System-Optimal Trip Scheduling and Routing in Commuting Networks

GANG-LEN CHANG, HANI S. MAHMASSANI, AND MICHAEL L. ENGQUIST

A time-space network formulation is presented for the system-optimal assignment to departure times and routes of traffic flows from multiple origins to a common destination. Time is discretized, and congestion is represented using simplified deterministic queuing stations. The solution minimizes total travel time in the system subject to arrivals at the destination taking place within a specified time interval. Alternatively, a formulation is presented for the minimization of a total cost measure consisting of a weighted sum of the users' travel time and schedule delay. The solution can be obtained using efficient and widely available pure network optimization algorithms. A numerical application is presented to illustrate the methodology, including a network generator developed for this purpose.

Peak-period congestion continues to be a severe daily annoyance in most metropolitan areas where large volumes of commuters desiring to arrive at their destinations within a narrow time interval compete for limited transportation system capacity. No major innovations for combating congestion seem to have emerged in the past 15 yr. Recently, the potential of advanced information and communication technology for congestion control appears to have rekindled interest and effort in this problem. However, the design and evaluation of various control strategies require deeper understanding of the systems' complex nature and methodologies with the capability to deal effectively with time-dependent flows in congested networks.

Several contributions have addressed the problem of finding a time-dependent flow pattern that satisfies dynamic user equilibrium conditions in an idealized system consisting of a single route containing a bottleneck and connecting a single origin-destination pair (*1–9*). Extensions have included multiple routes and alternative assumptions on the system's configuration or behavioral mechanisms underlying tripmakers' decisions (*10–14*). The day-to-day dynamics of the interaction between commuter decisions and congestion in a traffic system have also received some attention recently, using simulation experiments (*15*) and observational studies (*16–19*). Relatively little attention has been directed toward the problem of solving for time-dependent traffic patterns that are in some sense optimal from a total system cost standpoint.

Previous studies dealing with time-varying system-optimal traffic patterns have followed one of two lines: (a) optimizing the traffic-generation patterns in a given system with a single route (and one bottleneck) or (b) assigning known time-dependent flows from multiple origins to a single destination to the links of a network so as to minimize total system cost (travel time). Research along the first line consists of analytical derivations or discussions of system-optimal departure patterns, in connection with the aforementioned dynamic user equilibrium studies in an idealized system in which a number of commuters from the same origin are trying to arrive at a common destination at the same time (*6, 9*). Extension to the scenario of staggered work hours has been described elsewhere (*20*). Discussions of system-optimal departure patterns are also given by Hendrickson et al. (*21*), Fargier (*5*), and Newell (*14*).

Contributions along the second line of research are limited to situations where the time-dependent departures from multiple origins to a single destination are known, and congestion is modeled using link performance functions intended for static traffic assignment applications. As such, these are direct extensions of the standard system-optimal network assignment formulations. Merchant and Nemhauser (*22, 23*) formulated the problem as a discrete time, nonlinear but nonconvex math program where the objective is to minimize the total travel time spent by the given trips in the network. A recent paper by Carey (*24*) reformulates that problem as a convex nonlinear program, which is of course more attractive computationally than the previous formulation, and discusses possible extensions to more general situations.

The present paper addresses a more general system-optimal state, which includes not only the assignment of known time-varying flows to the links of a commuting network but also the determination of the corresponding optimal time-dependent traffic-generation patterns from the various origins, given constraints on desired arrivals at the destination. The paper presents a methodology for the system-optimal assignment of commuters to departure times and routes subject to specified constraints on acceptable arrivals. It consists of a time-space network formulation that can easily be solved using existing efficient network flow programming codes. The scope is still limited to commuting systems with a single major destination, such as a CBD or other large industrial or business employment center, but allows for multiple routes and multiple origins. It is intended primarily as a tool to explore the potential benefits that could be achieved from information-related and demand-side strategies aimed at reducing congestion.

The next section presents the conceptual framework and principal features of the proposed approach, followed by a detailed formulation for the time-space network of principal activities for a simple commuting system with a single route and a single origin. Extension of the formulation to more general situations with multiple routes and multiple origins is

G.-L. Chang, Department of Civil Engineering, University of Texas at Arlington; Arlington 76010. H. S. Mahmassani, Department of Civil Engineering, University of Texas at Austin, Austin 78712. M. L. Engquist, Management Science and Information Systems, University of Texas at Austin, Austin 78712.

discussed in a later section, followed by presentation of a numerical illustration. Finally, concluding comments and possible extensions are addressed.

MODEL FORMULATION

This section presents the key features of the time-space network formulation, including the representation of the traffic system. The context considered here is a commuting corridor surrounded by residential areas. For convenience and ease of presentation, we start with the simplest scenario, shown in Figure 1, where only one highway facility exists in the corridor for use by residents from adjoining areas in their daily commute to the same work destination. Concern here is primarily with the inbound, or home-to-work, direction.

For the purpose of analytical representation, the highway facility is conceptually divided into a number of sections with each including, at most, one entry ramp. The time spent in any section or ramp depends on the facility's service characteristics and the generated time-dependent flow patterns. In this formulation, the entire system is viewed as a network of queuing stations. Using a simplified representation adopted in several papers dealing with dynamic traffic assignment (*7, 10, 11, 25*), each highway section can be viewed essentially as a potential bottleneck with a given service rate (capacity). If the flow is less than this service rate, then only the free-flow travel time is incurred on the corresponding section; otherwise, a waiting time in queue is incurred, representing the excess travel time resulting from congestion. Likewise, entrance and exit ramps can also be modeled as typical, deterministic queuing stations with service rates depending on each ramp's physical capacity and control system (e.g., in the event of ramp metering). Further detail is given hereafter.

Because we are dealing with only one day's process at a time, the system is considered for a given duration that includes the earliest and latest possible (and meaningful) departure times. Time is discretized into equal intervals of a suitable small length Δt (in the order of a few minutes). The network formulation of the system-optimal dynamic assignment problem can be obtained by analogy to the time-space relation of individual vehicles traveling from the origin to the destination. The network is akin to a trans-shipment problem where it is desired to send flow units (tripmakers) from a set of supply points (origins) to a demand point (destination) at minimum cost using a network of arcs and nodes. The arcs correspond to activities, generally involving an expenditure of time or other cost, incurred on a per-flow unit basis; whereas the nodes correspond to the beginning and/or end of activities. The activities here include but are not limited to movement on physical highway links.

There are basically three categories of activities in the formulation: departure from origins, travel on links (including queuing at bottlenecks), and arrival at the common destination. Described are the formulation of each category as a separate subnetwork and then their integration to form the entire network for a given problem. The presentation is primarily graphical in nature; the notational convention used in the network formulation follows the work of Klingman and coworkers (*26*) and is summarized in Figure 2.

Formulation of the Departure Activity

For a given origin (supply) node, each discrete departure time alternative (of length Δt) is represented by a node, as shown in Figure 3, with the earliest and latest possible departure times denoted respectively by nodes DA_1 and DA_n, and the intermediate nodes labeled sequentially. The total number of commuters originating from location A constitutes the total supply (in trans-shipment terminology) for node A, which is connected to each of the associated departure nodes by a unique outgoing arc. The flow on arc (A, DA_k), $k = 1, \ldots, n$, obtained in the solution corresponds to the number of users that depart from A in the kth time slice; the set of these flows thus represents the optimal departure pattern of users at this location. The departure time here is taken at the entry of the highway facility. Thus arc (A, DA_k) corresponds to local travel from origin A to the facility. For simplicity, but without loss of generality, we assume that the time cost of this travel is a constant, T, associated with each arc (A, DA_k), $k = 1, \ldots, n$. This cost is not necessary from the perspective of model operation, however, because users at a given origin are uniquely assigned to an entry point. A more detailed formulation could let this assignment be determined in the optimal solution.

