Hierarchical Framework for Real-Time Traffic Control

K. LARRY HEAD, PITU B. MIRCHANDANI, AND DENNIS SHEPPARD

With the availability of faster computers and communications systems in the traffic control environment and more reliable monitoring and control hardware, better real-time control of traffic should be possible. The intelligent vehicle-highway system program now being proposed and implemented by transportation agencies, practitioners, and researchers will (a) need better real-time control methods for effectiveness in dealing with vehicle traffic and (b) allow the implementation and effectuation of better real-time traffic control. A framework for a hierarchical design of a real-time traffic control system is presented. The goal of the design concept is to respond to and monitor the various stochastic components of the traffic process with appropriate controls, frequencies, and sampling rates. The design is based on the decomposition of the traffic control problem into decision subproblems defined over different time and distance horizons. At the highest level of the hierarchy, the component process is considered that describes how over extended periods of time, travelers become aware of travel times and delays associated with the routes of a network and equilibrate into making routine route choices. At the middle level of the hierarchy, the faster dynamics are considered, for example, those dealing with traffic flows and queues during rush hours or traffic accidents. At the lowest level, the second-by-second dynamics in the traffic process are considered: the stochastic behavior of individual drivers and their responses to traffic controls at individual intersections. The conceptual design of RHODES, a prototype hierarchical traffic control system being developed for the city of Tucson, Arizona, and a comparison of its envisioned attributes with existing systems are described.

Advances in electronic control and communication technologies, coupled with significant increases in computer computational power and improvements in systems engineering and operations research methodologies, present an opportunity for significant improvements in traffic control systems. Computers have the ability to process information at rates that were only dreamed of 20 years ago. For example, telecommunications systems have utilized these technological and methodological advances to produce high-speed reliable communication between points separated by long distances. Integrated services digital network—or ISDN—technology allows the high-speed communication of voice, data, and video information over a single communication network among large groups of customers. Innovative large-scale distributed routing and flow control algorithms (based on advances in queuing theory, stochastic processes, optimization methods, and control theories) have been developed to address the problems associated with the utilization of modern telecommunication technologies.

Traffic control system design now benefits from these technological and theoretical advances in control and communication systems. Continued growth in travel demand without similar growth in new infrastructure has forced the traffic engineer to design traffic control systems that provide a higher level of performance without reducing safety and comfort. Modern communication systems allow the utilization of more information for traffic control than used by most existing traffic control systems. Synergistically, modern computers have the increased capacity required to process existing amounts of information. Furthermore, methodological advances in systems engineering and operations research can be used to design algorithms to improve system performance. The challenge for the traffic researcher is to design traffic control systems that integrate these advances.

Traditionally, advances in traffic control have resulted from extending existing models and control methodologies. This approach has been somewhat successful, but continued research in this direction does not use the available technological and methodological state of the art. The intelligent vehicle-highway system (IVHS) program proposed by transportation agencies, practitioners, and researchers presents a new structure for the traffic control system of the future (1). The integration of the advanced traffic management system (ATMS), the advanced traveler information system (ATIS), and the advanced vehicle control system (AVCS) components within IVHS will provide improved prediction of traffic volumes and flows and better control of the associated traffic. Under the IVHS umbrella, the solution of the traffic control problem requires new and innovative methods of information utilization and the generation of signal controls.

HIERARCHICAL DESIGN

The design of an ATMS requires a systems viewpoint of the problem, in which the entire system function must be considered. The function of a road network is to provide users a conduit for traveling from an origin to a destination. The function of the traffic control system is to manage the network so that travelers can traverse the network in a timely, safe, and efficient manner. Together these two functional components must satisfy the traffic demand placed on the system.

Within the framework of the IVHS structure, the ATMS must accept, as input, the available data on (a) travelers' origins and destinations, (b) the present and predicted traffic on the network, (c) the ATIS information provided, (d) the
AVCS signals suggested, and (e) the geometrics of the road network. In turn, it must produce, as output, control parameters that can be communicated to both the traffic control signals and AVCS. It is within this complex structure that a hierarchical system design concept for real-time traffic control is proposed. We focus only on the traffic control of an urban street network. The developed architecture will be extended to include freeway and corridor control later.

We refer to our traffic control concept as RHODES: a Real-Time, Hierarchical, Optimized, Distributed, and Effective System for traffic control. It is intended to provide a foundation that can be implemented independently and before the full realization of IVHS and allow for an evolution of an effective ATMS component within IVHS. The design concept for RHODES is based on the consideration of the characteristics of the traffic control problem. The direct synthesis of these considerations leads to a hierarchical control structure.

