On Real-Time Traffic-Signal Control

Meera Sampath
Raja Sengupta
Stephane Lafortune

University of Michigan
Ann Arbor

The University of Michigan
IVHS
Intelligent Vehicle-Highway Systems
College of Engineering • (313) 764-4332
Professor Kan Chen, 4112 EECS, Ann Arbor, MI 48109-2122, and
University of Michigan Transportation Research Institute • (313) 936-1066
Robert D. Ervin, 2901 Baxter Road, Ann Arbor, MI 48109-2150

IVHS Technical Report: #92-10

On Real-Time Traffic-Signal Control

Meera Sampath
Raja Sengupta
Stephane Lafortune

University of Michigan
Ann Arbor

IVHS Technical Report: #92-10

On Real-Time Traffic-Signal Control*

Meera Sampath, Raja Sengupta and Stéphane Lafortune
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109-2122

Abstract

Signal-control strategy has a major impact on the performance of traffic networks. We discuss in this report the problem of optimal traffic-signal control. The report is organised in two parts. Part I is an extension of the IVHS Technical Report 91-01. The dynamic model presented therein for optimal traffic assignment is extended to include optimal signal control. Part II of the report focusses on the problem of optimal control of an isolated intersection for the special case of time-invariant arrivals. We propose several algorithms for real-time control of a single intersection and compare these algorithms through interesting examples and counterexamples.

Keywords

intelligent vehicle-highway systems, traffic signal control,
state space model, dynamic programming

*Research supported in part by the IVHS Program at the University of Michigan and by the National Science Foundation under grant ECS-9057967.
July 22, 1992

PART I. Extensions to the dynamical system model to incorporate signalization

1. Introduction

A major goal of the Intelligent Vehicle Highway Systems (IVHS) is integration of Advanced Traveller Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS). IVHS envisions an optimal traffic-network system incorporating dynamic traffic assignment and real-time signal control, both of which would be operating in unison to produce an overall system-optimal traffic network. Optimal dynamic traffic assignment, obviously, depends on the signal control policy, while on the other hand, efficient signal control plans demand knowledge of the traffic flow pattern. Hence, an integrated approach to the problem of dynamic route guidance and optimal traffic control is called for. This report is an extension of the IVHS Technical report 91-01 titled, "A Dynamical System Model for Traffic Assignment in Networks," by Lafortune et al., which incorporates signalization in addition to route guidance.

2. Summary of the dynamical system model for traffic assignment

Report 91-01 focusses on the problem of dynamic traffic assignment in networks with multiple trip origins and destinations. The traffic network is modelled as a discrete time dynamical system. The system state consists of the sizes of all platoons, each of which represents all vehicles on a certain link with the same destination and the same earliest possible time of departing the link. A general form for the state transition function is provided giving the possible states at time $n+1$ as a function of the state at time n and of the feasible assignment or routing decisions for platoons that exit a link or join the network in the time interval $(n, n+1]$. Link travel times are modelled by either impedance functions or link outflow rate constraints. We assume that the network is subject to known time-varying demands from vehicles for travel between its origins and destinations during a finite time horizon. The objective is to assign the vehicles to links over time in order to minimize the total travel time experiencd by all vehicles using the network, resulting in a system-optimal assignment. Dynamic programming is used to solve the given optimization problem.

This model avoids complete microscopic detail by grouping vehicles in platoons irrespective of origin node and time of entry into the network, yet keeps track of a good level of information in terms of location of vehicles, at any given time, on the network.

One extension of the dynamical-system model, as presented in the report, is the use of stalling inputs. We make use of this input as the control input for signalization. (A red phase of the traffic signal is equivalent to applying a stalling input to the vehicles on that link.) The extended model considered here, thus, has two types of control inputs: one for routing and the other for signalization. We also present here a way to model phasing arrangement at any intersection by the use of binary variables.

3. Modelling of phasing arrangement at a signalized intersection

Consider a typical signalized intersection. We assign to each turn (including straight flows) a binary variable τ_i^j. These variables are broken up into disjoint sets where each set represents a phase. In addition to these sets representing distinct phases, we have another set of variables representing turns (or phases) that are always green.

The following rules are used to model phasing at an intersection:

1. Variables belonging to a particular set are made equal to each other.
2. Sum of variables, one from each set, is equal to one.
3. Variables corresponding to the always green phases are set to one.

