

Development of Advanced Traffic Signal Control Strategies for Intelligent Transportation Systems: Multilevel Design

NATHAN H. GARTNER, CHRONIS STAMATIADIS, AND PHILIP J. TARNOFF

The development of advanced traffic signal control strategies suitable for advanced traffic management within Intelligent Transportation Systems is described. The strategies consist of a multilevel design for the real-time, traffic-adaptive control of an urban signal network. This design permits the system to be built up gradually to offer varying degrees of responsiveness, depending on particular network and traffic characteristics. The more advanced control levels in the hierarchy incorporate the capabilities of the lower control levels. A principal goal of the multilevel design is to invoke a selected control strategy when it can provide the greatest benefits and thus maximize the overall effectiveness of the system.

In November 1991 the FHWA issued a solicitation for the development and evaluation of a real-time, traffic-adaptive signal control system (RT-TRACS) suitable for use in an Intelligent Vehicle Highway Systems (IVHS) environment (1). The following information was provided as background:

The FHWA's IVHS program consists of research and operational tests designed to combat traffic congestion. The thrust of the program is to develop and implement the technology necessary to mitigate the effects of congestion by maximizing the utility of existing transportation facilities. Some of the elements included in IVHS are transportation management, in-vehicle route guidance systems, integration of multi-modal transportation, integration of surface street and freeway control, traveler information systems, and incident management. In order for these elements to be integrated, a sophisticated traffic surveillance system must be deployed to provide the information necessary to enable the real-time traffic management.

One of the key elements to these systems is a real-time, traffic adaptive signal control logic that enables the implementation of the traffic management and control strategies specified above. This control logic needs to not only assess the current status of the network, it must include forecasting capabilities such that proactive, not reactive, control is provided. To the extent possible, the signal control logic must be interfaced with freeway performance data and provide integrated, network-wide control.

Current technology in real-time control is somewhat limited. In the United States, for example, there are no true real-time systems. Elsewhere, at least two systems have been developed and deployed at several locations; however, the applicability of these foreign systems to IVHS is only now being tested. This study is designed to answer this and many other similar questions from a technical perspective and be the focal point for the development of the signal control logic needed to support IVHS. (1) (Note: the term IVHS has since been changed to ITS.)

In this paper a multilevel design for RT-TRACS is presented. Each level incorporates a different methodology and has a different set of characteristics. The more advanced levels incorporate in a

nested fashion the capabilities of the lower levels. This kind of design enables the system to be built up gradually to offer varying degrees of responsiveness, depending on particular network and traffic characteristics. This design provides the best combination of features for the overall optimization of traffic performance. Among the potential features that may be included in the RT-TRACS design are

- Both distributed and centralized traffic control;
- Traffic-responsive, on-line optimization, as well as background fixed-cycle control;
- Capability to interact with dynamic traffic assignment to implement proactive control;
- Dynamic priority control on selected routes;
- Congestion avoidance and congestion relief strategies; and
- Artificial intelligence technology to optimize strategy selection.

In addition, the following features are also included:

- Effective use of existing resources in a community,
- Coexistence of different control generations within one system,
- Improved fallback capabilities in case of surveillance system failure, and most important
- Effective use of the accumulated experience with real-time control.

The traffic engineering profession has more than 30 years of experience with computer control of traffic signals and the development and testing of various types of adaptive control strategies. It is important to make maximum use of this experience in the development of any new systems. In the next section, this experience is briefly reviewed.

REAL-TIME SIGNAL CONTROL: PAST EXPERIENCE

A thorough understanding of past experiences with advanced traffic signal control strategies is critical to the development of effective RT-TRACS strategies for ITS. Failure to do so may cause past mistakes to be repeated and the same pitfalls as encountered by past developers (2).

After the introduction of computer-based traffic signal control systems in the 1960s, numerous experiments were conducted to develop more advanced (i.e., more responsive) control strategies. One of the most comprehensive studies was the (Urban Traffic Con-

trol System (UTCS) experiment in the 1970s by the FHWA. The UTCS project was directed toward developing and testing a variety of advanced network control concepts and strategies and lasted for almost a decade. Its results defined the state of the art in the United States to the present. The UTCS experiment was described in detail in a work by MacGowan and Fullerton (3).

Research and testing of control strategies in the UTCS project was divided into three generations, as shown in Table 1. The same nomenclature is used in this paper. The different generations are characterized as follows.

