedestrian traffic through appropriately allocating green times. Besides this safety consideration, a critical aspect of traffic signal analysis is modelling a signal plan's delay performance. In the idealized situation, traffic arrives and departs uniformly; a deterministic queue regime such as D/D/1 suffices to model the signal control system. To cater for stochastic traffic arrivals, Webster (1958) developed one of the most commonly used delay equations based on the M/D/1 queue regime. Webster's formula overestimates delay when the degree of saturation approaches 1. To rectify this problem, different sheared delay equations have been developed in the past, such as Allsop (1971). Their philosophy, however, remains the same – that the performance of a signalized junction in the presence of stochastic arrivals can be represented by a delay equation, in one form or another. It implies that the effect of stochastic arrivals is mainly handled by adding stochastic terms in delay formulae. Although using delay formulae for designing signal plans is convenient, it is somewhat indirect as certain arrival distributions are assumed in the formulae, which may be different from the actual traffic. This drawback can be addressed by incorporating the arrival distributions in a more direct manner.

For actuated control systems, a typical strategy is to switch the green time to the other approach whenever the queue on the approach being served vanishes (Newell, 1989). To fully implement this strategy, it will result in variable phase durations and cycle lengths. More recent studies focus on control schemes that can cater for over-saturated traffic, or spillback or blockage effects (e.g., Lo, 1999, 2001; Liu, Balke et al., 2008; Lo, 2004; Chow, 2007) via adaptive control methods (e.g. Lefeber and Rooda, 2006; Murat and Gedizlioglu, 2005). In these schemes, typically a fixed cycle length cannot be ensured; the resultant cycle time may be unreasonable short when the traffic is light, or exceedingly long when it is heavy. On the other hand, a stable cycle length is often required to allow for coordination between junctions along a corridor. It is recognized that delay reductions from coordination often contribute substantially as compared with other control elements (Newell, 1989).

Under the constraint of a fixed cycle, Lin and Lo (2009) proposed a quasi-dynamic control scheme that switches the green time according to the queue lengths between competing approaches. The goal of that scheme is to maintain the relative queue lengths of competing approaches to a certain desirable proportion. There are different ways of choosing this desirable proportion, such as using Miller's approach (Miller, 1963). Overall, the simulation result from Lin and Lo (2009) showed that their approach was consistently better than a fixed timing plan based on Webster's timing plan. In this study, we will extend the approach by connecting it to a formal reliability framework, as will be explained in detail later.

Lo (2006) developed a reliability-based framework to measure the performance of traffic signal. That study proposed the measure of phase clearance reliability (PCR), which describes the probability that the available green time of a given phase is able to clear the approach traffic. In fact, a similar measure was proposed by Haight (1959) but Lo (2006) extended the consideration to include the effects from previous cycles which he called the higher-order effect. Based on this reliability framework, a timing plan that satisfies a certain PCR requirement can be designed. As a fixed-time plan, in long run, it should achieve the specified PCR; but in the short run, due to random arrivals, some approaches may be overloaded temporarily. Therefore, there are benefits to be gained if the timing plan can be adaptively modified by shifting green times to approaches that are temporarily overloaded, while maintaining the cycle time to be unchanged. The questions are, of course, what the specific conditions are that would warrant this shift in green time, and how much green time we should relocate temporarily.

In this study, based on the reliability framework developed in Lo (2006), we develop a robust control scheme for general traffic arrival distributions while considering the effects of residual queue from a previous cycle. The proposed strategy incorporates the notion of a “queue-based signal switching rule” into the reliability framework. By varying the green time allocation, the strategy endeavours to contain queue lengths to be within certain permissible ranges based on an analysis of the phase clearance reliability of their corresponding phases. Within these permissible ranges, queues would be cleared within the next cycle with a specific desired probability, say 90%. Any queue states deviate from this permissible range will trigger green time switching. By doing so, instead of aiming to reduce the average system delay directly, the controller reallocates the green time to ensure its full utilization. In this way, the timing plans may change from cycle to cycle, depending on the queue state of the previous cycle, subject to the cycle time being fixed. The outline of this paper is as follows. Section 2 describes the logic of the proposed control scheme. Section 3 contains numerical studies to illustrate the performance of this control scheme. Section 4 provides some concluding remarks.

