ABSTRACT

This paper presents an overview of a new real-time traffic signal timing policy, named RT/IMPOST, that is designed to provide optimal signal timing over the entire range of traffic environments from undersaturated to oversaturated flow conditions. This paper describes the control policy design for the oversaturated traffic environment. For this traffic environment, a policy objective that seeks to maximize throughput (vehicle trips serviced) is suggested. To meet this objective, it is argued that managing queues to avoid intersection blockage and to fully exploit the available green time, constitutes a sound basis for such a control policy. A description of the analytical approach is presented and its implementation procedures in real time are detailed.

Simulation results are presented which support this policy design, demonstrating that this kind of control can provide good performance even under severe congested conditions.

KEY WORDS: Real-time traffic control, optimization, traffic congestion
1. INTRODUCTION

It is well known that fluid flow can transition rapidly from a laminar state to a turbulent one when some disturbance introduces instability into the flow process. In fluid flow analysis, a dimensionless parameter, the Reynolds Number, is an indicator of the division between these flow states. Turbulent flows are characterized by rapid random fluctuations in velocity. Analysis procedures for turbulent flow differ from those applied for laminar flow.

The flow of traffic in a signalized urban setting displays characteristics that are analogous to those of fluid flow. Undersaturated flow is stable: queues that may form on approaches to signalized intersections are generally serviced within one signal cycle and do not propagate onto other approaches. A transition into oversaturated flow, characterized by the unstable growth of queues over time can be rapid, and is associated with a dimensionless parameter known as the volume:capacity (v/c) ratio.

As is the case with the analysis of fluid flows, the analysis and control of traffic in an oversaturated flow environment differs from that for undersaturated flow. The RT/IMPOST policy therefore employs two distinct approaches depending on the flow regime:

- For undersaturated conditions, the policy is designed to minimize vehicle delay and stops and to deter queue spillback from short approaches. Sufficient capacity is provided by the signal control to service demand. The flow process is described by a macro-simulation model; this representation is used with an enumerative, iterative search procedure and a network decomposition algorithm, to optimize signal coordination [1].
- For oversaturated conditions, the policy is designed to maximize throughput by preempting queue growth to avoid system-wide instability, while assuring that all available capacity is fully utilized.

This paper describes the RT/IMPOST policy for oversaturated conditions only. While parts of this formulation have been integrated with the policy developed for undersaturated conditions for application in “mixed” environments (undersaturated networks with a limited number of oversaturated intersections), this policy for oversaturated environments is designed expressly for real-time application on systems with adequate surveillance and computer resources.

The over-saturated, signalized traffic environment is characterized by an excess of demand relative to capacity (v/c>1), with unstable queues growing over time causing intersection spillback and blockage, and the consequent spread of congestion over the network.

Historically, the formulation of traffic control policies was based on the assumption of an undersaturated traffic environment and focused on improving the efficiency of traffic operations by minimizing vehicle stops and/or delays. Most recent policies for congestion control tend to increase signal cycle length in the belief that reducing “signal switching losses” will materially increase capacity. However, a longer signal cycle generally increases delay and causes much longer queues which can block intersection and exacerbate congested conditions, while providing minimal increases in theoretical capacity.

The real-time RT/IMPOST (Real-Time Internal Metering Policy to Optimize Signal Timing) control policy was developed to expressly address these problems. This is an extension of the policy of the
IMPOST model developed under the U.S. National Cooperative Highway Research Program (NCHRP) in Project 3-38(3). The control objectives are: (1) maximize system throughput; (2) fully utilize intersection capacity; (3) preempt queue spillback that physically blocks intersections; and (4) provide equitable service. The policy consists mainly of Mixed-Integer Linear Programming (MILP) and Non-Linear Programming (NLP) formulations [2], [3].

A MILP tableau is formulated to compute the optimal signal offsets and the associated optimal queue lengths on all approaches along a two-way arterial system. This formulation depends on the policy maintaining an input-output flow balance on each saturated approach by adjusting signal phase durations every cycle. During every signal cycle, a NLP procedure adjusts the phase durations to stabilize the varying queue lengths so as to maximize system productivity. As long as the queue length is controlled to approximate its optimal value computed by the MILP, the occurrence of both “starvation” (which fails to deliver vehicles to the stop-bar in a timely manner and wastes green time) and spillback (which physically blocks intersections, thereby wasting green time) will be avoided, and throughput would be maximized.

2. DESIGN CONCEPTS

An over-saturated traffic environment displays unstable queues that grow over time. In an urban network of closely-spaced signalized intersections, these “overflow” queues can extend throughout the length of approaches and into the upstream intersections (“spillback”). The result is an intersection that is physically blocked denying access to cross street flow and delaying the discharge of traffic from the feeder approaches. Avoiding intersection blockage that reduces capacity is a control objective that specifically addresses the over-saturated traffic environment—such an objective is virtually irrelevant in an under-saturated traffic environment.