First-Generation Control (1-GC)—This mode of control uses prestored signal timing plans that are calculated off-line based on historical traffic data. The plan controlling the traffic system can be selected on the basis of time of day, by direct operator selection, or by matching from the existing library a plan best suited to recently measured traffic conditions (volumes and occupancies). This is named the traffic-responsive (TRSP) mode of plan selection. The mode of plan selection is determined by the operator. Frequency of update in the traffic-responsive mode is 15 min. 1-GC software also includes logic to enable a smooth transition between different signal timing plans, and a critical intersection control (CIC) feature that enables vehicle-actuated adjustment of green splits at selected signals. NCHRP research (4) has cast doubt on the efficacy of the CIC algorithm and proposed to consider Optimization Policies for Adaptive Control (OPAC) driven controllers as substitutes. Plans in 1-GC can be calculated by any off-line signal optimization method, such as TRANSYT-generated plans, or progression optimization methods such as MAXBAND. The 1½-GC is a strategy in which new timing plans are generated automatically when traffic conditions warrant it.

Second-Generation Control (2-GC)—This is an on-line strategy that computes in real-time and implements signal timing plans based on surveillance data and predicted values. The optimization process can be repeated at 5-min intervals; however, to avoid transition disturbances, new timing plans cannot be implemented more often than once every 10 min. 2-GC software contains an optimization algorithm (SIGOP), a traffic prediction model, subnetwork configuration models, CIC, and a transition model to minimize transition time between two plans.

Third-Generation Control (3-GC)—This strategy was designed to implement and evaluate a fully responsive, on-line traffic control system. Similar to 2-GC, it computed control plans to minimize a networkwide objective using predicted traffic conditions for input. The differences compared to 2-GC were that the period after which timing plans were revised was shortened to 3 to 5 min, and that cycle length was allowed to vary among the signals as well as the same signal during the control period (CP). This was accomplished by dividing the CP into an integral number of intersection-specific control intervals (CI), which were calculated on the basis of predicted volume and capacity ratios on each approach to the intersection using a Webster-like method. Thus, control intervals were approximately equal to the expected value of cycle length at each intersection. However, the switching points within each CI were also determined by the on-line optimization procedure. In this method, a simplified model of traffic flow scenarios was used to trace performance, in an iterative manner, through strings of mini-star networks composed of the signalized node and all its incoming and outgoing links (the CYRANO model, 5).

The dynamics of the control plan generation and implementation for the three UTCS strategies are illustrated in Figure 1. It is noteworthy that 3-GC is similar in concept to 2-GC. In both strategies the traffic data used in the timing plan being implemented are displaced by at least two control periods from the actual flow measurements. Therefore, the effectiveness of the control system response depends entirely on the quality of the prediction model. The different UTCS control strategies were designed to provide an increasing degree of traffic responsiveness, with an expectation to capitalize on the variability in traffic flows and to provide an improvement in urban street network performance. However, results of extensive field testing demonstrated that these expectations were not entirely fulfilled (3).

1-GC, in its various modes of operation, performed best overall and demonstrated that it can provide measurable reductions in total travel time over that which could be attained with a well-timed three-dial system (see Table 2). The traffic-responsive mode of 1-GC plan selection was generally more effective than the time-of-day mode. The 2-GC strategy was mixed but was overall inferior compared with 1-GC; it demonstrated some small improvements on

TABLE 1 Characteristics of UTCS Control Strategies

FEATURE	1-GC	2-GC	3-GC
Update interval	15 min	5-10 min	3-5 min
(Control Period)			(Variable)
Control plan generation	Off-line optimization selection from library by time-of-day, traffic responsive, or manual mode (7 plans used)	On-line optimization	On-line optimization
Traffic prediction	None	Historically based	Smoothed values
CIC	Fine tuning of splits	Fine tuning of splits and offsets	NA
Cycle length	Fixed within each section	Fixed within variable groups of intersections	Variable in time and space Predetermined for control period