Consider for example, the intersection depicted in Fig. 1. $\tau_l^{l'}$ = binary variable representing red or green state of the signal controlling flow from link l to link l'. A value of one indicates green and a value of zero, red. Following the rules outlined above, the phasing arrangement at this intersection can be modelled by the following set of equations:

$$\tau_5^4 = \tau_3^6 \tag{1}$$

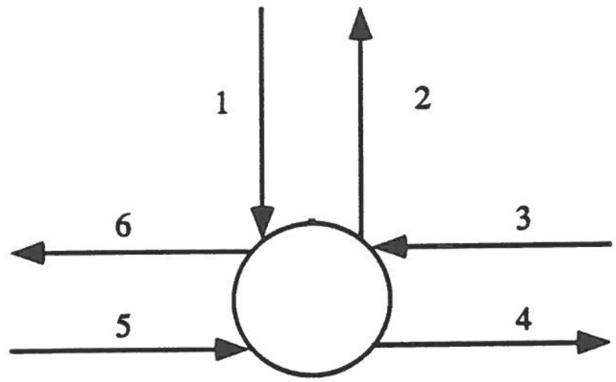

Figure 1: Three Phase Intersection

$$\tau_3^2 = \tau_1^6 = 1 \tag{2}$$

$$\tau_5^2 + \tau_1^4 + \tau_5^4 = 1 \ . \tag{3}$$

Following a similar procedure, any arbitrary phasing arrangement at any traffic intersection could be modelled.

4. Control inputs for signalization - Stalling inputs

In addition to the existing variables in the original dynamical system, we now define two new ones $v_l^{dl'}$ and $\tau_l^{l'}$. The following are the relevant variables to the signalization problem.

- u_s^{dl} - Stalling input (traffic going to destination d stalled on link l)
- $v_l^{dl'}$ - Traffic on link l which has been routed to link l' and whose destination is d
- u^{dl} - Traffic that actually gets routed into l, which is bound for destination d
- y^{dl} - Link output
- $\tau_l^{l'}$ - Binary variable reflecting green or red phase for traffic flowing from link l to l'.

Note that whereas, in the route-guidance problem, it was not necessary to keep track of the origin of the vehicles, the signal control problem requires knowledge of both the origin and destination of the vehicles at the intersections.

The example considered in the previous section has the following stalling inputs u_s^{d1},u_s^{d3} and u_s^{d5} for each destination d.

The following set of equations relates the above mentioned variables.

$$u_s^{dl} = \sum_{l'} (1 - \tau_l^{l'}) v_l^{dl'} \tag{4}$$

$$y^{dl} = \sum_{l'} v_l^{dl'} \tag{5}$$

$$u^{dl} = \sum_{l'} v_l^{dl'} - \sum_{l'} u_s^{dl'} \tag{6}$$

The last equation simply means that the number of vehicles that actually flow into l is the difference in number between those routed by the dynamic router and those that have been stalled on link l'.

It is evident from the above equations that the stalling inputs u_s^{dl} may be eliminated leaving equations for u and y in terms of v and τ. Thus, incorporation of signalization does not greatly increase the number of variables in the original model.

5. Modelling of the hypothetical link

We now introduce the hypothetical link - an artificial modelling device used to maintain a modular separation between the routing and the signalization parts. We associate with each turn, at an intersection, a hypothetical link. Each hypothetical link is fed by only one real link. The special behaviour of the hypothetical link models the queueing effects. The hypothetical links affect the capacity of the real links feeding them; i.e., the number of vehicles on a hypothetical link is accounted for in the link feeding it and when the capacity of a link feeding a hypothetical link is calculated, we add the number of vehicles on the hypothetical link also.

Introduction of the hypothetical link achieves the following objective. The effect of signalization is captured totally by the hypothetical link and the signal has no effect on the state equations of the real links. Vehicles leave the real links at times assigned by the impedance function and the stalling is essentially on the hypothetical links.

$v_l^{dl'}$, as defined in the previous section forms the input to the hypothetical links. The outflow rate of a hypothetical link is determined by its capacity. Now, on the hypothetical link, vehicles order themselves in a queue of packets, each of size equal to the outflow rate. At every instant of the green phase, one (and only one) packet (if present) leaves the link. If the number of vehicles coming out of a real link is not equal to the outflow rate of the hypothetical link, they merge into packets of size equal to the outflow rate.