Commuters may have to join a queue or be otherwise delayed at the entry point. The horizontal arc emanating from each departure node (see Figure 3) is designated to carry only those commuters actually entering the highway in that given interval. The upper bound of flow through each arc, denoted as $C1$, is used to control the maximum entry rate, reflecting either physical capacity restrictions or the effect of traffic control devices. The associated arc cost $T1$ represents the travel time to the next "state," a congested location in this case. Because of the preceding capacity constraint, commuters departing simultaneously (i.e., in the same time slice) may not all be allowed onto the facility at the same time. Thus each departure node DA_k is connected to the next departure alternative DA_{k+1} by arc (DA_k, DA_{k+1}), shown vertically in Figure 3. This arc will carry the excess number of commuters at DA_k who could not be served in a given time slice. The resulting waiting time is then captured by the arc cost,

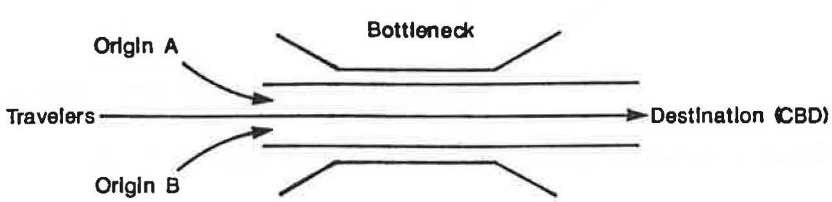

FIGURE 1 An idealized commuting system for analysis.

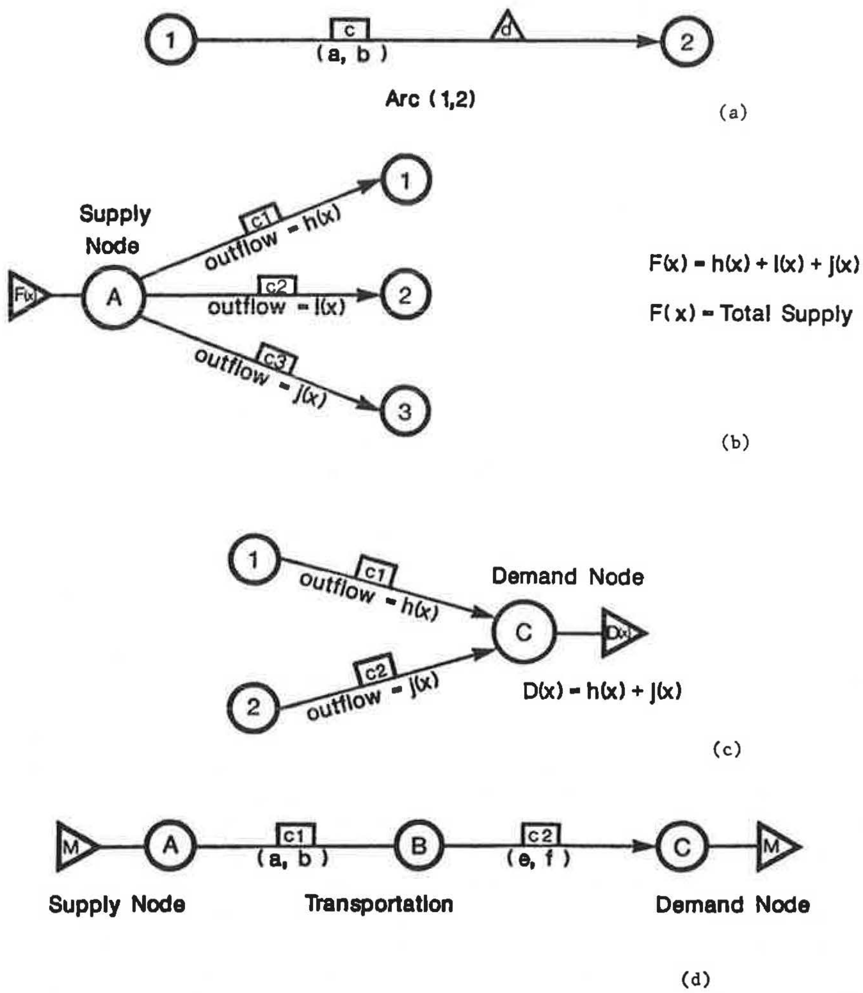

FIGURE 2 Notation for the network formulation.

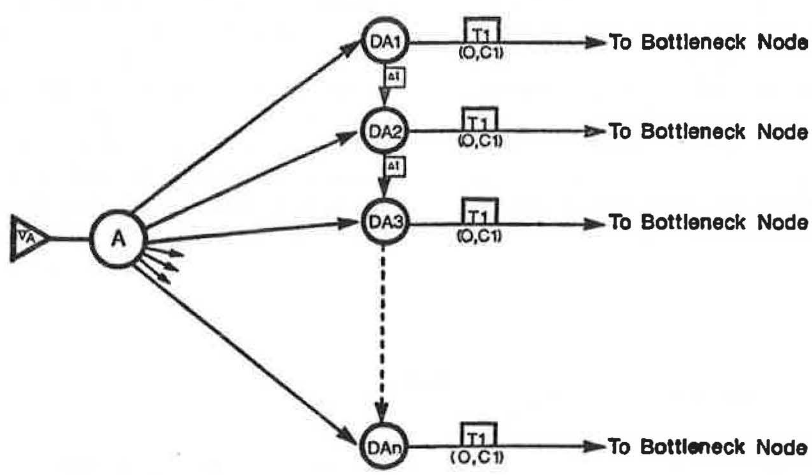

FIGURE 3 Graphical representation of the departure activity subnetwork.

Δt, equal in magnitude to the unit departure time slice. An upper bound on the flow on these waiting arcs can be specified to reflect physical storage capacity limitations or policy decisions.

Other trip generation sources, such as origin B in Figure 2, can also be represented in a similar manner.

Formulation of Congested Locations

It should first be noted that this representation is not intended to capture the details of the traffic flow phenomena taking place on the facility or the formation and dissipation of physical queues in the system. It is principally an approach to calculate realistic travel times under congested conditions for each link in the context of a pure network formulation of the system-optimal, time-varying assignment problem. There may not necessarily be a physical queue of stopped vehicles in the actual system, even if traffic is highly congested. Instead, users may be forced to slow down along some sections with particularly high concentrations. As noted in the previous section, such congested locations are modeled as queuing stations and are formulated as follows.

Two sets of nodes and associated connecting arcs are proposed to model congested locations. As illustrated in Figure 4, the first set of nodes, denoted as BI_k, $k = 1, \ldots, n$, is used to represent the arrival at the bottleneck, with nodes BI_1 and BI_n representing the earliest and latest arrival times, respectively. Each node in the set $\{\mathrm{BI}_k, k = 1, \ldots, n\}$ is connected by an arc $(\mathrm{BI}_k, \mathrm{BO}_k)$, shown horizontally in Figure 4, to a unique corresponding node in the second set $\{\mathrm{BO}_k, k = 1, \ldots, n\}$ designated to model the exit from the bottleneck. The upper bound on flow in each of these arcs is defined by the bottleneck's service rate S (i.e., the number of users allowed to go through in a given time slice Δt). The associated arc cost $T2$ is the time through the section in the absence of congestion. As in the formulation of the departure activity, vertical arcs (in Figure 4) are specified from each node BI_k to node BI_{k+1}, $k = 1, \ldots, n$, to carry the queuing flow, with arc cost again equal to the unit waiting time Δt. Note that no such arcs are shown in Figure 4 between consecutive BO_k nodes because queuing should not occur immediately upon exit from the bottleneck.