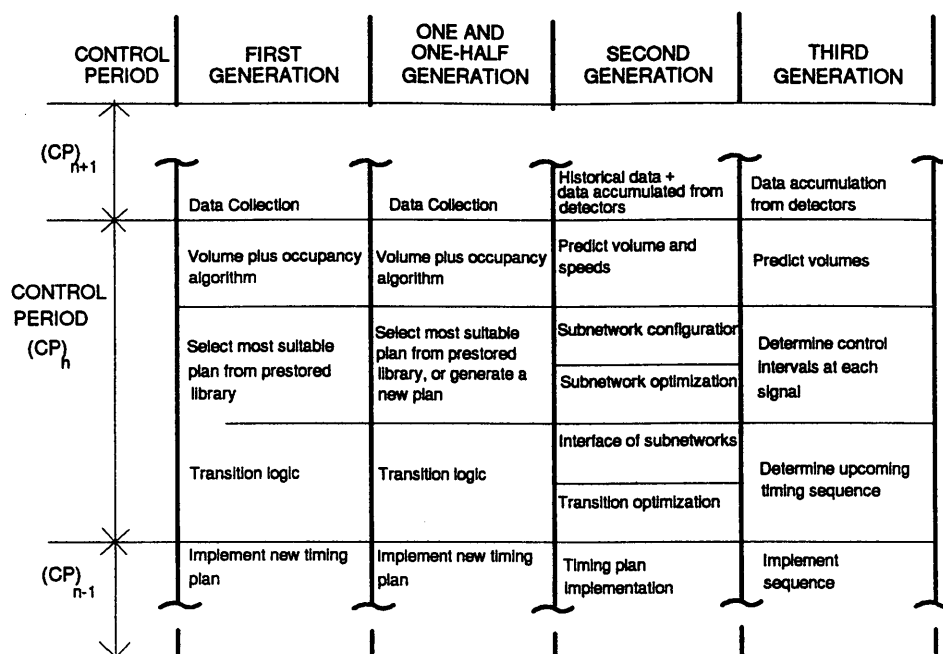


FIGURE 1 Dynamics of control plan generation and implementation in UTCS strategies.

the arterial but degraded traffic flow in the network. The 3-GC strategy, in the form tested in the UTCS system, was unsuccessful in responding to traffic flows and degraded performance under almost all conditions for which it was evaluated.

Thus, the more responsive strategies resulted in poorer performance than the fixed-cycle, nonresponsive strategies. This appeared to be counterintuitive. On the basis of these results, one might erroneously conclude that a library of timing plans generated off-line, based on historical data (from another day, 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 (the past 15, 5, or 3 min). However, a closer examination of the experiments reveals that the expectations 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 were used in the UTCS study to deliver the desired results.

Possible causes of the poor showing of the responsive UTCS strategies (6,7) follow.

- Because of the inherent inaccuracies in the measurement-prediction cycle, neither the 2-GC nor the 3-GC strategies could respond adequately to rapid changes in traffic flows.

- The frequent transition in signal timing may be more harmful than the adaptiveness sought by on-line optimization in 2-GC and in 3-GC. Considerable delays are incurred during the transition process.

- Although 3-GC had a variable-cycle feature, the entire signal switching sequence was predetermined for every control period and therefore not dynamically responsive to the actual traffic conditions on the street. In essence, it was a 2-GC strategy with an imposition of a different cycle length at each intersection. The benefits obtained from a locally optimal cycle length were insufficient to outweigh the loss of the benefits of synchronization among the intersections of the network.

- The control strategy used by 3-GC was centralized and the optimization procedure required a comparatively long time for convergence (much longer than 2-GC for the same size network). The

TABLE 2 Comparison of Results of UTCS Strategies

Traffic Responsive Strategy		% change in aggregate veh.-min. of travel with respect to base			
		AM Peak	Off Peak	PM Peak	All Day Average
1-GC	Arterial	-2.6	-4.0	-12.2	NA
	Network	-3.2	+1.9	-1.6	NA
2-GC	Arterial	-1.3	-3.8	+0.5	-2.1
	Network	+4.4	+1.9	+10.7	+5.2
3-GC	Arterial	+9.2	+24.0	+21	+16.9
	Network	+14.1	-0.5	+7.0	+8.2

time allotted for this procedure was insufficient for reaching a good optimum.

The failure of the more responsive strategies to accomplish their stated goals was not limited to the UTCS experiment. Similar results were experienced by the British Transportation Research Laboratory (8) and by Metropolitan Toronto Traffic Department (9). The basic premise has always been that on-line traffic control strategies should be capable of providing results that are better than those produced by the off-line methods. Because this premise was not accomplished in the experiments that were conducted during the 1970s, it was clear that new strategies had to be developed to be able to implement it successfully. To achieve this goal, the following prescription for the development of an effective demand-responsive traffic control system was offered in 1982 (10):

1. The system must be designed to provide better performance than off-line methods. Although this may appear self-evident, it was not always recognized explicitly in the development of responsive strategies in the past.
2. Development of new concepts is needed and not merely the extension of existing concepts. Effective responsiveness is not achieved by implementing off-line methods at an increased frequency. New methods that are better suited to the variability of traffic must be developed.
3. The system must be truly demand-responsive, that is, to adapt to actual traffic conditions and not to predicted values that may be

far off from the actual conditions. As a corollary to this principle, if the traffic conditions cannot be adequately sensed or predicted, then lower-generation strategies may be more advantageous than higher-generation strategies.

4. It should not be arbitrarily restricted to control periods of a specified length, but be capable of frequent updating of plans as necessary.

A significant advance toward these goals was achieved during the 1980s with the introduction of SCOOT in the United Kingdom (11), which may be considered a 2½-GC strategy, and by SCATS in Australia (12), which is considered a 1-GC/TRSP variant. These strategies are noteworthy, especially in their ability to generate timings on-line. However, a close examination of field test results reveals that the methods score both successes and failures compared with traditional fixed-time control. For example, floating car surveys that were conducted in the course of the evaluation of the SCOOT method produced results of the type shown in Figure 2 (11). It is evident that the advanced method does not always perform better than the base method. Sometimes its performance is inferior or indistinguishable.