LOGIC OF THE PROPOSED CONTROL SCHEME

The goal of the proposed control scheme is to keep the queuin state to be within certain range. We refer to this range as the permissible range, defined as a set of queue states that can be cleared in the next green phase with a specified probability. Without loss of generality, we explain the logic of the proposed control scheme with an example for a single intersection with two competing approaches in the north and westbound directions (or approaches 1 and 2) without turning movements.

We begin with a fixed-time plan (Figure 1), designed for the intersection based on the prevailing traffic condition together with some specific considerations (signal coordination, for instance) according to the Webster approach or a reliability-based timing plan with PCR=0.9 (refer to Lo, 2006 for details).

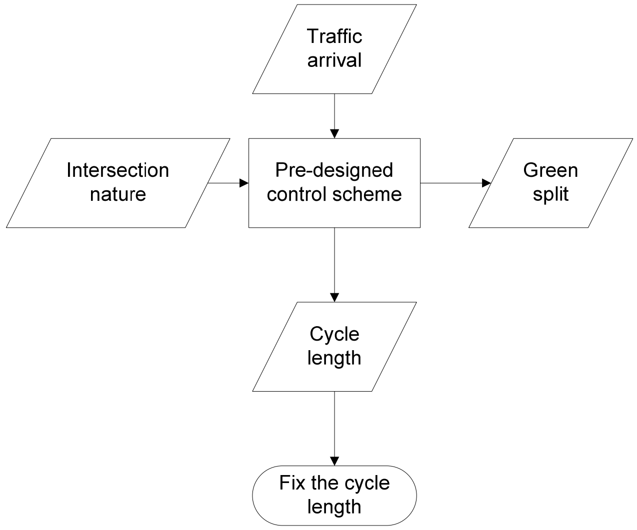

Figure 1 – A pre-designed fixed-time plan

Based on the current timing plan, the approach volume, and a desired PCR (to be explained in more detail later), the controller defines the permissible range for the intersection. At the end of each red phase, the sensors record the queue length on both approaches. Then the controller determines whether the queue lengths have exceeded the boundaries of the permissible ranges. If so, the controller reallocates the green times between different approaches so as to bring the queuing states back to the permissible range; otherwise, the pre-determined fixed-time plan continues without any changes. The logic of how the controller works is shown schematically in Figure 2.

In the following subsections, we will define exactly how the permissible ranges are determined and the green time relocation procedure.

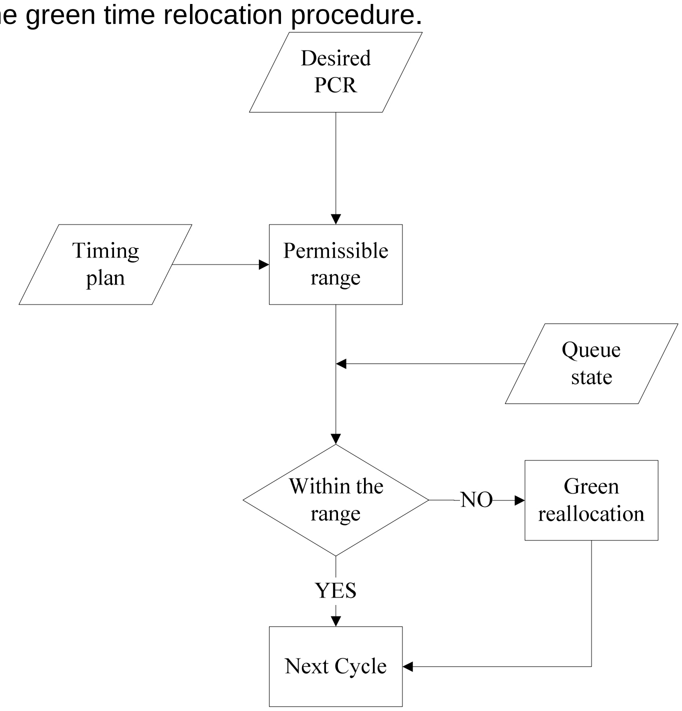

Figure 2 – The schematic diagram of the control logic

Permissible range

The objective of the adaptive signal control proposed here is to ensure that each approach can accomplish a phase clearance reliability (PCR) of α, or that the queue for each approach would be cleared in its next green phase with a probability of at least α. Mathematically, this requirement for a single approach is expressed as:

$$P\{q+\lambda g \le \mu g\} \ge \alpha \tag{1}$$

where q is the residual queue of the approach from the previous cycle; λ, μ are arrival and departure rates (in vehicles per second, veh/s) respectively; g is green time. The term $q+\lambda g$ gives the total queue (the residual queue plus current arrivals) to be cleared in the current cycle; μg gives the amount of traffic that can be cleared within the available green time.

The left hand side of (1) can be transformed to:

$$P\left\{\lambda \le \mu - \frac{q}{g}\right\} \ge \alpha \tag{2}$$

Let the arrival rate of traffic follows a certain probability distribution, then one can find a specific value λ_0 such that $P\{\lambda \le \lambda_0\} = \alpha$. Comparing with (2), we can say that (2) holds as long as:

$$\lambda_0 \le \mu - \frac{q}{g}\text{, or rearranging,} \tag{3}$$

$$q \le (\mu - \lambda_0) g \tag{4}$$

In order to achieve (2), the residual queue from the previous cycle must satisfy (4), which defines a range. Any residual queue lengths satisfy (4) have a probability of at least α to be cleared in the next green phase. This range is referred to as the permissible range in this study. In other words, the permissible range defines a set of residual queuing states that possess a certain specified probability of being cleared in the next cycle. The boundary of the permissible range for a single approach is defined as:

$$H = (\mu - \lambda_0)\, g$$

As an example, suppose traffic arrivals at an approach follow a log-normal distribution with mean m and variance σ. The cumulative probability distribution of λ can be expressed as:

$$P\{\lambda \le \lambda_0\} = \frac{1}{2} erfc\left[-\frac{\ln \lambda_0 - m}{\sigma\sqrt{2}}\right] \tag{5}$$

where $erfc(x)$ is the complementary error function. λ_0 can be derived as:

$$P\{\lambda \le \lambda_0\} = \frac{1}{2} erfc\left[-\frac{\ln\lambda_0 - m}{\sigma\sqrt{2}}\right] = \alpha$$
$$\Rightarrow erf\left[-\frac{\ln\lambda_0 - m}{\sigma\sqrt{2}}\right] = 1 - 2\alpha$$
$$\Rightarrow -\frac{\ln\lambda_0 - m}{\sigma\sqrt{2}} = erf^{-1}(1-2\alpha)$$

where $erf(x)$ and $erf^{-1}(x)$ are the error function and inverse error functions, respectively. Rearranging, we obtain:

$$\lambda_0 = \exp\left[m - \sqrt{2}\sigma \cdot erf^{-1}(1-2\alpha)\right] \quad (6)$$

The permissible range on this approach thus can be expressed as:

$$H = (\mu - \lambda_0)\, g = \left\{\mu - \exp\left[m - \sqrt{2}\sigma \cdot erf^{-1}(1-2\alpha)\right]\right\} g \quad (7)$$

Since residual queues may appear on different approaches of the intersection, we repeat the above analysis to all approaches and define the permissible ranges with two boundaries:

$$\begin{cases} H_1 = (\mu_1 - \lambda_0^1) g_1 = \left\{\mu_1 - \exp\left[m_1 - \sqrt{2}\sigma_1 \cdot erf^{-1}(1-2\alpha)\right]\right\} g_1 \\ H_2 = (\mu_2 - \lambda_0^2) g_2 = \left\{\mu_2 - \exp\left[m_2 - \sqrt{2}\sigma_2 \cdot erf^{-1}(1-2\alpha)\right]\right\} g_2 \end{cases} \quad (8)$$

where the variable subscripts 1 and 2 in (8) denote their respective approaches, as are the superscripts 1 and 2 in λ_0^1 and λ_0^2.