The goal of RHODES is to respond to the natural stochastic behavior of traffic. This stochasticity, which is both spatial and temporal, results from independent trip generations between spatially distributed origins and destinations, driver route selections, transit traffic, pedestrians, distribution and fluctuation of vehicle speeds, network events (accidents, road closures, etc.), and driver and vehicle characteristics (headway, speed, size, etc.). The spatial and temporal response characteristics of these stochastic sources are best described on different time and distance scales. Generation of origin-destination trips and response to network events such as road closures for construction evolve in time periods of days, weeks, and months. Transit traffic and transients in traffic due to accidents and scheduled events (e.g., rush hours and sporting events) affect the network within hours and minutes. Drivers and vehicles respond in time scales of minutes and seconds to events such as phase changes at signals, moving vehicle traffic, pedestrians, and queues at intersections. Together all of these sources result in the evolution of a complex stochastic system. A real-time traffic control system must respond to the various stochastic events in the system with appropriate time constants.

Identifying the appropriate performance criteria and response time constants to events is crucial in structuring the traffic control system. When the traffic network is lightly to moderately loaded, it may be more appropriate to control traffic so that vehicles are allowed to flow as freely as possible, without stops—that is, with the objective of accommodating individual vehicles. Under heavy loads (congestion) it may be more appropriate to control traffic for better network performance, that is, to make vehicles flow to accommodate the entire network traffic instead of individual travelers.

Most existing traffic control approaches respond to the stochastic nature of the traffic by attempting to statistically smooth the data and respond to average characteristics. Whereas such approaches may be appropriate for responding to long-term, slowly varying characteristics, they fail to realize that the data also represent actual traffic fluctuations that statistical computations cannot smooth. This reduction of information use can be understood by considering an analogous problem in speech processing. Speech is a stochastic process that when measured, recorded, or coded, is corrupted with electronic and sampling noise. Both the speech and the noise that affects it are processes that contain a high degree of variance. The sound made by a single letter “s” and a sample realization of Gaussian noise are indistinguishable. The goal in speech processing is to separate the information in the speech from the measurement noise. If statistical smoothing is used, the natural variance in the speech and in the noise will be reduced. But it is the variance in the speech—the words, the notes, the tones, the silent periods—that contains the useful information. Speech-processing methods, such as linear predictive coding and adaptive noise cancellation, have been developed to address the problem of eliminating noise from the signal. Furthermore, the choice of sampling rate used in speech processing depends on the variance (frequency content) of both the signal and the noise. Similarly, for the time/distance scales corresponding to the traffic characterization at each level of the RHODES hierarchy, the sampling rates need to be chosen and estimation methods developed that eliminate the measurement noise from the corresponding signal (traffic characterization) at each level. At the lower levels in the hierarchy, at which decision time scales are in seconds and minutes, estimation problems are more difficult because both the signals and the noise may have large variances.

Figure 1 shows the functional block diagram for RHODES. This hierarchical control architecture consists of four levels of control and real-time monitoring of vehicle flow. The hierarchical decomposition of the total traffic control problem considers the problem at the highest level in an aggregate fashion as well as with a long-term systems perspective. At the lowest level, the problem is decomposed, spatially and temporally, with the resulting subproblems considering short-term details with a local (intersection) viewpoint (see Figure 2). The highest-level problem is referred to as the “network loading problem,” the second level as the “network flow control problem,” the third as the “intersection control problem,” and the fourth as the “traffic signal actuation problem.”

At the highest level of the hierarchy we envision a stochastic traffic equilibrium module for network loading, in which the decision time horizons are in hours, days, and weeks. The premise for this model is that over this period, travelers become somewhat aware of the travel times, delays, and the associated statistical characteristics of links and routes (e.g., “during peak periods it takes between 15 to 20 min to go from the intersection of Swan Road and Sunrise Drive to the intersection of Campbell Avenue and Speedway Boulevard”), and they make route choices accordingly. This results in a stochastic equilibrium, which in essence provides an estimate, in a probabilistic sense, of the predicted loads on the links of the network. Mirchandani and Soroush provide a detailed discussion of this model and the approaches to find this stochastic equilibrium (2). Changes in network design and land-use patterns, trends in traffic flow, and ATIS information provided to the travelers may be fed back to this decision-making function to adjust the model parameters and predict near-future loads. Essentially, this planning level of the hierarchy provides (a) a priori estimates of link loads and (b) a posterior prediction of the trends in the change of loads from real-time data. This process constitutes the outer feedback loop for RHODES, as shown in Figure 1.