These principles serve as a basis for the development of a new family of highly advanced adaptive control strategies: OPAC (13), PROLYN (14), and UTOPIA (15). After the OPAC lead, all strategies adopted a dynamic programming optimization methodology within a rolling horizon framework. This approach shows great promise as an element of the RT-TRACS design but has not yet

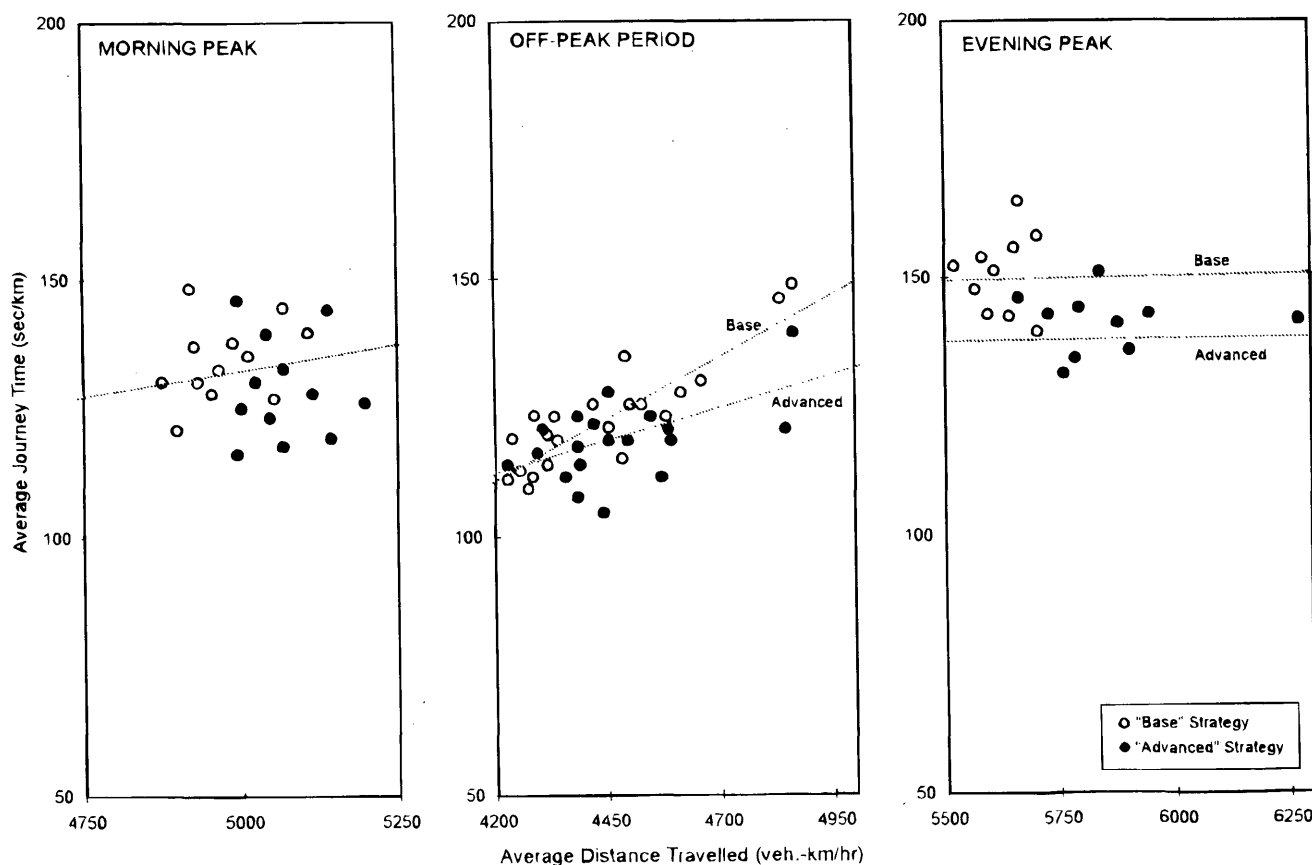


FIGURE 2 Examples of floating car surveys in evaluation of advanced strategies.

been fully developed and tested. Further discussion of this approach is given in the next section, which provides new framework, or architecture for the progressive development of the RT-TRACS strategies to optimize system performance.

ADVANCED CONTROL STRATEGIES: MULTILEVEL DESIGN

In the new framework for advanced traffic control, a multilevel design is offered in which each level in the hierarchy encompasses, in a nested fashion, the capabilities of the lower levels. The control levels correspond, in some respects, to the different UTCS control generations; however, they also contain significant differences and enhancements. Descriptions of the envisioned levels of control follow.