Graphically, the permissible range of the intersection is a rectangle bounded by H_1 (wide) and H_2 (high).

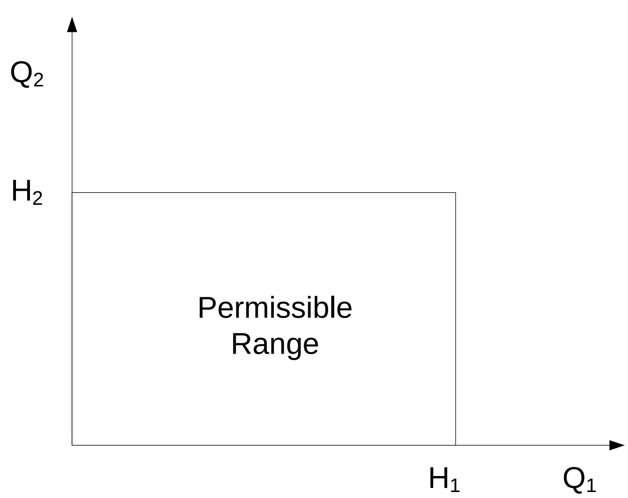

Figure 3 – The permissible range

In the following subsection, we use q_i to denote the queue length on approach *i* and H_i denotes the boundary of permissible range in *i*th approach.

Green reallocation

At the end of each cycle, the controller evaluates whether the queue lengths on both approaches deviate from the permissible ranges defined above. Obviously, the result must be one of the following four scenarios:

Scenario 1: $q_i \le H_i$

Scenario 2: $q_1 > H_1$ and $q_2 \le H_2$

Scenario 3: $q_1 \le H_1$ and $q_2 > H_2$

Scenario 4: $q_i > H_i$

The controller will adjust the green durations to cater for these four different scenarios, as explained below.

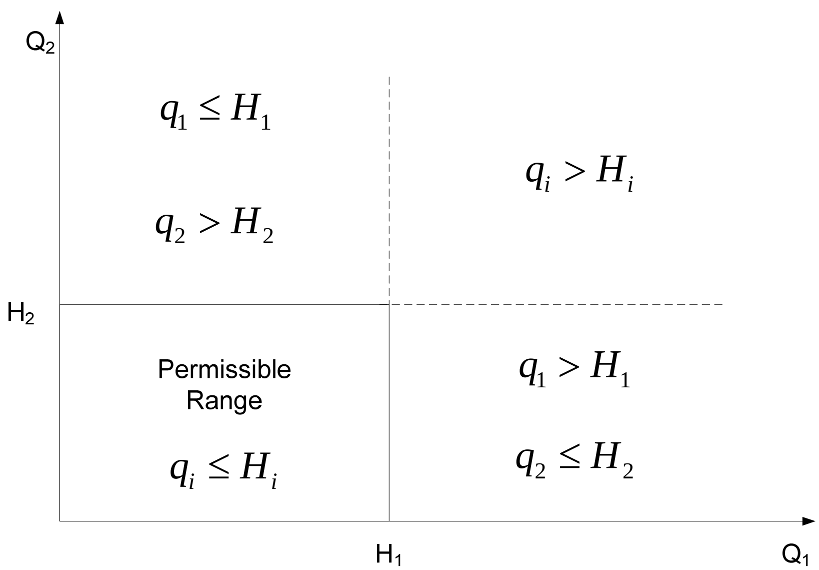

Figure 4 – Four scenarios of residual queuing states

For Scenario 1, when $q_i \le H_i$, the residual queues on both approaches are within the permissible range. There is no need to change the green split.

For Scenario 2 or 3, the queue on only one of the two approaches falls outside the permissible range. The controller will reallocate green time in the next cycle from the approach with a shorter queue to one with a longer queue. As the two scenarios are symmetric, the analysis for Scenario 2 can be readily modified for Scenario 3. In the following, we focus the discussion for Scenario 2.