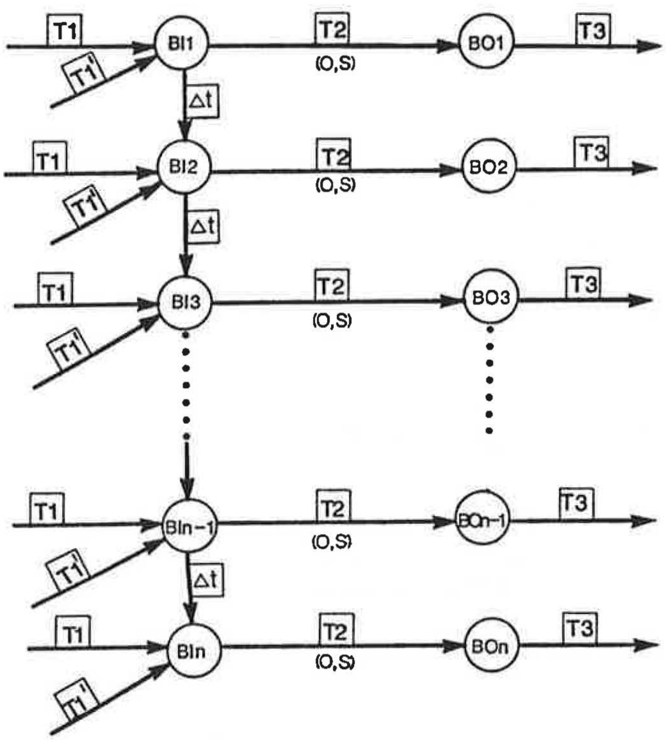

FIGURE 4 Graphical representation of the bottleneck area subnetwork.

Finally, all arcs incident to the BI_k nodes, except for the queue-carrying vertical arcs, are intended to carry commuters arriving at the bottleneck. Because each of these arcs originates at a departure node, the specification of the associated arc cost and upper bound on flow must be consistent with that specified in the departure formulation. Likewise, the arcs leaving each of the BO_k nodes carry the flow departing from the bottleneck. No upper bounds are shown for these arcs in Figure 4 because the flow has already been regulated by the service rate S of the bottleneck, although we may want to specify such upper bounds for more general systems.

Formulation of the Arrival Process

The formulation of the arrival process subnetwork depends on the explicit definition of the system optimum sought. So far, we have implied that the desired solution would minimize total system cost, calculated as the sum of all arc costs incurred by the assigned flows. These arc costs have in turn been specified as either uncongested travel times or delays due to congestion at bottlenecks. We need to address further the costs contributed to the objective function incurred in conjunction with the arrival process, as well as the constraints that need to be satisfied by this process. Now considered here are two basic alternative formulations reflecting different assumptions about the users' preferences or cost function: a satisfying formulation and a utility maximization one. We also describe how variants can be modeled.

Before describing these two formulations and the underlying assumptions, it is useful to consider, qualitatively, the nature of the departure patterns that can be expected in the solution. First, it must be recognized that it is generally not feasible for all users to arrive simultaneously (in a single time interval Δt) at the desired destination. There is a minimum duration for the arrival period that is governed by the capacity of the bottlenecks. If users were allowed to arrive at any time before the official work start time, then one can almost always find a solution that minimizes the total travel time in the system and that involves absolutely no queuing (i.e., all the vertical arcs in the network formulation would have zero flows). Unfortunately, such a solution would likely exhibit so much spread in the departure (and arrival) pattern that it would be meaningless. In other words, we would have a trivial problem if there were no constraints on either the range of possible departures or the range of possible arrivals and if travel time were the sole consideration in the objective function.

The first meaningful formulation we consider here constrains all arrivals to take place within a specified time band. It is consistent with empirical evidence that workers like to allow some extra time prior to the official work start time (*18,* 27); as such they may be indifferent to arrivals if they are within a reasonable time band. Referring to a recent paper by Mahmassani and Chang (12), it can also be noted that this formulation would yield the "best" departure pattern among the multiple patterns that satisfy boundedly rational user equilibrium conditions for a given value of the indifference band, assumed to be identical across users.

The second formulation places a penalty on the time between actual arrival at the destination and the work start time, also referred to as schedule delay. Thus the user's utility function would include both the travel time and the schedule delay, the latter multiplied by a weight reflecting its valuation relative to travel time. This type of function would be consistent with the classical microeconomic view of this problem, as presented by Vickrey (9) and by Hendrickson and Kocur (7). The solution would involve a trade-off between travel time and schedule delay, which would lead to spread-out departures and arrivals and thus high schedule delays. We next describe the network representation of the two cases, starting with the satisfying formulation.

In all cases, we define a set of arrival nodes D_r, $r = 1, \ldots, n$, that define the arrival time alternatives, generally corresponding to the departure nodes DA_1 through DA_n. The satisfying feature is included in the formulation by specifying the subset of consecutive nodes from D_k to D_n as the acceptable range of arrival times, as shown in Figure 5. All commuters are supposed to traverse at least one of those nodes to end their trips. Each of these nodes is connected to a supersink (or total demand) node DE, the common destination, by arcs (D_k, DE), with upper bound on flow denoted by C3 in Figure 5. This value may be the same across these arcs, representing the physical constraint for the arrival rate, or may vary across arcs to reflect the operation of traffic control devices. Each of the feasible arrival nodes is connected to the next one by a vertical arc with cost Δt to convey the queuing flow.

Unlike nodes D_k through D_n, nodes D_1 through D_{k-1} are not connected to the supersink. Vertical arcs with very high costs (M) are specified between each pair (D_i, D_{i+1}), $i = 1, \ldots, k - 1$. This will prevent flows in the network from taking paths ending in an unsatisfactory arrival time (i.e., outside the band) unless there is no feasible solution for the specified arrival time band. Obviously, nonzero flow on any of the "big M" arcs in the final solution will be a sign of unfeasibility, which could be resolved by widening the acceptable arrival band to include additional arrival nodes.

Given the foregoing formulations of the three principal activities, the network for the entire system can be constructed through careful integration of the three subnetworks, as shown in Figure 6 for the idealized commuting system of Figure 1. We next describe how the formulation of the arrival process can be modified to represent the utility maximization case.

Utility Maximization Formulation

As noted previously, the total trip cost of commuters depends, under this rule, on the specification of the utility function. A commonly used specification in this context involves a trade-off between trip time and schedule delay, of the form:

$$TC_{i,t} = (a \cdot TR_{i,t}) + (\delta \cdot b \cdot SDE_{i,t}) + (1 - \delta) \cdot c \cdot SDL_{i,t} \quad (1)$$