0-LC: Base Control—0-LC: Base Control is the most basic type of signal control and is a mix of fixed-time and traffic-actuated controllers and some arterial coordinated systems. The traffic engineer determines timing patterns manually, or by some PC-based programs, using available historical volume data. Various arterial progression schemes may be used. Manual adjustment of timings in the field is used because of inaccurate or inadequate data. Timings evolve by trial and error. There is infrequent updating of plans, typical of small-town operation or isolated sections of urban areas with limited technical support.

1-LC: First-Level Control—This level is similar to the UTCS/1-GC system. There is centralized control with limited availability of traffic surveillance and communications. Timing plans are calculated off-line using historical data and stored in the computer's data base. A variety of arterial and network optimization packages may be used, such as TRANSYT, PASSER II, MAXBAND, MULTIBAND, and so forth. The plans can be implemented by time of day, manual activation, special events, or traffic-responsive modes. SCATS is an advanced version of the latter mode. **1½-LC** is available as an option to enable automatic updating of timing plans for changing traffic and network conditions.

2-LC: Second Level Control—A basic 2-GC system follows the UTCS designation: a centralized control with on-line optimization was a fixed, common cycle time for dynamic subnetwork configurations. Typical timeframes for the optimization are between 5 and 10 min. Traffic volumes are predicted for the upcoming interval during which the new timing is to be implemented. Optimization is performed by an off-line strategy that was adapted for on-line control, such as SIGOP, RTOP, or PRINET (PRiority NETWORK optimization). It requires more advanced computational, surveillance, and communication capabilities than 1-LC. Advanced versions of 2-LC systems (perhaps designated as **2½-LC**) include SCOOT.

3-LC: Third Level Control. This is a fully adaptive traffic signal control system that incorporates the capabilities of all the previous levels, yet may relinquish their restrictive characteristics in favor of improved traffic performance; for example, a common cycle time is not required for coordination, although it is not necessarily excluded from the control parameter set. Optimization of phase sequence capabilities may be available at selected locations. It consists of on-line, dynamic optimization that can be performed centrally or and distributively, or both, through a network of smart controllers. Subnetworks can be configured dynamically to carry out optimal policies as required by the overall optimization objective.

An advanced surveillance system is required for this control level. It meets specifications of the original UTCS/3-GC.

An example of a strategy that meets the requirements of 3-LC is the OPAC strategy originally developed by Gartner (13). OPAC is based on dynamic programming (DP) optimization, which generates optimal signal switching sequences in real time. The OPAC strategy was initially implemented for individual, distributed intersection control and has exhibited good performance in field testing (16). It is now being extended for network operation under the RT-TRACS research program. A bilevel hierarchy is being used, where the upper level ensures the optimal coordinated operation of the individual smart controllers at the lower level. A schematic of the operation is shown in Figure 3. The primary signal is a smart controller that runs the OPAC algorithm and interacts with the neighboring satellite controllers to optimize local performance. Each signal is, in turn, a primary signal in its own mininetwork. In this way, the entire network becomes interconnected and coordinated.

4-LC: Fourth Level Control—This system level includes all the capabilities of 3-LC with additional intelligence as follows:

- Can interact with a dynamic traffic assignment module in the ATMS to implement proactive control (*i.e.*, in anticipation of the projected traffic volumes and routes),
- Can provide dynamic priority control on selected routes,
- Can implement congestion avoidance strategies, and
- Can implement congestion relief strategies.

5-LC: Fifth Level Control—This is a super level that incorporates the capabilities of all the previous levels. Most important, it makes the most efficient use of the control strategies in those systems based on accumulated expertise and experience under local conditions. Selection of the appropriate control strategy for the particular conditions is done by artificial intelligence technology. An analysis and explanation of the underlying basis for this design is given in the next section.

INTELLIGENT STRATEGY SELECTION

The hierarchical framework described will offer the possibility to tailor the system's capabilities to particular needs and means. Advanced strategies can be deployed gradually and can coexist with lower-level strategies. The full spectrum of capabilities can be built up gradually in terms of deployment of sensors, surveillance equipment, controllers, and communications, as well as central control hardware and software.

It has been common wisdom that increasing responsiveness contributes to improved traffic performance. As indicated previously, this is not always true. Consider the following facts based on evidence obtained from carefully conducted field experiments that were reported in the literature (10).