Since the queue on approach 2 is shorter and within the permissible range, its default green time would not be fully utilized in the next cycle; hence part of it can be reallocated to approach 1. We use δ to denote this green time reallocated to approach 1. Then the new green time allocation should be:

$$\begin{cases} g_1^{'} = g_1 + \delta \\ g_2^{'} = g_2 - \delta \end{cases}$$

Obviously, with the additional green time, it is expected that the performance of approach 1 would improve, while maintaining the same cycle time. However, we cannot subtract too much green time from the competing approach; otherwise, the queue or delay there may increase sharply. The amount of green time that can be swapped from approach 2 to approach 1 can be estimated by the following analysis. The bottom line is that approach 2 should not be harmed by the swap. That is, approach 2 should remain in the permissible range, expressed as:

$$\max\{q_2 + \lambda_2 g_2 - \mu_2 g_2, 0\} + \lambda_2 (g_1 + \delta) \le H_2 \tag{9}$$

In the term $\max\{q_2 + \lambda_2 g_2 - \mu_2 g_2, 0\}$, $q_2 + \lambda_2 g_2$ represents the sum of the residual queue from the previous cycle and the arrivals within the current green time; $\mu_2 g_2$ represents the capacity of the current green. Therefore, $\max\{q_2 + \lambda_2 g_2 - \mu_2 g_2, 0\}$ represents the residual queue at the end of the current green. And $\lambda_2 (g_1 + \delta)$ represents the amount of arrivals at approach 2 during its red phase after swapping green time δ to approach 1. Thus, the LHS of (9) requires that the residual queue at the end of the next red phase to be within the permissible range H_2.

Since approach 2 currently falls within the permissible range, i.e. from the definition of phase clearance reliability, we have

$$P\left\{\max\{q_2 + \lambda_2 g_2 - \mu_2 g_2, 0\} = 0\right\} \ge \alpha .$$

By selecting a high α, which is probably the case for signal control, one can approximate $\max\{q_2 + \lambda_2 g_2 - \mu_2 g_2, 0\} = 0$. Thus, (9) can be simplified to:

$$\lambda_2 (g_1 + \delta) \le H_2 ,$$

with the resulting δ stated as: $\delta \le \frac{H_2}{\lambda_2} - g_1$.

Similarly, repeating the analysis for Scenario 3, the amount of green time that can be added to approach 2 is $\delta \le \frac{H_1}{\lambda_1} - g_2$.

The situation is more complicated when both queues fall outside the permissible ranges since we cannot simply swap green times between the approaches. Actually, this scenario indicates that the junction as a whole is temporarily overloaded. In this case, the method described in Lin et al (2009) applies, which aims at maintaining the queue lengths between the different approaches at a balanced proportion.

Lin et al (2009)'s control scheme first defines a desired proportion for queue lengths on the two approaches based on a stochastic analysis of the relative queuing states. This proportion is shown as the solid line in Figure 5, with a slope of s. Any queue state deviates from the desired proportion will be brought back to the line by adopting different green splits. Basically, three plans will be applied; we refer them as Plans A, B and C. Plan A is the dominant one and will be adopted when queue states do not deviate from the desired proportion too much. Plan B and C are considered as transitional plans and should be in operation whenever the queue states changes substantially. The condition for switching between different plans is based on the queue states as shown in Figure 5. Regions A, B and C are separated by two parallel dashed lines. Plans A, B and C are executed when the current queue state falls into the corresponding region.