This merging of vehicles does not give rise to any problem in the single destination case. But in the case of multiple destinations, when vehicles bound for two different destinations compete for the same packet, some sort of priority scheme is necessary to resolve conflicts. In order to take care of this problem, we split vehicles according to the ratio of the number of vehicles bound for each destination and, based on this ratio, decide the number of vehicles of each destination in a packet.

With these additional modifications, the dynamic equations of the hypothetical link are developed on the same lines as those of the real links. It is to be noted that the dynamic equations for the hypothetical links are different only in the case where the associated τ variable is zero (which implies stalling). Otherwise, when $\tau = 1$, the equations for the hypothetical link are the same as those of the real links. (Equations omitted for brevity.)

6. Conclusion

We have presented here an integrated approach to the problem of route guidance and signal control. The dynamical system model, originally developed for the traffic assignment problem, has been extended to incorporate signalization. The generality of the original model allowed easy incorporation of optimal signal planning with little modification. Areas of future investigation outlined in report 91-01, namely (1) study of the computational limitations of this approach and (2) investigation of the real time feasibility of this approach, remain open for investigation. To the best of our knowledge, this is the first report on an integrated approach to the problem of dynamic route guidance and optimal signal control.

PART II. Real-time control of an isolated intersection in the case of time-invariant arrivals

1. Introduction

We consider the problem of modelling and optimally controlling signalized intersections. In contrast to Part I of the report, we deal with the problem of signal control in isolation, i.e., assuming that the routing is fixed. In addition, we consider the case of constant, steady flow of traffic at a single isolated intersection. Detailed analysis of this special case would allow us to get an understanding of the fundamental issues involved in signal control, which in turn would enable us to extend the results to the case of time-varying flows and networks of intersections.

2. Modelling of an isolated signalized intersection

A platoon of vehicles leaving a signalized intersection will gradually spread as vehicles pass each other, or adjust to more cautious headways. Eventually, the platoons from successive signal cycles will spread enough as to overlap to form a steady traffic stream. If this traffic stream should now approach a second traffic signal, the behaviour of the traffic at the second signal should be independent of the timing of the first traffic signal. Hence we consider interesections at which the approaching traffic flow is uniform. This problem, while being interesting in its own right, also forms the basis for the case of time-varying flows.

When a traffic stream approaches a signalized intersection, a queue is formed during the red phase of the signal. This queue is cleared in part or fully during the subsequent green phase. The following equation models the signalized intersection as a dynamic system.

$$q_i(t+1) = q_i(t) + r_i(t) - (\min\{q_i(t) + r_i(t), k_i\}\tau_i(t)) \quad (1)$$

$$q_i(t_0) = q_i^0 \qquad i = 1, \ldots n.$$

This is the state equation for an n phase traffic intersection where

1. $q_i(t)$ - Queue (number of vehicles) in the i^{th} phase at time t
2. $r_i(t)$ - Arrival rate in the i^{th} phase at time t
3. k_i - The outflow rate of the i^{th} phase (also termed as capacity/saturation flow)
4. $\tau_i(t)$ - Decision variable which takes on a value of 1 or a value of zero. A value of 1 indicates that the i^{th} phase has a green signal and the queue can be served. When $\tau_i(t)$ is zero, the signal is red and the i^{th} phase is not served.

Time invariant case: In the case of a constant, steady arrival pattern, the above equation simplifies to

$$q_i(t+1) = q_i(t) + r_i - (\min\{q_i(t) + r_i, k_i\}\tau_i(t)) \tag{2}$$

Note that when the i^{th} phase is served, a maximum of k_i vehicles leave the intersection (k_i being the capacity of the i^{th} phase). In this case, the queue might not be completely cleared if there are more than k_i vehicles in the queue. On the other hand, if the total number of vehicles in queue at time t (including those arriving at time t) is less than the capacity k_i, all vehicles get cleared when the signal turns green.

The ensuing discussion is based on the valid assumption that $r_i < k_i$, i.e., the arrival rate is less than the capacity of the link (undersaturated condition).

The optimal control problem is then to determine the switching sequence $\tau_i^*(t)$, $\;t = t_o, \ldots T_h$, where $\tau_i^*(t)$ is the optimal control policy. An optimal policy is one that (a) minimizes the queue length or (b) maximizes the throughput of the intersection over the optimization time horizon T_h.