- **FACT 1.** TRANSYT settings are frequently used as the base for comparison of advanced strategies (such as SCOOT). But it has been demonstrated that, in some cases, other off-line methods can provide significant improvements in performance over TRANSYT. Therefore, when an advantage of a responsive method is claimed, it should be analyzed whether a more suitable off-line method would have given the same, or better results.
- **FACT 2.** Experiments with 2-GC methods have shown swings in both directions: improvement in some cases, degradation in others (3,9). The situations in which 2-GC methods are advantageous

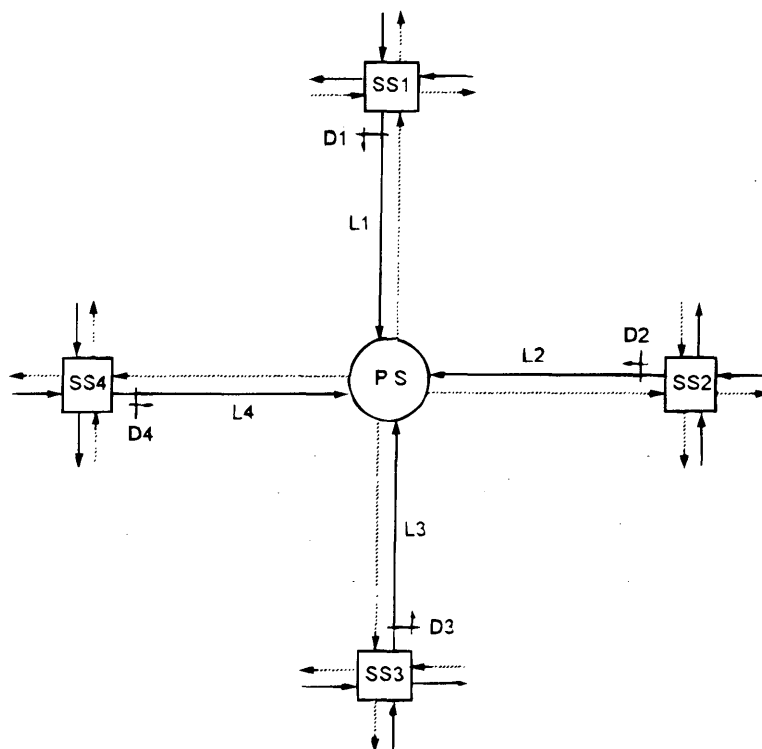


FIGURE 3 Architecture of OPAC network operation. (PS = primary signal, SS = satellite signal, D = detector, L = link)

should be characterized and not used when they are disadvantageous.

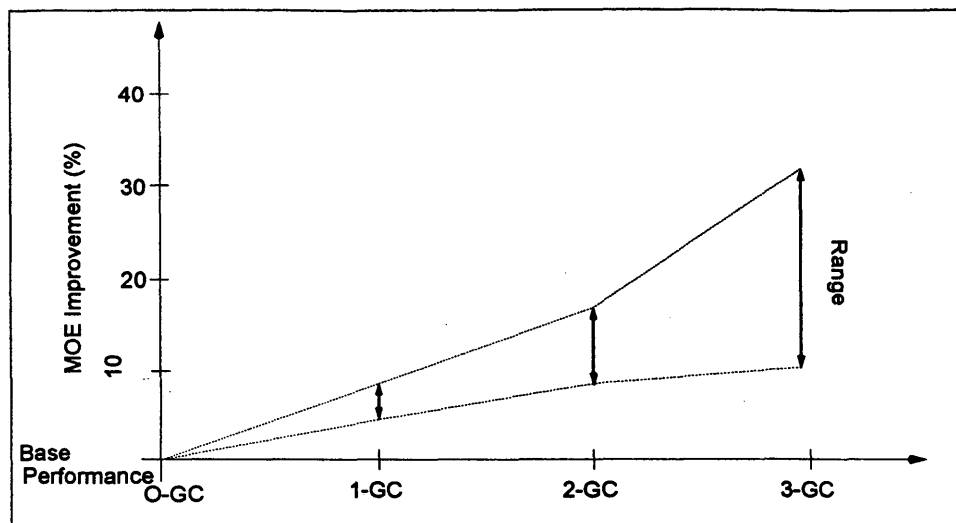
- **FACT 3.** Traffic-responsive methods such as SCOOT or OPAC also have shown considerable swings in performance under different traffic conditions. An example of this behavior is shown in Fig. 2, which illustrates the field evaluation of a responsive method relative to a base fixed-time plan. It is evident that (a) in some cases, no overall advantages are realized compared to nonresponsive methods and (b) in most cases, the performance data have a wide spread around the average. Thus, although responsive methods can provide substantial benefits compared with nonresponsive methods, they are also likely to degenerate into poor performance if not properly applied.