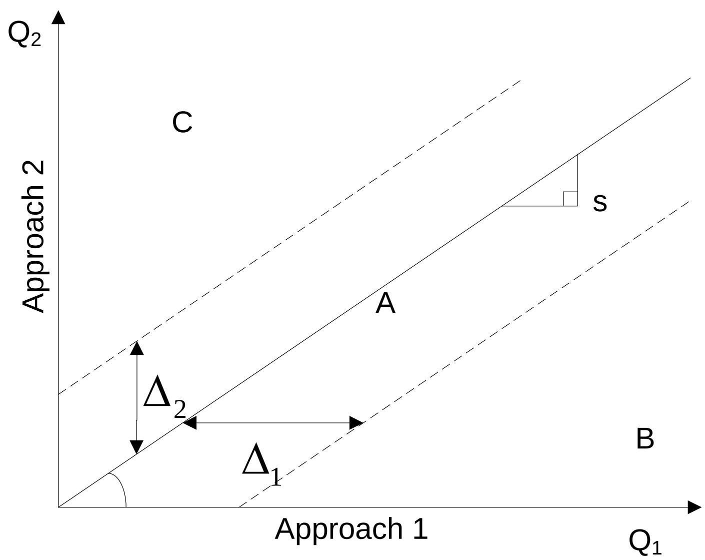

Figure 5 – Condition for signal switching

Suppose that the green splits in Plans A, B and C are $\left(g_1^A, g_2^A\right)$, $\left(g_1^B, g_2^B\right)$ and $\left(g_1^C, g_2^C\right)$, respectively. Plan B gives more green time to approach 1 while plan C gives more to approach 2. With plan A determined by the prevailing traffic condition, the green time for plan B must satisfy $g_1^B = g_1^A + \delta^B$ and $g_2^B = g_2^A - \delta^B$; the green time for plan C must satisfy $g_1^C = g_1^A - \delta^C$ and $g_2^C = g_2^A + \delta^C$, while maintaining a common cycle length.

According to Lin et al (2009)'s analysis, the parameters in Figure 5 are chosen as:

$$\delta^B = C - g_1^A - L - \frac{\lambda_2}{\mu_2} C$$

$$\delta^C = C - g_2^A - L - \frac{\lambda_1}{\mu_1} C$$

$$\Delta_1 = \mu_1 g_1^B - \lambda_1 C + \left(\lambda_2 C - \mu_2 g_2^B\right) s$$

$$\Delta_2 = \mu_2 g_2^C - \lambda_2 C + \left(\lambda_1 C - \mu_1 g_1^C\right) / s \qquad \text{where } s = \frac{\lambda_1}{\lambda_2}.$$

The signal will be switched to B when $q_1 - q_2 s > \Delta_1$ or to C when $q_2 - q_1 / s > \Delta_2$. In the interest of space, we omit the detail of the analysis here, which can be found in Lin et al (2009).

NUMERICAL STUDIES

Simulation tests are conducted to investigate the performance of the proposed control scheme. The result (average delay and average queue length) generated from the proposed scheme is compared with the result obtained from the Webster's optimal timing plan and a timing plan with a PCR of 0.9 (see Lo, 2006 for details of this approach).

We study an intersection with two competing approaches in the northbound and westbound directions (or approach 1 and 2), each with the same capacity. Each approach allows for only through traffic and vehicle arrivals follow a log-normal distribution. The log-normal distribution is appropriate for traffic arrivals in a cycle considered as a multiplicative product of many independent random arrivals. The degree of saturation of the entire intersection is set to increase by seven levels, from 0.6, via 0.65, 0.7, 0.75, 0.8, 0.85, to 0.9, to study how the proposed control scheme performs in different levels of congestion. Furthermore, for each degree of saturation, two cases are constructed: first, the total arrivals are distributed equally between the two approaches; second, one approach carries 80% of the traffic while the other 20%. In total, 14 cases (seven times two) are evaluated for comparison purposes. For each case, 10 iterations are simulated, with 20 cycles in each iteration.

First, we investigate how the controller performs by plotting the residual queues during the control period, as compared with the results obtained from a fixed-time plan with PCR=0.9. The residual queue for each cycle is represented by a single point as shown in Figure 6, which includes the results for 60 cycles when the degree of saturation is set at 0.75. The figure also shows the permissible range, the rectangle in the lower left corner, which allows for the residual queue of 13 vehicles for approach 1 and 15 for approach 2. Figure 6 illustrates that the proposed controller is capable of pulling the residual queue length back to the permissible range, with a much larger proportion of pink dots falling within the permissible range as compared with the dark blue dots. In another words, the proposed control scheme allows for a much higher probability of clearing traffic in the intersection.