Consistent with the objective of avoiding intersection blockage, is the objective, maximize throughput. Clearly in a capacity-constrained environment, making full use of the available capacity is a primary objective.

To satisfy these objectives, it is necessary to identify bounds on both queue lengths and on the signal offsets. Consider an approach defined by the link, (B, A) as shown:

\[
\text{B} \quad \rightarrow \quad \text{A}
\]

The control policy determines the minimum offset, \(\Delta_{\text{min}}\), that can just avoid starvation and the maximum offset, \(\Delta_{\text{max}}\), that can just avoid allowing the queue to spill-back into intersection, B. It can be shown that these relations are as follows:

\[
\begin{align*}
[\Delta]_{\text{max}} &= \frac{L}{v_d} \left[1 - r(1 + \frac{v_L}{w})\right] + \min \left[\frac{G_d - s}{1 - p_B}, \frac{L - W}{W}\right] \left[1 - \frac{w}{u}\right] \\
[\Delta]_{\text{min}} &= \frac{L}{v_L} \left[1 - \frac{rhv_L}{L_v}\right]
\end{align*}
\]

where
\([\Delta]_{\text{max}}, [\Delta]_{\text{min}}\) = [Maximum, Minimum] value of relative signal offset: the elapsed time between the beginning of the green phase of intersection, A and that at intersection, B in seconds.

\(G_A\) = Duration of green phase servicing approach at A, sec.

\(h\) = Mean queue discharge heading, sec./veh.

\(L\) = Approach length, ft.

\(L_v\) = Mean length of vehicles, ft.

\(p_a\) = Left + Right turn percentages of traffic turning from the feeder approach

\(r\) = Queue length: Approach length, ratio

\(s\) = Start-up lost time for queue discharge, sec

\(u\) = Stopping wave speed, fps

\(v_L\) = Speed of lead vehicle in platoon, fps

\(v_d\) = Mean speed of vehicles in platoon, fps

\(W\) = Width of intersection at B, ft.

\(w\) = Discharge wave speed, fps

See Figure 1 for a graphical representation of these parameters.

Given the need to “meter” the flow of vehicles into every approach in order to control their queue lengths, it is necessary to establish acceptable bounds on queue length ratio. The minimum queue length ratio, \([r_o]_{\text{min}}\), is usually formed by the turn-in flow, if any, at the upstream intersection, B, and by the need to provide a sufficient supply of vehicles at the stop-bar to avoid starvation; the maximum ratio, \([r_o]_{\text{max}}\), is limited by the approach length and influenced by the desire to fully utilize the available green phase duration. The resulting relations are:

\([r_o]_{\text{min}} = \max \left[ \frac{L_v}{L} N_c P_e \left( \frac{L}{L} \right)_c, 1.1 \frac{2L_v}{L} \right] \]

\([r_o]_{\text{max}} = \min \left[ 1 - \frac{W + F}{L}, 1.1 \frac{(G_A - s) L_v}{hL} \right] \]

where

\(F\) = Safety factor to guard against spill-back, ft.

\(G_A\) = Green phase duration at the downstream intersection, sec.

\(h\) = Mean queue discharge headway, sec./veh.

\(L\) = Approach length, ft., stop-bar to stop-bar

\(L_v\) = Mean vehicle space occupied in a standing queue, m/veh.

\([r_o]\) = Queue length ratio: queue length/approach length

\(s\) = Start-up lost time, sec.

\(W\) = Width of upstream intersection, ft.
These bounds on signal offsets and on queue length ratios form a “Solution Space” that defines the desired conditions for control, as shown in Figure 2:

- Any point within the shaded area is an optimal solution for that link.
- It is seen that there is an intrinsic “coupling” between signal offset (coordination) and queue length. Of course, the size and shape of this solution space varies from approach to approach and is also dependent on other signal control parameters: cycle and green phase durations. On this basis, it was possible to develop analytical models in the form of mathematical programming tableaus, discussed below.

3. DESCRIPTION OF THE POLICY
The policy consists of the following components:

1. An algorithm to estimate queue length on all detectorized approaches. The algorithm relies on data provided by at least one detector per lane, located approximately 250-500 feet upstream of the stop-bar, as well as information describing the state of the signal control at the downstream intersection.
2. A Mixed-Integer Linear Programming (MILP) tableau that provides a network-wide solution expressed in terms of signal control parameters (offsets, phase durations) and queue length. The objective function is formulated to maximize throughput, the constraint relations bound the physical lengths of queues to ensure that (a) there is always an adequate supply of general vehicles at the stop-bar at the start of the green phase to fully utilize the service capacity; and (b) there will be no queues “spilling-back” into the upstream intersection.