Before we address the problem of determining an optimal policy, we take up the issue of stability of a signalized intersection.

3. Stabilizability of a signalized traffic intersection

Stability deals with the following issue: given the inflow rate r_i and the outflow rate k_i, can one ensure, through an appropriate choice of signal timing plan, that the queues in all the phases do not grow unbounded?

Definition 1: The system described by Equation (2) is said to be stabilizable if $\exists M \in R$, and $\tau_i(t), t = t_0, \ldots T_h$ s.t. $q_i(t) \leq M \quad \forall t \geq t_0, \quad \forall_i$. ■

Definition 2: Stabilizing periodic control (SPC)

A control policy $\tau_i(t)$ is called a stabilizing periodic control if it

1. bounds the queue in every phase, i.e., $q_i(t) \leq M \quad \forall t \geq \ t_0, \forall i$
2. is periodic with period T, i.e., $\tau_i(t) = \tau_i(t+T) \quad \forall t > t', \ t' \geq t_0$. ■

It follows from the above definition that a stabilizing periodic control ensures that the total arrivals over the period T are equal to the total number of vehicles cleared over the period. This implies that the system settles into a "steady-state" of inflow and outflow, which are constant over a period, i.e.,

$$T r_i = \sum_{k=t}^{t+T} \tau_i(t) c_i(t) \quad \forall i \quad \forall t \geq t'$$

where $c_i(t)$ is the number of vehicles cleared at time t in the i^{th} phase, which is equal to $(\min\{q_i(t)+r_i, k_i\})$. Our definition also allows for an initial transient period $[t_0, t']$ after which the steady state is attained.

Theorem 1: The system described by the dynamic equation (2) is stabilizable iff $\exists$ a stabilizing periodic control. In other words, equation (2) stabilizable $\Longleftrightarrow \exists \tau_i(t), \quad t', \quad T$ s.t.

$$\begin{aligned} \tau_i(t) &= \tau_i(t+T) \\ q_i(t) &= q_i(t+T) \quad \forall t \geq t', \forall i \end{aligned}$$

(Proof Omitted.) ■

The above theorem simply states that the system would be stable, i.e., the queues would remain bounded for a given set of arrival-headway values if and only if there is some periodic control policy that stabilizes the system.

It is well known that all existing traffic controllers are periodic. This theorem justifies the use of such periodic control policies.

Theorem 2: The System (2) is stabilizable iff $\exists T_i > 0, i = 1, \ldots n$ s.t.

$$\sum_{i=1}^{n} T_i = T \quad \text{and} \quad T r_i \leq T_i k_i \quad i = 1, \ldots n \,. \tag{3}$$

For a two-phase intersection, the above simplifies to

$$\begin{aligned} T_1 + T_2 &= T & (4) \\ (T_1 + T_2) r_1 &\leq T_1 k_1 & (5) \\ (T_1 + T_2) r_2 &\leq T_2 k_2 \,. & (6) \end{aligned}$$

(Proof Omitted) ■

In the above set of equations, T_1 stands for the number of time instants phase 1 is served , while T_2 stands for the number of time instants phase 2 is served. The period of the policy T is equal to $T_1 + T_2$. $(T_1 + T_2)r_i$ is the total number of arrivals over a period T, while $T_i k_i$ is the maximum number of vehicles that could be cleared from the i^{th} phase if it is served for T_i time instants. Henceforth, we shall restrict the discussion to a two-phase isolated intersection.

Note that the above theorem simply means that if the conditions of the theorem are satisfied, the total number of arrivals over a period is less than or equal to the total number of vehicles that could be cleared over a period. This would ensure boundedness of the queues.

Corollary: System (2) is stabilizable iff

$$\frac{r_2}{k_2 - r_2} \leq \frac{k_1 - r_1}{r_1} \,. \tag{7}$$

■

This corollary follows directly from the above theorem. This is an important corollary in the sense that given a set of inflow and outflow rates for any two phase intersection, one can immediately conclude whether it is possible to develop a signal-control policy for the system that would assure that the queues do not grow unbounded. Alternately, given the arrival pattern at an intersection, the designer can immediately decide upon the minimum required capacity of the road links at the intersection.