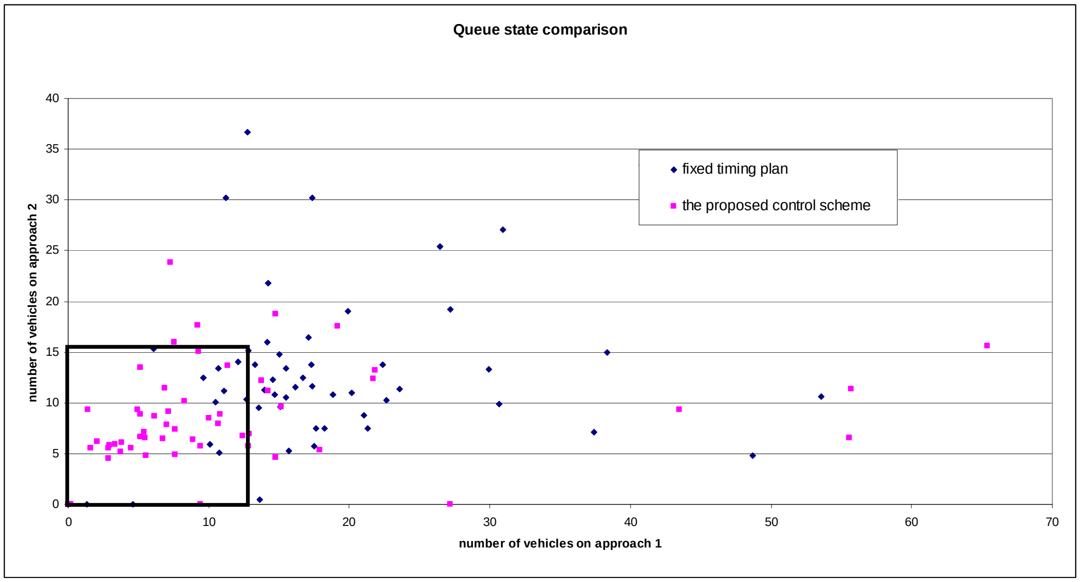

Figure 6 – The resultant queue lengths on the two approaches

Furthermore, the results in terms of percentage reduction in average vehicle queue length and average vehicle delay are displayed in Tables 1 and 2. The left-most column shows the 14 scenarios, with the degree of saturation gradually decrease from 0.9 to 0.6, as shown in the second left-most column. Also in the second left-most column, the notations "b" and "ub", respectively, refer to the case with equal volume on both approaches and the case with 80/20 volume split on the two approaches. Overall, from Tables 1 and 2, it is evident that the proposed control scheme, in addition to allowing a higher probability of clearing traffic in the intersection, also reduces both the average queue length and vehicle delay, as compared with the Webster optimal timing plan and the timing plan with a PCR of 0.9.

Table 1 – Compared with Webster's timing plan

simulation runs	Degree of saturation	percentage reduction	
		queue length	vehicle delay
1	0.9b	5.5	7.9
2	0.9ub	1.1	1.7
3	0.85b	5.9	8.9
4	0.85ub	1.8	2.7
5	0.8b	3.5	5.5
6	0.8ub	1.8	3.2
7	0.75b	5.1	8.1
8	0.75ub	4.8	6.7
9	0.7b	4.7	7.3
10	0.7ub	5.4	9.6
11	0.65b	3.0	4.8
12	0.65ub	3.1	5.5
13	0.6b	7.5	13.6
14	0.6ub	2.4	7.0

Table 2 – Compared with a fixed timing plan with PCR=0.9

simulation runs	Degree of saturation	percentage reduction	
		queue length	vehicle delay
1	0.9b	3.7	5.5
2	0.9ub	10.2	13.3
3	0.85b	4.2	6.4
4	0.85ub	6.6	9.7
5	0.8b	4.8	7.8
6	0.8ub	2.8	3.6
7	0.75b	3.5	5.6
8	0.75ub	3.8	5.1
9	0.7b	3.7	5.9
10	0.7ub	5.0	8.1
11	0.65b	11.6	17.2
12	0.65ub	3.8	7.4
13	0.6b	1.9	3.6
14	0.6ub	2.0	4.8

Overall, the proposed control scheme can achieve a reduction of about 5% in average queue length, and a reduction of about 10% for average vehicle delay. Although the overall delay is not an objective to be explicitly minimized – the approach merely aims to bring the queue states back to the permissible range, queue delay is reduced substantially.