The argument is best illustrated by Figure 4. It shows the spreads in performance that are perceived to exist among the different control generations (Figure 4a), as well as the spreads in performance that were actually measured in the field (Figure 4b). The points shown in the figure were collected from published reports in the literature (3,8,9,11). These figures illustrate an interesting phenomenon: the more advanced (ie., responsive) strategies do not lead to improved performance all the time, notwithstanding the common perception that they do. More likely, they lead to a wider spread in performance results compared with a common basis. In some cases, because of reasons that are not completely explained, traditional off-line methods perform better than responsive methods. Therefore, one of the principal objectives of any intelligent control system would be to invoke a particular control strategy that will be

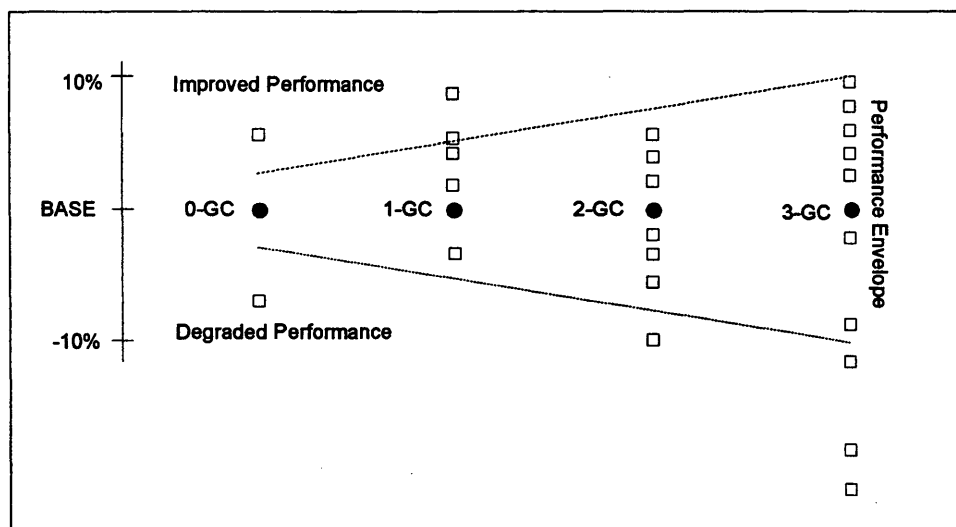
most suitable for existing conditions so that overall performance of the system is optimized.

It is envisioned that an expert system can be developed to ensure that a particular generation of control strategies will be implemented when it will provide the maximum benefits. Development of the system will require a careful characterization of signal networks and the identification of the particular traffic flow patterns that would be most amenable to benefit from a particular control strategy. The underlying proposition is that lower-level strategies may often be as good or better than higher level more advanced strategies. By recognizing the conditions under which the particular strategies perform best, overall system performance can be optimized. This would be the role of the 5-LC control level.

An overview of the operational flow diagram of the 5-LC strategy selection process is shown in Figure 5. This is a dynamic system in which the first step is the evaluation of the initial traffic conditions, as well as evaluation of the impact of external factors like weather conditions, type of day, and so forth. Based on this information, an initial control strategy is selected, best suited for the identified conditions. The strategy is then implemented and its performance is continually evaluated. The performance of the strategy, coupled with a dynamic traffic assignment model and the expected response of the users of the system to the control strategy, is used to predict traffic characteristics in the next control period. On the basis of these predictions, the system determines whether a new control policy is warranted. In case a new control policy is required, the most beneficial one is selected. If no change is needed, the performance continues to be monitored. In this way the performance



(a)



(b)

FIGURE 4 (a) Expected relative performance of control generations; (b) reported relative performance of control generations.

of the system is continuously evaluated and control strategies are updated only when conditions warrant. On-line simulation can be used as an aid in evaluating the performance of the system.

CONCLUSIONS

Although advanced technologies enable development of ever more sophisticated strategies, experience has taught that such strategies do not always result in improved performance. A major contribution of the new technologies will be to enable identification and recognition of the particular characteristics of the traffic system and selec-

tion of the most appropriate strategy for any existing situation. Frequently, such strategy might be a traditional fixed-cycle network optimized signal pattern or an arterial progression scheme. It does not necessarily have to be a real-time optimized control strategy. In other words, by providing a menu of strategies, one is likely to do better than by tying oneself to one particular strategy. The conclusion to be drawn from examining previous studies is the following: let us not throw out the experience gained in the past 30 years, but let us build on it gradually to develop improved traffic signal control operations. This is the basis for the framework proposed in this paper. Much of the development will depend on experimentation with alternative strategies and learning from experience in the field.

