Abridgment

Prescription for Demand-Responsive Urban Traffic Control

NATHAN H. GARTNER

State-of-the-art traffic control strategies in urban networks are calculated off-line, stored in a computer's memory, and selected for implementation on the street by various real-time criteria such as time-of-day, level of congestion, or special events. Many research efforts have been directed toward the development of new strategies that would relieve the traffic engineer of the constant burden of data collection and strategy revision and, at the same time, provide an improved level of service. This paper reviews results from past studies and analyzes their implications with respect to the development of improved generations of urban traffic control strategies. It then proposes a prescription for demand-responsive control that has the potential for overcoming many of the deficiencies of past efforts and leading to a significant improvement in urban traffic performance.

The large variety of control hardware and strategy software now available to the traffic engineer and system designer is changing continuously (1). The last decade has seen the introduction of computer-based traffic control systems in ever-increasing numbers. Several hundred such systems have already been installed and many more are under development throughout the world.

Strategies are commonly calculated off-line by arterial or network optimization techniques and are then stored in the computer's memory for implementation by various on-line criteria. Attempts have also been made to develop strategies that are calculated on-line in response to prevailing traffic conditions. The goal has been to improve traffic performance through adaptive control as well as to relieve the traffic engineer from the constant burden of data collection and strategy revision. These attempts have met with mixed success.

The emergence of new microprocessor technologies has given new impetus and new opportunities for the development of such strategies. The purpose of this paper is to assess the current status of strategy development and to offer a prescription for future developments.

STRATEGY DEVELOPMENT AND TESTING

Foremost among the computer-control strategies conducted during the past 15 years is the Urban Traffic Control System (UTCS) research project, which was conducted by the U.S. Department of Transportation (2).

The project was directed toward the development and testing of a variety of network control concepts and strategies, divided into three generations of control, as shown in Table 1. The different generations can briefly be characterized as follows.

First-Generation Control

First-generation control (1-GC) uses prestored signal timing plans that are calculated off-line based on historical traffic data. The plan that controls the traffic system can be selected on the basis of time-of-day (TOD), by direct operator selection or by matching from the existing library a plan best suited to recently measured traffic conditions (TRSP). The matching criterion is based on a network threshold value composed of volumes and occupancies. Frequency of update is 15 min. One-GC software also includes logic to enable a smooth transition between different signal-timing plans, a critical intersection control (CIC) feature that enables vehicle-actuated adjustment of green splits at selected signals, and a bus priority system (BPS) at specially instrumented intersections. Plans can be calculated by an off-line signal optimization method; traffic network study tool (TRANSYT)-generated plans were selected for testing in UTCS.

Second-Generation Control

Second-generation control (2-GC) is an on-line strategy that computes and implements in real-time signal timing plans based on surveillance data and predicted volumes. The optimization process [an on-line version of the traffic signal optimization program (SIGOP)] is repeated at 5-min intervals; however, to avoid transition disturbances, new timing plans cannot be implemented more often than every 10 min.

Third-Generation Control

The third-generation control (3-GC) strategy was conceived to implement and evaluate a fully responsive, on-line traffic control system. Similar to 2-GC, it computes control plans to minimize a network-wide objective by using predicted traffic conditions for input. The differences compared with 2-GC are that the period after which timing plans are revised is shorter (3-5 min), and cycle length is required (a priori) to vary among the signals as well as at the same signal during the control period (CP).

Analysis of the dynamics of control plan generation and implementation for the three UTCS strategies is important. In 1-GC the traffic pattern (volume and occupancy) during interval n-1 is used to make a decision whether a new plan should be called from the library for interval n. No prediction is used. Two-GC and 3-GC are similar in concept. Detector measurements are accumulated up to, and including, interval n-1. These data are used during interval n to predict volumes and speeds and to generate the timing plans that are then implemented in interval n+1. In both strategies the traffic data used in the timing plan that is being implemented are displaced by at least one interval from the corresponding measured flows. The different UTCS control strategies were designed to provide an increasing degree of traffic responsiveness through a reduction of the update interval, with a view to improving urban street network performance. However, results of extensive field testing showed that the expectations were not entirely fulfilled (2,3).