Level 2 of the hierarchy represents the high-level decision making for setting signal timings to optimize vehicle flow in the network. If flows were perfectly uniform and predictable, optimal timing plans could be downloaded in an open-loop
fashion. The computational requirements for such a fixed-
time control system are not stringent; timing plans can be
generated off-line using, for example, TRANSYT. This is the
assumption and the process by which many current systems
are set in the United States. However, flows are stochastic,
and to be real-time responsive, trends in traffic volumes must
be monitored, traffic volume time profiles must be estimated,
and, if necessary, new timing decisions must be implemented.
It is envisioned that the network flow control function at Level
2 will continually update the estimates of the traffic volumes
and flow profiles with a decision horizon in the range of sev­
eral minutes. Because of the potential computational com­
plexity of this problem, it is necessary to apply methodological
advances in (a) problem decomposition, (b) parallel com­
putation and, (c) good heuristics and approximations to
develop solution methods. The network flow control function
forms the middle feedback loop for RHODES, as shown in
Figure 1.

Conceptually, the network flow control problem can be
further decomposed into two sublevels, Levels 21 and 22. The
framework of the decision problem at every level is depicted
in Figure 3. Here the estimator and optimizer functions ex­
licitly take into consideration a dynamic traffic model of the form
\[ x(t + 1) = f(x(t), u(t)) + \xi(t) \]  \hspace{1cm} (1)
which states that \( x(t + 1) \), the state of the system (volumes,
queues, travel times) at time \( t + 1 \), is a function of \( x(t) \), the
state at time \( t \), \( u(t) \) the controls at time \( t \), and a stochastic
exogenous noise \( \xi(t) \). (Equation 1 represents a discrete dy­
namic traffic model; a corresponding differential equation
represents a continuous model.)

At Level 21 of the network flow control, the decision sub­
problem (referred to as the capacity allocation problem) is as
follows:

Given, at time \( t_n \), the predicted exogenous inputs \( \lambda_n(t) \) and
outputs \( \gamma_n(t) \), at each node \( i \), the capacities \( c_i \) on flows from node
\( i \) to node \( j \), the current travel times \( t_f \) on link \( [i,j] \), determine the
fraction of time that "green light" should be allocated to each
flow movement.

This problem can be solved as a linear programming model
of the decision problem.

At Level 22, the decision subproblem (referred to as the network
coordination problem) is as follows:

Given, at time \( t_n \), the platoon movements within the network,
and approximate target allocations of green time, what should
be the phase sequences and approximate green and red periods
for each flow movement?

This problem could be modeled as a discrete network flow
problem and should be solvable in 2 to 3 min.

In describing the subproblems at Levels 21 and 22, the
performance criteria for the corresponding optimization models
FIGURE 3 Framework of network flow control decision model at each level of hierarchy.

have been purposefully left out. The dominant optimization criterion (or criteria) at each level depends on the state of the network. Perhaps, it is most appropriate to minimize average queues when the network is saturated and congested and to minimize stops when the network use is very low. The RHODES framework allows the use of different criteria for different traffic conditions. The criteria that are most suitable must be determined through experimentation, field testing, and experience.

The intersection control at the third level can also be decomposed into the two Sublevels 31 and 32. The decision subproblem at Level 31 pertains to the determination of target timings at each intersection. The corresponding estimation-optimization subproblem (referred to as the signal scheduling problem) may be stated as

Given, at time \( t_0 \), the traffic flow profiles entering the intersection, the phase sequences and approximate green and red periods for each flow movement, what are the optimal light change epochs for the next phase sequence?

This problem can be solved as a dynamic program in a distributed fashion (for each intersection). These local estimation-optimization problems should be solved within a 1-min time frame.

Levels 21, 22, and 31 provide target timings (phase sequences, phase times, splits, offsets) and allowable variances to the Level 32 subproblem. The allowable variances are to inform the local controllers of the sensitivity of the network flow to variations in the actual timings. These allowable variances will generally decrease as intersection saturation increases. The intersection controllers will use these timings and variances to respond to the stochastic fluctuations in traffic flow. The decision subproblem (referred to as the intersection dispatching problem) at Level 32 is a simple one:

Based on the observable traffic on the approaches to the intersection and vehicles in the queues, should the current phase be shortened or extended?