CONCLUDING REMARKS

This paper developed a traffic signal control scheme for a single signalized intersection. By combining the concepts of queue-based signal switching and phase clearance probability, the proposed scheme has the advantages of allowing for a higher reliability of clearing traffic in the intersection, as well as more robust performance for overflow traffic. Meanwhile, the average queue length and overall delay of the intersection are also reduced substantially.

In actual implementation, the cycle time may be determined by other factors, i.e., signal coordination and capacity considerations. With this cycle length, a timing plan designed based on the prevailing traffic condition should be able to handle the arrivals in the long run. However, in the short run, due to the stochastic nature of arrivals, vehicle arrivals on one approach may spike up suddenly. If no action is taken, the residual queue there will take a long time to get cleared. In this proposed approach, while maintaining the cycle time to be unchanged, we swap the green times between different approaches based on their current queuing situation. The goal is to fully utilize the green times of the intersection as a whole, and avoid the situation that one approach has unused green time, whereas the other is short. Under this condition, the application of the proposed control scheme can substantially improve the performance of the intersection.

The present proposed control scheme is developed for a single intersection. Our current research is on how to extend the proposed scheme for coordinated intersections. We will need to extend the definition of phase clearance reliability (PCR) to a form that can cater for coordinated intersections. We hope to be able to report new findings of this extension in the future.

ACKNOWLEDGEMENT

This study is supported by the Competitive Earmarked Research Grants from the Research Grants Council of the Hong Kong Special Administrative Region (Fund number: HKUST 6283/04E and 617007).

REFERENCE:

Allsop, R.E. (1971). Delay-minimizing Settings for Fixed-time Traffic Signals at a Single Road Junction. IMA Journal of Applied Mathematics, 8(2), 164-185.

Chow, A.H.F. and Lo, H.K. (2007). Sensitivity analysis of signal control with physical queuing: Delay derivatives and an application. Transportation Research Part B: Methodological, 41(4), 462-477.

Haight, F.A. (1959). Overflow At a Traffic Light. Biometrika, 46(3-4), 420-424.

Lefeber, E. and Rooda, J.E. (2006). Controller design for switched linear systems with setups. Physica A: Statistical Mechanics and its Applications, 363(1), 48-61.

Lin, W., Lo, H.K. and Xiao, L. (2009). Robust Quasi-Dynamic Traffic Signal Control, 2009, .

Liu, H., Balke, K.N. and Lin, W., (2008). A reverse causal-effect modeling approach for signal control of an oversaturated intersection. Transportation Research Part C: Emerging Technologies, 16(6), 742-754.

Lo, H.K. (2006). A reliability framework for traffic signal control., IEEE Transactions on Intelligent Transportation Systems, 7; 7(2), 250-260.

Lo, H.K. (2001). A Cell-Based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plans. Transportation Science, 35(2), 148-164.

Lo, H.K. (1999). A novel traffic signal control formulation. Transportation Research Part A: Policy and Practice, 33(6), 433-448.

Lo, H.K. and Chow, A.H.F., (2004). Control Strategies for Oversaturated Traffic. Journal of Transportation Engineering, 130(4), 466-478.

Miller, A.J. (1963). Settings for Fixed-Cycle Traffic Signals. OR, 14(4), 373-386.

Murat, Y.S. and Gedizlioglu, E. (2005). A fuzzy logic multi-phased signal control model for isolated junctions. Transportation Research Part C: Emerging Technologies, 13(1), 19-36.

Newell, G.F. (1989). Theory of Highway Traffic Signals. University of California, Berkeley, CA.: Institute of Transportation Studies, Course Notes.

Webster, F.V. (1958), Traffic Signal Settings. Road Research Technical Paper, 39.