One-GC, in its various modes of operation, performed overall best and demonstrated that it can provide some measurable reductions in total travel time over that which could be attained with a well-timed three-dial system. Two-GC had a mixed bag, but was overall inferior compared with 1-GC. Three-GC, in the form tested in the UTCS system, seriously degraded traffic flow under almost all the conditions for which it was evaluated. A summary of the results is given in Table 2 (3). Similar results were experienced in the Glasgow (4) and Toronto (5-7) experiments.

DISCUSSION OF RESULTS

From the studies cited above one may erroneously

Table 1. Characteristics of UTCS strategies.

Feature	First Generation	Second Generation	Third Generation
Update interval, control period	15 min	5-10 min	3-5 min, variable
Control plan generation	Off-line optimization, selection from library by time-of-day, traffic responsive, or manual mode	On-line optimization	On-line optimization
Traffic prediction	None	Historically based	Smoothed values
Critical intersection control	Fine tuning, splits	Fine tuning, splits and offsets	NA
Cycle length	Fixed within each section	Fixed within groups of intersections	Variable in time and space, predetermined for control period

Table 2. Comparison of results of UTCS strategies.

Generation	Traffic Responsive Strategy	Change in Aggregate Vehicle Minutes of Travel with Respect to Base (%)			
		Morning Peak	Off Peak	Evening Peak	All Day Avg
First	Arterial	-2.6	-4.0	-12.2	-5.7
	Network	-3.2	+1.9	-1.6	-1.3
Second	Arterial	-1.3	-3.8	+0.5	-2.1
	Network	+4.4	+1.9	+10.7	+5.2
Third	Arterial	+9.2	+24.0	+21.2	+16.9
	Network	+14.1	-0.5	+7.0	+8.2

conclude that a library of timing plans generated off-line, based on historical data (from another month, perhaps another year but for the same time period of the day), is more effective than timing plans generated on-line, based on very recent data (past 15, 5, or 3 min). However, a closer examination of those studies indicates that the expectations of the researchers were not fulfilled--not because their rationale was wrong (that traffic-responsive control should provide benefits over fixed-time control) but because of a failure of the models and procedures that they implemented to deliver the desired results. A major cause for this failure appears to be in the measurement-prediction cycle used by the procedures.

Most available traffic control methods claim to be traffic responsive in some sense. Even 1-GC strategies are traffic responsive to a certain extent--plans may be replaced at 15-min intervals in response to predicted traffic volume changes. But, these methods are not truly responsive: They do not respond to actual traffic conditions but to hypothetical conditions--the hypothesis is only as good as the model and the predictions used in the optimization. This is the most critical aspect in all the responsive strategies listed above. The traffic-flow process and the optimization procedure form an inseparable closed-loop control system. The control values can only be effective if an accurate model is used in the optimization. However, all the different generation-control strategies do not have an accurate model; they use an abstract model that is calibrated by predicted (thus inherently inaccurate) smoothed volume data. Such a model cannot take account of short-term fluctuations; in essence, by aggregating and smoothing the data, the information content that is most important for on-line demand-responsive control is destroyed.

Large discrepancies were observed (sometimes in excess of 50 percent) when comparing the performance of 2-GC and 3-GC predictors with actual volumes over successive 5-min intervals (8). When aggregated over shorter than 5-min periods, the discrepancies can be even larger. Moreover, suppose one could predict the volume in each cycle with complete accuracy (i.e., with a zero mean error value). Even then the resulting real-time control strategy might be ineffective. For example, the following numbers represent vehicle arrivals for two cycles, grouped into 5-s intervals, on a signal-controlled approach with a 60-s cycle time:

1 1 2 1 1 2 0 2 0 0 1 1

0 1 0 0 1 1 2 1 2 1 1 2

During both cycles the flow is the same (12 vehicles), yet the optimal control strategy for each should be entirely different because of the different distribution of the arrivals within the cycle.