Since the enormous number of factors that produce the short-term fluctuations in the observed traffic either are unknown or cannot be modeled, a model-based exact optimization method is not suitable for Level 32. However, concepts from artificial intelligence and learning theory may be used to develop a solution method that learns the characteristics of individual intersections and responds in real-time to the short-term traffic fluctuations that occur in a time frame of seconds and minutes.

Developments of the intersection control module and the establishment of appropriate vehicle detectors provides the inner feedback loop for RHODES for real-time local (distributed) control (see Figure 1).

The fourth control level, referred to as traffic signal actuation, is the interface with the local controller equipment. At this level, phase sequences, phase times, and offsets are passed to the controller. The key element to the success of these decisions is that field data regarding vehicles on all approaches are provided with sufficient advance notice to affect local timing decisions through control actuation.

The RHODES hierarchical structure provides a general framework for the design of a real-time traffic control system to react and affect the stochastic nature of vehicle traffic on a network. The hierarchical structure addresses the different decision and estimation problems that have different time-distance scales and different response time characteristics at each hierarchical level. Existing systems for traffic control address issues at one or two levels of the hierarchy, but none directly addresses the entire traffic control problem. In the following section several existing traffic control systems are discussed within the RHODES hierarchical framework.

COMPARISON WITH EXISTING APPROACHES

Existing traffic control systems include signal timings based on both fixed- and real-time control. The following categorization, used by many researchers and practitioners, distinguishes the mechanisms whereby signal timing adjustments are made (3):

- First-generation control (1-GC) involves off-line optimization and subsequent manual–or time-of-day–based implementation of new signal timing plans.
Second-generation control (2-GC) involves generation of timing plans based on predicted trends in traffic condition and stepwise transitions among timing plans.

Third-generation control (3-GC) involves on-line optimization (i.e., in real time) with very short (1- to 2-min) sampling periods between updates. The cycle lengths, offsets, and splits change continuously.

One-and-a-half-generation control (1.5-GC) is a strategy with some of the features of both 1-GC and 2-GC. It involves automatic development of signal timing plans, but implementation requires operator approval.

First- and second-generation systems can be found throughout the world (3). In the United States, the most commonly used control system has been the Urban Traffic Control System (UTCS). Developed by FHWA during the 1970s, UTCS is capable of 1-GC (4) and 2-GC (5) control but not 3-GC control. The UTCS structure depends on time-of-day (TOD) plans that are developed off-line and on the basis of average conditions on the network during corresponding time periods and downloaded automatically at the corresponding time of day. To develop time-of-day plans for UTCS, a number of signal optimization programs have been developed and enhanced over the years, such as SIGRID, TRANSYT-7F, SIGOP, and PASSER II.

SCOOT is a notable example of 3-GC control (6–11). From available literature and personal communications, it appears that SCOOT makes incremental adjustments to the current signal timing plan (including cycle lengths, phase lengths, and offsets) for the next cycle, in response to changing traffic demands and suggestions by TRANSYT optimizations that are continually being performed “in the background.” Red Deer (Alberta) was the first North American installation of SCOOT. Currently, installations are underway in Toronto, Ontario; Halifax, Nova Scotia; and Oxnard, California. The original prototype installations were made in Glasgow, Scotland, and Coventry, England, in 1984 (12); the associated evaluations indicate that it performed better than fixed-time control.

There are two real-time network control schemes developed and implemented in Australia: SCATS (Sydney Coordinated Adaptive Traffic System), and TRAC (Traffic Response Area Control). The more widely used system is SCATS, originally developed by Sims (13) for Sydney but now implemented in Melbourne, Adelaide and other cities in Australia, as well as in New Zealand and several major cities in Asia. As it is for SCOOT, very few technical details are published on SCATS. From the data available and personal communications, it appears that SCATS uses a hierarchical control architecture. At the local level, each subsystem (a set of intersections prespecified by a traffic engineer) makes independent decisions on its timing parameters (cycle, offsets, and phase lengths) on the basis of the degree of saturation in the subsystem. Adjacent subsystems “marry” and get coordinated by a higher-level regional computer when their cycle times are equal or nearly equal. Likewise, when the degrees of saturation and the consequent desired cycle lengths become different, the two subsystems “divorce.” It is not clear how the timing parameters are adjusted on-line, but from observing SCATS’ operations it is clear that the parameters are incrementally adjusted to varying traffic conditions to provide stability and damping in the overall control system.