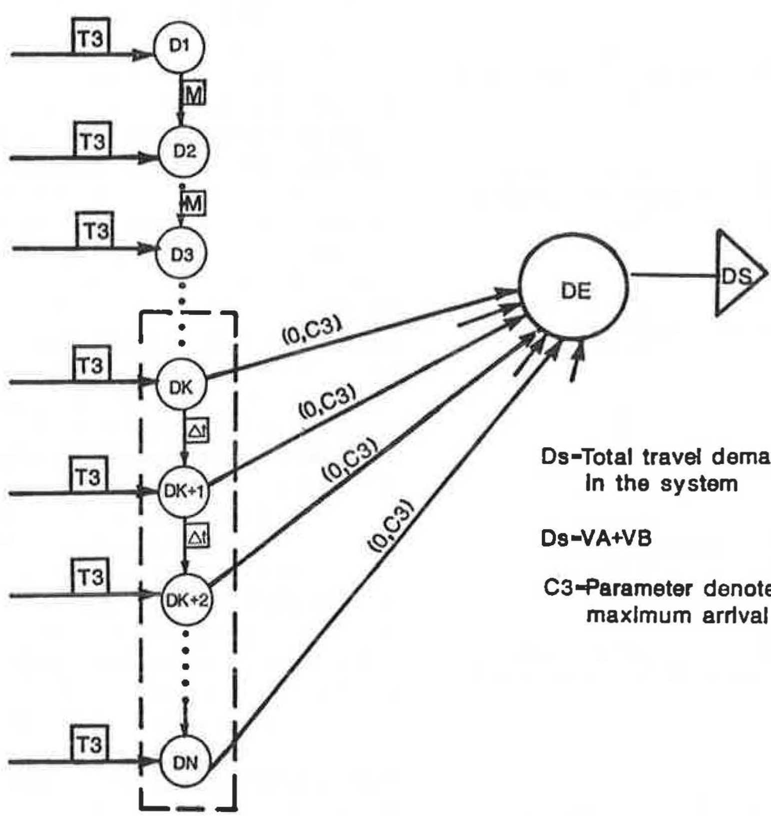

FIGURE 5 Graphical representation of the satisfying formulation for the arrival process subnetwork.

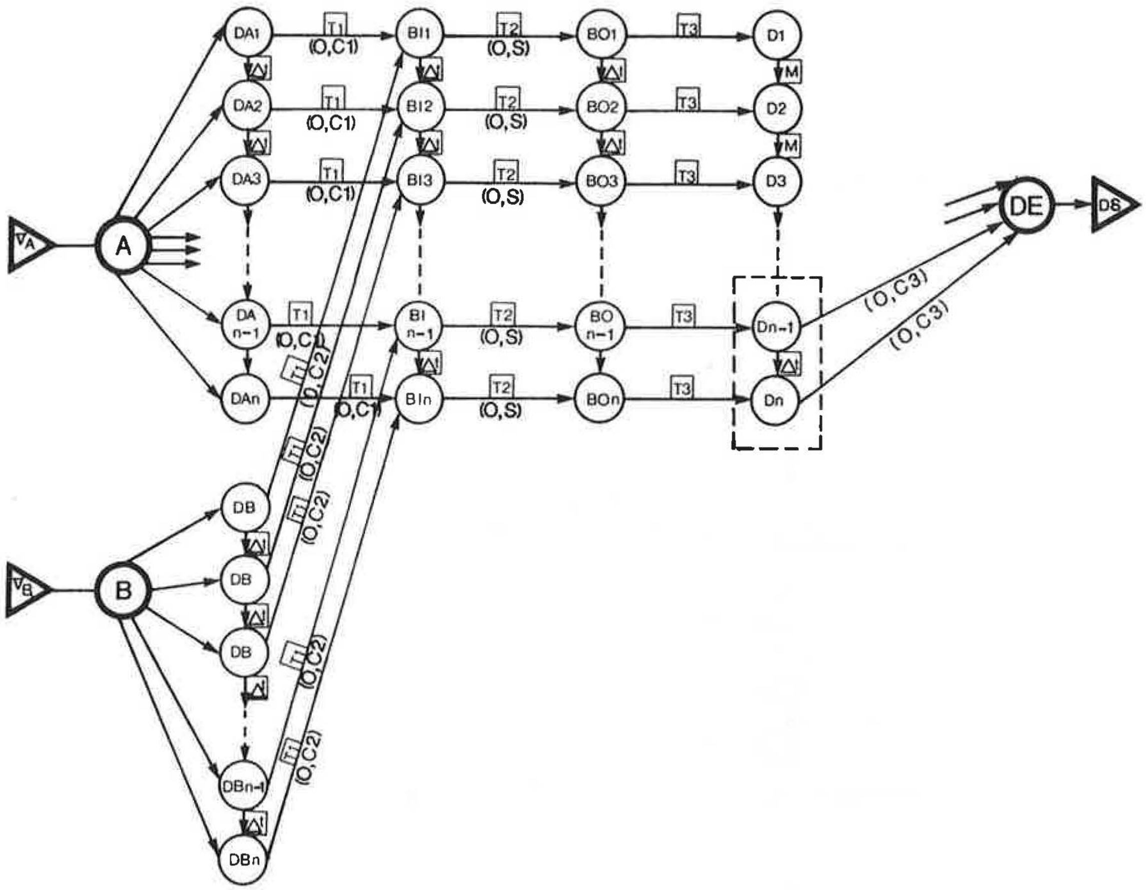

FIGURE 6 Example network formulation for the idealized commuting system.

where

$TC_{i,t}$ and $TR_{i,t}$ = the total travel cost and travel time, respectively, incurred by flow unit i departing at time t;

$SDE_{i,t}$ and $SDL_{i,t}$ = the schedule delay for early and late arrival, respectively, relative to the desired arrival time;

a, b, and c = parameters capturing the disutility of a unit of travel time, schedule delay for early and late arrival, respectively (it is convenient to set $a = 1$ and scale b and c accordingly); and

δ = a binary variable equal to 1 for early arrival and to 0 for late arrival.

We assume hereafter that all users are identical in terms of the parameters of the preceding function. To capture this trade-off between schedule delay and travel time, the network formulation of the arrival process can be modified as shown in Figure 7. Let node AR_n denote the work starting time; the other arrival time nodes form two groups: AR_1 to AR_{n-1}, and LAR_1 to LAR_m, which correspond to early and late arrivals, respectively. Only node AR_n is connected to the total demand node DE to force all flows, except those arriving at the AR_n node via a horizontal travel arc, to traverse the needed number of queuing arcs to reach AR_n before they can terminate their trips. The summation of the costs incurred on these arcs yields the schedule delay cost. In this formulation, the specification of the arc cost consists of the time slice Δt multiplied by an appropriate factor consistent with the underlying utility function (Equation 1); for instance, in Figure 7, the multipliers EC and LC are equal to b/a and c/a, respectively.

The solution of the minimum cost trans-shipment problem under the preceding specification of the arc costs will thus be optimal for the system in terms of minimizing the total disutility of system users. Several variants are possible here, such as constraining all arrivals to occur within a particular time band. In this case, a large number M can be imposed on all vertical queuing arcs with at least one end outside the band and the schedule delay costs on those entirely within the band (still only node AR_n would be connected to DE). Alternatively, one can represent a utility function combining the behavioral features of both the satisfying and utility maximizing formulation. In particular, an indifference band of acceptable arrivals can be specified where all nodes in the indifference subset are connected to DE and no cost is associated with the vertical arcs connecting nodes in that subset. Vertical arcs outside this band will, however, be assigned a cost equal to the schedule delay disutility (but not large M).

The values of the relative weights of the various cost components would of course have to be determined outside this particular methodology. One use of this formulation is that is allows the systematic investigation of the impact of these relative valuations on the character of the optimal solution and the associated total system costs. However, the assumption of identical valuation across users may be too strong for practical applications.