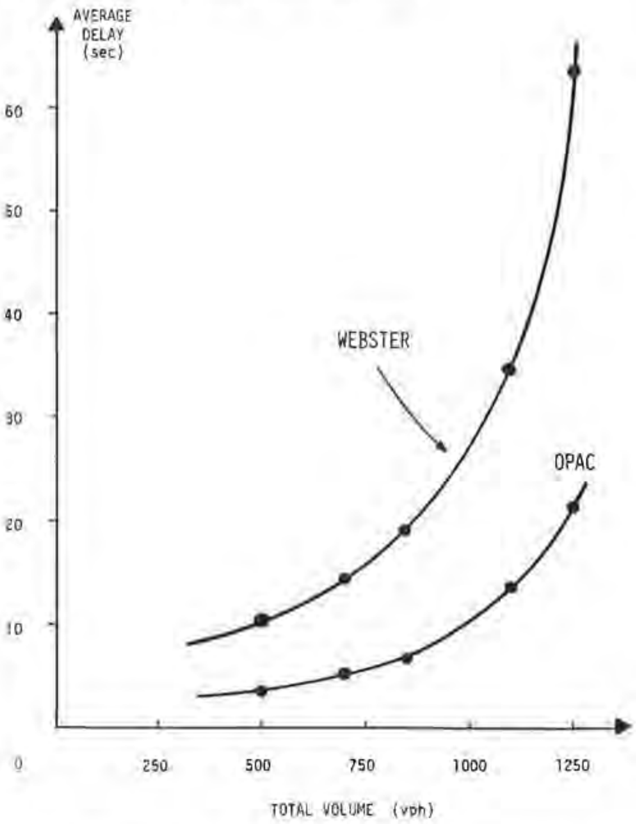
Clearly, an effective demand-responsive traffic control system requires the development of new concepts and not merely the extension of existing concepts toward shorter time frames (i.e., going down from hourly intervals to 15-, 5-, or 3-min intervals, or a cycle time) and using predicted values that are less and less reliable. Data from detectors provide information about past traffic behavior, but a traffic-responsive system must make decisions that result in good control in the future. Ways must be devised to predict future traffic behavior from past detector measurements.

PRESCRIPTION

On-line traffic control strategies should be capable of providing results that are better than those produced by the off-line methods. Simulation studies have indicated that, if under ideal conditions complete information on vehicle arrivals was available, responsive control strategies could reduce as much as 50 percent of the delay incurred by using existing nonresponsive strategies (9,10). To achieve this goal, the following requirements for the development of an effective demand-responsive traffic control system are proposed.

1. The system shall provide better performance than off-line methods. This is the primary criterion, everything else is secondary. Although it may seem self-evident, it was not always explicitly recognized in the development of responsive strategies. In some cases it was superseded by less relevant criteria such as mainstreet platoon progression or variable cycle time.
2. Development of new concepts is needed and not merely the extension of existing concepts. As demonstrated by the experiments reviewed in this paper, effective responsiveness is not achieved by implementing off-line methods at an increased frequency. New methods have to be developed.
3. The system must be truly demand-responsive; i.e., adapt to actual traffic conditions and not to historical or predicted values that may be far off from the actual.
4. It should not be arbitrarily restricted to control periods of any length but should be capable of updating plans at any time, at any location.

Figure 1. Comparison of average delay per vehicle.



5. The system should not be encumbered by a network model structure that requires extensive centralized computer capability. The model should be decentralized in its decisionmaking and use only those data that are directly pertinent to the decisions it has to reach. Decentralization increases the overall computation power, simplifies the data requirements and processing, and enhances the effectiveness of the control strategies that are generated.

6. The system should obviate the conventional notions of offset, split, and cycle time, which are inherent in all existing signal-optimization methods. The pattern of any individual signal should consist of a continuously varying, demand-responsive, sequence of on (effective green) and off (effective red) times that are only subjected to appropriate lower and upper bounds.