Graphically, the above theorem can be represented as follows: Note that any set of $\{T_1, T_2\}$ in

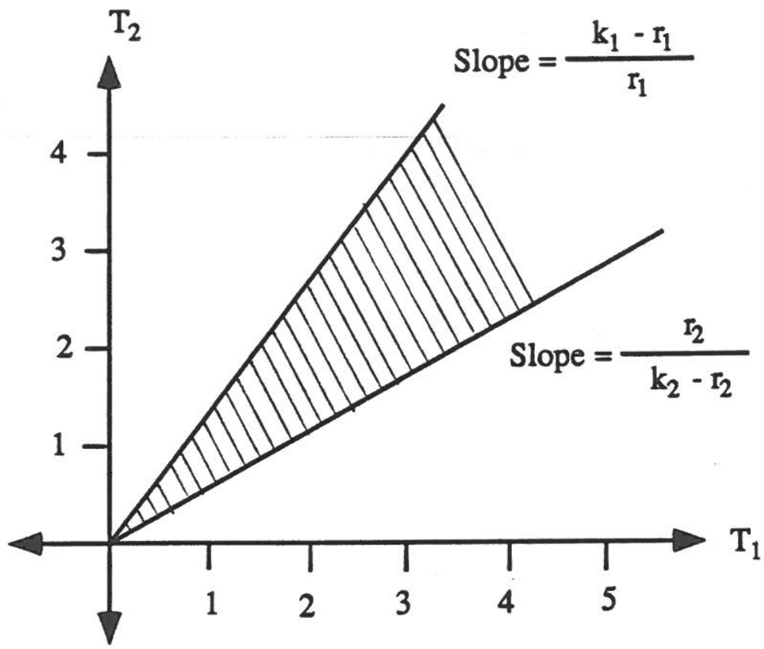

Figure 2: Graphical Representation of Theorem 2

the shaded region would lead to an SPC.

4. Optimal signal timing plans

The notion of stabilizability was introduced in the previous section. We shall now take up the problem of optimal control of an isolated intersection. In view of what was discussed in the previous section, we shall restrict our attention to periodic policies only. In other words, we try to find out an optimal policy from among the class of periodic policies.

The performance index that we shall consider is the following: $J = \frac{1}{T_h} \sum_{t=t_0}^{t_0+T_h} (q_1(t) + q_2(t))$ where T_h is the optimization horizon.

For the class of periodic policies, this reduces to

$$J = \frac{1}{T} \sum_{t=0}^{T} (q_1(t) + q_2(t)) \tag{8}$$

i.e., our objective is to minimize the average queue length over a period.

We shall now look into several candidate optimal control policies and compare these various policies through examples.

The following four policies are examined:

1. Serve the longest queue policy
2. Maximum instantaneous throughput policy
3. Bang-bang policy
4. Mixed-phase policy

4.1. Serve the longest queue policy

At any time t, this policy serves that phase which has the longest queue or the maximum number of vehicles waiting (this includes the vehicles that arrive at time t).

Example 1:

Let the arrival-headway values for a single isolated two-phase intersection be as follows:

Arrival in phase 1 $\quad r_1 = 4$

Arrival in phase 2 $\quad r_2 = 5$

Capacity of phase 1 (link 1) $\quad k_1 = 6$

Capacity of phase 2 (link 2) $\quad k_2 = 15.$

$\tau_1(t)$ and $\tau_2(t)$ are 1-0 variables for phase 1 and phase 2 respectively. These reflect the signal-control policy. $\tau_i(t) = 1$ implies that phase i is green at time $[t, t+1)$.

q_1 and q_2 stand for the queue lengths in phase 1 and phase 2, respectively. Define the variables

$$\begin{aligned} Q_1(t) &= q_1(t) + r_1 \\ Q_2(t) &= q_2(t) + r_2 \end{aligned}$$

i.e., $Q_i(t)$ = sum of queues at time $t+$ the arrivals at time t. The optimal policy decision is based on $Q_i(t)$. If at any time $t, Q_1(t) > Q_2(t)$, phase 1 is served, and vice-versa. This leads to the following table:

t	r_1	r_2	q_1	q_2	Q_1	Q_2
0	0	1	0	0	4	5
1	1	0	4	0	8	5
2	0	1	2	5	6	10
3	1	0	6	0	10	5
4	0	1	4	5	8	10
5	1	0	8	0	12	5
6	1	0	6	5	10	10
7	0	1	4	10	8	15
8	1	0	8	0	12	5
9	1	0	6	5	10	10
10	0	1	4	10	8	15
11	1	0	8	0	12	5
12	1	0	6	5	10	10

We see that the above control policy leads to a "steady-state" operation of the system.