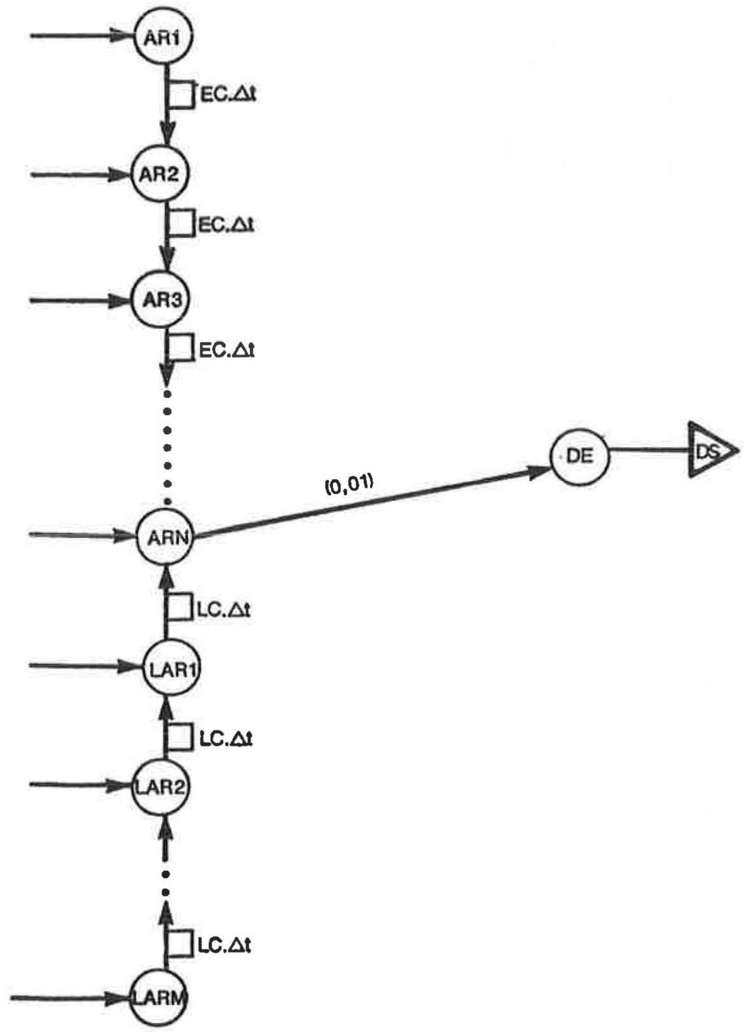

FIGURE 7 Network formulation of the arrival activity under the utility-maximizing decision rule.

EXAMPLES OF MORE GENERAL SYSTEMS

In this section we describe the representation, in the context of the preceding network modeling framework, of more general situations encountered in commuting systems. Still dealing with multiple origins, single destination systems, we first consider multiple bottlenecks (in series) along a single route, then multiple parallel routes. These types of systems have also been considered by Ben-Akiva et al. (*10, 11*) in their study of stochastic user equilibrium time-dependent flows.

Case 1: Multiple Bottlenecks Along a Single Route

Figure 8 depicts an example commuting system with two congested sections, BA and BB, where commuters departing from origin A have to traverse both sections, whereas those from downstream origin B encounter only the second bottleneck, BB. The network formulation for this problem is shown in Figure 9. Two sets of nodes $\{DA_i, i = 1, \ldots, n\}$ and $\{DB_j, j = 1, \ldots, n\}$, as defined previously, represent the feasible departure time alternatives of commuters from origins A and B, respectively. The first bottleneck BA is modeled by a set of node pairs, with each pair $\{(BA_k$ and $BA_k^*), k = 1, \ldots, n\}$ as described in the previous section. In the same manner, activities in the second bottleneck are represented by the set of node pairs, $\{(BB_k, BB_k^*), k = 1, \ldots, n\}$. The cost and upper bound associated with each arc are defined as shown in Figure 9, in a manner similar to the basic model of the previous section. Note that the set of arcs $\{(BA_k^*, BB_k), k = 1, \ldots, n\}$ corresponds to travel between the end of the first bottleneck section and the beginning of the second; no upper bounds on flow on these arcs need to be specified as these flows are regulated by the upstream bottleneck and no additional generation takes place in that sector. For the same reason, no vertical arcs connect the BA_k^* nodes. Finally, the arrival process follows the satisfying formulation illustrated in Figure 5, where the set of nodes $\{AR_t, t = 1, \ldots, n\}$ corresponds to the array of possible arrival times and the subset of those connected to the total demand node represents the presumed acceptable arrival interval. It should be mentioned that the possible departure periods for the two origins A and B are assumed to have an identical length and thus an equal number of nodes, for clarity of presentation. This is not

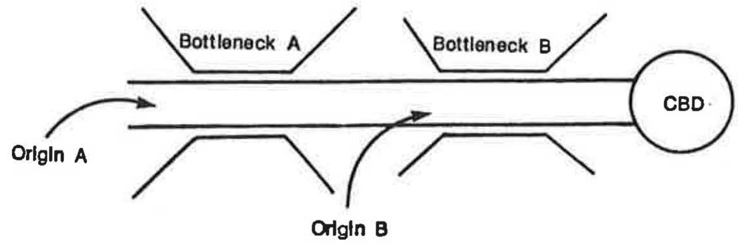

FIGURE 8 Commuting system with two bottlenecks along the single route.

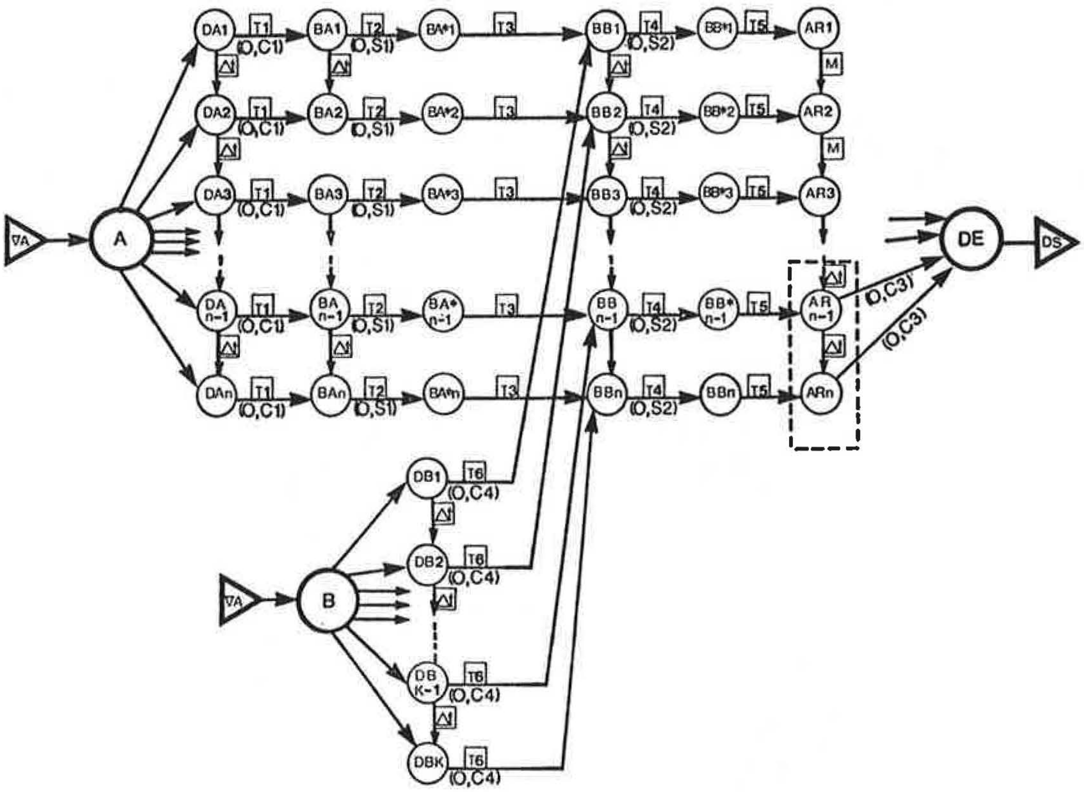

FIGURE 9 Network formulation for commuting system of Figure 8.

necessary, however, as long as they are properly connected to the rest of the network.