Can such a system be realized? The likelihood of its development is greatly enhanced by the continuous improvements in microprocessor technologies. By combining the potential capability of the microprocessor with demand-responsive strategies such as those proposed by Miller (11) or those used in SCOOT (12), SCAT (13), or the optimization policies for adaptive control (OPAC) programs (14), it is to be expected that substantial advances in the state of the art can be achieved.

An example of the potential benefits that can be expected from truly demand-responsive strategies is shown in Figure 1 (14). It compares the average delay at a two-phase signal-controlled intersection when timings are determined by Webster's method and by an OPAC strategy. OPAC is a demand-responsive strategy that dynamically optimizes signal timings. It uses a rolling horizon concept based on a combination of measured and calculated arrival patterns and can be implemented by a microprocessor. This strategy can provide, under ideal conditions, up to 60 percent reduction in delay with respect to a fixed-time strategy. Although such performance may be hard to expect in real life, this result is an indication of the tremendous opportunities that microprocessor-based demand-responsive strategies can offer. Undoubtedly, much more research and experimentation would be needed to take advantage of these opportunities.

ACKNOWLEDGMENT

This paper is based, in part, on a research project conducted under sponsorship of the Office of University Research, U.S. Department of Transportation. The opinions expressed in the paper, however, are mine and not necessarily those of the sponsoring agency.

Discussion

K. Todd

The UTCS evaluation study (3) reported a statistically significant reduction in delay through TRSP versus three-dial, averaging 3.9 percent throughout the test area during the evening peak period. On the strength of this result, the author, as others have done before him, considers TRSP more effective than a well-timed three-dial system.

The listed improvement was primarily brought about through delay reductions for outbound traffic on sections 1 and 2. For example, on outbound Wisconsin Avenue (route 11, section 2), delay was reduced by 15.3 percent (15). Without such improvements, TRSP would have given little or no advantage over three-dial, possibly a degradation.

Note that the evaluation study omitted to measure whether this reduction in delay might not have been accompanied by longer delays at critical downstream intersections beyond the test area. Was this 15 percent cut in delay merely a faster way of getting to the nearest bottleneck, where cars had to wait that much longer? The situation is similar to that often encountered after construction of an overpass: Congestion is transferred to another location.

To be complete, an evaluation study must include losses caused by transitions between timing plans, transitions between test sections operating on different timing plans, and transitions between the test area and outlying areas. In the absence of information on whether delay reductions within the UTCS test area might not have been accompanied by longer delays elsewhere, claims that UTCS and other computer-based control systems can bring an improvement over a well-timed three-dial system should be treated with reserve, just as evaluation studies should be treated with reserve if they apply results obtained from only a certain percentage of the intersections within the system to the entire system.

Can real-time control dispense with historical data? The computer not only has to receive correct information on vehicle arrivals, it must also be able to determine the subsequent action that will bring the desired result; e.g., the least delay throughout the entire system.

Assume the most simple example: A single side street vehicle arrives at a red signal on an arterial. Without additional information as to future side street arrivals, the least delay would be produced by giving this side street vehicle green immediately or as soon as a convenient gap is detected on the arterial. If historical data showed a side street arrival rate of, say, 60 vehicles/h, a different decision would have to be made in order to hold the side street red until the interruption of main street traffic produces no greater delay than the cumulative side street delay. (The example takes no account of stops or stop penalties.) If additional information on side street arrival distribution and main street volume fluctuations could be known, a different strategy would have to be devised; however, this information cannot exist, nor can the effect of each control decision on the remainder of the street system be foreseen. The example can be expanded to the far more complex situations found in systems that comprise more than one side street approach.

Real-time control can measure a variety of parameters, but it cannot, without historical data, predict the number of turning movements; nor can a computer know the delay that turning traffic will produce, because the delay will depend on the turning vehicle's position in the platoon, the gap distribution in opposing traffic, and the presence of pedestrians walking with the green.