It is interesting to note that, in the case of "serve the longest queue policy," though we do not restrict ourselves to the class of periodic policies, the policy does indeed turn out to be periodic.

For the above example, the period $T = 3$ with $T_1 = 2$ and $T_2 = 1$. Note that this $\{T_1, T_2\}$ satisfies conditions (5) and (6):

$$\begin{aligned}(T_1 + T_2)r_1 &= (2+1)4 \leq T_1 k_1 = 2(6)\\ (T_1 + T_2)r_2 &= (2+1)5 \leq T_2 k_2 = 1(15)\end{aligned}$$

The average delay over a period is $J = 33/3$.

4.2. Maximum instantaneous throughput policy

This policy considers not only the queues at any instant of time t, but also the capacity of each phase and serves that phase which would lead to a maximum number of vehicles being cleared from the system at every time instant t.

Contrary to what one might expect, examples reveal that this policy, far from being the optimal one, can actually lead to an unstable system. This is illustrated in the following example.

Example 2: The arrival and headway values are the same as those in the previous example.

t	r_1	r_2	q_1	q_2	Q_1	Q_2
0	0	1	0	0	4	5
1	1	0	4	0	8	5
2	0	1	2	5	6	10
3	1	0	6	0	10	5
4	0	1	4	5	8	10
5	1	0	8	0	12	5
6	1	0	6	5	10	10
7	0	1	10	0	14	5
8	1	0	8	5	12	10
9	1	0	12	0	16	5
10	0	1	10	5	14	10
11	1	0	14	0	18	5

Examination of the queue pattern in phase 1 reveals that this system is unstable.

Going back to the set of conditions (5) and (6), we see that

$$\begin{aligned}(T_1 + T_2)4 &\leq 6T_1 \\ (T_1 + T_2)5 &\leq 15T_2 .\end{aligned}$$

The smallest set of $\{T_1, T_2\}$ that satisfies these inequalities is $\{2, 1\}$; i.e., the minimum number of time instants that phase 1 has to be served per period is 2. The maximum throughput policy, on the other hand, serves phase 1 only once per period, thus leading to instability.

We conclude that the maximum throughput policy leads to instability due to its "greedy" nature of always trying to send the maximum number of vehicles out of the system, which results in insufficient service to certain phases.

4.3. Bang-bang policy

The bang-bang control policy serves one phase continuously, allowing the queue in the other to build up onto a certain maximum limit. It then switches over to the other phase, letting the queue in the former phase build up again to a maximum. (Hence, the term "*bang-bang*".) The maximum limit in a phase is determined based on the maximum number of vehicles that could be cleared by serving that phase a sufficient number of times subsequently.

For the two phase system we let the queues build up as follows:

Phase 1 - Maximum queue length $T_2 r_1$.

Phase 2 - Maximum queue length $T_1 r_2$.

Assume that the queue in phase 1 is allowed to build up to T_2r_1 (this would mean serving phase 2 for T_2 time instants). If the next T_1 time instants are devoted to phase 1, the total arrivals over the period $T = T_1 + T_2$ will be $(T_1 + T_2)r_1$. If condition (5) is satisfied, i.e., $(T_1 + T_2)r_1 < T_1k_1$, then by dispatching up to k_1 vehicles every time instant for the next T_1 instants, the queue in phase 1 is guaranteed to be brought to zero.

On the other hand, the queue in phase 2 would build up to T_1r_2 while phase 1 is being served. If, at this point, we switch over to phase 2 and continue serving it for the next T_2 instants, we are again guaranteed to have complete clearance of phase 2 queues at the end of T_2 instants, if the condition $(T_1 + T_2)r_2 < T_2k_2$ is satisfied. Thus, a stable system ensues.

Example 3: Consider the same example as before.

$$\begin{aligned} r_1 &= 4 \qquad k_1 = 6 \\ r_2 &= 5 \qquad k_2 = 15 \end{aligned}$$

$$\begin{aligned} (T_1 + T_2)4 &\leq 6T_1 \\ (T_1 + T_2)5 &\leq 15T_2 \end{aligned}$$

The smallest $\{T_1, T_2\}$ pair satisfying above is $\{2, 1\}$. Hence, set $T_1 = 2, T_2 = 1$. Then $T = T_1 + T_2 = 3$, $T_2r_1 = 4$ and $T_1r_2 = 10$. This leads to the following table:

t	r_1	r_2	q_1	q_2
0	0	1	0	0
1	1	0	4	0
2	1	0	2	5
3	0	1	0	10
4	1	0	4	0
5	1	0	2	5
6	0	1	0	10
7	1	0	4	0

The average delay $= \frac{21}{3} = 7$. This beats the "serve the longest queue" policy, which had an average delay of 11 units.