With the formulation of Figure 9, the system-optimal departure distribution patterns can be solved using any existing minimum cost, linear network code that implements the network simplex algorithm or its variants. See Kennington and Helgason (*28*) for a discussion of these algorithms.

Case 2: Multiple Parallel Routes

In the commuting system of Figure 10, there are three parallel routes, each containing two bottleneck sections, and commuters can choose their departure time as well as their route. To construct the network formulation of such a system, we can essentially follow the same procedure as in Case 1, with each route being formulated independently as one subnetwork. Then all subnetworks are tied together at both the common supply nodes and arrival nodes.

Figure 11 illustrates the resulting network formulation for this system. Nodes DA_k, DB_k, and DC_k, $k = 1, \ldots, n$ denote the feasible departure period of commuters from location A to travel through Routes A, B, and C, respectively. Node pairs ($A1_k$, $A1_k^*$) and ($A2_k$, $A2_k^*$), $k = 1, \ldots, n$, represent Bottlenecks 1 and 2, respectively, on Route A. Node pairs ($A3_k$, $A3_k^*$) and ($A4_k$, $A4_k^*$), $k = 1, \ldots, n$ represent Bottlenecks 3 and 4, respectively, on Route B. Node pairs ($A5_k$, $A5_k^*$) and ($A6_k$, $A6_k^*$), $k = 1, \ldots, n$ correspond to Bottlenecks 5 and 6, respectively, on Route C. The arrival process is represented as before with a common set of nodes D_1 to D_n for the arrival period, with the subset D_k to D_n defining the acceptable arrival time band.

Again, it should be noted that, for convenience of presentation, the feasible departure periods for the three routes in Figure 11 are assumed to consist of the same number n of time intervals. It is possible to let the feasible departure period vary from route to route. However, attention should be given to the formulation of the arrival period if the length of departure period is specified differently for different routes or if travel times in the absence of congestion on each route are not identical. Then the arrival time nodes D_1 and D_n should correspond to the earliest and latest possible arrival times, respectively, for any of the possible departure alternatives, on any route and from any origin.

The commuting activities from origin B can also be formulated in the same manner, but the complete graphical representation is not incorporated in Figure 11 for clarity. With the complete network thus formulated, the minimum cost flow pattern in the network will yield the system-optimal assignment to both routes and departures times. In this example, we have considered only nonoverlapping routes. However, a more general transport network can also be modeled in this framework, although the clarity of the graphical presentation would suffer markedly. A numerical example is presented in the next section along with some comments on implementation.

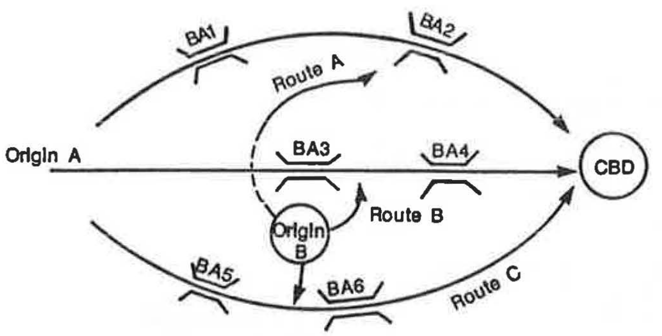

FIGURE 10 Commuting system with multiple bottlenecks on parallel routes.

NUMERICAL APPLICATION

To illustrate some of the issues involved in the application of the methodological framework discussed in this paper and the type of results one can expect, we describe an application to the commuting system shown in Figure 12a. The system is similar to that in Figure 10 in that it consists of two origins (A and B) with access to two parallel highway facilities to the common CBD destination. Each route contains two "bottleneck" sections, the first of which is traversed only by tripmakers from origin A. A constant access time of 5 min is assumed from each origin to the corresponding entry point on the highway facility. Figure 12b shows the characteristics (travel time, capacity per Δt) of each spatial link. Each node

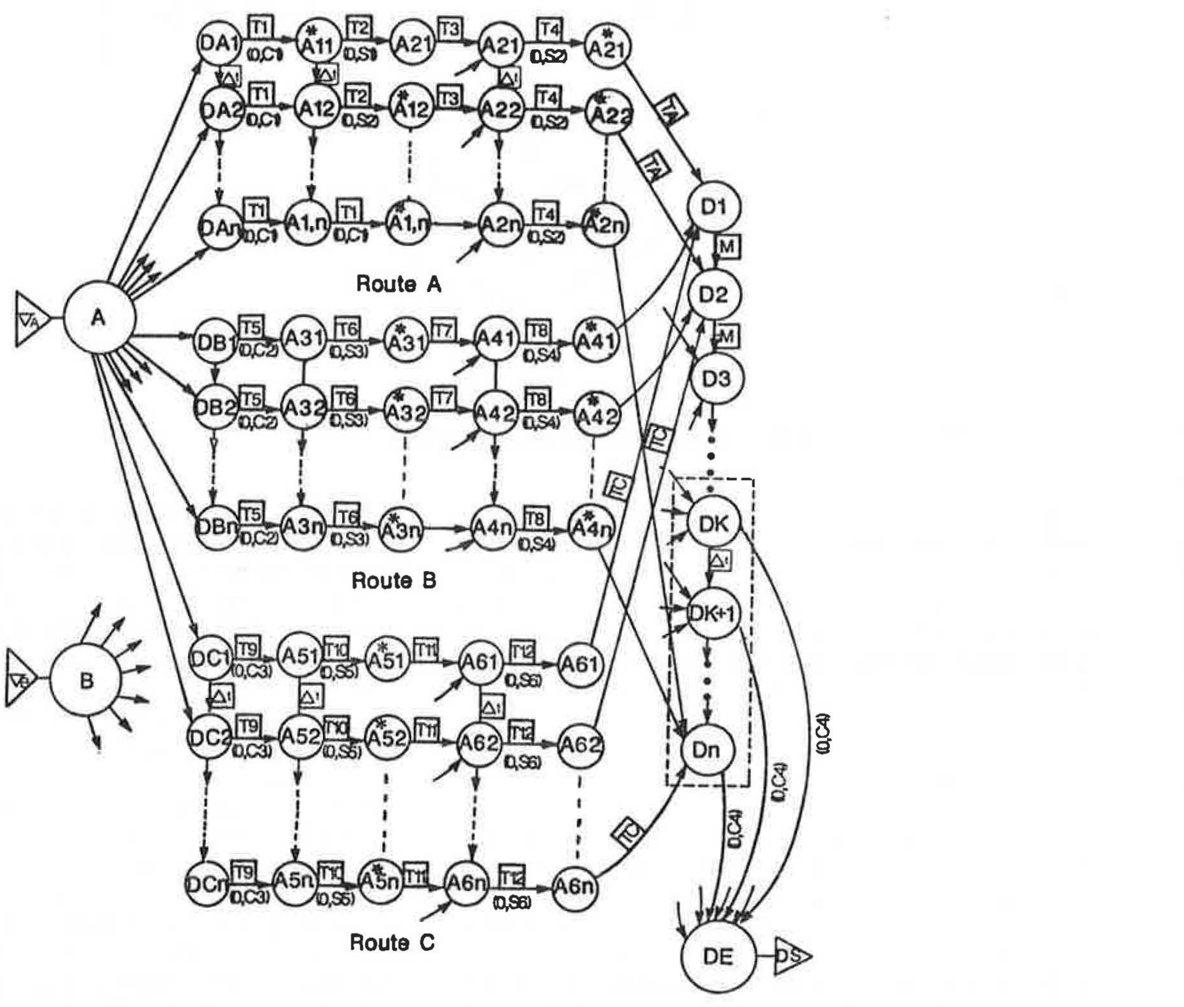

FIGURE 11 Network formulation for commuting system of Figure 10.