From these and more complex examples it may be concluded that real-time control has to be enhanced by historical data, that historical data can produce only a coarse prediction, and that factors needed for truly effective real-time control are not available and cannot be predicted correctly, nor can the effect of the computer's control decisions on subsequent traffic movement be assessed. An attempt to predict future traffic behavior from past detector measurements means trying to predict the unpredictable. Many mathematicians are confident it can be done.

Author's Closure

Todd addresses two issues--one concerns the validity of the UTCS evaluation study and the other is on the use of historical data in real-time control.

Regarding the UTCS evaluation study, I believe that it merits much wider analysis and discussion than has appeared so far in the literature or than I have done in my brief comments. But the issues raised by Todd are not germane to the subject matter of my paper. They should be directed to those who have conducted the evaluation study. Whether Todd's hypotheses are true or false would have no effect on the paper's analysis. In any case, I cannot support them in lieu of scientific evidence.

Concerning the second issue, the use of historical data, I would like to point out that my paper is in the nature of a review and analysis. It does not present any methodological details. Those are described elsewhere (12) and will be included in forthcoming papers. Therefore, any discussion on this topic is merely an expression of opinion.

Historical data have an important role in real-time control, but not in the way they were used in the 2-GC or 3-GC strategies that were implemented in UTCS. A great deal of useful information about future traffic behavior can be derived from detector measurements and effectively used in controlling traffic.

REFERENCES

1. JHK and Associates. An Approach for Selecting Traffic Control Systems. NCHRP, Manual, March 1977.
2. J. MacGowan and I.J. Fullerton. Development and Testing of Advanced Control Strategies in the Urban Traffic Control System. Public Roads, Vol. 43, Nos. 2, 3, and 4, 1979-1980.
3. R.D. Henry, R.A. Ferlis, and J.L. Kay. Evaluation of UTCS Control Strategies--Executive Summary. FHWA, Rept. FHWA-RD-76-149, Aug. 1976.
4. J. Holroyd and D.I. Robertson. Strategies for Area Traffic Control Systems Present and Future. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR569, 1973.
5. Improved Operation of Urban Transportation Systems. Corporation of Metropolitan Toronto, Toronto, Vol. 1, March 1974.
6. Improved Operation of Urban Transportation Systems. Corporation of Metropolitan Toronto, Toronto, Vol. 2, Nov. 1975.
7. Improved Operation of Urban Transportation Systems. Corporation of Metropolitan Toronto, Toronto, Vol. 3, Nov. 1976.
8. W.R. McShane, E.B. Lieberman, and R. Goldblatt. Developing a Predictor for Highly Responsive System-Based Control. TRB, Transportation Research Record 596, 1976, pp. 1-2.
9. D.I. Robertson. Traffic Models and Optimum Strategies: A Review. Proc., International Symposium on Traffic Control Systems, Berkeley, CA, Aug. 1979, pp. 262-288.
10. P.J. Tarnoff. Concepts and Strategies--Urban Street Systems. Proc., International Symposium on Traffic Control Systems, Berkeley, CA, Aug. 1979, pp. 1-12.
11. A.J. Miller. A Computer Control System for Traffic Networks. Proc., 2nd International Symposium on the Theory of Traffic Flow, London, 1963.
12. SCOOT: A Traffic Responsive Method of Coordinating Signals. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Leaflet LF679 Dec. 1977 and Leaflet LF930 March 1981.
13. A.G. Sims. The Sydney Coordinated Adaptive Traffic System. Proc., Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems, ASCE, New York, 1979, pp. 12-27.
14. N.H. Gartner. Demand-Responsive Decentralized Urban Traffic Control. Office of University Research, U.S. Department of Transportation, Dec. 1980.
15. J.L. Kay, J.C. Allen, and J.M. Bruggeman. Evaluation of First Generation UTCS/BPS Control Strategy. FHWA, Executive Summary, Rept. FHWA-RD-75-26, 1975.

Publication of this paper sponsored by Task Force on Traffic Signal Systems.