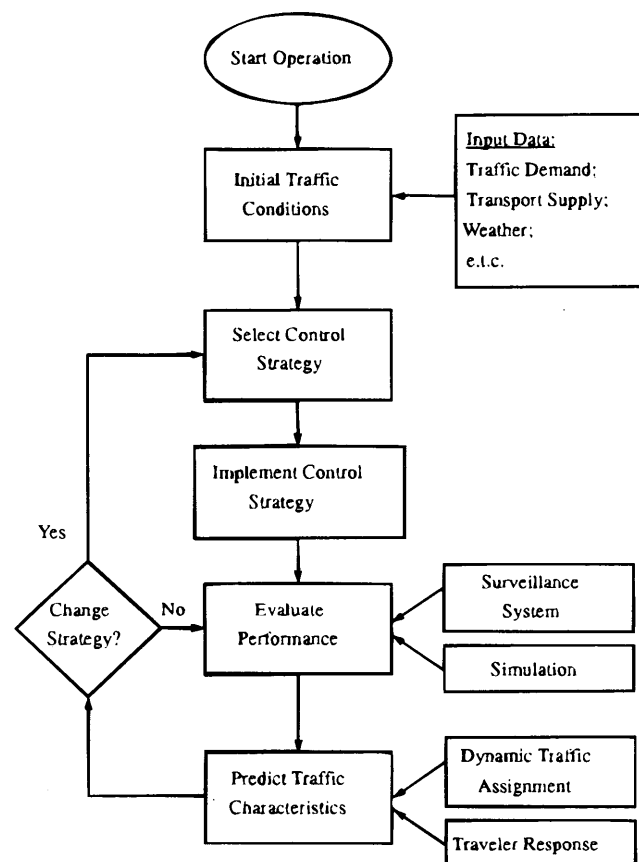


FIGURE 5 Operational flow diagram for strategy selection process.

ACKNOWLEDGMENTS

The paper is based, in part, on a research project sponsored by FHWA of the U.S. Department of Transportation with Farradyne Systems, Inc., as prime contractor and the University of Massachusetts, Lowell, as subcontractor.

REFERENCES

1. *Real-Time, Traffic Adaptive Control for IVHS*. Solicitation No. DTFH61-92-R-00001, FHWA, U.S. Dept. of Transportation, November 1991.
2. Tarnoff, P. J., and N. H. Gartner. Real-Time, Traffic Adaptive Signal Control. *Proc., Advanced Traffic Management Conference*, St. Petersburg, Fla., Oct. 1993.
3. MacGowan, J., and I. J. Fullerton. Development and Testing of Advanced Control Strategies in the Urban Traffic Control System (three articles). *Public Roads*, Vol. 43 Nos. 2-4, 1979-1980.
4. *Traffic Adaptive Control, Phase I: Critical Intersection Control Strategies*. Farradyne Systems, Inc., NCHRP Project 3-38, Vol. 3, Sept. 1989.
5. *Variable Cycle Signal Timing Program*. National Technical Information Service, KLD Associates, Inc., May 1974.
6. Gartner, N. H. Urban Traffic Control Strategies: The Generation Gap. *Proc., 2nd International ATEC Congress on Traffic and Transportation Control Systems*, Paris, France, 1980.
7. Gartner, N. H. Demand-Responsive Traffic Signal Control Research. *Transportation Res. 19A*, 1985.
8. Robertson, D. I. Traffic Models and Optimum Strategies: A Review. *Proc., International Symposium on Traffic Control Systems*, Berkeley, Calif., Aug. 1979.
9. *Improved Operation of Urban Transportation Systems*, Vol. 1-3, Metropolitan Toronto Corp., Toronto, Canada, 1974-1976.
10. Gartner, N. H. Prescription for Demand-Responsive Urban Traffic Control. In *Transport Research Record 881*, TRB, National Research Council, Washington, D.C., 1982, pp. 73-75.
11. Hunt, P. B., D. I. Robertson, R. D. Bretherton, and R. I. Winton. *SCOOT: A Traffic Responsive Method of Coordinating Signals*. TRRL Report LR 1014, United Kingdom, 1981.
12. Lowrie, P. R. SCATS: The Sydney Co-Ordinated Adaptive Traffic System. *International Conference on Road Traffic Signalling*. IEE, London, England, 1982.
13. Gartner, N. H. OPAC: A Demand-Responsive Strategy for Traffic Signal Control. In *Transportation Research Record 906*, TRB, National Research Council, Washington, D.C., 1983.
14. Henry, J. J., and J. L. Farges. PROLYN. *Proc., 6th IFAC/IFIP/IFORS Symposium on Transportation*, Paris, France, 1989.
15. Di Taranto, C., and V. Mauro. UTOPIA. *Proc., 6th IFAC/IFIP/IFORS Symposium on Transportation*, Paris, France, 1989.
16. Gartner, N. H., P. J. Tarnoff, and C. M. Andrews. Evaluation of Optimized Policies for Adaptive Control Strategy. In *Transportation Research Record 1324*, TRB, National Research Council, Washington, D.C., 1991, pp. 105-114.

Publication of this paper sponsored by Committee on Traffic Signal Systems.