3. A Non-Linear Programming (NLP) tableau that adjusts phase durations every signal cycle to control queue lengths on every approach throughout the network.

4. IMPLEMENTATION OF REAL-TIME CONTROL
The control paradigm is illustrated in Figure 3. As shown, there are two levels of control, corresponding to two different update intervals:

- Surveillance data is acquired from detector actuations and organized to provide updates every signal cycle.
- The queue estimation algorithm accesses these detector data and estimates the queue lengths at the beginning of green, for all approaches, every cycle.
- The NLP model accesses these queue estimates and computes the optimal signal phase durations.
- An algorithm determines every cycle whether there is a need to update signal timing based on detector data. If so (and at least 8 cycles have elapsed since the last MILP update), the MILP model is activated to compute updated optimal signal offsets, using current phase durations and volumes.
- A signal transition algorithm implements these offset adjustments rapidly and with minimum disruption of traffic operations.

These control activities continue over time, yielding signal timing patterns that change continually while the traffic environment is over-saturated.
5. TESTING RT/IMPOST

The RT/IMPOST policy was integrated with KLD’s WATSim microscopic simulation model to simulate the performance of the policy in a real-time environment. This was an invaluable tool that was used to detect flaws in the RT/IMPOST design or software, and then to compare its performance with other models.

Comparison with other models must take into account that RT/IMPOST is a real-time traffic-responsive control policy, while the other models provide pre-timed (i.e., “fixed-time”) signal timing plans. As a result, these comparisons identify the benefits of real-time control and of RT/IMPOST, vs. the other models cited.

Table 1 tabulates the results obtained for a test arterial network where the traffic environment was originally under-saturated, became over-saturated due to rising demand, then returned to under-saturation as demand fell. As indicated, RT/IMPOST provided impressive reductions in delay relative to the other models.

Figure 4 contains two computer animation frames generated by WATSim comparing RT/IMPOST with SYNCHRO. Notice the spill-back condition in (b) and the exceedingly long queues on the cross street approaches reflecting the effects of intermittent intersection blockage along the artery. In contrast, Figure 4(a) shows that all intersections are clear and that queues on the cross streets are minimal.

Figure 5 illustrates the outcome of another simulation study [4]. Here, results produced by RT/IMPOST, when integrated in an operating real-time signal control system named, COSMOS, are compared with COSMOS results operating without RT/IMPOST. As shown in Figure 5.1, COSMOS with RT/IMPOST provides about 25% higher throughput (vehicle trips serviced) for the southbound and cross street approaches along an arterial section in Seoul, South Korea, than does COSMOS, alone. Results for the northbound approach are comparable.

The comparison of delay per vehicle (seconds) is more dramatic. With RT/IMPOST, COSMOS produces about 25% less delay on the cross streets, 10% less delay in the northbound direction and 75% less delay in the southbound direction, than was the case for COSMOS without RT/IMPOST.

6. CONCLUSION

RT/IMPOST includes a real-time congestion-control signal timing policy that incorporates new principles designed expressly for over-saturated traffic environments. Simulation testing has yielded positive comparisons with other policies. We seek to test the policy in the field to confirm its value.

7. REFERENCES


Table 1. Model Statistics for Test Artery at the End of Two Hours (WATSim simulation results)

<table>
<thead>
<tr>
<th>Policy</th>
<th>Delay (hrs.)</th>
<th>Speed (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT/IMPOST</td>
<td>388</td>
<td>11.0</td>
</tr>
<tr>
<td>PASSER</td>
<td>570 (47%)</td>
<td>8.2 (-25%)</td>
</tr>
<tr>
<td>TRANSYT</td>
<td>596 (54%)</td>
<td>7.9 (-28%)</td>
</tr>
<tr>
<td>SYNCHRO</td>
<td>668 (72%)</td>
<td>7.2 (-35%)</td>
</tr>
</tbody>
</table>

Figures in () are percent differences: (Policy – RT/IMPOST)/(RT/IMPOST)
Figure 1. Congested Traffic Environment
Figure 2. Solution Space \([\Delta, r]\) for a Representative Link
Figure 3. Overview of Control Policy Implementation
Figure 4. Comparison of Queue Formation: (a) RT/IMPOST; (b) SYNCHRO
Figure 5-1. Comparison of Throughputs with/without RT/IMPOST

Figure 5-2. Comparison of Per Vehicle Delay with/without RT/IMPOST