TRAC is a system developed by the Main Roads Department, Queensland, that combines aspects of SCOOT and SCATS but does not perform incremental optimization (14). Each subsystem can have up to 12 stored plans, and the best plan is downloaded by a regional computer for implementation. The plan may be selected by average detector occupancy, time of day, manually, or, in principal, by any performance measure observable by detector data. The plans stored in each subsystem may be developed off-line, using TRANSYT for example. In personal communication, Lees indicated that plans may be continually updated depending on recent detector data and associated derived measures (14).

Each of the existing systems works well and addresses some of the issues that RHODES design attempts to address. The major drawback is that these systems are not proactive and, therefore, cannot easily accommodate the commonly occurring significant transients. The stochastic traffic equilibrium component at top of the hierarchy, as well as the model-based traffic predictions at each hierarchical level, allows RHODES to be proactive. A proactive system attempts to predict future demand to be placed on the network and to accommodate this demand as it evolves. Typically, control signals at each level respond to predictions over several time constants for the level. For example, at Level 31 of the intersection control, we consider predictions over time periods that may be equivalent to several cycles, as opposed to a single cycle generally used in most systems.

The intersection dispatching component at the bottom of the hierarchy is intended to make RHODES reactive to second-by-second random fluctuations in traffic and is implemented as a distributed control system. A reactive system responds to both predicted and unpredicted demand as it evolves. A distributed control system allows local control decisions at spatially separated locations.

For proper decision making at local controllers and at higher levels of the RHODES hierarchy, appropriate interconnection and communication is necessary among the levels, both through hardware-connecting processing and monitoring units and software for passing inputs and outputs to various decision-making algorithms. We cannot overemphasize the importance of modern computer and communication technologies and novel algorithmic methods to make the hierarchy of RHODES work effectively.

Table 1 presents a comparison of the preceding systems in terms of whether the system is (a) proactive, (b) reactive, (c) distributed, or (d) hierarchical. In addition, each of the systems is classified according to the timing decisions method used.

Implemented 1-GC UTCSs generally use fixed-time plans and sometimes allow for time-of-day plan selection. Implemented 1.5-GC and 2-GC UTCSs allow on-line selection of timing plans responding to time of day or detected traffic conditions, or plans are generated on-line on the basis of predicted smooth traffic flows. SCOOT and TRAC are closer to 3-GC control as characterized by McShane (3), in which plans can be either selected on-line or generated (and incrementally adjusted) on-line, in a time scale of a few minutes.

It is important to note that the preceding systems try to come up with timing plans for the whole network in terms of cycle times, offsets, phase lengths, and so forth. In considering whether such a strategy could be optimum, an optimum timing plan implicitly assumes the existence of steady-state condi-
DISCUSSION OF RESULTS

This paper has introduced a control hierarchy obtained from a system perspective of the real-time traffic control problem. Although the proposed system has not been fully developed for simulation and demonstration purposes, it has developed a conceptual design that responds to deficiencies in existing and available real-time control systems. This approach allows for an evolution of hardware and software developments. For example, constraints posed by existing signal controllers may be incorporated in the decision algorithms for determining optimal timings.

A research team at the University of Arizona and the city of Tucson, Arizona, together with the Pima County Association of Governments, is further investigating the viability of the RHODES concept. The Arizona Department of Transportation has provided research funding, and the city of Tucson has agreed to consider the implementation and field testing of the RHODES system once it has been developed and tested through computer simulation.

The hierarchical control architecture developed for RHODES parallels similar approaches used in modern manufacturing and production control systems. A factory is loaded with jobs; jobs are routed through processing centers, scheduled at the centers, and dispatched to the processing units in a hierarchical fashion to optimize appropriate measures of performance (17). The processes within a factory are also stochastic, and the production control problem is complex. The hierarchical decomposition of the production control problem, on the basis of time and distance scales for the various manufacturing processes, allows us to understand and solve this problem—and it is hoped that it will also aid in solving the real-time vehicle traffic control problem.

ACKNOWLEDGMENTS

The research discussed has been supported by the Arizona Department of Transportation. The authors wish to acknowledge stimulating discussions with their colleagues who have been working on this research effort: Jeff Goldberg, Julia Higle, Paul Sanchez, Suvrajeet Sen, and Bob Wortman of the University of Arizona; George List of Rensselaer Polytechnic Institute; Tom Buick and David Wolfson from the Transportation Planning Division of the Pima Association of Governments; and Louis Schmitt from the Arizona Department of Transportation. These discussions have contributed greatly in crystallizing the RHODES architecture.

REFERENCES


The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Arizona Department of Transportation or FHWA. This report does not constitute a standard, specification, or regulation.

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