We now discuss the choice of T_1, T_2 values. Inspection of figure 2 reveals that the choice of T_1, T_2 values satisfying conditions (5) and (6) is not unique. Any set of $\{T_1, T_2\}$ lying within the shaded sector would lead to a stable system. We conjecture that the smallest T_1 and the smallest T_2 satisfying conditions (5) and (6) are the optimal values among the entire set of $\{T_1, T_2\}$ values.

Before illustrating this with an example, we shall state the following theorem:

Theorem 3: Let $S_1 = \{T_1^1, T_1^2, \ldots T_1^n, \ldots\}$ and $S_2 = \{T_2^1, T_2^2, T_2^3, \ldots T_2^n, \ldots\}$ be indexed sets of T_1 and T_2, values such that the pair $\{T_1^n, T_2^n\}$ satisfies (5) and (6) for all n. Then the pair $\{T_1^*, T_2^*\}$ also satisfies (5) and (6), where

$$T_i^* = \min\{T_i : T_i \in S_i\} \quad i = 1, 2 .$$

■

Example 4: Consider the following parameter values:

$$\begin{aligned} r_1 &= 2 \qquad k_1 = 10 \\ r_2 &= 3 \qquad k_2 = 5 \end{aligned}$$

Applying conditions (5) and (6) to the above set of values, we get $T_1 = 1, T_2 = 2$ as the smallest possible values of T_1 and T_2. Applying the bang-bang policy with the above set of values, we obtain

$$\begin{aligned} T_2 r_1 &= 4 \\ T_1 r_2 &= 3 \end{aligned}$$

t	r_1	r_2	q_1	q_2
0	1	0	0	0
1	0	1	0	3
2	0	1	2	1
3	1	0	4	0
4	0	1	0	3

with average cost $J = \frac{10}{3}$. Now consider another set of valid $\{T_1, T_2\}$ values, $T_1 = 2, T_2 = 5$. These values lead to

$$\begin{aligned} T_2 r_1 &= 10 \\ T_1 r_2 &= 6 \end{aligned}$$

t	r_1	r_2	q_1	q_2
0	1	0	0	0
1	1	0	0	3
2	0	1	0	6
3	0	1	2	4
4	0	1	4	2
5	0	1	6	0
6	0	1	8	0
7	1	0	10	0
8	1	0	2	0
9	0	1	0	6

with average cost $J = \frac{44}{7}$.

The above example reveals that choosing the smallest set of T_1, T_2 values leads to a lower cost. The use of the minimum T_1, T_2 values ensures that the switchover from phase A to phase B takes place as soon as the queue in phase A reduces to zero, thus minimizing the extent to which the queue in phase B is allowed to build up. In other words, the bang-bang policy with minimun T_1, T_2 leads to a more effective use of the capacity of the system.

A point worth noting at this stage is that we have not considered switching costs in our analysis. If switching cost constraints were included, the minimum T_1, T_2 values might not be the optimum set of values.

4.4. Mixed-phase policy (Split-phase policy)

Assuming that the minimum values of T_1 and T_2 are chosen, we now investigate whether the bang-bang policy is optimal.

Example 5: Let

$$\begin{aligned} r_1 &= 4 \qquad k_1 = 11 \\ r_2 &= 5 \qquad k_2 = 9 \\ T_1 &= 2 \qquad T_2 = 3 \end{aligned}$$

Consider, first, the bang-bang policy:

t	τ_1	τ_2	q_1	q_2
0	1	0	0	0
1	1	0	0	5
2	0	1	0	10
3	0	1	4	6
4	0	1	8	2
5	1	0	12	0
6	1	0	5	5
7	0	1	0	10
8	0	1	4	6
9	0	1	8	2

The average cost per period is $J = \frac{52}{5}$.

For the same set of $\{T_1, T_2\}$ values, now consider a different control policy wherein every period does not have a red phase followed by a green phase but rather a mixed phase; i.e., each period consists of more than one red phase or green phase.