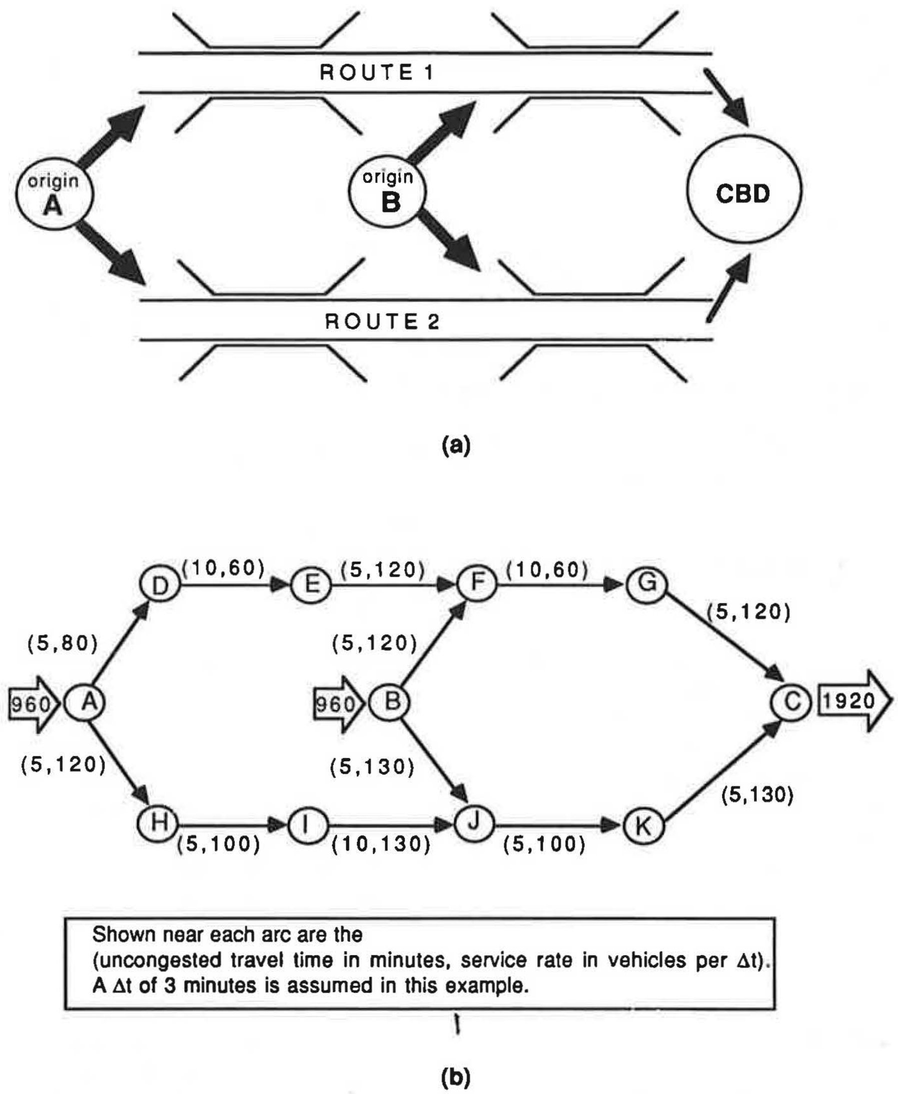

FIGURE 12 Commuting context and data for numerical example.

is assumed to generate a total of 960 vehicle trips during the commuting period.

The network formulation involves adding the time dimension to model waiting times due to congestion and formulating the departure and arrival processes. It should be apparent that developing and coding the time-space network can be a rather time-consuming task. As this network exhibits an obvious repetitive structure, however, this task can be very effectively supported by a network generating code. We have developed such a network generator for commuting systems involving multiple parallel routes with multiple origins. The program is interactive and requires simple input on the number of origins, number of routes, number of spatial nodes, operational characteristics of each spatial arc (i.e., the highway sections and ramps), total trips from each origin, size of the time slice Δt, as well as the range of possible departure times from each origin and acceptable arrival time band (for the satisfying formulation described earlier). This obviously greatly simplifies the practical use of this formulation, as the network can now be generated in an interactive session that requires only a few minutes. The network is then ready for solution by any pure network optimization code. These codes are known for their efficient execution and can easily handle networks with tens of thousands of arcs, thereby alleviating concern about the size of the network needed to model even relatively small physical commuting networks.

For the example under consideration, we have executed the algorithm for three different lengths (in minutes) of the acceptable arrival band: 15, 36, and unconstrained (i.e., all arrival time nodes in the range considered are connected to the total demand sink node). The latter case is included to provide a benchmark for comparing the effect of tightening or relaxing the size of the acceptable arrival (indifference) band on the departure patterns. It was assumed that 8:00 was the common work start time, thus the indifference band would correspond to 7:45–8:00 A.M., 7:24–8:00 A.M., and anytime before 8:00 A.M., respectively. The case with 15 min is not feasible, because that would imply a combined arrival rate much in excess of the capacity of the bottlenecks on the two routes. Actually, 36 min is the minimum feasible arrival period, yielding a total system cost of 52,800 min and a uniform arrival pattern of 160 arrivals per Δt (equal to 3 min in this example; see Figure 12). Because this solution involves no queuing, it cannot be improved on, as evidenced by the solution for the

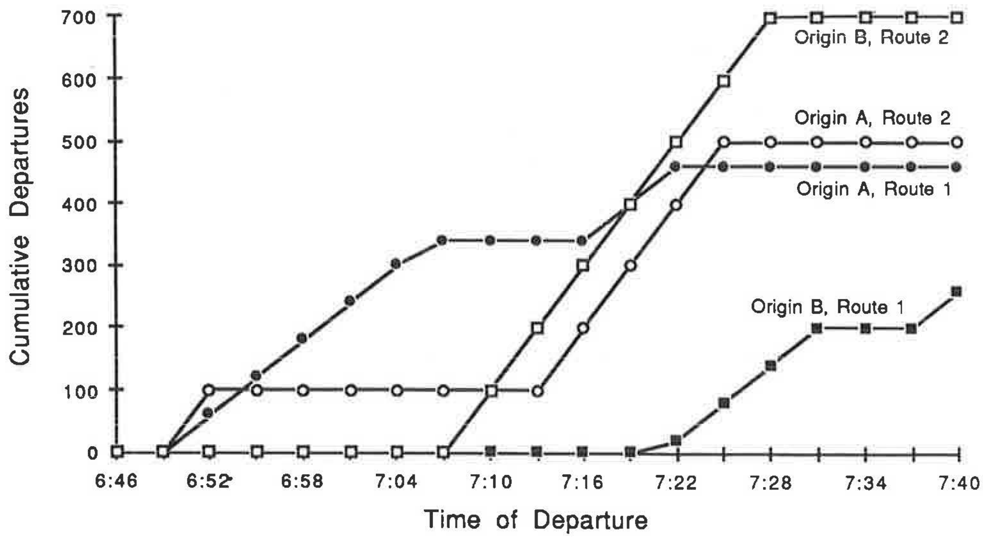

FIGURE 13 Cumulative number of departures from each origin by route.

unconstrained case, which yields more spread out departure and arrival patterns but at the same system cost.

The solution of the network optimization problem also includes the departure pattern from each node (i.e., the set of flows on the arcs connecting each origin to the possible departure time alternatives on each route), the arrival pattern at the CBD, the flows on the vertical queuing links, as well as all the link flows, in addition to the value of the objective function at optimality. The departure patterns from both origin A and origin B for each route are illustrated for the 36-min arrival period in Figure 13.