In the example considered below, phase 1 has a control policy $0-1-0-0-1$, i.e., red-green-red-red-green every period of 5 time instants.

t	r_1	r_2	q_1	q_2
0	0	1	0	0
1	1	0	4	0
2	0	1	0	5
3	0	1	4	1
4	1	0	8	0
5	0	1	1	5
6	1	0	5	1
7	0	1	0	6
8	0	1	4	2
9	1	0	8	0
10	0	1	1	5
11	1	0	5	1

Here, the average cost per period is $J = \frac{32}{5}$. This policy beats the bang-bang policy for the given set of inflow, capacity values!

The above example shows that for the same splits (% of each period alloted to phase) the interleaving of service-time instants affects optimality. It is well known that conventional traffic control does not use this idea of a mixed phase; each phase extends continuously for a certain number of time instants followed by change of phase, which again lasts for a set of consecutive time intervals, this pattern being repeated periodically. The above example shows that such a "continuous-phase" approach might not be optimal always in the case of (time-invariant) steady flow, when there are no switching costs.

The problem now is to determine the "optimal interleaving." We introduce at this point a procedure for determining the interleaving and it is our conjecture that this leads to the optimal policy.

Procedure for determining the optimal interleaving:

1) Given a set of r_i and k_i values, obtain the minimum T_1, T_2 values that satisfy conditions (5) and (6).

2) Find $R = \left\lfloor \frac{\max\{T_1,T_2\}}{\min\{T_1,T_2\}} \right\rfloor$

Assume, without loss of generality that $T_2 > T_1$, then, $R = \left\lfloor \frac{T_2}{T_1} \right\rfloor$

The policy that we propose is as follows:

Serve phase 2 R times; switch; serve phase 1 once; switch; serve phase 2 R times; switch and so on. After T_1 such switches, phase 1 is served the required number of T_1 times while phase 2 is served RT_1 times. Now, serve phase 2 the additional number of times required to ensure that it is served T_2 time instants in all (note that T_2/T_1 is not generally an integer).

The idea behind the above policy is that it ensures lesser build-up of queues in either phase and leads to a more "uniform" flow of traffic out of the intersection.

Example 6: Let $T_1 = 3$ and $T_2 = 7$. Then $R = \lfloor T_2/T_1 \rfloor = \lfloor 7/3 \rfloor = 2$. The optimal switching sequence is:

$$
\begin{array}{llll}
(1,0) & (0,1) & (0,1) & \\
(1,0) & (0,1) & (0,1) & \\
(1,0) & (0,1) & (0,1) & (0,1)
\end{array}
$$

Examination of example 5 reveals that the choice of interleaving that leads to a better policy than the bang-bang policy does indeed arise out of the above procedure.

To summarize, the solution strategy to determine the conjectured optimal control policy is as follows:

1. Given arrival rates r_i and capacity values k_i, check if condition (7) holds. If not, system is not stabilizable.
2. If (7) is satisfied, find minimum values of T_1 and T_2 satisfying constraints (5) and (6).
3. Determine the interleaving based on the procedure outlined above.

We see that this solution procedure involves *no search* at all, which is indeed the major advantage of this procedure. Unlike most existing signal-optimization schemes, which involve the use of search techniques, (OPAC, etc.) we have a closed-form solution based on the conjecture outlined above.

5. Conclusion

In this report, we have discussed in detail the problem of optimal control of signalized intersections. A dynamic model for traffic intersections was presented and the idea of stabilizability was introduced. Several different control policies were examined for optimality. The emphasis was primarily on developing a control strategy for signalized intersections that involves little or no search.

In this report, the focus was primarily on a single isolated intersection with a steady flow of traffic. The analysis presented here could be extended to handle the following situations:

1. Multiphase intersections where we have more than two phases and a network of intersections, as opposed to a single isolated intersection.

2. Time-varying flows: The case of time-varying arrival patterns could be dealt considering average traffic flows over a period and implementing "rolling horizon type" strategies for optimal control.

3. Additional constraints on minimum and maximum allowed green times could be imposed to make the optimization problem more meaningful and practical. Since switching involves additional expenses one might include switching constraints too.

It remains to validate the conjecture on the optimal control policy proposed herein for a single intersection. Further analysis is needed to see if it is possible to arrive at such closed-form solutions for more complicated situations involving the extensions listed above.