CONCLUDING COMMENTS

In this paper, a network formulation framework was proposed to solve for the system-optimal time-varying flows in urban commuting networks, yielding optimal departure patterns from each origin on each route as well as the dynamic assignment of traffic to the network's components. The solution of the formulated problem can take advantage of state-of-the-art, large-scale network optimization algorithms. Of course, the representation of the traffic phenomena that may be occurring on the facilities is admittedly crude and simplified; however, this has been a problem in much of the network traffic assignment work, for the static case and particularly for the time-varying formulations. We feel that some compromises in representation, when applied judiciously to preserve the character of the system insofar as the phenomena of interest are concerned, are worth the resulting relative ease of the solution procedure and thus the ability to explore and gain insight into the various aspects of this problem.

It should further be noted that the work presented here is not motivated by a desire to force people to leave at specified times and on preset routes, or by a naive presumption that they would comply if told to do so. Rather, it is intended to generate a benchmark, an "ultimate" solution against which to compare the effectiveness of various strategies, such as, for example, flexible work arrival times. Furthermore, it can be a useful tool to examine the potential of information-related strategies, whereby users could be guided toward the optimal solution. Of course, economists hold the view that one could approach the desired state through pricing; this strategy is not a particularly strong motivator for this work. Another appropriate application of this methodology is the development of contingency evacuation plans for use during some emergency, such as a hurricane or an incident at a nuclear power plant, or for military purposes.

Several improvements and extensions of the methodological framework can be considered. In terms of system representation, extension to the many origins to many destinations case would be most desirable. However, the penalty is rather severe as the problem would then exhibit the features of a capacitated multicommodity problem, which requires additional assumptions for proper resolution, in addition to the obvious increase in the level of complexity required in the representation. Improvements in terms of traffic modeling are certainly possible, but one would then have to sacrifice the easy-to-solve pure network formulation.

ACKNOWLEDGMENTS

This paper is based on research funded by National Science Foundation grant to the University of Texas at Austin. The authors are grateful to Robert Herman for his encouragement and contribution to their efforts in this general problem area. This work has also profited from useful discussion with Andre de Palma of Northwestern University.

REFERENCES

1. A. S. Alfa and D. L. Minh. A Stochastic Model for the Temporal Distribution of Traffic Demand—the Peak Hour Problem. *Transportation Science*, Vol. 13, No. 4, 1979, pp. 315–324.
2. M. Ben-Akiva, M. Cyna and A. de Palma. Dynamic Model of Peak Period Congestion. *Transportation Research*, Vol. 18B, 1984, pp. 339–355.
3. C. F. Daganzo. The Uniqueness of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck. *Transportation Science*, Vol. 19, No. 1, 1985, pp. 29–37.

4. A. de Palma, M. Ben-Akiva, C. Lefevre, and N. Litinas. Stochastic Equilibrium Model of Peak Period Traffic Congestion. *Transportation Science*, Vol. 17, No. 4, 1983, pp. 430–453.
5. P. N. Fargier. Effects of the Choice of Departure Time on Road Traffic Congestion: Theoretical Approach. In *Proc., Eighth International Symposium on Transportation and Traffic Theory* (V. F. Hurdle et al., eds.), University of Toronto Press, Toronto, 1983, pp. 223–263.
6. J. V. Henderson. Road Congestion: A Reconsideration of Pricing Theory. *Journal of Urban Economics*, Vol. 1, 1974, pp. 346–365.
7. C. Hendrickson and G. Kocur. Schedule Delay and Departure Time Decisions in a Deterministic Model. *Transportation Science*, Vol. 15, 1981, pp. 62–77.
8. M. J. Smith. The Existence of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck. *Transportation Science*, Vol. 18, No. 4, 1984, pp. 385–394.
9. W. S. Vickrey. Congestion Theory and Transport Investment. *American Economic Review*, Vol. 59, 1969, pp. 251–261.
10. M. Ben-Akiva, A. de Palma, and P. Kanaroglou. Effects of Capacity Constraints on Peak-Period Traffic Congestion. *Transportation Research Record 1085*, TRB, National Research Council, Washington, D.C., 1986, pp. 16–26.
11. M. Ben-Akiva, A. de Palma, and P. Kanaroglou. Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates. *Transportation Science*, Vol. 20, No. 3, 1986, pp. 164–181.
12. H. S. Mahmassani and G. L. Chang. On Boundedly Rational User Equilibrium in Transportation Systems. *Transportation Science*, Vol. 21, No. 2, 1987, pp. 89–99.
13. H. S. Mahmassani and R. Herman. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. *Transportation Science*, Vol. 18, No. 4, 1984, pp. 362–384.
14. G. F. Newell. The Morning Commute for Nonidentical Travelers. *Transportation Science*, Vol. 21, No. 2, 1987, pp. 74–88.
15. H. S. Mahmassani and G. L. Chang. Experiments with Departure Time Choice Dynamics of Urban Commuters. *Transportation Research B*, Vol. 20B, No. 4, 1986, pp. 297–320.
16. G. L. Chang. Departure Time Choice Dynamics in Urban Transportation Networks. Ph.D. dissertation. Department of Civil Engineering, University of Texas at Austin, 1985.
17. H. S. Mahmassani and G. L. Chang. Dynamic Aspects of Departure Time Choice Behavior in a Commuting System: Theoretical Framework and Experimental Analysis. *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 88–101.
18. H. S. Mahmassani, G. L. Chang, and R. Herman. Individual Decisions and Collective Effects in a Simulated Traffic System. *Transportation Science*, Vol. 20, No. 4, 1986, pp. 258–271.
19. H. S. Mahmassani and C.-C. Tong. Availability of Information and Dynamics of Departure Time Choice: Experimental Investigation. *Transportation Research Record 1085*, TRB, National Research Council, Washington, D.C., 1986, pp. 33–46.
20. J. V. Henderson. The Economics of Staggered Work Hours. *Journal of Urban Economics*, Vol. 9, 1981, pp. 349–364.
21. C. Hendrickson, D. Nagin, and E. Plank. Characteristics of Travel Time and Dynamic User Equilibrium for Travel-to-Work. In *Proc., Eighth International Symposium on Transportation and Traffic Theory*, (V. F. Hurdle et al., eds.), University of Toronto Press, Toronto, 1983, pp. 321–347.
22. D. K. Merchant and G. L. Nemhauser. A Model and an Algorithm for the Dynamic Traffic Assignment Problem. *Transportation Science*, Vol. 12, 1978, pp. 183–199.
23. D. K. Merchant and G. L. Nemhauser. Optimality Conditions for a Dynamic Traffic Assignment Model. *Transportation Science*, Vol. 12, 1978, pp. 200–207.
24. M. Carey. Optimal Time-Varying Flows on Congested Networks. *Operations Research*. Vol. 35, No. 1, 1987, pp. 58–69.
25. V. F. Hurdle. Equilibrium Flows on Urban Freeways. *Transportation Science*, Vol. 15, No. 3, 1981, pp. 255–293.
26. D. Klingman, N. Phillips, D. Steiger, R. Wirth, R. Padman, and R. Krishnan. An Optimization-Based Integrated Short-Term Refined Petroleum Product Planning System. *CBDA* 123. Center for Business Decision Analysis, Unversity of Texas at Austin, 1985.
27. C. Hendrickson and E. Plank. The Flexibility of Departure Times for Work Trips. *Transportation Research A*, Vol. 18A, No. 1, 1984, p. 25–36.
28. J. L. Kennington and R. V. Helgason. Algorithms for Network Programming. John Wiley & Sons, New York, 1980.